Evaluation of Combined Spectral Enhancement and Phase Adjustment in Hearing-Aids

EVALUATION OF COMBINED SPECTRAL ENHANCEMENT AND PHASE ADJUSTMENT IN HEARING-AIDS

BY
TIMOTHY J. ZEYL, B.Eng.

A THESIS
SUBMITTED TO THE DEPARTMENT OF ELECTRICAL & COMPUTER ENGINEERING
AND THE SCHOOL OF GRADUATE STUDIES
OF MCMASTER UNIVERSITY
IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF APPLIED SCIENCE

Master of Applied Science (2010)
(Electrical & Computer Engineering)

McMaster University
Hamilton, Ontario, Canada

TITLE:	Evaluation of Combined Spectral Enhancement and Phase Adjustment in Hearing-Aids
AUTHOR:	Timothy J. Zeyl B.Eng, (Electrical & Biomedical Engineering) McMaster University, Hamilton, Canada
SUPERVISOR:	Dr. Ian C. Bruce
NUMBER OF PAGES:	xiii, 174

Abstract

Auditory nerve fibers in an ear with outer hair cell damage can be conceptualized as filters having a broadened frequency response area, a shallower phase response and a shorter group delay with respect to a healthy fiber, particularly at low stimulus presentation levels. As well, the presence of inner hair cell damage requires increased stimulus presentation levels for restoration of fiber discharge rates, which results in broad auditory filters with shallow phase response and short group delay. As a consequence, the discharge times in the impaired ear in response to a tone stimulus are more coincident across a population of fibers with a range of characteristic frequencies. This behaviour resembles the spatiotemporal response pattern in a healthy auditory periphery in response to loud stimuli and has been postulated as a potential correlate to loudness recruitment. Commercial hearing aids do not address any changes in the phase characteristics of impaired auditory nerve fibers.

The present study evaluates the potential for correction of the altered temporal relationships in the neural firing pattern of the impaired ear by a hearing aid. An improved version of the spatiotemporal pattern correction (SPC) scheme (Shi *et al.*, 2006) is presented, which measures the instantaneous difference in group delay between a bank of model healthy and impaired auditory nerve fibers and inserts the corresponding delays into an analysis-synthesis gammatone filterbank in the hearing aid. Human testing of the original processing scheme showed that listeners preferred unprocessed sounds over processed sounds and that no systematic improvement in speech intelligibility was provided by the processed speech (Shi *et al.*, 2006; Calandruccio *et al.*, 2007). However, no computational analysis of this scheme was reported and no attempts were made to restore synchrony to the formants of voiced speech in the neural response. To address this issue, the current work pre-processes SPC with an updated version of the multiband and improved, contrast-enhanced frequency shaping (MICEFS) algorithm (Harte *et al.*, 2006) to restore formant synchrony.

These processing schemes are evaluated with a computational model of the auditory periphery (Zilany and Bruce, 2006, 2007) in response to a synthesized vowel for a number of hearing loss types. Analysis indicates that SPC disrupts the spatiotemporal response, but some synchrony can be preserved if the signal is pre-processed with MICEFS and if the SPC delays are applied at the appropriate time. There are some

technical problems associated with the processing scheme discovered in this study that remain unresolved. These include: i) a non-flat frequency response through the analysis-synthesis filterbank due to changes in the relative temporal alignment of filterbank channels, ii) group delay corrections that are based on potentially incorrect frequencies due to the spread of synchrony in auditory nerve responses, and iii) frequency modulations in the processed signal created by the insertion of time-varying delays, which results in noise that is audible to normal listeners. Despite these issues, evaluation with an error metric derived from auditory nerve response cross-correlations shows that this processing scheme improves some features of the spatiotemporal response, even though it degrades others.

Acknowledgements

There are many people who have helped me throughout the course of this thesis work, and I am very grateful.

I thank my supervisor Dr. Ian Bruce, who has been a real mentor to me. His guidance has been invaluable throughout this thesis and he has helped me to grow as a student. I am also thankful for his friendship and hospitality, especially during the six months of this research spent in Melbourne.

Thanks to all my colleagues at McMaster and to my defence committee members, Dr. Suzanna Becker and Dr. Hubert de Bruin, who have both been good teachers in the years leading up to this thesis. I would also like to thank Prof. Tony Burkitt and all my colleagues in Melbourne for welcoming me into their lab and providing a stimulating research atmosphere.

Thanks to the Eng family for inviting me into their home on many weekends and feeding me well. Thanks especially to Dorothy Eng for her love and support throughout this thesis and for encouraging me to study in Melbourne, even when it meant sacrifice for her. Special thanks to my family for their love and support throughout all of my schooling, and especially in the last two years. Thanks to Mom and Dad for their unfailing encouragement in all of my endeavours. For all of this I am truly blessed, and I thank God for these gifts.

This work was funded in part by NSERC Discovery Grant 261736 and by an NSERC Canadian Graduate Scholarship.

Notation and Abbreviations

AN	=	Auditory Nerve
BF	=	Best Frequency
BM	=	Basilar Membrane
CF	=	Characteristic Frequency
dB	=	deciBel
DFT	=	Discrete Fourier Transform
DSP	=	Digital Signal Processing
F1	=	First Formant
F2	=	Second Formant
F3	=	Third Formant
F4	=	Fourth Formant
IHC	=	Inner Hair Cell
NAL-R	=	National Acoustic Laboratories - Revision 1
NAL-RP	=	National Acoustic Laboratories - Revised, Profound
NAL-NL1	=	National Acoustic Laboratories - Non-linear, version 1.1
OHC	=	Outer Hair Cell
SPL	=	Sound Pressure Level
STFT	=	Short-Time Fourier Transform

Contents

List of Figures

Chapter 1

Introduction

The sensation of sound involves the transmission of electrical activity, in the form of action potentials (or spikes), from the inner ear to the brain via the auditory nerve (AN). Information describing the sound at the level of the auditory nerve is thought to be encoded in: i) the rate at which action potentials are firing, ii) the specific auditory nerve fibers that are firing, and iii) the timing of action potentials. These three properties of the nerve fiber response can be conceptualized by the transfer function of an auditory filter with a specific gain, bandwidth and phase response.

In an ear with damage to the outer hair cells in the organ of Corti, which play a role in the frequency specificity of individual auditory nerve fibers, the auditory filters often have a broadened frequency response area, a shallower phase response and a shorter group delay than in healthy ears. An ear with inner hair cell loss can be approximated by a filter with lower gain or a higher threshold of response, which requires the stimulus to be amplified for sufficient audibility. This indirectly results in auditory filters with broad bandwidths, shallow phase response and short group delay, because these changes occur as sound level increases. Therefore, both inner and outer hair cell impairments mean that, not only is there increased level thresholds at which AN fibers begin to fire, there is also decreased frequency selectivity and errors in the timing of action potentials in AN fibers. All of these can affect the neural information encoding the sound.

Commercial hearing-aids have accounted for decreased audibility by applying frequency and level specific gain to the incoming sounds. However, they have not been able to account for the flattened phase response and shortened group delay exhibited by auditory nerve fibers in impaired ears, which results in abnormal timing of action potentials. Shi *et al.* (2006) present a spatiotemporal pattern correction (SPC) scheme that attempts to correct for errors in the timing of action potentials by correcting for impaired auditory filter group delays. Their scheme measures the instantaneous difference in group delay between a bank of model healthy and impaired AN fibers,

then inserts the corresponding delays into an analysis-synthesis gammatone filterbank in the hearing aid. However, they tested their scheme on human listeners, and found that it performed poorly, with listeners in general preferring unprocessed sounds.

The original SPC scheme did not attempt to restore normal synchronous activity in the auditory nerve before attempting to correct for the impaired auditory filter group delays. In high-frequency sensorineural hearing impairments, there is decreased synchrony in the AN response to the higher frequency components of the stimulus. As well, AN fibers that would normally be synchronized to high frequency components show increased synchrony to lower frequency components. Because the auditory model's group delay is influenced by the input stimulus, this could have resulted in group delay calculations that were based on the wrong synchronization frequency in the original SPC scheme.

This thesis makes several improvements on the SPC processing scheme and attempts to restore normal synchrony in the auditory nerve prior to SPC filtering by pre-processing the sound with an updated version of the Multiband and Improved Contrast-Enhanced Frequency Shaping (MICEFS) scheme introduced by Ansari (2005) and Harte *et al.* (2006). MICEFS combines multiband compression with a spectral enhancement scheme that amplifies the second and third formants of voiced speech to restore synchrony to those formants in the AN response pattern. A computational model of the auditory periphery is employed (Zilany and Bruce, 2006, 2007) to evaluate whether pre-processing the SPC scheme with MICEFS can improve its performance and to determine if the generally poor results of human testing of the original SPC can be explained in its AN response pattern.

1.1 Scope of Work

The goal of this work is to analyze the effectiveness of hearing aid signal processing schemes at restoring the spatiotemporal response pattern in an impaired auditory nerve and to interpret the results in order to make improvements on these schemes. The results of such improvements are presented in this thesis and compared to the performance of the original processors. Analysis of the AN spatiotemporal response pattern is done to determine which features are restored and how this may be related to sound perception.

Hearing-aid signal processing schemes were simulated with a computer, and their development was achieved in the MATLAB® programming language. It was not in the scope of this thesis to implement the processing schemes developed here on a real hearing aid. As such, memory and speed limitations that are inherent in hearing-aid digital signal processing (DSP) chips were not applied to the development of the SPC processing scheme. However, the MICEFS algorithm was implemented in Simulink®, which will allow for future programming of real-time implementation on a DSP chip,

although this last step has not been implemented and no memory limitations were accounted for in this case.

1.2 Contributions of this Work

The MICEFS algorithm was used as a pre-processor for the SPC scheme to determine if restoration of synchrony to the formants of voiced speech would improve SPC's ability to correct for impaired group delay. In addition to this, several improvements were made to the MICEFS algorithm. First, it was implemented in Simulink®, which will allow for a real-time implementation on a DSP chip. Second, parameters of the multi-band compression and spectral enhancement processing blocks that appear in MICEFS were adjusted to improve its performance across hearing-loss types and stimulus sound pressure levels. It was determined that the spread of synchrony to individual formants is sensitive to the amount of gain that is applied to each formant, which can be dependent on the hearing-loss profile. This could mean that derivation of the MICEFS contrast enhancement scheme must be dependent on the individual hearing-loss profile, or prescribed non-linearly as a function of hearing loss.

The SPC algorithm originally presented by Shi *et al.* (2006) has been updated to incorporate the auditory model of Zilany and Bruce (2006, 2007) to prescribe the insertion delays at each channel in the filterbank. As well, the SPC scheme is updated to account for the different latencies in the modelling path and the processing path; this ensures that group delays prescribed at a specific point in time in the modelling path are applied at the corresponding time-point in the processing path. Accounting for latencies in this way is shown to be important for retaining some formant synchrony after SPC processing, and this version of SPC demonstrated the best overall performance on the error metrics used in this thesis.

In addition to the development and evaluation of these hearing-aid signal processing schemes, a novel error metric was introduced for comparison of model healthy and impaired auditory nerve response. This metric is based on the cross-correlation of the discharge patterns between AN fiber pairs with neighbouring characteristic frequencies. A version of this metric that compares the lags at which these cross-correlations are maximized is used in this thesis to quantify the relative phase of firing between fibers without, in principle, requiring any specific knowledge about the stimulus frequency. Further work can be done on this error metric so that it incorporates the value of cross-correlation so that information about spiking rate is preserved.

1.3 Thesis Layout

The next chapter, Chapter 2, introduces some background on current knowledge of the mammalian auditory system. Focus is given to the human ear, but much of the

data derived from electrophysiological recordings comes from various mammals. This chapter introduces the concept of the auditory filter and describes how its properties affect auditory nerve response. Next, Chapter 3 gives a background on the computational model of the auditory periphery of Zilany and Bruce that is used in both the implementation of the SPC processing scheme and in the evaluation of hearing-aid schemes developed in this thesis. Following this, Chapter 4 provides motivation for the present work by describing current hearing-aid schemes and points out where they lack effectiveness at restoring some of the features of impaired auditory nerve response presented in Chapter 2. Chapter 5 describes the signal processing methods used in this thesis; it describes the implementation of the MICEFS and SPC schemes used here and explains the improvements that were necessary. Following this, Chapter 6 presents some of the methods used for evaluating the effectiveness of hearing-aids at restoring a near-normal auditory nerve response. As well, this chapter describes the development of the neighbouring AN fiber cross-correlation metric. Chapter 7 presents the results of evaluating various combinations of MICEFS and SPC and Chapter 8 discusses these results. Then Chapter 9 concludes this thesis with a summary of findings and suggestions for future research.

Chapter 2

Background i. Normal and Impaired Hearing

This chapter describes the nature of hearing as it pertains to the peripheral auditory system, from the outer ear to the auditory nerve, and is particularly concerned with the neural coding of sound at the level of the auditory nerve. It begins with a description of the physical nature of sound then describes the processes involved in translating that sound from a mechanical vibration into a neural code. Functional and anatomical descriptions of various processing stages in the ear are presented here in the same order that they are processed in the auditory periphery. Auditory nerve fibers are conceptualized as bandpass filters, and their properties are described from a signal processing standpoint. Hearing impairment is described in terms of the changes it induces on these filter properties, and some thought is given to the effects of these changes on the perception of sound.

2.1 The Nature of Sound

Sound can be described in physical terms as a pattern of pressure changes in a medium such as air that originates from the vibration of an object (Yost, 2007). In air, pressure fluctuations are the result of changes in density of the air molecules in a finite region. When the air molecules are squeezed together it is called condensation, and when the air molecules are more sparse than normal it is called rarefaction (Moore, 2003). A sound wave describes a pattern of changes in air condensation and rarefaction that propagates through space.

Sound can also be described in perceptual terms. This means that people are sensitive to physical sound waves and through the act of hearing they can perceive the sound. For example, when someone says a word, such as 'hello', they produce a pattern of pressure fluctuations in the air which can be heard by a listener. The

listener is sensitive to this sound, and can identify the word, likely extracting certain features from this pressure waveform that are important for perception. Even more, a listener can identify the word 'hello' when spoken from different talkers, each of whom will generate a slightly different pressure waveform when speaking this word. In addition to word identification, other perceptual features can be recognized like the loudness of the sound, the pitch and emotional quality; all of which can add to the meaning of this sound. This translation of sound from a physical pressure waveform to a mental percept is complex and its description is not fully known. The current chapter describes some aspects of the translation of the sound waveform into a neural response in the auditory nerve (AN) and identifies characteristics of this response that correlate with certain perceptual features.

One of the first stages in hearing is the collection of sound by the outer ear and subsequent vibration of the eardrum. Pressure changes in the physical sound wave result in pressure differences between the air outside the eardrum and the air inside the eardrum, which causes the eardrum to vibrate. This describes the physical transfer of sound from the vibration of an object in the outside world to a vibration inside our ears. Processing of this sound in the auditory periphery, as well as the anatomy of the ear, will be described in the next section.

2.2 Anatomy and Physiology of the Auditory Periphery

2.2.1 Outer and Middle Ears

The auditory periphery refers to the portion of the auditory system which is located peripherally to the auditory midbrain brainstem and includes the ear and the auditory nerve[1]. A depiction of the human ear is shown in Figure 2.1 in which three physically distinct sections can be identified. These sections are called the outer ear, the middle ear and the inner ear.

The outer ear includes the pinna, external auditory canal (meatus) and the tympanic membrane (eardrum) (Tortora and Derrickson, 2006). The outer ear provides a frequency dependent gain to the sound pressure waves that it collects (Pickles, 2008). This is a consequence of frequency dependent resonances that occur when a sound is collected due to the the physical shape of the outer ear. In addition to providing frequency dependent gain, the outer ear plays a role in sound localization. The pinna is shaped like a concave disk aimed towards the front hemisphere of an individual's auditory space and this horizontal asymmetry causes sounds originating from the

[1]Some sources inlcude only the cochlea and the auditory nerve in the definition of the auditory periphery as in Pressnitzer *et al.* (2008).

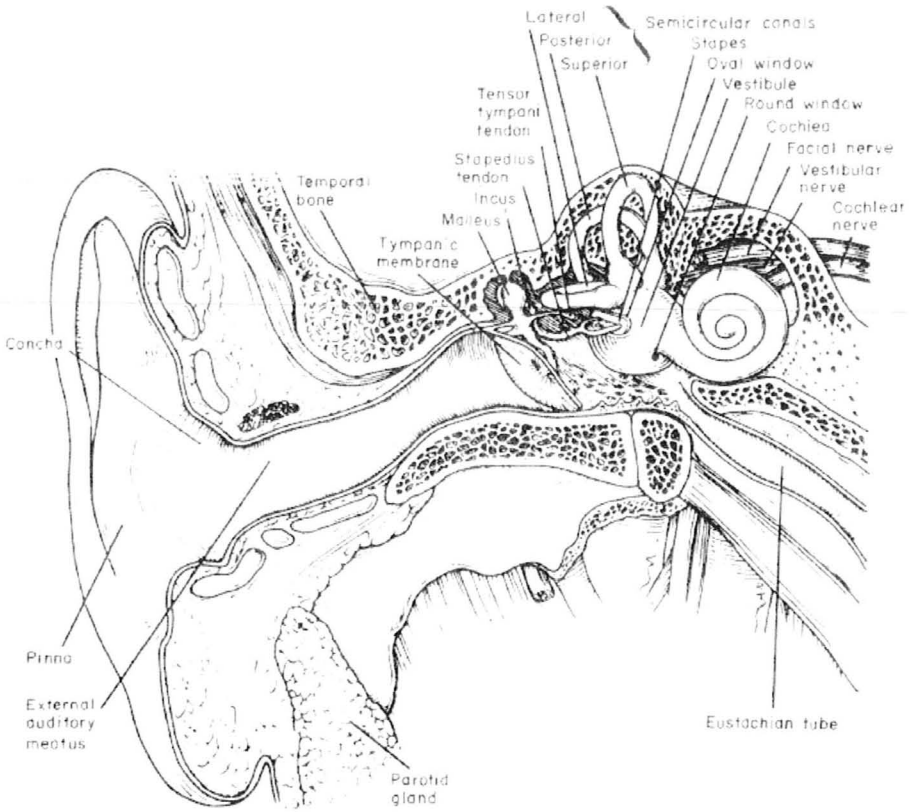

Figure 2.1: A schematic diagram of the human ear demonstrating its three compartments: the outer ear, the middle ear, and the inner ear. See text for a description of these compartments. (Reprinted from Kessel and Kardon, 1979)

front to be amplified differently than those originating from behind. This means that two identical sounds, one coming from in front and one from behind will have distinct characteristics at further stages of auditory processing, allowing the brain to differentiate between the front and the back auditory space. The pinna is also vertically asymmetrical, thereby providing unique signal shaping as a function of relative height of the sound source (Pickles, 2008).

The middle ear is an air-filled cavity that lies between the outer and inner ears. It is bounded to the outer ear by the tympanic membrane and to the inner ear by the round window and the oval window, two membranous openings in the bony structure of the inner ear. The middle ear connects the outer and inner ears with a chain of small bones, called ossicles, that are held in place with ligaments. The malleus is attached to the tympanic membrane which initiates a chain reaction of movement in the ossicles when it vibrates. The malleus connects at its other end to the incus, which connects to the stapes. When sound is collected by the outer ear, the vibration

of the ossicles is set into motion and the stapes pounds on the membrane-covered oval window of the inner ear. This causes the fluid inside the bony capsule of the inner ear to be set into motion (Tortora and Derrickson, 2006).

The middle ear acts as an impedance transformer, coupling the low impedance of the tympanic membrane to the high impedance of the oval window. The surface area of the tympanic membrane is about 20 times larger than that of the oval window (Geisler, 1998) and this concentrates the forces acting on the eardrum onto a smaller area in the oval window, increasing the pressure transmitted (Pickles, 2008).

The middle ear also provides some frequency dependent shaping; its transfer function is defined as the pressure gain across the middle ear from the ear canal to just inside the cochlear vestibule (refer to Figure 2.2) (Aibara *et al.*, 2001). The shape of the pressure gain as a function of frequency can be described as a sloping bandpass filter with a peak pressure gain of around 23.5 dB at 1.2 kHz, with a slope of around +/- 6 dB/octave surrounding 1.2 kHz (Aibara *et al.*, 2001). The reduction in gain at low frequencies can, at least in part, be attributed to the stiffness of the tympanic membrane. At low frequencies, the displacement of air is large compared to higher frequencies so the tympanic membrane must be displaced to a greater degree, requiring more force (or a higher sound pressure level (SPL)) to accommodate that displacement (Pickles, 2008). At high frequencies the reduction in middle ear gain may be caused in part by an increase in the required amount of acceleration on the ossicles in order to maintain a constant SPL (Pickles, 2008).

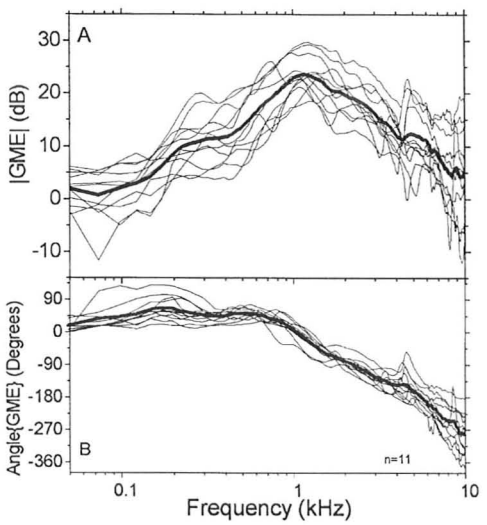

Figure 2.2: Middle ear sound pressure gain, GME, as a function of frequency measured from the outer ear to the inner ear in 11 human cadavers. *Top Panel*: Magnitude response, and *Bottom Panel*: phase response. (Reprinted from Aibara *et al.*, 2001)

2.3 The Cochlea

The human inner ear is a bony labyrinth that can be separated into two anatomically and functionally distinct parts, one part dedicated to the vestibular system[2] and the other dedicated to the auditory system. The cochlea is the auditory portion of the inner ear. It is a small spiral apparatus, about 1 cm wide and 5 mm high (from base to apex), containing three fluid filled canals, called scalae, running side by side throughout the spiral (Pickles, 2008). These scalae, called the scala vestibuli, scala media and scala tympani are depicted in a cross-section of one turn of the cochlea in Figure 2.3. Each of these canals is filled with fluid; the scala vestibuli and the scala tympani contain perilymph, a fluid with an ionic composition similar to extracellular fluid, and the scala media contains endolymph, a fluid with ionic composition similar to intracellular fluid. The scala vestibuli and the scala tympani are continuous at the apex of the cochlea, so fluid is able flow from one space to the other. The scala media is isolated from the other scalae throughout the length of the cochlear spiral by two membranes. It is separated from the scala vestibuli by Reissner's membrane and from the scala tympani by the basilar membrane (BM).

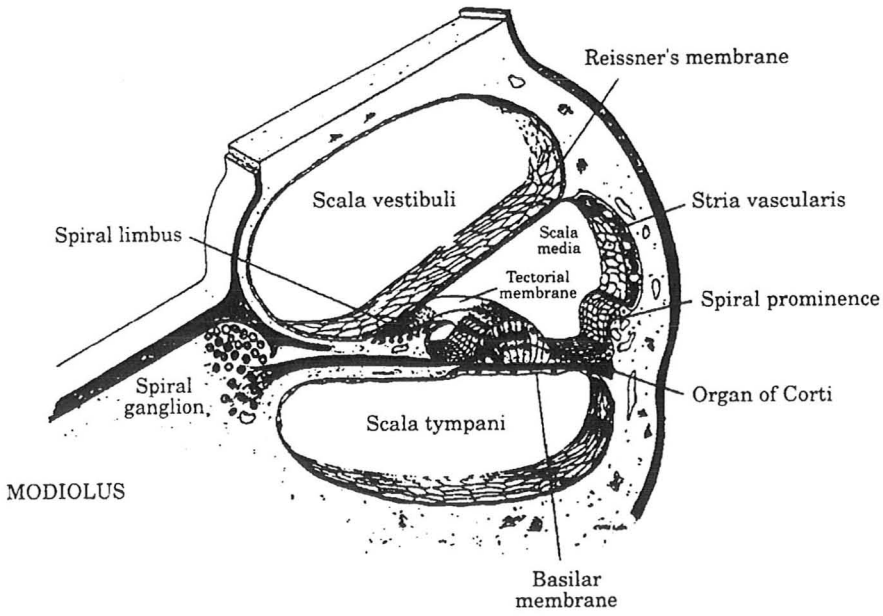

Figure 2.3: Illustration of a cross section of the cochlear spiral. The three scala compartments, the scala vestibuli, scala media and scala tympani are shown. (Reprinted from Bloom and Fawcett, 1975)

[2]The vestibular system is important for our sense of balance and spatial orientation.

The middle ear communicates with the cochlea through the oval window; the stapes pushes on the oval window when it vibrates causing fluid to flow within the cochlea. The perilymph in the scala vestibuli is thus set into motion by the movement of the oval window, and because it is an incompressible fluid, this causes displacement in the round window (Geisler, 1998). When the stapes pushes forward on the oval window, the direction of the flow of perilymph is from base to apex along the length of the scala vestibuli, then around the helicotrema and from apex to base along the length of the scala tympani; this is depicted in Figure 2.4. As the sound wave propagates through the perilymph along the cochlea, some energy is absorbed by vibration of the BM (Geisler, 1998).

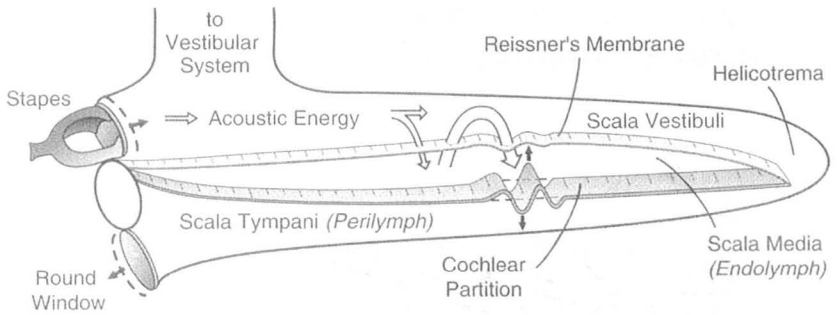

Figure 2.4: Illustration of the cochlear spiral uncoiled to demonstrate the absorption of acoustic energy in the vibration of the basilar membrane. Not drawn to scale. (Reprinted from Geisler, 1998)

2.3.1 Basilar Membrane

The basilar membrane is a thin, flat tissue separating the scala tympani from the scala media and running along the length of the cochlea. It is narrow at the base of the cochlea and gradually gets wider towards the apex, this has a profound effect on its elastic properties. As a wave propagates in the perilymph, the basilar membrane is displaced back and forth in a direction perpendicular to its surface. It is generally thought that there are two components contributing to the vibration of the basilar membrane in response to a stimulus: one is passive and is derived from the physical characteristics of the basilar membrane; the other is an active feed forward process, which enhances the vibration of the basilar membrane. The passive response provides baseline tuning properties of the basilar membrane and it is described here. The active response enhances the tuning of the passive response and will be described further in Section 2.3.4.

In the passive BM response, the position of maximum vibration is dependent on the tissue's mass, stiffness, and the frequency of the stimulus. Since the basilar membrane is narrow at the cochlea's base, its elastic properties are such that it resonates with high frequency stimuli, and therefore high frequencies are encoded at the base of the cochlea. Likewise, low frequencies are encoded at the apex of the cochlea because the basilar membrane is wider and resonates to low frequency energy. Due to the gradual widening of the membrane from base to apex, there is a gradient of frequency dependent resonance from high to low frequencies along the basilar membrane, with each place tuned to a specific frequency. Figure 2.5 shows an uncoiled, stylized basilar membrane resonating to a mid-frequency stimulus along with auditory nerve fibers attached to the basilar membrane at different frequency places. The place of AN fiber attachment and the tuning of the basilar membrane determine the frequency at which an AN fiber is most sensitive, termed the fiber best frequency (BF) or characteristic frequency (CF).

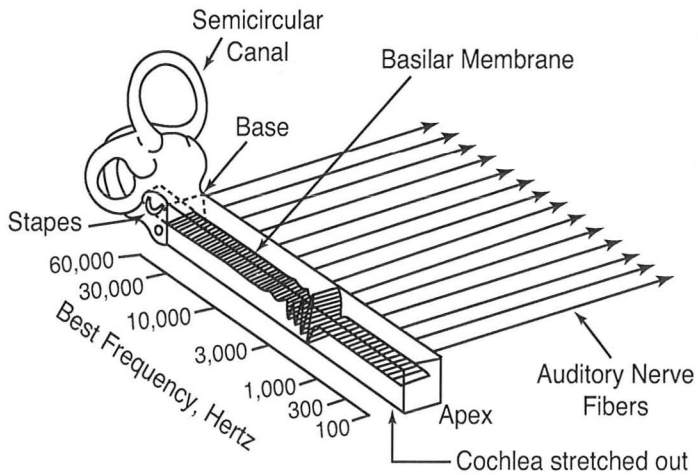

Figure 2.5: Illustration of the basilar membrane uncoiled to demonstrate its frequency tuning properties. High frequencies are encoded at the base and low frequencies are encoded at the apex. (Reprinted from Sachs *et al.*, 2002)

Travelling Wave

In cadavers, there is no active component to basilar membrane vibration present so the passive response can be measured. The patterns of basilar membrane vibration in response to tones were studied by von Békésy (1947). His description of the travelling wave in the basilar membrane is of particular importance and Figure 2.6 demonstrates some of his findings. The travelling wave theory describes a gradual

growth in amplitude of basilar membrane vibration when measured from the stapes to the place of resonance, then a relatively quick decrease in vibration amplitude further along the basilar membrane. The envelopes of the vibration amplitude can be seen in the top panel of Figure 2.6 in response to a variety of tones. The phase angle associated with each envelope demonstrates that the wave travels in the direction from the base to the apex, or from the stapes to the helicotrema. The phase of basilar membrane vibration in response to tones is particularly important as it helps explain the relative phase of activation of auditory nerve fibers, where the response of fibers that are more sensitive to higher frequency components leads the response of fibers that are more sensitive to lower frequencies. This is consistent with the frequency decomposition in the basilar membrane shown in Figure 2.5 and its phase of activation described in the bottom panel of Figure 2.6.

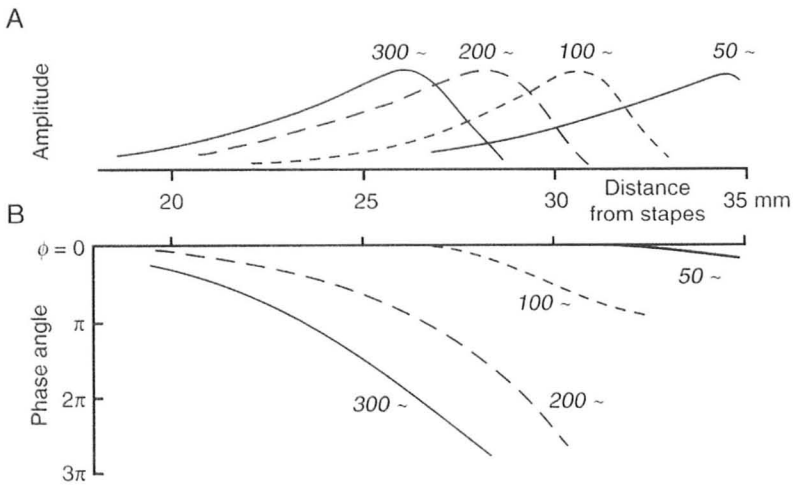

Figure 2.6: Illustration of the travelling wave effect observed in the basilar membranes of human cadavers. *Top Panel*: Envelopes of basilar membrane displacement in response to tones whose frequencies are labelled in Hz. *Bottom Panel*: Relative phase of basilar membrane displacement. (Reprinted from Pickles, 2008; Original figure from von Békésy, 1947)

2.3.2 Organ of Corti

The organ of Corti sits on top of the basilar membrane inside the scala media and runs along the entire length of the basilar membrane. It contains the auditory sensory cells, called the inner hair cells (IHCs) and outer hair cells (OHCs). These cells perform the transduction of mechanical energy (stored in the vibration of the basilar membrane) to electrical energy (stored in the potentials created by ions within the hair cell). Panel

A of Figure 2.7 depicts a cross-section of the organ of Corti at one place along the basilar membrane, and shows the placement of the inner and outer hair cells. There are three to five rows of outer hair cells and one row of inner hair cells throughout the cochlear spiral (Pickles, 2008). The inner hair cells are located proximal to the central axis of the cochlear spiral (modiolus) with respect to the outer hair cells, which are located distally.

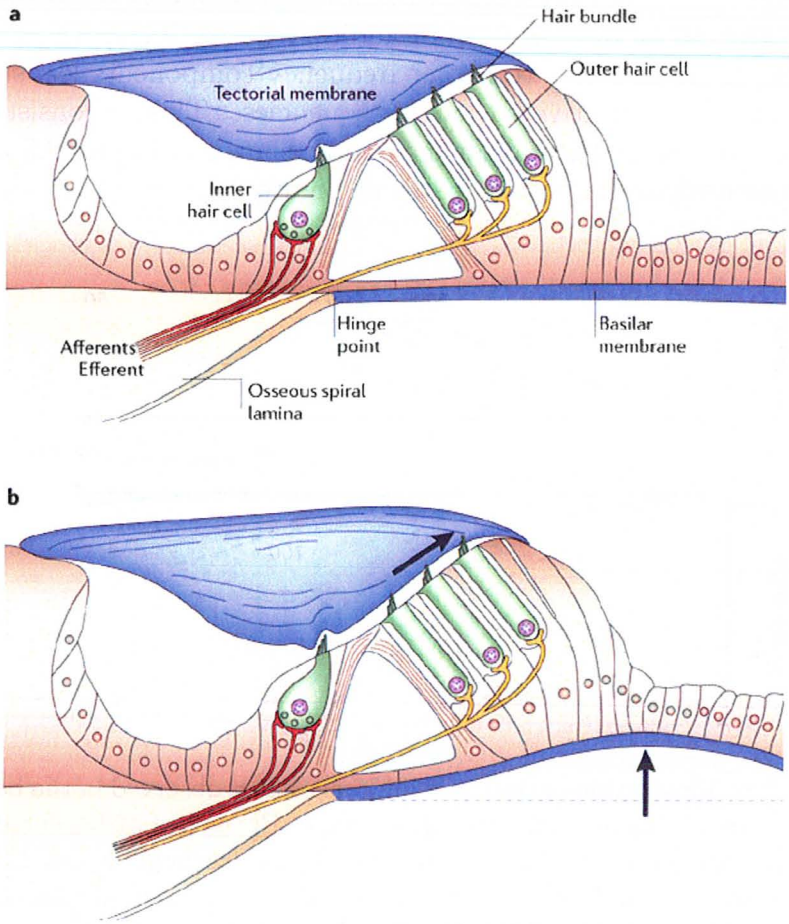

Figure 2.7: *Top Panel*: Diagram of a cross section of the organ of Corti. *Bottom Panel*: Demonstration of the direction of hair cell displacement by the tectorial membrane as a result of upward displacement of the basilar membrane. (Reprinted from Fettiplace and Hackney, 2006)

Sitting on top of the organ of Corti is a fibrous flap called the tectorial membrane. The long hairs of the outer hair cells are embedded in the tectorial membrane, while the inner hair cells are not attached. As the basilar membrane vibrates perpendicularly to its surface, the tectorial membrane bends the hairs of the inner and outer

hair cells back and forth in the proximal-distal dimension. The stereocilia (hairs) of the outer hair cell move rigidly with the tectorial membrane but the movement of inner hair cell stereocilia are bent by the flow of endolymph as the tectorial membrane slides back and forth across the organ of Corti (Pickles, 2008).

2.3.3 Mechanoelectrial Transduction

As the basilar membrane vibrates, the stereocilia of the hair cells are bent through the movement of the tectorial membrane. Panel B of Figure 2.7 demonstrates that as the basilar membrane is displaced upward, the tectorial membrane pushes the stereocilia of both the outer and, indirectly, the inner hair cells towards the tallest stereocilia. Panel A of Figure 2.8 shows a cross-section of the top of an outer hair cell with a hair bundle containing three stereocilia. Adjacent stereocilia are attached by side links and tip links; these links can be seen in Panel B of Figure 2.8. Movements of the hair bundle towards the tallest stereocilia cause the mechanoelectrical transduction (MET) channel, which is found on the tip of the shorter hairs in the bundle, to be activated. Movements in the opposite directions close the MET channels (Shotwell *et al.*, 1981).

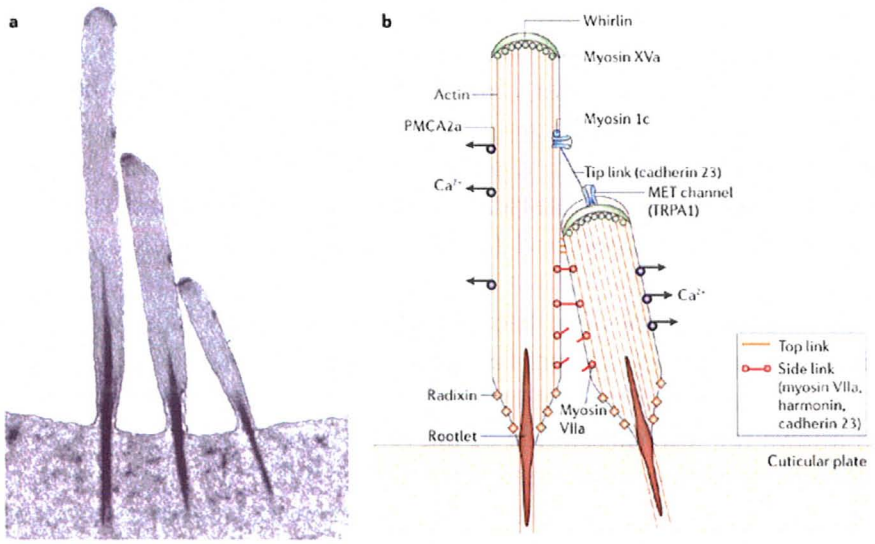

Figure 2.8: Cross section of a bundle of hair cell stereocilia. *Left Panel*: Transmission electron micrograph image of the stereocilia in an outer hair cell of a guinea pig. *Right Panel*: Diagram of a generalized stereociliary bundle indicating the tip links and MET channels. (Reprinted from Fettiplace and Hackney, 2006)

The MET channel is a cation channel permeable to positive ions including Ca^{2+} and K^{+} (Fettiplace and Hackney, 2006; Pickles, 2008). Activation of the MET channels by deflection of stereocilia results in an influx of cations and depolarization of the hair cell. In a generalized hair cell with afferent innervation, this depolarization triggers release of neurotransmitter and subsequent generation of action potentials in the auditory nerve fibers that innervate the hair cell[3] (Pickles, 2008). Action potential generation in afferent auditory nerve fibers will be described in more detail in Section 2.3.6.

Around 90-95% of the auditory nerve fibers are afferent fibers and most of these innervate the inner hair cells, whose role is generally recognized as the electrical signal pathway (Fettiplace and Hackney, 2006). The remaining nerve fibers are efferent fibers, mostly innervating the outer hair cells, whose role is thought to be part of an active, feedforward cochlear amplifier (Dallos, 1992), which will be described next.

2.3.4 Cochlear Amplifier and Outer Hair Cell Electromotility

The cochlear amplifier is a concept stating that there is some active component to basilar membrane vibration, a mechanical amplification that enables an increase in vibration amplitude as well as enhancement of frequency selectivity. Section 2.3.1 introduces the idea of two modes of basilar membrane vibration, one passive and one active. The concept of the cochlear amplifier came about because measurements in living cochleas showed that basilar membrane vibration has an enhanced and sharper peak than would be predicted from a purely passive model. So far, no passive model has been able to account for the large amplitude of vibration and the narrow region of resonance that has been measured (Pickles, 2008), and therefore an active model has been necessary.

The amplified peak of vibration brought about by the cochlear amplifier increases the sensitivity of the auditory system to low intensity sounds; that is, a soft sound will produce larger basilar membrane vibrations than would be expected from a passive model. This allows for low thresholds of nerve fiber activation. The sharper peak of basilar membrane vibration associated with the cochlear amplifier enhances the frequency resolution. This means that the amplification of basilar membrane vibration is frequency selective, and its effects are localized near the region of basilar membrane resonance (Geisler, 1998).

The cochlear amplifier is thought to be produced by the expansion and contraction of outer hair cells (Fettiplace and Hackney, 2006) but the exact mechanical process

[3]Afferent auditory nerve fibers form neural synapses to the base of inner hair cells. Release of neurotransmitter into the synapse causes neurotransmitter binding to the ligand-gated channels on the auditory nerve fiber and depolarization of the fiber.

is not definitive (Pickles, 2008). The outer hair cells have a mechanical response to electrical depolarization and hyperpolarization, both of which occur in response to different directions of stereocilia deflection. In response to depolarization, the outer hair cells contract along their long axis, and are shortened; conversely, in response to hyperpolarization, the body of the hair cells expand.

One hypothesis about how outer hair cell contraction and expansion could produce the effects of the cochlear amplifier is as follows: an upwards displacement in the basilar membrane will cause positive deflection of the outer hair cell stereocilia (see Figure 2.7), causing depolarization in the hair cell. This would cause the outer hair cells to contract, and might be able to pull the basilar membrane upwards even further, thereby introducing an active component to the vibration (Pickles, 2008).

2.3.5 Compressive Non-linearity

The amplitude of basilar membrane vibration, at the place of resonance, grows non-linearly in response to stimuli of increasing loudness. There is a region corresponding to low-level sounds that responds linearly to stimuli with great sensitivity; then the amplitude of vibration begins to be compressed with respect to sound level at around 30-40 dB SPL (Sound Pressure Level). Figure 2.9 shows the basilar membrane non-linear response to increasing stimuli levels, as well it includes hypothetical curves corresponding to the passive component and active component of basilar membrane response. This figure reinforces the theory of two modes of vibration in addition to describing the compressive non-linearity of the basilar membrane. As the stimuli levels increase even further, the levels of vibration amplitude begin to grow linearly, in correspondence with passive membrane vibration (Johnstone *et al.*, 1986).

It is important to note that this compressive non-linearity is only seen in response to tones with frequencies near the resonant frequency of the current position on the basilar membrane (Ruggero *et al.*, 1997). This is likely because of the localization of the cochlear amplifier. Figure 2.10 shows that one point on the basilar membrane in the chinchilla tuned to 10 kHz will respond somewhat linearly to tones whose frequency is not near the characteristic frequency, but has a compressive non-linearity for tones near CF.

2.3.6 Action Potentials in AN and Higher Processing

The afferent auditory nerve fibers innervating the inner hair cells are activated by depolarization of the inner hair cells. Activation of the MET channel described in Section 2.3.3 causes depolarization of the the inner hair cell. This depolarization creates an electrical gradient for positive ions to flow into the cell and causes an influx of calcium ions (Ca^{2+}) into the body of the inner hair cell. Influx of calcium ions causes

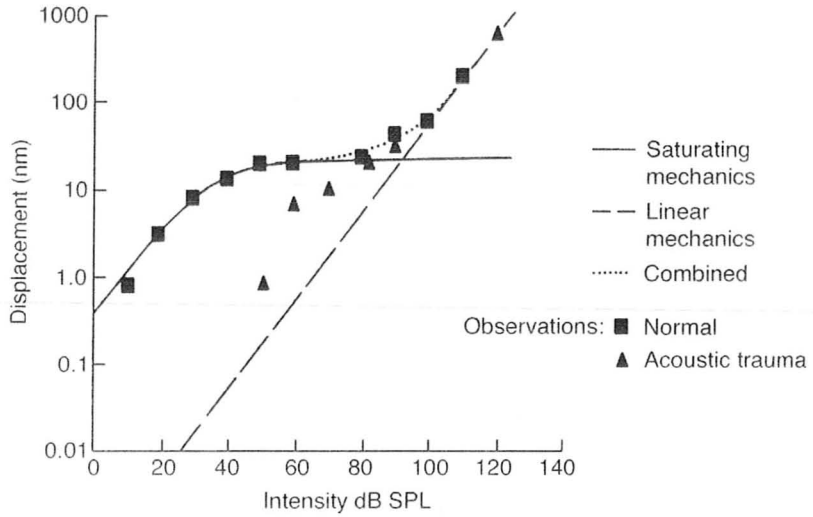

Figure 2.9: Input-Output curves of basilar membrane response amplitude as a function of sound level. Square markers represent observations in a healthy cochlea and triangle markers represent observations in a traumatized cochlea. The solid line represents the saturating mechanics of the active cochlear process, the dashed line represents the linear mechanics associated with an absence of cochlear amplifier, and the dotted line delineates the predicted behaviour if these mechanisms were combined. (Reprinted from Pickles, 2008; Original figure from Johnstone *et al.*, 1986)

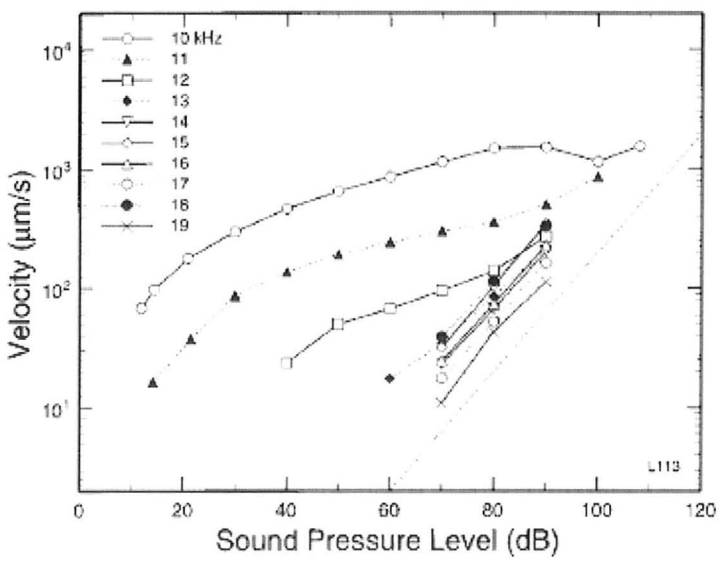

Figure 2.10: Velocity-Intensity curves measured in the chinchilla basilar membrane in response to tones with a variety of frequencies. The characteristic frequency of this basilar membrane position is 10 kHz. (Reprinted from Ruggero *et al.*, 1997)

release of neurotransmitter that is contained in vesicles near the synaptic cleft. These vesicles surround a synaptic bar, or ribbon, in the inner hair cell as shown in Figure 2.11, and neurotransmitter is released into the synaptic cleft through the process of exocytosis. The neurotransmitter, likely to be glutamate (Ottersen *et al.*, 1998), binds to the post-synaptic receptors on the afferent nerve fiber causing channels to open, and subsequent depolarization of the afferent nerve fiber. Significant depolarization causes action potential (or "spike") generation in the axons of the afferent fibers which propagate to the cochlear nucleus where the auditory nerve terminates.

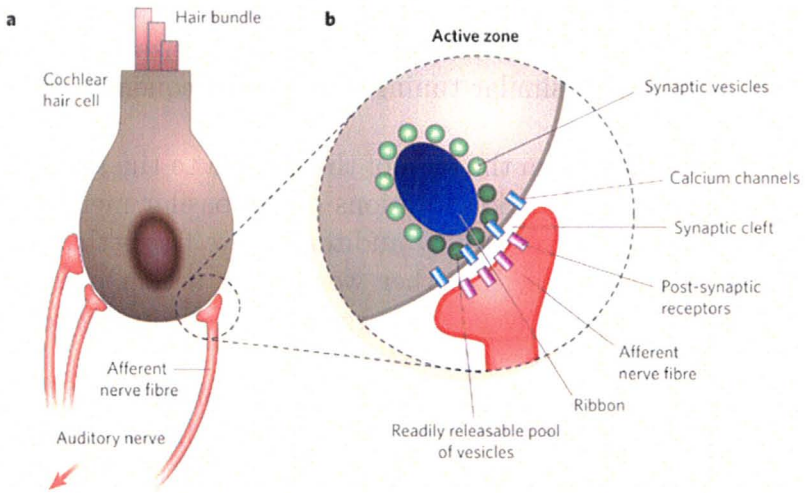

Figure 2.11: *Left Panel*: Illustration of an inner hair cell profile. *Right Panel*: Close up of the synapse made between the inner hair cell and an afferent auditory nerve fiber. (Reprinted from Parsons, 2006)

The cochlear nucleus is the first stage in central auditory processing, and here the afferent signal pathway is split up into two main streams, one for intelligibility and one for sound localization (Pickles, 2008). Certain features are extracted or processed for purposes like sound localization and spectral enhancement before the two streams are recombined in the inferior colliculus (Pickles, 2008). In addition to these features, Carney (1994) suggests that loudness could be encoded in the cochlear nucleus through coincidence detection of spikes in convergent auditory nerves that are sensitive to a range of frequencies. This is important because it could mean that firing rates in the cochlear nucleus are directly affected by the phase of auditory nerve response; the restoration of which by a hearing aid is studied in this thesis. This will be described in further detail in the next section which is concerned with auditory nerve response.

2.4 Auditory Nerve Response

As described above, the inner hair cell can be conceptualized as part of the signal path in the inner ear, transducing the acoustic waveform collected from the outer ear into an electrical waveform for further processing by higher auditory structures. This section will focus on the afferent auditory nerve fibers that innervate the inner hair cells and encode the auditory stimulus.

There are about 30,000 auditory nerve fibers in a typical human ear, of which about 90 to 95% innervate the inner hair cell. Any one inner hair cell is innervated by around 20 afferent nerve fibers, and each afferent nerve fiber only innervates one inner hair cell (Pickles, 2008). This structure of innervation means that activity in one afferent auditory nerve will be related to the depolarization of one particular inner hair cell and will have a similar tuning curve and frequency selectivity to that inner hair cell.

All of the inner hair cells, from the base of the cochlea to the apex, are innervated by afferent auditory nerve fibers. So vibrations of the basilar membrane at any one particular place are encoded primarily by auditory nerve fibers that are attached to the organ of Corti at that place. In other words, the action potential firing rate in response to a tone will typically be the largest for auditory nerve fibers that are located nearest to the place of resonance for that tone.

Every one of these auditory nerve fibers can be conceptualized as a signal processing bandpass filter, called an auditory filter, where the auditory stimulus is the input and the instantaneous firing rate of the fiber is the output[4]. Every auditory filter will have some magnitude response, phase response and group delay—characteristics which are inherent in the definition of a filter. The next few sections will describe these characteristics of the auditory nerve response in the framework of an auditory filter.

2.4.1 Magnitude Response

Each auditory nerve fiber has a certain frequency selectivity, on account of the place of attachment to the organ of corti, and the tuning properties of the basilar membrane. It has the most sensitivity to one particular frequency, known as the fiber's characteristic frequency (CF), to which the center frequency of the corresponding bandpass auditory filter is equated. The sensitivity of an auditory nerve fiber decreases as the stimulating frequency deviates from the fiber's CF. This can be depicted with a tuning curve, as shown in Figure 2.12. A tuning curve, or frequency threshold curve, plots the intensity of a tone that is required to induce a certain threshold of response

[4]The output could also be defined as the transmembrane voltage of the auditory nerve fiber or the probability of firing.

as a function of frequency; in this case a specified threshold of firing rate in the fiber is of interest (specific thresholds of BM velocity and BM displacement are also plotted). Another way that the magnitude response can be characterized is with an iso-intensity curve, as shown in Figure 2.13. Here the firing rate of the relevant fiber is plotted in response to input tones of fixed level as a function of frequency. Figure 2.13 plots iso-intensity curves for a number of fixed stimulus input levels. These resemble the magnitude response functions that are commonly defined for filters.

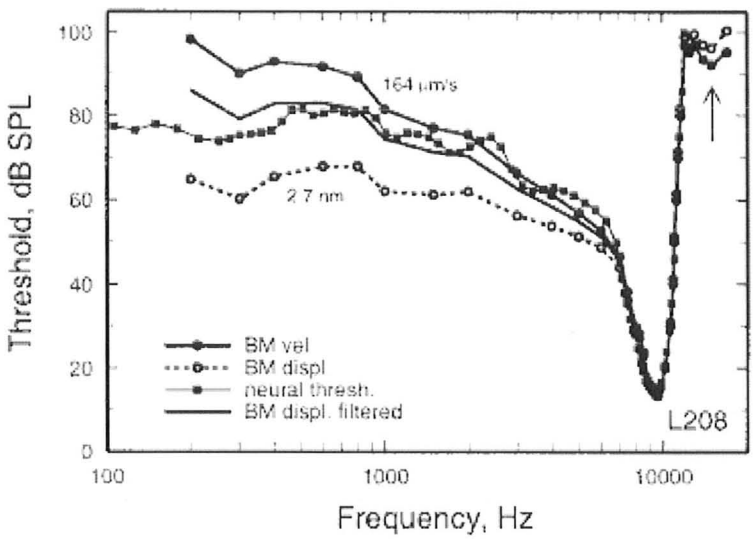

Figure 2.12: Auditory nerve fiber tuning curve (labelled as 'neural threshold') for a particular place of attachment to the chinchilla basilar membrane. Thresholds of basilar membrane velocity and amplitude are also plotted. (Reprinted from Ruggero *et al.*, 2000).

Figures 2.12 and 2.13 demonstrate the bandpass nature of a typical auditory filter. The filters are tuned broadly towards low frequency stimuli, but exhibit a sharp attenuation of high frequency stimuli. This filter shape can be understood in terms of the physical vibration of the basilar membrane in response to a tone of a specific frequency. The direction of the traveling wave in the basilar membrane is from base to apex, or high fiber CF to low fiber CF. As the pressure wave travels in this direction, more and more energy is dissipated in the vibration of the basilar membrane until the place of resonance is reached. There, most of the energy in the wave is dissipated so the vibration does not continue much further along the basilar membrane. Since low frequency tones resonate near the apex, each fiber with a CF higher than this tone will have some amount of response, accounting for the greater sensitivity that auditory filters exhibit to low frequencies than to high frequencies.

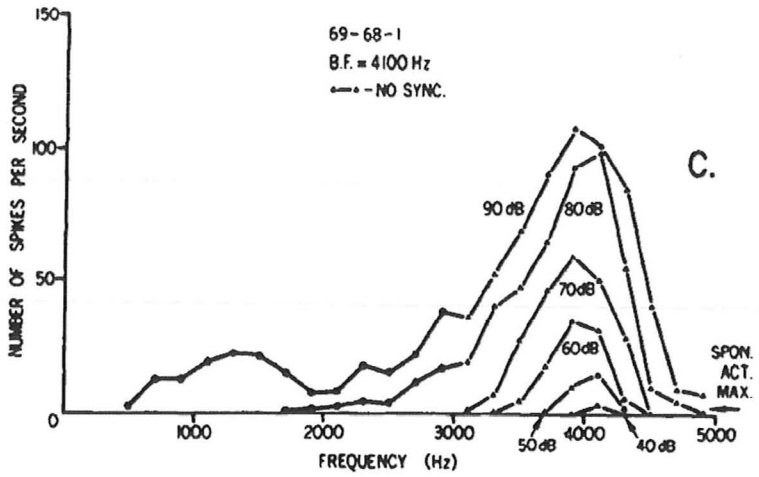

Figure 2.13: Isointensity curves for a squirrel monkey auditory nerve fiber with characteristic frequency of 4000 Hz. (Reprinted from Rose *et al.*, 1971)

Magnitude Response as a Function of Stimulus Intensity

It is interesting to examine the behaviour of auditory nerve rate response as a function of increasing stimulus intensity, or loudness. This can be done in a couple of ways, one is with the aid of a rate-level curve which plots the average firing rate of a fiber in response to stimuli of increasing intensity levels. Example rate-level curves corresponding to one model auditory nerve in response to a variety of tones are shown in Figure 2.14. One important feature of this figure is the non-zero firing rate that exists when the stimulus is inaudible (roughly ≤ 0 dB SPL) which is attributable to the spontaneous activity of each auditory nerve fiber (roughly 50 spikes/second in this model fiber). Other important features to note include the saturation of firing rate as the stimulus reaches a certain intensity and the slope of the growth of firing rate with intensity.

Another way to examine the response to increasing stimulus levels is to look at the growth of iso-intensity curves with level. As is indicated from the iso-intensity curves in Figure 2.13, the bandwidth of each auditory filter is broadened in response to increasing levels of input. This is attributable to a spread of vibration in the basilar membrane as the intensity of the stimulus increases, which drives more fibers with a higher rate. As the stimulus level increases, vibration in the BM will naturally spread out due to its passive characteristics; as well, saturation of the cochlear amplifier will result in decreased frequency selectivity. The broadening of the auditory filter's bandwidth has important consequences on the behaviour of its phase response, which will be outlined in the next section.

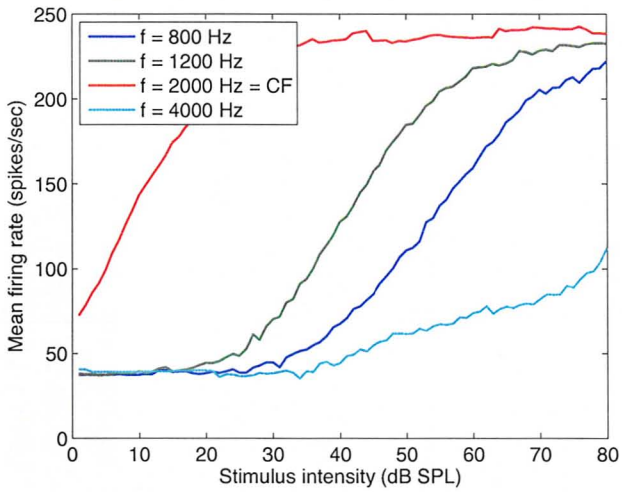

Figure 2.14: Four rate-level curves for a model auditory nerve fiber with a characteristic frequency 2000 Hz in response to a variety of tones. The rate is given as a firing rate in spikes/sec averaged over 50 msec. The auditory nerve model of Zilany and Bruce (2006, 2007) was used to generate these curves.

2.4.2 Phase Response

In order to quantify the phase response of auditory filters, some degree of periodicity in the discharge pattern of a single auditory nerve must be preserved. For this reason, before the phase response of an auditory filter is described, synchronization in response to periodic stimuli is discussed.

Synchrony

In response to tones of low frequency, the timing of action potentials generated in the auditory nerve is phase-locked to a particular phase in the cycle of the stimulating tone. This does not mean that a spike will occur at every cycle of the stimulus[5], but that when spikes occur, they will likely occur at the same phase in the stimuli's cycle (Pickles, 2008). The left panel in Figure 2.15 demonstrates a model AN fiber spiking in response to a CF tone[6]; spikes were summed over 500 stimulus repetitions. The degree of phase locking can be modelled with a probability distribution over one period of the stimulus, and can be easily approximated with a period histogram of the spiking events in correspondence to the stimuli's period. The right panel of Figure

[5]Pickles (2008) suggests spikes in a single auditory nerve could be as rare as once every 100 cycles.
[6]i.e. a tone whose center frequency is set to the characteristic frequency of the fiber of interest.

2.15 shows a period histogram of model discharge times to a CF tone; here all the spikes were counted in response to 500 repetitions of a 50 msec long stimulus. Note that spikes tend to occur at a phase of π radians.

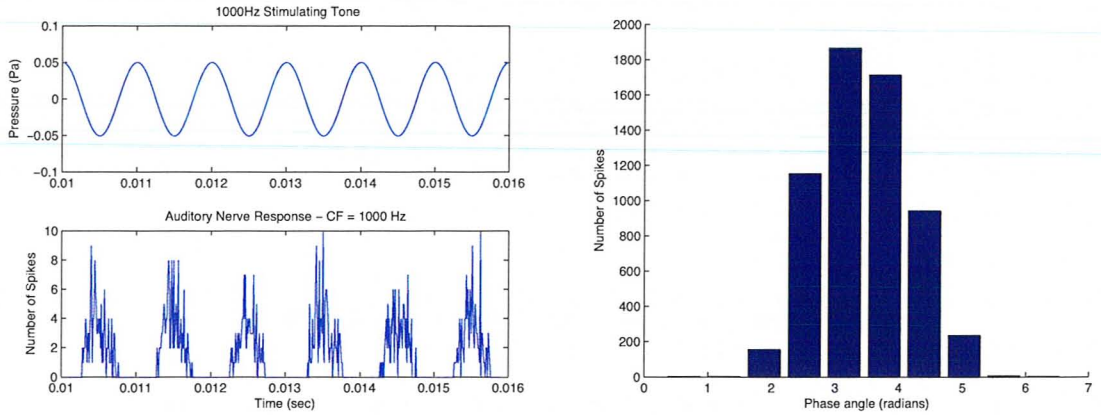

Figure 2.15: Figures illustrating the phenomenon of phase-locking, or synchrony in the auditory nerve. *Left Panel*: A sinusoidal stimulus (shown on the top) is presented to the computational auditory nerve model of Zilany and Bruce (2006, 2007). The peri-stimulus time histogram (PSTH) of the spiking response is shown on the bottom. Spikes occur at regular intervals corresponding to the frequency of the stimulus. *Right Panel*: Period Histogram derived from the PSTH in the left panel. The phase of each spiking event is given relative to one cycle of the stimulus and then spikes within each phase range are counted.

Although phase-locking occurs at low frequencies and is preserved at high intensities, it is not observed in response to high frequency stimuli. A contributing factor for the absence of phase-locking at high stimulus frequencies could be that the a.c. response of the hair cells becomes attenuated at high frequencies because of the capacitance of the hair cell wall[7] (Pickles, 2008; Russell and Sellick, 1978). In this case spikes would be generated in response to the d.c. depolarization of the hair cell, as opposed to the a.c. response. With enough attenuation of the a.c. component, no synchronization of spikes to one phase in the cycle of the a.c. depolarization would be possible. Pickles (2008) suggests that the degree of phase locking reduces above 2 kHz, and that 5 kHz is often cited as the upper limit to phase-locking. For this reason, phase analysis is usually restricted to low stimulus frequencies.

It is important to note here that although phase locking only occurs in response

[7]A capacitor offers little impedance to high frequency electrical signals, as such it behaves as a short circuit as frequency increases. This short circuit attenuates the a.c. depolarization of the cell.

to low stimulus frequencies, it can still occur in fibers with high characteristic frequencies. As long as these fibers are responding to stimuli with low frequency, they can exhibit synchrony. In addition to this, it has been shown that AN fibers can phase lock to the low frequency envelope of complex signals, such as in the case of amplitude modulation where AN response synchronizes to the modulation frequency (Joris and Yin, 1992).

Characteristics of AN Phase

The phenomenon of synchrony in the auditory nerve response allows for the measurement of phase angle relative to the phase of the stimulating input, an important characteristic of auditory filters.

The phase of action potential generation in afferent auditory nerve fibers should intuitively be in phase with the velocity of basilar membrane deflections upwards towards the scala vestibuli. This direction of basilar membrane deflection causes depolarization in the inner hair cells which causes neurotransmitter release, and activation of the auditory nerve fiber (also see Figure 2.3). In the apex of the cochlea, where auditory nerve fibers with low characteristic frequencies are attached, this is indeed observed (Pickles, 2008; Cheatham and Dallos, 1999).

However, in the base of the cochlea, where fibers with high characteristic frequencies are attached, there is a greater degree of complexity regarding the phase of action potential generation. At low stimulus intensity levels, the phase of AN response in this cochlear region is 180 degrees out of phase from what is expected. That is, the phase of the AN response is in phase with the deflection of the BM towards the scala tympani as opposed to the scala vestibuli. Then, at high stimulus intensity levels ($\sim$ 90 dB SPL and above), there is a 90-180 degree phase shift so that responses are now in phase with the deflection of the BM towards the scala vestibuli (Ruggero *et al.*, 2000; Pickles, 2008). This has been described as a component-1/component-2 (C1/C2) transition where C1 corresponds to the AN response at low stimulus intensities, and C2 corresponds to the AN response at high intensities (Kiang and Moxon, 1972; Zilany and Bruce, 2006; Heinz and Young, 2004). Zilany and Bruce (2006) suggest that the C1/C2 transition might arise in the IHC, which is evidenced to have two transduction mechanisms (Sewell, 1984) that could correspond to two separate modes of excitation by the basilar membrane.

Phase Response as a Function of Stimulus Intensity

One can also analyze the characteristics of the phase response of an auditory filter as a function of stimulus level. As described in Section 2.4.1, the bandwidth of an auditory filter increases with stimulus level. As a result, the auditory filter has a flattened phase versus frequency response in response to loud sounds (Robles and

Ruggero, 2001). The broadening of the auditory filter, and flattening of its phase response is depicted in the left panel of Figure 2.16. An auditory filter's flattened phase manifests itself in a fiber population response with increased coincidence of spike times in response to a stimulating tone. This behavior is demonstrated in the right panel of Figure 2.16.

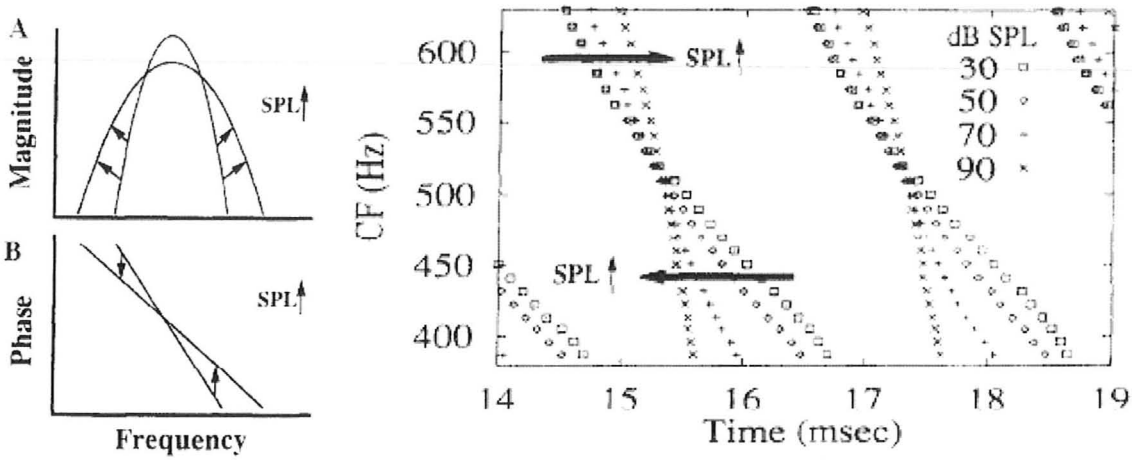

Figure 2.16: The effects of increasing stimulus intensity on the auditory filter transfer function and spatiotemporal response. *Left Panel*: (A) Schematic of the magnitude response of an auditory filter. As the SPL increases the filter's bandwidth increases. (B) Schematic of the phase response of an auditory filter. As the SPL increases, the slope of the phase transfer function becomes shallower. *Right Panel*: Spatiotemporal response pattern to a 500 Hz tone at increasing levels. The symbols represent the times of peak discharge probability. As SPL increases, spiking events become more coincident across the population. (Reprinted from Carney, 1994)

2.4.3 Group Delay

Another important characteristic of an auditory filter is its group delay, which accounts for the overall delay that a signal will experience when passing through the filter (Shi *et al.*, 2006). Group delay is defined by Oppenheim and Schafer (1999) as the negative, instantaneous rate of change of the system's phase response as a function of frequency. This can be formalized as

$$\text{grd}[H(e^{j\omega})] = -\frac{d}{d\omega}\{\arg[H(e^{j\omega})]\}, \tag{2.1}$$

where ω is the variable frequency in radians, $H(e^{j\omega})$ is the filter's transfer function,

and $\arg[H(e^{j\omega})]$ is the continuous phase of $H(e^{j\omega})$. Oppenheim and Schafer (1989) go on to show in Problem 5.3 of their text that for a narrowband signal of the form $x[n] = s[n]\cos(\omega_o n)$, where $s[n]$ is a low frequency envelope, the output of a filter's transfer function $H(e^{j\omega})$ is given by:

$$y[n] = s[n - \tau_{\text{grd}}(\omega_o)]\cos\{\omega_o[n - \tau_{\text{ph}}(\omega_o)]\}, \tag{2.2}$$

where τ_{grd} is the filter's group delay and $\tau_{\text{ph}} = -\angle H(e^{j\omega})/\omega$ is the filter's phase delay. This equation indicates that the envelope of a narrowband signal centered about ω_o is delayed by the group delay $\tau_{\text{grd}}(\omega_o)$. In other words, the group delay can be thought of as the delay imposed on the temporal envelope of a narrowband signal as it passes through the filter.

Group Delay as a Function of Stimulus Intensity

Since the group delay is proportional to the derivative of auditory filter's phase response, as the intensity of a stimulating sound increases the group delay through the filter will decrease. The effects of increasing stimulus presentation levels on the characteristics of an auditory filter can be summarized as follows: i) the bandwidth of the filter is increased, ii) the phase response is flattened, and iii) the group delay is shortened. This behaviour is important to understand when analyzing the auditory nerve response in an impaired ear.

2.5 Impaired Auditory Nerve Response

There are two main categories of hearing loss called conductive loss and sensorineural loss. Conductive loss arises out of some irregularity or damage to the outer or middle ears and can be caused by blockage, damage to the eardrum, or immobility of the middle ear ossicles (Pickles, 2008). This type of loss may involve some frequency shaping, but it is usually a simple type of loss that can be corrected quite well with a hearing aid.

Sensorineural hearing loss, on the other hand, arises in the inner ear, usually with damage to the hair cells. It can be caused by acoustic trauma (i.e. by exposure to very loud sounds), ototoxic drugs, or infections. There is also sensorineural hearing loss that occurs with age, called presbyacusis, which is a degenerative process. This category of hearing loss is quite complex, causing problems with perception as well as audibility (Moore, 2003), and hearing aids have not been able to fully correct for it.

2.5.1 Physiological Changes Associated with Sensorineural Loss

Within the sensorineural hearing loss category, there can be impairment to either the inner hair cells or the outer hair cells, or both. Impairment to these hair cell types has ill effects. Outer hair cells act as a cochlear amplifier which increases the sensitivity of auditory nerve fibers and provides enhanced frequency selectivity (see Section 2.3.4). When there is damage to the outer hair cells, this functionality is degraded, resulting in fiber tuning curves with decreased sensitivity and broadened bandwidth. As a result, sometimes the frequency to which the fiber is most sensitive becomes lower than the characteristic frequency of its healthy counterpart. The right panel of Figure 2.17 compares a healthy auditory nerve fiber tuning curve with a tuning curve corresponding to a fiber that has outer hair cell damage. In this case the sensitive peak near the fiber's characteristic frequency is absent, resulting in a decreased characteristic frequency compared to the healthy fiber.

Damage to the inner hair cells causes attenuation of the signal transduction. If the outer hair cells remain healthy, then inner hair cell impairment will result in a fiber having a tuning curve with the same basic shape, but with decreased sensitivity. This is shown in the left panel of Figure 2.17. In this case, amplification of the sound is required for the impaired listener to retain sufficient audibility. In addition to an increased threshold, inner hair cell impairment indirectly causes broadening of the auditory filter; the increased thresholds require higher stimulus presentation levels which has already been shown to result in increased filter bandwidths (see Section 2.4.1).

When there is mixed hair cell impairment, the effects of damage to both hair cell types are combined. In general, all these types of hair cell impairment will result in auditory nerve fibers having broadened frequency selectivity and decreased sensitivity. This translates into a conceptual impaired auditory filter providing less gain and having a broader bandwidth. Referring to Section 2.4, it is apparent that the broadened bandwidth of the impaired auditory filter will result in a flattened phase response and shorter group delay compared to the healthy auditory filter. This holds true for all sound presentation levels.

2.6 Perceptual Correlates to Auditory Nerve Response

Abnormal response in the auditory nerve caused by sensorineural hearing loss leads to deficiencies in the hearing impaired person's perception of sounds. It is important to identify which aspects of the impaired auditory nerve response lead to perceptual

deficits when using a hearing aid design framework that attempts to restore healthy auditory nerve response. This will enable hearing aid design targeted to restore important perceptual features of the auditory nerve response. This section will outline perceptual correlates of auditory nerve responses as they pertain to the work of this thesis.

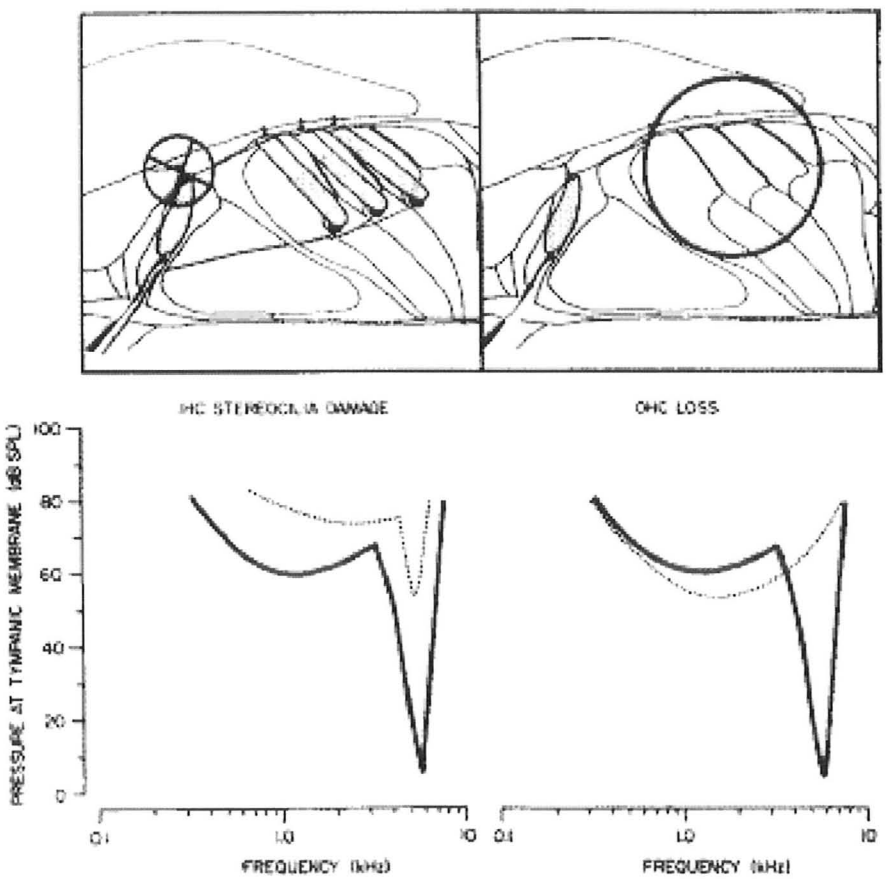

Figure 2.17: Comparison of the tuning curves associated with damage to either the IHCs or the OHCs. The left hand column illustrates damage to the inner hair cell stereocilia, and gives a typical tuning curve that would result from IHC impairment (dotted curve). A healthy tuning curve is included as a reference (solid curve). The right hand column illustrates complete loss of the outer hair cells and gives a typical tuning curve that would result from this loss (dotted curve). (Reprinted from Kiang *et al.*, 1986)

2.6.1 Vowel Formant Identification

Voiced speech[8], including vowels, has frequency spectra that are composed of several prominent peaks called formants, which originate from resonances in the vocal tract. Figure 2.18 shows the envelope of two example synthetic vowels along with a portion of their temporal waveform; the formants are identified as F1 to F5. Formants vary in their frequency composition and their relative amplitude, both of which contribute in some way to the perception of a vowel. It is believed that the first two or three formants are the most important for vowel perception and that vowels can, in general, be identified by the relative frequency positioning of the first two formants (Peterson and Barney, 1952; Kiefte *et al.*, 2010). The relative formant amplitude, or spectral tilt, has also been shown to affect the identification of vowels; Kiefte and Kluender (2005) demonstrate that manipulation of relative formant amplitude affects the discriminability of synthetic vowels [*u*] and [*i*].

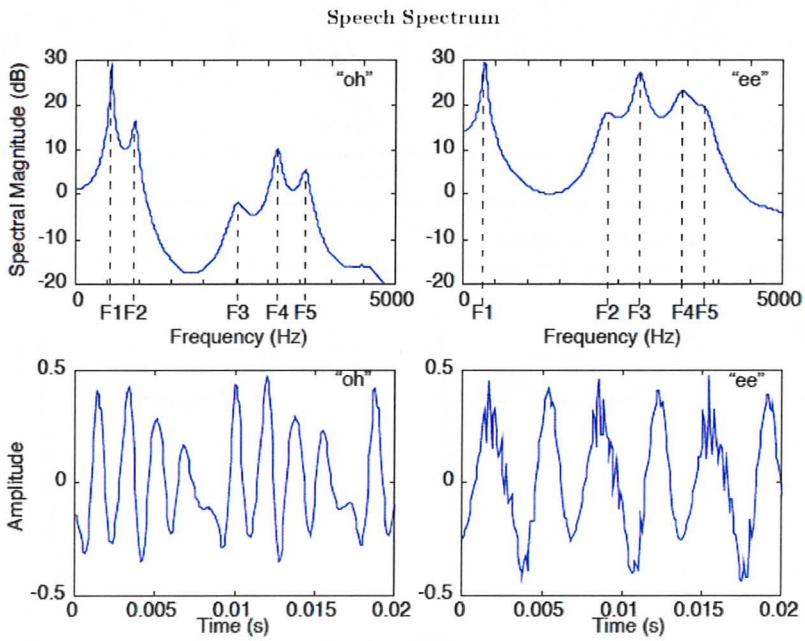

Figure 2.18: Example vowel spectrums illustrating the formants and the corresponding temporal waveforms. *Left Column*: The magnitude spectrum of the synthetic vowel 'oh' is shown on top of the corresponding temporal pressure waveform. *Right Column*: The synthetic vowel 'ee'. (Reprinted from Johnson, 2009)

[8]Voiced speech is produced when vocal cords alternate between open and closed positions periodically (Warren, 2008).

The importance of the composition of formants in the identification of vowels suggests that the representation of formants in the auditory nerve response is also important. Formants could be represented in the auditory nerve in a number of ways including in the discharge rate of fibers with characteristic frequencies near the formant frequencies and in the timing of fiber discharges, which often demonstrate synchrony to the formant frequencies. In a healthy ear, auditory nerve fibers with center frequencies near the formant frequencies become captured by the formants as sound level increases, that is they are synchronized to the formants (Young and Sachs, 1979; Miller *et al.*, 1997). In an impaired ear, fibers will respond to the input over a broader frequency range and an upward spread of synchrony to F1 is often seen (Miller *et al.*, 1997). This means fibers that would normally be synchronized to frequency components higher than F1 are now synchronizing to F1. Miller *et al.* (1999) show that manipulation of the relative formant amplitudes can help to minimize the upward spread of synchrony that is seen in impaired auditory nerve response. This provides motivation for developing contrast enhancement schemes to restore synchrony in the impaired ear, and will be discussed in Chapters 4 and 5.

2.6.2 Loudness

The loudness percept can be defined as a feature of the sense of hearing that can be ordered on a scale from quiet to loud (Moore, 2003). This is a subjective quantity and is quite variant between individuals, especially in the case of hearing impairment. Despite the difficulties that arise due to its subjective nature, an important motivation for the quantification of loudness is in the assessment of hearing impairment. To this end, attempts have been made to quantify the perception of loudness, which is often done by asking listeners to match the level of a reference tone to a sound of interest. Using this method, the intensity of the reference tone can give some measure of loudness.

An interesting area of research is the development of loudness models that are based on activity in the auditory nerve. It seems reasonable to assume that loudness is completely encoded in the activity of the auditory nerve, because the auditory nerve is the only pathway from the ear to higher auditory structures in the brain. The problem then arises of decoding this information into a measure of loudness.

One common idea is that loudness should increase with the total amount of auditory nerve activity. This could be in the form of an increase in the total number of action potentials in response to a sound (e.g., Relkin and Doucet, 1997) or in the spread of fiber excitation across a range of characteristic frequencies (e.g., Moore, 2003). The ability of these models to predict loudness perception data varies, but an important standard is how well hearing impairment can be modelled, a situation where some loudness models have difficulty.

2.6.3 Recruitment of Loudness

Listeners with sensorineural hearing impairment have an abnormal perception of loudness, termed loudness recruitment. This abnormality can be described as a rapid increase in the perception of loudness as the sound level increases (Joris, 2009). Loudness recruitment arises out of the decreased dynamic range in which sounds are both audible and tolerable for a listener with hearing impairment. Figure 2.19 demonstrates the perception of loudness in impaired listeners that characterizes loudness recruitment; Moore *et al.* (1985) asked listeners with unilateral hearing loss[9] to match the loudness of a tone presented to their healthy ear with a tone of fixed level in their impaired ear. The results indicate a rapid growth of the perception of loudness in the impaired ear.

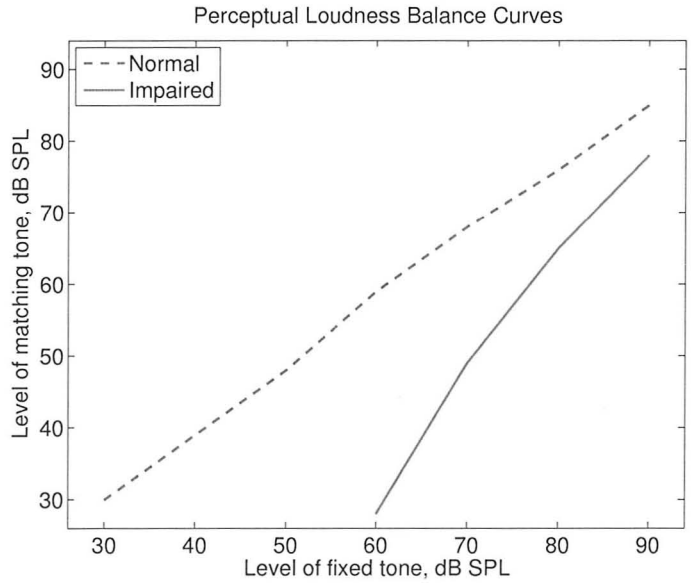

Figure 2.19: Average perceptual loudness balance curves for persons with normal, and unilateral hearing loss. Moore (2003) asked five listeners with relatively uniform loss of around 45-60 dB HL to match the loudness of a fixed-level tone in their impaired ear, with a variable-level tone in their healthy ear. (Data interpreted from Moore, 2003)

There have been a number of attempts to correlate auditory nerve response to loudness recruitment (Heinz *et al.*, 2005), and three hypotheses are depicted in Figure 2.20. The first hypothesis that could explain recruitment, presented in the figure's top row, shows that if the rate-level curves of auditory nerve fibers following acoustic

[9]Unilateral loss is hearing loss in one ear, healthy hearing in the other ear.

trauma were steepened, loudness recruitment might result. This would seem reasonable given that there is a loss of compressive nonlinearity associated with sensorineural impairment that steepens the amplitude-level curve of basilar membrane vibration (Oxenham and Bacon, 2003). However, in an auditory nerve study, Heinz and Young (2004) found that the rate-level curves of auditory nerve fibers were not steepened as might be expected from the steepened amplitude-level curve of the basilar membrane.

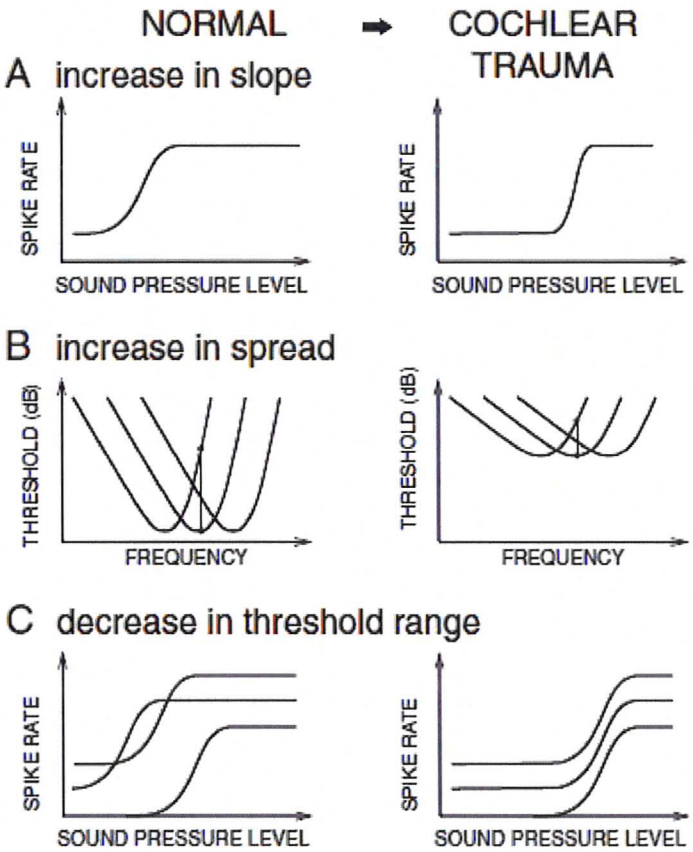

Figure 2.20: Illustrations of several hypotheses to explain loudness recruitment with changes in auditory nerve response. (A) Increased slope of AN fiber rate-level curve. (B) Spread of excitation due to increased response thresholds and broadened bandwidths. (C) Decreased range of individual fiber thresholds. (Reprinted from Joris, 2009)

A second explanation for loudness recruitment is derived from the broadening of filters that occurs following acoustic trauma (shown in the second row of Figure 2.20). Due to the broadened filters, a small increase in the stimulus intensity level should recruit fibers with a wider range of characteristic frequencies. This would correspond to a wider spread of excitation in response to an increase in intensity, which is in

agreement with the psychophysical loudness perception data. However, Heinz *et al.* (2005) found that widening of AN response occurs at intensities above the recruitment and does not occur for the intensities where recruitment is observed.

A third feature of AN response that could explain loudness recruitment, shown in Figure 2.20, would be if there was a decreased range of AN fiber thresholds (Moore *et al.*, 1985). In healthy ears there is a range of thresholds for AN fibers in a narrow characteristic frequency region that could be important for loudness perception. But again, recent auditory nerve data did not find a decreased range of AN fiber thresholds following acoustic trauma (Heinz *et al.*, 2005).

Heinz *et al.* (2005) offer an alternative hypothesis for a physiological correlate to loudness recruitment based on changes in basilar membrane response which ultimately leads to steeper rate-level curves in the response of neurons in higher auditory structures, such as the cochlear nucleus. Impaired basilar membrane response leads to auditory filters with increased bandwidth and flattened phase response (recall Section 2.5). An increase in the bandwidth across a population means that the population will respond more similarly in response to complex stimuli such as speech, because the population now has similar frequency content in their passbands. As well, Carney (1994) demonstrates that the flattened phase response means an AN fiber population will respond more coincidently to narrow-band periodic stimuli (see Figure 2.16). These attributes of the impaired auditory filter could lead to recruitment-like behaviour in the cochlear nucleus if its neurons behave as spiking coincidence detectors across a range of AN characteristic frequencies (Carney, 1994). As well, recent data from Cai *et al.* (2009) shows that certain types of neurons in the cochlear nucleus demonstrate recruitment like behaviour in their rate-level curves following acoustic trauma.

This provides motivation for efforts to correct for the spatiotemporal response properties of impaired auditory filters with a hearing aid. Attempts are made to account for the shortened group delay of impaired auditory filters by using a version of spatiotemporal pattern correction (SPC), which will be described in Chapters 4 and 5. By correcting for impaired group delays, the relative timing of AN discharges across a population could be restored which may aid in loudness perception. The auditory model used in this thesis will be described next, as familiarity with it is required to understand the SPC implementation.

Chapter 3

Background ii. A Model of the Auditory Periphery

The work of this thesis relies heavily on the use of computational models of the auditory periphery, both in the development of signal processing schemes, as well as in their evaluation. Specifically, the model employed here is that of Zilany and Bruce (2006, 2007), which improves on previous models that have been developed with a number of collaborators[1] (e.g., Carney, 1993; Zhang *et al.*, 2001; Bruce *et al.*, 2003). This model provides a reliably accurate description of the spiking pattern of an auditory nerve fiber in response to complex stimuli across a range of stimulus presentation levels (Zilany and Bruce, 2007). This is an important consideration in the employment of models for use with hearing-aid algorithm design.

The utilization of an auditory model in this thesis follows a recent trend in hearing aid design. Edwards (2007) explains, in a work describing the future of hearing aid technology, that the use of auditory models is becoming important for design of signal processing algorithms. This is evidenced by a number of hearing aids that make use of this design principle (e.g., Bondy *et al.*, 2004; Chen *et al.*, 2005; Shi *et al.*, 2006). The basic concept is that accurate models of a normal and impaired auditory system can be used to develop a hearing aid that, when included in the impaired model, can predict the ability of the hearing aid to restore the auditory response to normal. A schematic of this design framework is shown in Figure 3.1.

Limitations arise in this method for hearing aid development based on the accuracy of the models utilized. A model is always a simplified representation of the object of interest, as such it is important to test the design with human listeners. However, testing with human listeners does not allow for analysis of auditory nerve response and is much costlier than using a model. One specific limitation of using of the Zilany and Bruce model arises out of the differences between cat and human cochleas. The

[1] A new model of the auditory periphery has been released during the course of this thesis; it is not used here, but the reader is referred to Zilany *et al.* (2009).

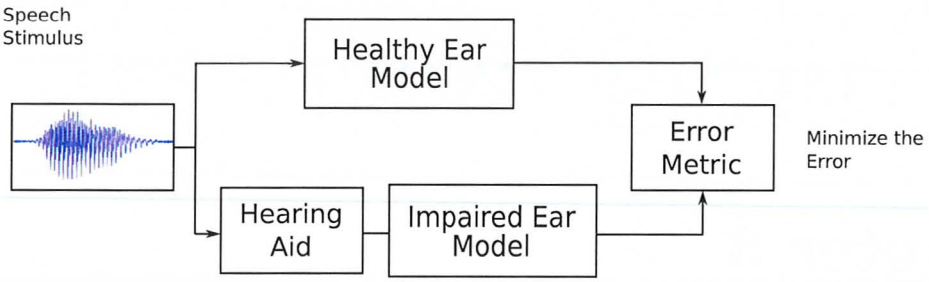

Figure 3.1: Illustration of the use of healthy and impaired auditory periphery models in hearing-aid design. The impaired model is individualized for each impairment.

model is designed to replicate auditory nerve responses in a cat, whereas the hearing-aids are applied to humans. An uncertainty in using the cat model is that the upward spread of synchrony to F1 observed in cat data may not be as important in a human cochlea because it has a longer length relative to its frequency range, likely providing better frequency resolution (Bruce *et al.*, 2003; Recio *et al.*, 2002). However, the model is thought to correspond well to human phsyiology (Bondy *et al.*, 2004), and the anatomy of the cat's cochlea is generally similar to that of a human (Nadol, 1988), but proportionally smaller.

3.1 Head Related Transfer Function

To model the frequency shaping done by the outer ear, the head-related transfer function described by Wiener and Ross (1946) is used. This transfer function appears in Figure 3.2 and is used as a pre-processor to the Zilany and Bruce model which begins at the eardrum.

3.2 Peripheral Model Description

The Zilany and Bruce model emulates the processing done in the auditory periphery from the middle ear through to the auditory nerve. It is a phenomenological model designed to replicate data obtained from *in vivo* electro-physiological recordings in the auditory nerve of cats in response to complex stimuli. That is, the model does not attempt to replicate every mechanical and biological process involved in the ear, but instead uses filter design theory to predict the effects that these processes have on the auditory signal at various stages in the periphery. Each stage of processing in the auditory periphery, as described in Chapter 2, is represented by a processing block in the model. A depiction of the model processing blocks is given in Figure 3.3.

The input to the model is a computational representation of an acoustic waveform

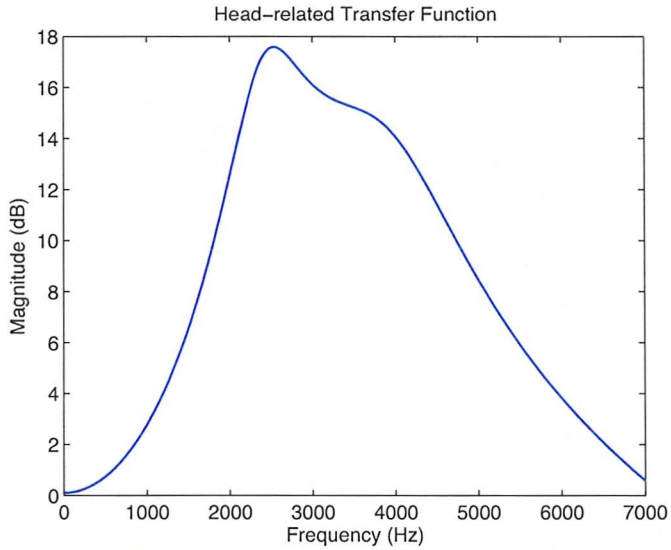

Figure 3.2: Head-related transfer function of Wiener and Ross (1946) used to model the outer ear.

(in units of Pascal) presented to the eardrum. The signal then passes through a middle ear filter and is presented to several blocks representing the cochlea. The signal path is separated into two parts, the C1 and C2 filters. The C1 filter and the C2 filter can be conceptualized as providing the tuning of the basilar membrane and representing the signal path for two modes of transduction in the C1/C2 transition hypothesis (recall Section 2.4.2 under the subheading 'Characteristics of AN Phase'). There is a control path that also contributes to the tuning of the basilar membrane and represents the effects of the cochlear amplifier and outer hair cell. The outputs of the C1 and C2 filter are combined in the inner hair cell processing block so that C1 dominates at low levels and C2 dominates at very high levels. Subsequent processing blocks represent the auditory nerve synapse and action potential generation in the auditory nerve.

For the purposes of this thesis, the neural response is interpreted as the output of the synapse model $s(n)$, which represents the synapse between the inner hair cell and the auditory nerve. The synaptic output, $s(n)$, is the discharge rate, in units of spikes/second in the auditory nerve excluding refractory effects of action potential generation[2] (Zhang *et al.*, 2001). Refractory effects are subsequently included in the model's spike generator, which uses the synaptic output as a driving function for a simulated non-homogeneous poisson process to generate individual spikes (see Zhang

[2]The refractory period refers to the period of time immediately following an action potential during which no subsequent spiking is possible.

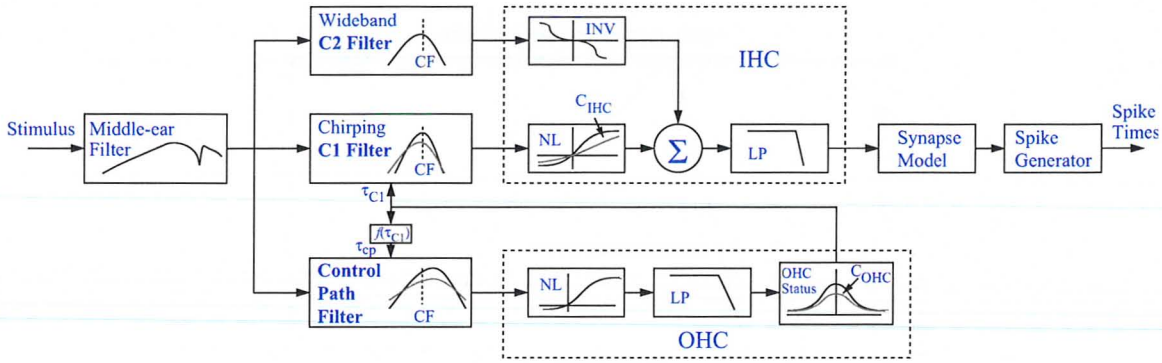

Figure 3.3: Schematic of the auditory periphery model of Zilany and Bruce (2006, 2007).

et al. (2001) for a description). This thesis uses the synaptic output to describe the activity in the auditory nerve in favour of the spike times because this method requires less computation and $s(n)$ is a deterministic function of the stimulus. This approach is justified by the idea that if the driving function of spike times can be restored, then so can the spike times themselves. This philosophy is resilient to potential limitations on account of excluding refractory effects because the model does not represent impairment to the IHC-AN synapse or to the AN itself.

Impairment in the ear can be modelled by specifying levels of damage for inner and outer hair cells, which is useful for analyzing either mixed impairment or impairment to a specific hair cell type. An individual's hearing loss profile can also be represented by the model; currently 2/3 of this loss is attributable to outer hair cell damage, while the other 1/3 is attributable inner hair cell damage. This is consistent with histological assessment of inner and outer hair cell damage following acoustic trauma (Zilany and Bruce, 2007; Liberman and Dodds, 1984). Further individualization of an impaired auditory model could be done if the degree of inner and outer hair cell damage is assessed in each case.

This auditory model is used not only in the evaluation of simulated hearing aids but also in the design of signal processing methods that are used in those hearing aids. In particular the group delay through the C1 filter of healthy and impaired models is used in an attempt to restore healthy group delays in auditory filters. The calculation of group delay through this model's C1 filter will be described in Chapter 5. The original MICEFS and SPC processing schemes will be introduced in the following chapter to provide motivation for the work of this thesis.

Chapter 4

Background iii. Hearing Aids

A hearing-aid is an electronic device whose main purpose is to amplify a sound before presentation to the eardrum of a hearing impaired person in order to restore audibility (Bruce, 2006). Hearing aids traditionally provide frequency dependent amplification that is related to the patient's hearing loss versus frequency profile, or audiogram[1]. This strategy attempts to restore audibility as a function of frequency and has been employed for cases of conductive hearing loss and sensorineural hearing loss. While this strategy is effective at treating conductive loss, its performance in cases of sensorineural hearing loss is sub-optimal due to non-linear changes in the inner ear associated with this type of loss. Audibility may be restored in the case of sensorineural hearing loss, but other perceptual deficits related to the impaired auditory nerve response remain.

Modern hearing aids use digital filtering techniques to employ sophisticated signal processing algorithms that can be used to target non-linear deficits associated with sensorineural hearing loss. Some improvements in perception have been made with advanced signal processing methods, as will be described below. However, there are still perceptual deficits that have not been fully accounted for. The work of this thesis is in the development and analysis of signal processing strategies in hearing aids to restore speech perception for persons with sensorineural hearing loss due to hair cell damage. A background of hearing aid algorithms describing their advantages and disadvantages is given below to provide a framework for this study.

4.1 Linear Prescriptions

Some simple, and early gain prescriptions, in the development of hearing aid algorithms apply linear amplification to the stimulus as a function of sound level. Linear

[1]The audiogram measures listeners' thresholds of audibility to a range of tones and is given in terms of a threshold shift from normal hearing in dB HL (dB Hearing Level).

prescriptions can provide frequency shaping, but by definition, the amount of gain that is applied to each frequency does not change with the intensity of the input sound. One notable linear prescription for sensorineural hearing loss is the 'half-gain rule', first introduced by Lybarger in 1944, who observed that hearing loss sufferers tended to choose an amplification level at roughly half of their audiogram (Dillon, 2001). In this prescription, the gain at each frequency is given as half of the loss in the audiogram in dB.

Many linear gain prescriptions are closely related to the half-gain rule. These include the NAL-R (National Acoustics Laboratory - Revised) (Byrne and Dillon, 1986), and NAL-RP (NAL - Revised, Profound) (Byrne *et al.*, 1990) prescriptions which have been used extensively in hearing-aids. The NAL-R prescription modifies the half-gain rule so that there is less gain at the low frequencies, because most of the spectral energy in speech lies in the low frequencies. The NAL-RP makes further modifications to account for profound hearing losses where sound discomfort levels are significantly elevated by providing more than half of the audiogram in these cases. An example audiogram, along with several linear prescriptions are presented in Figure 4.1.

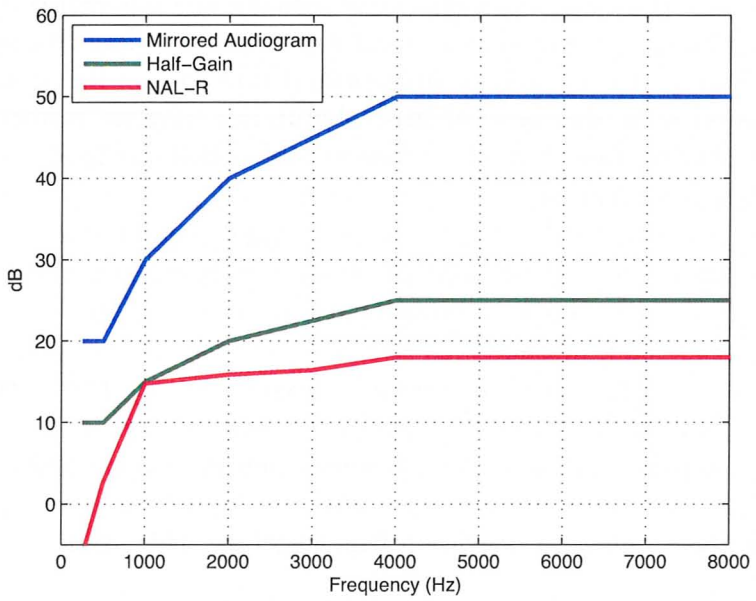

Figure 4.1: An example mirrored audiogram is given in terms of threshold shifts in dB. The gain vs. frequency profiles of the linear hearing aid prescriptions NAL-R and the half-gain rule are shown for this audiogram.

Dinath and Bruce (2008) analyzed several linear prescriptions with a computational model of the auditory periphery and demonstrated that they seem to find a balance between restoring the normal average spiking rate and the normal temporally precise auditory nerve response. As well, linear prescriptions can restore audibility in most cases, but they fail to account for the recruitment of loudness that is associated with sensorineural hearing loss. This means that they will over-amplify loud sounds beyond the hearing impaired person's threshold for discomfort. It is important for hearing aids to address this problem, which will be discussed next.

4.2 Compression Schemes

To address the problem of loudness recruitment, most modern hearing aids vary the amount of gain applied as a function of the input sound level. This is done with a compression scheme that accounts for the decreased range in hearing impaired listeners where sounds are both audible and tolerable (recall Figure 2.19). The gain applied in a compression scheme is structured so that low level sounds are amplified maximally, then as the input level grows in the compression region the amount of amplification decreases. This is demonstrated in Figure 4.2 which shows a simple input/output curve where the compression region spans the input levels 40 to 100 dB SPL.

Most of these schemes are Wide-Dynamic Range Compression (WDRC) schemes, which means their compressive regions span a broad range of sound pressure levels. They can operate on a single-band or multiple-bands; single-band compression is demonstrated in Figure 4.2. In a single band compression scheme, the input level is taken as the overall level of the input sound, summed across all frequency bands, and the gain is applied uniformly across frequency. In this way, the gain-frequency profile of the hearing aid processing scheme will retain its shape at all input levels, but the overall level of the gain curve will be adjusted based on the input level. In multi-band compression schemes, multiple frequency channels can be compressed independently. That is, the gain prescribed in a specific frequency band will be compressed based on the power of the input sound in that frequency band alone. In this way, the resultant gain-frequency profile across all frequencies can change shape dynamically, depending on the frequency spectra of the incoming sound.

Important parameters that affect the degree of compression applied are shown in Figure 4.2. These include the compression threshold or kneepoint, which determines the level at which compression is initiated; the compression ratio, which is given by the slope of the input-output curve in the compression region and determines the degree of compression applied; and the gain below threshold, which sets the gain at low input levels. In addition to these parameters there are factors that affect the speed of compression, determined by the period of time the compressor uses to detect

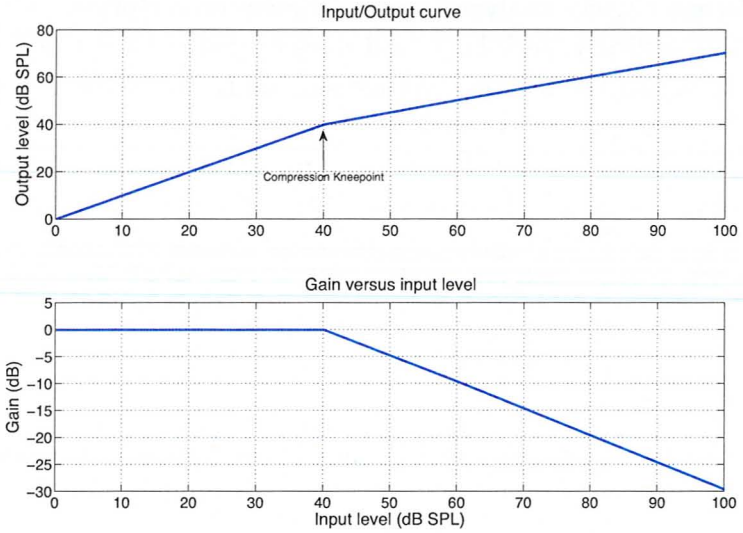

Figure 4.2: *Top Panel*: Input-Output curve shown for a compressive gain with a kneepoint of 40 dB SPL and a compression ratio of 2:1. *Bottom Panel*: The compressive gain associated with the Input-Output curve above.

changes in the signal level and adjust its gain accordingly. The attack time determines the speed at which the gain will decrease in response to an instantaneous increase in sound pressure level. Conversely, the release time determines the speed at which the gain will increase in response to an instantaneous decrease in sound pressure level (Dillon, 2001).

The duration of the attack and release times in a compression scheme determine whether it will be fast acting or slow acting. Short attack and release times will result in rapid changes in the gain applied, due to rapid changes in the level of the speech waveform. These attack and release times can be short enough to change the level between speech phonemes so this is sometimes called *phonemic compression*. One benefit of fast compression is that if a sound becomes rapidly intense, the compressor will quickly attenuate the sound, avoiding discomfort for the hearing aid wearer (Dillon, 2001). However, this may cause inter-phoneme level differences to be altered, which could be an intelligibility cue for some speech (Dillon, 2001). As well, if the fast acting compressor has many channels, then the magnitude spectrum may be flattened at any particular time (Moore, 2008). This may interfere with the relative level of formant frequencies in voiced speech, which could be important for vowel identification (Kiefte *et al.*, 2010).

A slow compression scheme, with long attack and release times, will result in slow

changes to the gain applied. This behaviour works well when the hearing aid wearer transitions between listening environments of varying intensity, however brief changes in intensity may cause discomfort (Moore, 2008). Slow-acting compressions schemes with multiple channels do not distort the short-term signal spectra, but they cannot account for rapid changes in frequency dependent level, which may be important to address frequency dependent loudness recruitment (Moore, 2008). In addition, slow compression schemes may not perform well when alternating talkers have large differences in the loudness of their voice (Moore, 2008).

In general, WDRC amplification schemes have advantages and disadvantages, depending on the hearing loss of the hearing aid wearer and the listening environment. They can provide benefit to the wearer, but there is still room for improvement in terms of customer satisfaction, especially for those with severe hearing loss (Kochkin, 2005). Some of the disadvantages of WDRC provide motivation for making use of recent processing schemes that attempt to restore a near normal auditory nerve response pattern, these include multiband and improved contrast-enhanced frequency shaping (MICEFS) and spatiotemporal pattern correction (SPC). MICEFS addresses the potential flattening of the magnitude spectrum caused by multiband compression by combining it with a contrast enhancement scheme that attempts to restore the neural representation of formant frequencies in the auditory nerve. SPC addresses the problem of impaired compression in the basilar membrane with a different approach: it attempts to account for the changes affecting the timing of AN response that occur with sensorineural hearing loss as opposed to compressing the gain. These schemes will be described in the following two sections.

4.3 MICEFS - Multiband and Improved Contrast-Enhanced Frequency Shaping

The MICEFS processing scheme developed by Ansari (2005) and Harte *et al.* (2006) is an extension of the contrast-enhanced frequency shaping (CEFS) algorithm that was originally proposed by Miller *et al.* (1999). It combines fast-acting multiband compression with a contrast enhancement scheme that attempts to restore synchrony in the auditory nerve to the first three formants of voiced speech. The original CEFS scheme suppresses the first formant in a vowel compared to the second and third formants in an effort to reduce an upward spread of synchrony to F1 that has been observed in impaired auditory nerve fibers in response to a synthesized vowel (Miller *et al.*, 1997). This was implemented with some success, however as the synchrony to F1 was localized, there was an abnormal upward spread of synchrony to F2 and no synchrony to F3 was restored (Miller *et al.*, 1999). An improved version of the contrast enhancement scheme was developed by Ansari (2005) that amplified the F2

and F3 regions using the half gain rule at those frequencies, while providing relatively less gain in the trough between F2 and F3. This provided some success in limiting the upward spread of synchrony to F2 that was observed by Miller *et al.* (1999) and restoring some synchrony to F3. A depiction of the gain versus frequency profiles provided by CEFS and MICEFS are shown in the top and bottom panels of Figure 4.3, respectively.

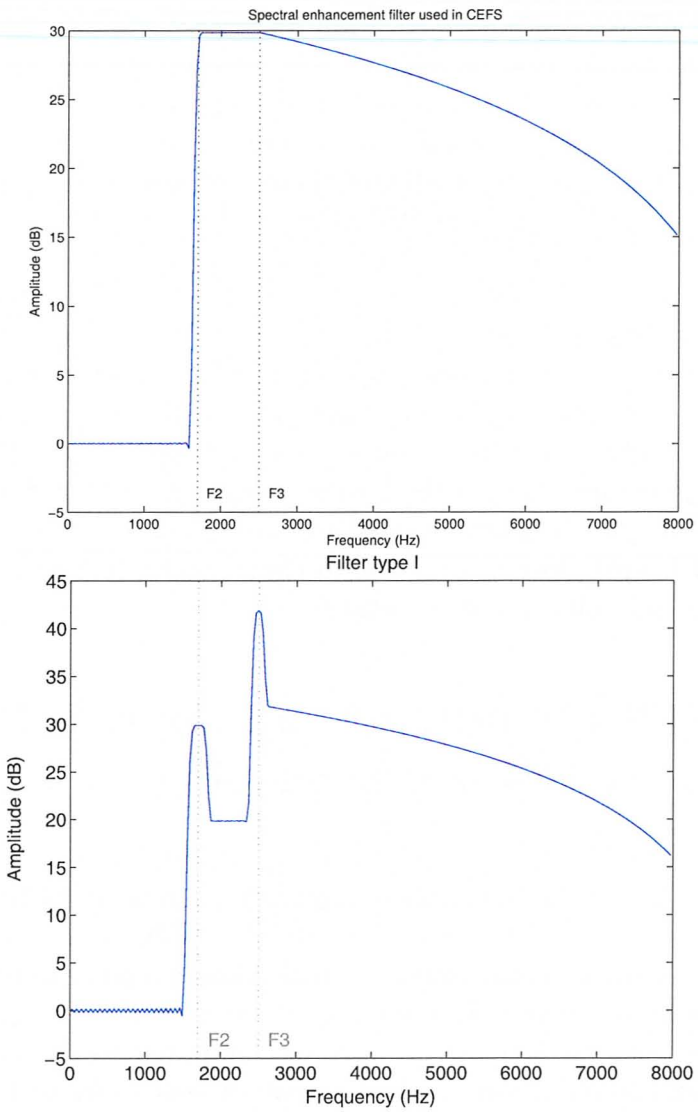

Figure 4.3: *Top Panel*: Example gain versus frequency profile provided by the CEFS processing scheme of Miller *et al.* (1997). *Bottom Panel*: Example gain versus frequency profile for the contrast enhancement scheme of the original MICEFS. (Reprinted from Ansari, 2005)

The MICEFS algorithm is able to apply a contrast enhancement scheme in running speech by using the formant tracker of Mustafa and Bruce (2006), which enables continuous evaluation of the formants required for calculating the gain-frequency profile. An illustration of the MICEFS processing scheme is shown in Figure 4.4, where the top row of processing blocks implements the multiband compression and the bottom part of the figure represents the formant tracking and contrast enhancement.

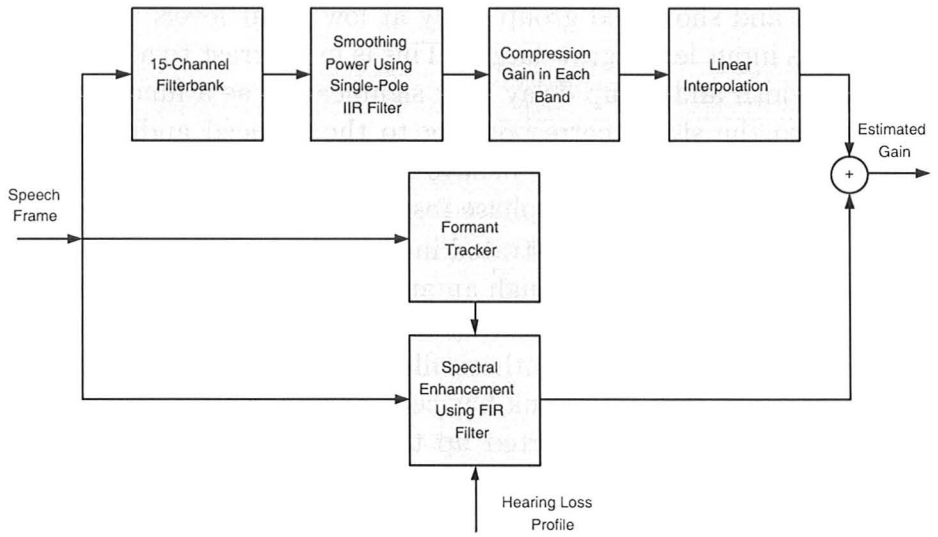

Figure 4.4: Block diagram of the processing done on a single speech frame in the STFT. The top row of processing blocks perform mulitiband compression, the middle row tracks the formants, and the bottom row performs the spectral enhancement. (Reprinted from Ansari, 2005)

Bruce (2004) demonstrated that adding this type of multiband compression to the CEFS scheme did not degrade the model neural representation of formants in running speech, and Ansari (2005) showed that MICEFS also improved synchrony to F1, F2 and F3 in an impaired ear model. The inclusion of multiband compression in MICEFS should address loudness recruitment in a similar manner that WDRC schemes do, by limiting the output level to the hearing-aid wearers dynamic range. As well, it will be interesting to combine the MICEFS scheme with a spatiotemporal pattern correction scheme that attempts to restore temporal aspects of the auditory nerve discharge pattern, which could also be related to loudness recruitment (Carney, 1994).

4.4 SPC - Spatiotemporal Pattern Correction

Spatiotemporal pattern correction (SPC) is a hearing aid signal processing algorithm originally proposed by Shi *et al.* (2006) and later modified and tested by Calandruccio *et al.* (2007). The motivation behind the development of this scheme was to correct abnormal temporal properties of the auditory nerve discharge pattern that are associated with sensorineural hearing loss and loudness recruitment. In particular, it attempts to account for the shortened group delay exhibited by auditory filters in the impaired auditory nerve, as described in Section 2.5. An impaired auditory filter has a broad bandwidth and shortened group delay at low input levels which only varies by a small degree as input levels grow larger. This is in contrast to a healthy auditory filter, whose bandwidth and group delay vary significantly as a function of level. By inserting delays into the signal corresponding to the reduced auditory filter group delays, the impaired AN response may behave as if its group delays were restored to normal and exhibit improved relative phase response across AN fiber populations.

The SPC processing scheme is illustrated in Figure 4.5. It has two pathways, one pathway that passes the stimulus through an auditory model to determine the group delays of the healthy and impaired auditory filters, and another pathway that processes the stimulus with an analysis-synthesis filterbank and inserts the corresponding delays into each channel. The filterbank has center frequencies ranging from 100 Hz to 5000 Hz, but delays are only inserted up to 2000 Hz because SPC benefit was deemed to decrease as frequency increases further[2] (Shi *et al.*, 2006). For each center frequency included in the filterbank of the processing path, there is a corresponding channel in the modelling path whose characteristic frequency is equal to this center frequency.

The processing scheme of Shi *et al.* (2006) uses the auditory model of Heinz *et al.* (2001) to describe the healthy and impaired auditory filters and to calculate their corresponding group delays. This model uses gammatone bandpass filters[3] to represent the tuning of auditory filters along the basilar membrane; these filters change their bandwidths as a function of level, which results in time-varying group delay. Healthy and impaired group delays through the gammatone filters are calculated for each frequency channel in the SPC scheme, and the time-varying difference in group delay between the two auditory models is inserted into the corresponding frequency channel of the filterbank in the processing path. The appropriate delay is inserted into each frequency channel by time shifting these channels by an integer number of points (Shi *et al.*, 2006).

The motivation behind the SPC scheme suggests that prior insertion of delays into

[2]2000 Hz is sometimes cited as the frequency at which phase-locking begins to roll off (Pickles, 2008), which can explain the decreased SPC benefit at high frequencies.

[3]Gammatone filters will be described in more detail in Section 5.2.2.

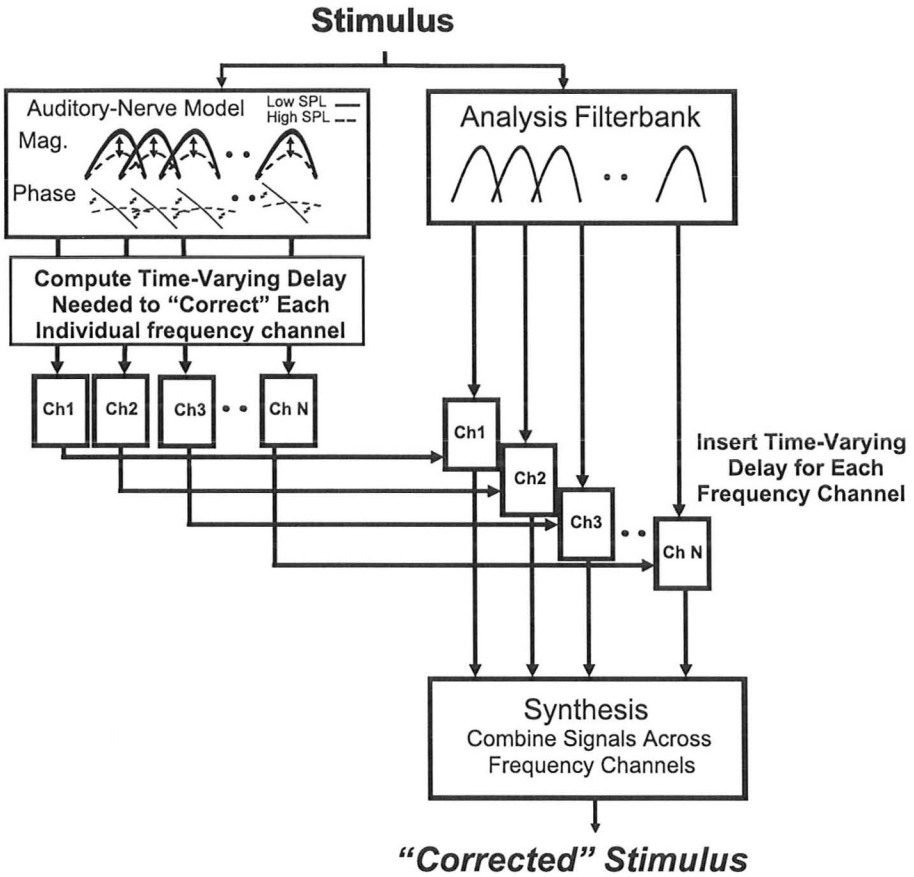

Figure 4.5: Block diagram of the processing done in spatiotemporal pattern correction scheme. The modelling path is shown on the left side and the processing path is shown on the right side. (Reprinted from Shi *et al.*, 2006)

each frequency channel of the signal might be able to compensate for the abnormal group delays of the impaired auditory filters as the real basilar membrane decomposes the time-domain waveform into its various frequency components. In this way, the temporal variations in the impaired auditory filter's output in response to an SPC processed signal might be similar to the temporal variations in the output of the healthy auditory filter in response to an unprocessed signal. This might then correct the impaired spatiotemporal response pattern, in the hope that this correction could help restore speech perception, in particular the perception of loudness in the hearing impaired.

This line of thought seems reasonable, however testing of the SPC scheme with hearing impaired listeners showed that unprocessed sounds were preferred over processed sounds in general (Calandruccio *et al.*, 2007). There were some cases where

a small amount of SPC processing was preferred, but no systematic improvement in speech intelligibility was provided by the processed speech (Shi *et al.*, 2006; Calandruccio *et al.*, 2007).

Shi *et al.* (2006) did not study the effects of SPC with a computational model. The reasonable motivation behind SPC and its poor results in human listening tests provides incentive for analysis with a computational model of the impaired ear. Systematic evaluation of the SPC scheme with such a model should provide insight into the poor performance in human listeners and could result in modifications to the SPC scheme that provide some benefit.

As well, Shi *et al.* (2006) did not attempt to restore normal synchrony to the formants in voiced speech when prescribing the insertion delays with a computational model. This could be an important correction because the group delay computed from the computational model is dependent on a feed-forward control path that can incorporate hearing impairment. Since hearing loss can result in abnormal synchrony to formant frequencies, the insertion delays that SPC prescribes might be derived from healthy and impaired group delays that are driven by two different synchronization frequencies. This would result in erroneous insertion delays prescribed by SPC. The current work pre-processes SPC with MICEFS to determine if more benefit can be achieved if near-normal synchrony to the speech formants is first restored.

In addition, MICEFS and SPC both attempt to account for the reduced non-linear compression observed in the sensorineural impaired ear using different strategies. MICEFS does this with its compression scheme targeted to limiting the discharge rates in the auditory nerve. SPC does this by attempting to correct for abnormal group delays that are associated with loss of compression in the impaired ear. Since the loss of non-linear compression in the basilar membrane has been associated with loudness recruitment, it will be interesting to determine if a combination of these two approaches can be used effectively to restore loudness perception.

This motivates the work of this thesis, which is in the analysis and modification of both the MICEFS and the SPC scheme, as well as their combination. The implementations of these schemes that are developed in this study is described next.

Chapter 5

Methods i. Signal Processing

This chapter describes the implementation of the MICEFS and the SPC processing schemes used in this thesis. Particular emphasis is given to methods that differ from the original implementations by Ansari (2005) and Shi *et al.* (2006), respectively. Several variations of these processing schemes will be developed here and evaluated in Chapter 7, to determine which versions generate the best results. The results will be discussed in Chapter 8.

5.1 MICEFS

The basic framework for the MICEFS processing used in this thesis remains the same as it has been presented by Ansari (2005) and summarized in Section 4.3 of this thesis. The gain applied by MICEFS is implemented in the frequency domain using the STFT (Short-Time Fourier Transform) with a sampling rate of 16 kHz and a 128 sample long DFT (Discrete Fourier Transform). The formants of voiced speech are tracked continuously using the formant tracker of Mustafa and Bruce (2006). This information is combined with the individual's hearing loss profile and used to prescribe a specific contrast enhancement scheme. A 15-channel compression scheme works in parallel to the contrast enhancement, and the gains prescribed from both paths are combined to give the MICEFS gain.

A block diagram of the processing framework for the MICEFS algorithm appears in Figure 5.1, which demonstrates its use of the STFT. The time domain speech signal is buffered into 8 msec frames (128 samples/16 kHz sampling rate), each frame overlapping by 50% and multiplied by an 8 msec Hanning window, as demonstrated in Figure 5.2. Each frame is then zero-padded by 384 samples distributed evenly at the beginning and end of each frame. This method of zero-padding avoids phase distortion, which is important for correct operation of the SPC scheme that affects the relative phase response of auditory nerve fibers. Using the STFT in this way,

MICEFS operates on a frame-by-frame basis and applies its contrast enhancement and compression gain to each frame by direct multiplication prior to conversion to the time domain.

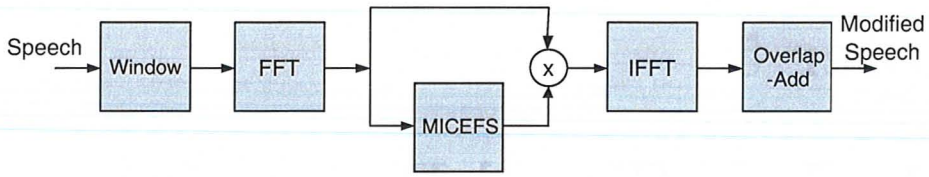

Figure 5.1: Block diagram of the use of the short-time Fourier transform (STFT) in MICEFS. (Reprinted from Ansari, 2005)

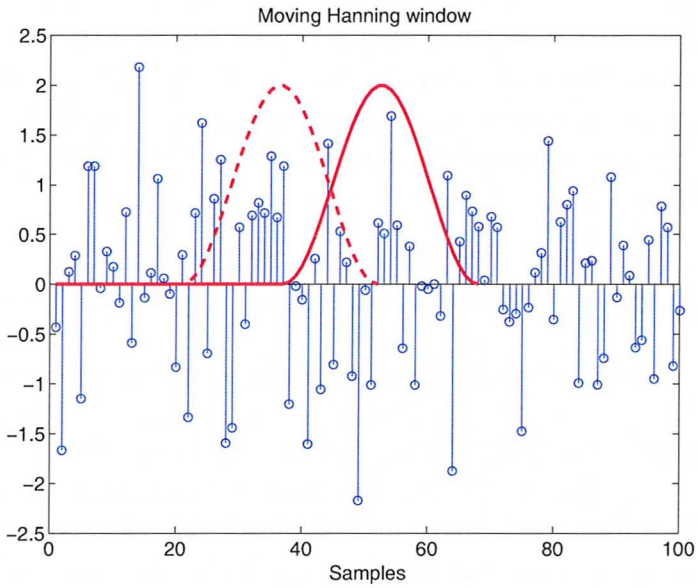

Figure 5.2: Demonstration of the windowing of the speech waveform with a Hanning window that is advanced by 50% of the window length. (Reprinted from Ansari, 2005)

5.1.1 Formant Tracker

The contrast enhancement filter in the MICEFS processing scheme depends on accurate predictions of the formant frequencies in order to amplify those formants accordingly. The formant tracker of Mustafa and Bruce (2006), used in MICEFS, finds estimates of the first four formants in continuous speech using an adaptive bandpass filterbank to emphasize each of the formants prior to estimation. Each of the formant

filters in the filterbank adapt dynamically using feedback by positioning their poles at the previous estimate of the desired formant frequency and placing zeros at the other formant frequencies. A snapshot in time of the magnitude response of these four filters can be seen in Figure 5.3, which clearly demonstrates a pole at the filter's formant estimate and zeros at the other formant estimates. The current formant frequency is then extracted from the output of each of these formant filters using first-order linear prediction to fit a single pole to each of these signals (Mustafa and Bruce, 2006).

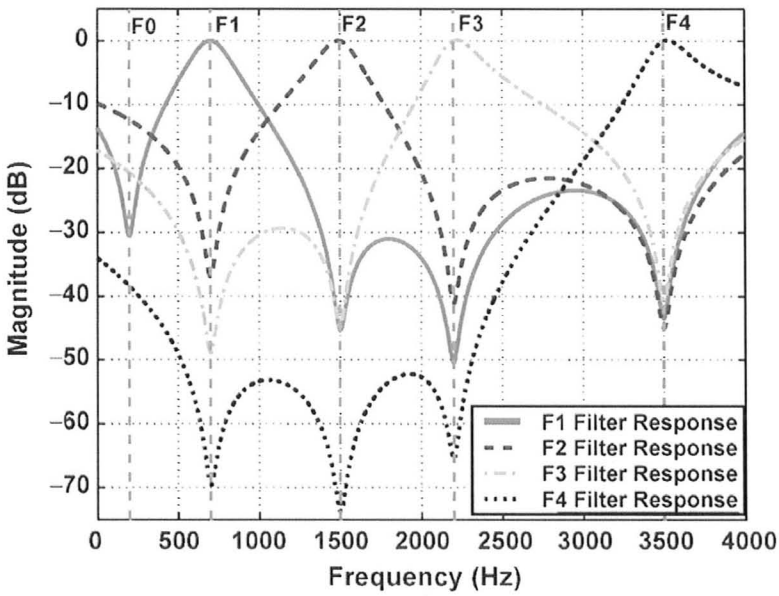

Figure 5.3: An example gain-frequency response of the four formant filters in the Mustafa and Bruce (2006) formant tracker at one time instant. (Reprinted from Mustafa and Bruce, 2006)

5.1.2 Contrast Enhancement

The original MICEFS[1], denoted MICEFS-0 in this thesis, uses the half-gain rule to prescribe the gain at the second and third formant frequencies in the contrast enhancement filter, while no gain is applied to the first formant; this can be seen on the left hand side of Figure 5.4, which plots an example gain-frequency profile for the vowel /ɛ/. The gain in the trough between F2 and F3 is 10 dB below the gain at F2. Above F3, the gain decreases from the value of the gain at F3 minus 10, to the

[1]Filter type I in Ansari (2005) is used in this implementation.

square root of the gain at F3 minus 10 at the Nyquist frequency. The width of the passbands at F2 and F3 are approximately 200 Hz and 100 Hz respectively, but this varies as the formant frequencies change.

MICEFS-0 sets the gains according to the formant frequencies as in Table 5.1, where HG(f) is half the hearing loss in the audiogram interpolated to the frequency f, and HL(f) is the hearing loss at f. The gain determined in Table 5.1 is found for all frequency bins that appear in the DFT by linear interpolation, and this results in a gain-frequency profile as shown on the left hand side of Figure 5.4.

Table 5.1: MICEFS contrast enhancment gains.

Frequency (Hz)	MICEFS-0 Gain (dB)	MICEFS-1 Gain (dB)
0	0	0
$(7 \times \text{F1})/16$	0	0
$(10 \times \text{F1})/16$	0	HG(F1)
$(13 \times \text{F1} + 3 \times \text{F2})/16$	0	HG(F1)
$(11 \times \text{F1} + 5 \times \text{F2})/16$	0	HG(F1)-10
$(2 \times \text{F1} + 14 \times \text{F2})/16$	0	HG(F1) -10
$(\text{F1} + 15 \times \text{F2})/16$	HG(F2)	HG(F2)
$(14 \times \text{F2} + 2 \times \text{F3})/16$	HG(F2)	HG(F2)
$(13 \times \text{F2} + 3 \times \text{F3})/16$	HG(F2) $-$ 10	HG(F2) $-$ 10
$(2 \times \text{F2} + 14 \times \text{F3})/16$	HG(F2) $-$ 10	HG(F2) $-$ 10
$(\text{F2} + 15 \times \text{F3})/16$	HG(F3)	0.52×HL(F3)
$(15 \times \text{F3} + F4)/16$	HG(F3)	0.52× HL(F3)
$(14 \times \text{F3} + 2 \times F4)/16$	HG(F2) $-$ 10	0.52× HL(F3) -10
$F_s/2$	(HG(F2)-10)/2	(0.52× HL(F3))/2

Multiple updates to the MICEFS processing scheme were developed throughout the course of this thesis, however only the version that produced the best results will be presented here and evaluated in Chapter 7. This version is denoted 'MICEFS-1' and its contrast enhancement scheme is calculated in the right-hand column of Table 5.1. The contrast enhancement scheme of MICEFS-1 provides some gain at F1 so that synchrony to F1 is not suppressed in impaired-aided auditory nerve responses; the benefits of this approach are demonstrated in Chapter 7. MICEFS-1 also provides slightly more gain at the third formant to ensure that auditory nerve response to F3 is not suppressed by F2. Figure 5.4 shows the gain-frequency profile provided by the contrast enhancement scheme of MICEFS-1 to highlight the differences in the contrast enhancement schemes of the two MICEFS versions.

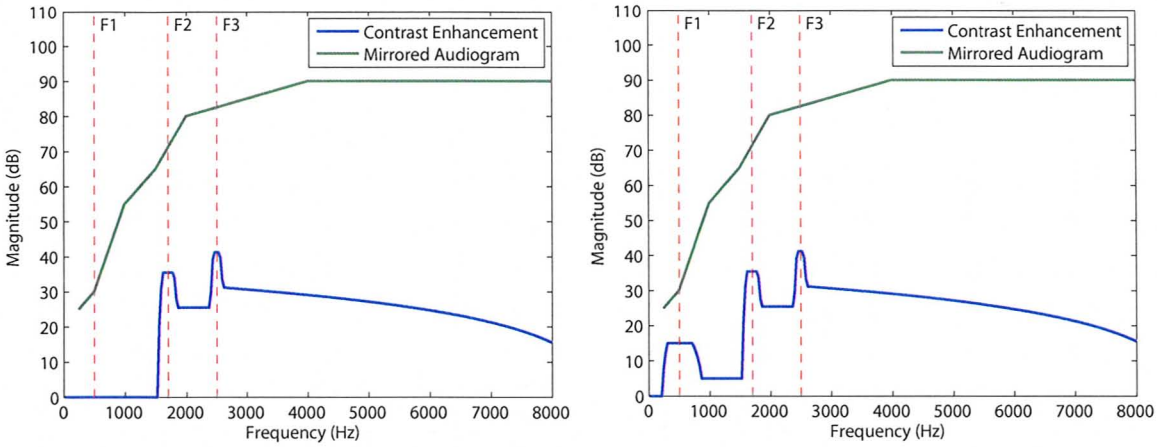

Figure 5.4: *Left Panel*: Gain-frequency profile of the contrast enhancement scheme for MICEFS-0 in response to a synthesized vowel for a moderate-to-severe audiogram. *Right Panel*: Contrast enhancement for MICEFS-1.

5.1.3 Multiband Compression

The framework used to implement multiband compression in MICEFS-1 is the same as it was used in Ansari's version of MICEFS (MICEFS-0). It is a 15-channel compression scheme with filters that have center frequencies spaced at 1/3 octaves beginning from 250 Hz. The first filter is a lowpass filter with a unity passband and a cutoff frequency of 250 Hz, while the remaining filters are bandpass filters with unity gain. Their passbands have a width of 4 auditory filter ERBs (See Appendix A, Section A for calculation of the ERB), 2 ERBs to either side of the center frequency, and a roll off from the passband of 30 dB/octave.

The multiband compression works by calculating the power in each frequency band, then adjusts the gain at each filter's center frequency accordingly and interpolates this gain into all the frequency bins present in the DFT. The power is calculated in each band as follows:

$$P_{\text{band}} = \sum_{k=0}^{N-1} |H[k]|^2 . |X[k]|^2 \tag{5.1}$$

where $|H[k]|$ is the absolute value of one compression filter's frequency response, $|X[k]|$ is the magnitude of the DFT for the speech frame of interest, frequency is indexed by k, and $N = 512$ is the number of frequency bins in the DFT. This power is smoothed to implement fast acting compression with an instantaneous attack time, and a release time of around 60 ms (Ansari, 2005).

The gain at the compression filter's center frequency, Gain(f_c), is prescribed by

$$\text{Gain}(f_c) = G_{\text{th}}(f_c) - (P_{\text{band}} - C_{\text{th}}(f_c)) \times (1 - \frac{1}{\text{CR}(f_c)}), \tag{5.2}$$

where G_{th} is the gain below the compression threshold, C_{th} is the compression threshold, and CR is the compression ratio. Then the gain at each of the 15 center frequencies is linearly interpolated to all the frequency bins in the DFT.

In addition to an improved contrast enhancement scheme, MICEFS-1 also has a different set of multiband compression parameters. The parameters that are adjusted include the compression threshold, the gain below the compression threshold, and the compression ratio. The differences between the compression parameters for MICEFS-0 and MICEFS-1 are shown in Table 5.2. In MICEFS-1 the compression ratio, CR, has a sigmoidal shape that increases from a value of 1.4 to 2 as a function of frequency.

The motivation for altering the compression parameters was derived from a comparison of the compression gains that resulted from processing with MICEFS-0 and those that were prescribed by NAL-NL1 (NAL-Non-linear version 1.1), a standard compressive fitting procedure. This comparison appears in Chapter 7. The compression parameters were initially changed so the compression gains at the formant frequencies would be similar to those prescribed by NAL-NL1. However, the final values of the compression parameters were determined on an ad hoc basis by their ability to restore synchrony in the neural response. In this regard, MICEFS-1 results in less compression at low frequencies in order to preserve synchrony to F1 at high intensity levels. As well, the threshold of compression, C_{th}, is increased relative to MICEFS-0 so that adequate synchrony remains at low intensity levels.

Table 5.2: MICEFS compression parameters

Parameter	MICEFS-0	MICEFS-1
CR(f_c)	2:1	CR(f_c) = $(2 - 1.4)(1 - s(f_c)) + 1.4$, where $s(f_c) = 1 - \text{sigmoid}(9 \times 10^{-3}(f_c - 800))$ and $\text{sigmoid}(x) = 1/(1 + \exp(-x))$
$C_{th}(f_c)$	40 dB SPL	45 dB SPL
$G_{th}(f_c)$	0 dB	0

5.2 Spatiotemporal Pattern Correction

This thesis develops several versions of the spatiotemporal pattern correction (SPC) processing scheme presented in Shi *et al.* (2006) that was designed to correct for shortened group delays in a population of impaired auditory nerve fibers. The SPC scheme has been introduced in Chapter 4 and its implementation is described here, particularly where it differs from the original processor. A flow diagram of the processing scheme is reprinted in Figure 5.5 for convenience. The SPC scheme has two pathways: one pathway that processes the stimulus with an auditory model to determine the group delays through the healthy and impaired auditory filters; and another pathway that processes the stimulus with a filterbank and inserts the corresponding difference in group delay into each channel before channel synthesis.

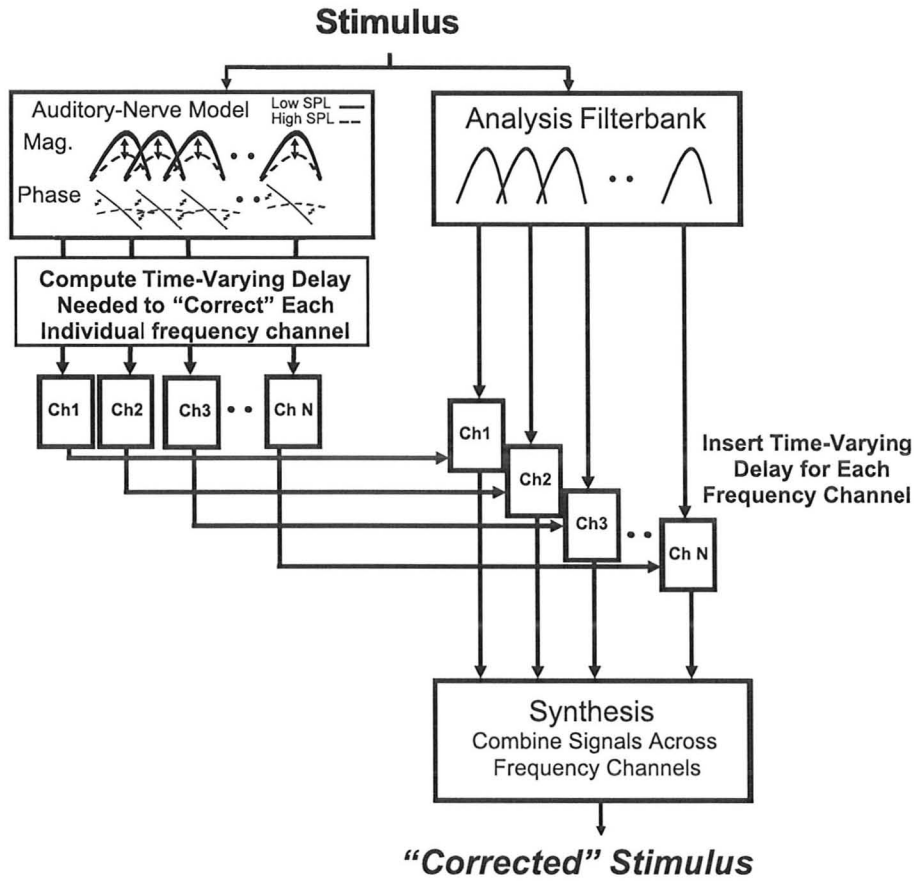

Figure 5.5: A block diagram of the processing done in the spatiotemporal pattern correction scheme. The modelling path is on the left, and the processing path is on the right. (Reprinted from Shi *et al.*, 2006)

5.2.1 Modelling path

The modelling path of the the SPC scheme consists of two auditory models, one healthy and one impaired, each representing 52 auditory nerve fibers. The fiber characteristic frequencies are evenly spaced on an ERB scale[2], with 2 filters per ERB from 100 Hz to 5000 Hz; however, delays are only inserted up to 2000 Hz. Individual fiber characteristic frequencies can be calculated with Equation A.3, Appendix A. The purpose of this modelling path is to find the differences in the group delay of auditory filters in the impaired ear compared to the healthy ear.

Group delay is a measure of the overall delay that a signal experiences when passing through a filter and is generally shorter for impaired auditory filters due to their broadened bandwidths (Shi *et al.*, 2006). For this reason, the difference in instantaneous group delay between the healthy and impaired model auditory filters is measured in the modelling path and subsequently inserted into the corresponding frequency channels in the processing path of the SPC scheme. The group delay through the auditory filters is computed in all versions of the SPC scheme that appear in this thesis as the delay through the C1 filter in the Zilany and Bruce (2007) model. This is opposed to the gammatone filters that were used in the original implementation of SPC by Shi *et al.* (2006), which employed the auditory model of Heinz *et al.* (2001). The C1 filter can be viewed in the context of the auditory model in Figure 3.3 as the signal path corresponding to stimulation of the inner hair cell by basilar membrane displacement. It describes the tuning of a particular point on the basilar membrane which is affected greatly by outer hair cell impairment. The bandwidth of the C1 filter increases with outer hair cell impairment but also varies as a function of time and is dependent on the input stimulus, including the sound pressure level.

The auditory model of Zilany and Bruce (2006, 2007) is used in favor of the model of Heinz *et al.* (2001) that was used in the original SPC scheme because it provides an improved description of auditory nerve response in cats. Particularly, the 2006, 2007 model represents the auditory nerve responses at high sound pressure levels more accurately. This is important when employing a model for use with the SPC scheme because the temporal response of AN fibers varies as a function of level, and this may have some effect on the perception of loudness for the individual with hearing loss (Carney, 1994). It also includes a middle ear filtering section which provides spectral shaping that can effect the level of input to a particular point on the basilar membrane, determining the bandwidth and group delay for that auditory filter.

[2]See Appendix A for an explanation of ERB.

C1 Filter

The C1 filter in the Zilany and Bruce model is a 10^{th} order digital filter with two complex second-order poles, one complex first-order pole, their complex conjugates, and a real 5^{th} order zero. The order of this filter is important as it determines the sharpness of its tuning, which is also affected by OHC impairment and sound level (Zilany and Bruce, 2006). The positioning of the poles and zeros determine the C1 filter's frequency response, which reflects known auditory nerve physiology such as its tuning curve and magnitude response. The poles and zeros vary as a function of time depending on the output of the control path, which accounts for the tuning effects of signal sound pressure level and OHC impairment. These time-varying poles and zeros are then used to determine the frequency response and transfer function of the healthy and impaired auditory filters, including their group delay which is used in the processing path of SPC.

The poles and zeros of the C1 filter are first defined in the continuous-time (or analog) domain and then a digital filter is implemented as a difference equation using the bilinear transformation to obtain the necessary coefficients. The bilinear transformation allows for translation of poles and zeros from the analog domain to the digital domain and inherently enforces a frequency warping described in Oppenheim and Schafer (1999). The 10^{th} order filter is implemented as a serial cascade of 5 second-order difference equations, each representing a filter with a single pole and a single zero. The second-order difference equation can be written as follows:

$$\begin{aligned} a_0 \cdot y[n] = b_0 \cdot x[n] + b_1 \cdot x[n-1] + b_2 \cdot x[n-2]... \\ - a_1 \cdot y[n-1] - a_2 \cdot y[n-2] \end{aligned} \tag{5.3}$$

where the a's and b's are the coefficients determined using the bilinear transform.

C1 Filter Group Delay

The group delay through the C1 filter cannot be calculated using the digital filter coefficients because of a numerical error that results on account of large discrepancies in the magnitudes of the filter coefficients. For example, the denominator coefficients for the C1 filter at a characteristic frequency of 500 Hz in response to a CF-tone at 50 dB SPL are as follows:

$$\begin{aligned} a_0 &= 1 \\ a_1 &= -8 \times 10^{10} \\ a_2 &= 3.97 \times 10^{10}. \end{aligned}$$

To address this problem, the group delay through the C1 filter is instead calculated from the analog pole-zero definition of the filter. This is found using Equations 9-15 of Zilany and Bruce (2006). The group delay for an arbitrary analog system function with one pole and one zero can be derived by direct application of the formal definition of group delay, in Equation 2.1. An arbitrary analog pole-zero system function of the complex variable $s = \sigma + j\omega$ is described by:

$$H(s) = \frac{(s-q)}{(s-p)(s-p^*)} \tag{5.4}$$

where q is a real zero, $p = \sigma_p + j\omega_p$ is a complex pole, and $p^* = \sigma_p - j\omega_p$ is its complex conjugate. The angle is then

$$\angle H(s) = \angle(s-q) - \angle(s-p) - \angle(s-p^*), \tag{5.5}$$

which becomes:

$$\angle H(\sigma + j\omega) = \arctan(\frac{\omega}{\sigma - \sigma_q}) - \arctan(\frac{\omega - \omega_p}{\sigma - \sigma_p}) - \arctan(\frac{\omega + \omega_p}{\sigma - \sigma_p}). \tag{5.6}$$

The system's frequency response is the system transfer function $H(s)$ evaluated along the imaginary axis (i.e., $s = j\omega$) so the system's phase response can be written as

$$\angle H(j\omega) = \arctan(\frac{\omega}{-\sigma_q}) - \arctan(\frac{\omega - \omega_p}{-\sigma_p}) - \arctan(\frac{\omega + \omega_p}{-\sigma_p}). \tag{5.7}$$

Recall from Equation 2.1 that the group delay is $-\frac{d}{d\omega}\angle H(j\omega)$, so taking the derivative of Equation 5.7[3], we get:

$$grd[H(j\omega)] = \frac{1/\sigma_q}{1 + (\frac{\omega}{-\sigma_q})^2} - \frac{1/\sigma_p}{1 + (\frac{\omega-\omega_p}{-\sigma_p})^2} - \frac{1/\sigma_p}{1 + (\frac{\omega+\omega_p}{-\sigma_p})^2}. \tag{5.8}$$

Equation 5.8 is used to find the group delay through a transfer function with one real zero and a complex pole with its complex conjugate as described by Equation 5.4. The implementation of the C1 filter in Zilany and Bruce (2006) is a cascade of five such transfer functions, where the difference equation coefficients have been found using the bilinear transform. The group delay of Equation 5.8 for each of these five transfer functions can be added to approximate the group delay through the digital filter that appears in the model[4].

[3] $\frac{d}{dw}\arctan x = \frac{1}{1+x^2}$

[4] These group delay calculations are tested thoroughly in Appendix C.

An example group delay versus frequency curve is shown in Figure 5.6 for a healthy and a mildly impaired auditory filter with a characteristic frequency of 1500 Hz and a stimulus with a sound pressure level of 60 dB SPL. The group delay through these filters varies as a function of stimulus frequency, where narrowband stimuli with a frequency close to CF will experience the most group delay. In the modelling path the frequency content of the stimulus is not specifically known, for this reason the instantaneous group delay through the C1 filter is taken as the group delay for a stimulus at its characteristic frequency.

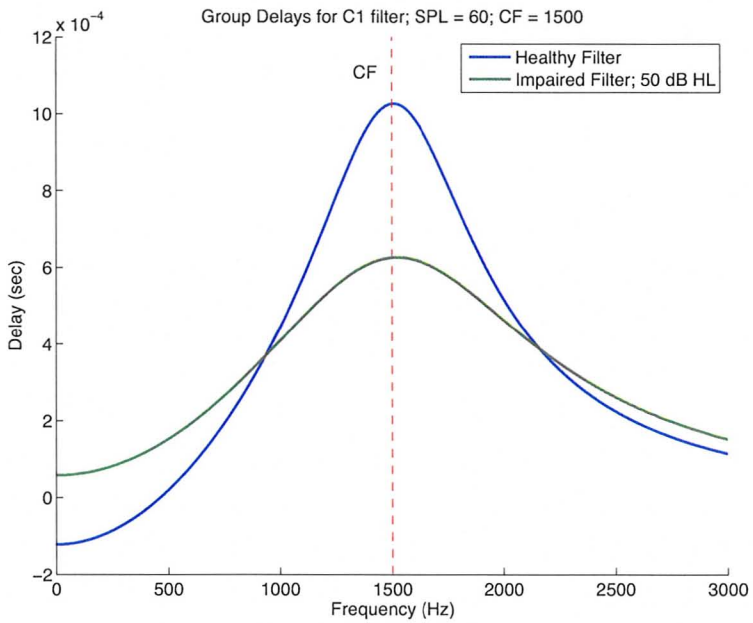

Figure 5.6: Group delays calculated through the auditory model's C1 filter for a fiber tuned to 1500 Hz in response to a CF-tone. Group delays are shown for a healthy fiber and an impaired fiber with a hearing loss of 50 dB at CF.

Group delay is found for each model auditory nerve fiber in this manner for both the healthy and impaired models and the difference between them is taken. The group delay through the healthy auditory nerve fibers is generally longer than the group delay through the impaired auditory nerve fibers because they have sharper tuning. This means that the difference in group delay is generally positive, which corresponds to the requirement of a delay being inserted into the analysis-synthesis gammatone processing path.

5.2.2 Processing path

The delays calculated in the modeling path of the SPC algorithm are inserted into the stimulus in the processing path, which can be seen on the right side of Figure 5.5. This is done by separating the stimulus into individual frequency channels with an analysis-synthesis filterbank, inserting the time dependent delays for each channel, then linearly combining the filtered channels into one audio signal. The analysis-synthesis filterbank is a gammatone filterbank presented by Hohmann (2002) and is the same that is used in the SPC scheme by Shi *et al.* (2006).

The analysis-synthesis filterbank consists of gammatone filters whose center frequencies are equated to the characteristic frequencies of the auditory nerve fibers evaluated in the modeling path. In this way, the difference between healthy and impaired group delays at individual auditory filters can be applied to filterbank channels of corresponding frequency.

Gammatone Filterbank

The gammatone filterbank presented by Hohmann (2002) and used by Shi *et al.* (2006) consists of 52 linear 4^{th} order gammatone filters that are implemented by cascading four first order filters. The difference equation of these first order filters is given by Hohmann as

$$\tilde{y}_n = \tilde{x}_n + \tilde{a} \cdot \tilde{y}_{n-1} \tag{5.9}$$

where $\tilde{x}$ and $\tilde{y}$ are the complex input and output respectively and $\tilde{a}$ is a complex coefficient defined for each filter in the filterbank. The actual input to this filter is real valued, but it is cascaded so subsequent iterations of this equation include a complex valued input. The coefficient $\tilde{a}$ is determined as follows:

$$\begin{aligned}
\tilde{a} &= \lambda e^{j\beta} && (5.10)\\
\text{with} \quad \lambda &= \exp(-\frac{2\pi b}{F_s}) && (5.11)\\
\text{and} \quad b &= \text{ERB}_{\text{aud}}(f_c)/a_\gamma && (5.12)\\
\text{and} \quad a_\gamma &= \frac{\pi(2\gamma-2)!2^{-(2\gamma-2)}}{(\gamma-1)!^2} && (5.13)\\
\text{and} \quad \beta &= 2\pi\frac{f_c}{F_s} && (5.14)
\end{aligned}$$

where $\gamma = 4$ is the filter's order, f_c is the filter's center frequency, F_s is the sampling frequency of 32 kHz, and ERB_{aud} is given in Equation A.2, Appendix A. The

magnitude response for an example gammatone filter with a center frequency of 1344 Hz appears in Figure 5.7. Each gammatone filter is a bandpass filter with a 3-dB bandwidth that is given by Hohmann as

$$\text{BW} = \frac{c_\gamma}{a_\gamma}\text{ERB}_{\text{aud}} \tag{5.15}$$

$$\text{with} \quad c_\gamma = 2\sqrt{2^{1/\gamma} - 1}. \tag{5.16}$$

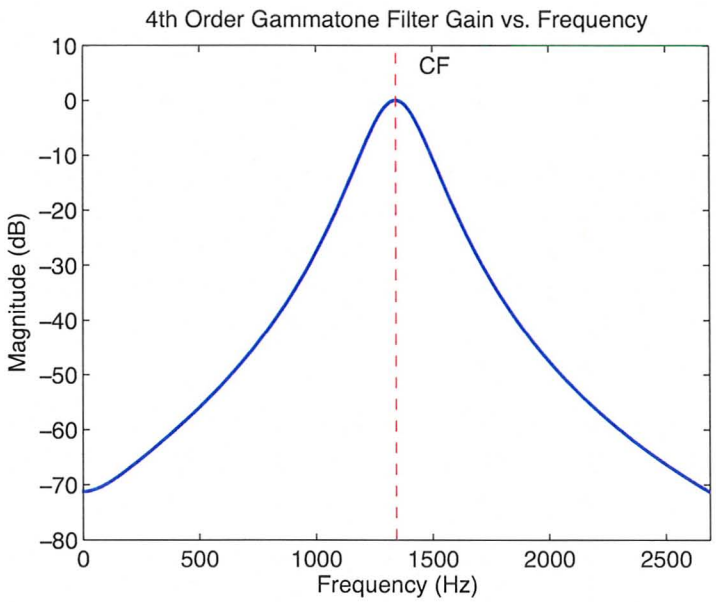

Figure 5.7: Gain-Frequency profile for a single 4th order gammatone filter with center frequency 1344 Hz.

Synthesis of filterbank channels requires temporal alignment because each filter's bandwidth increases as its center frequency increases. As the bandwidth of the filter increases, the group delay through the filter decreases, causing each filter's impulse response to have its peak response at a different time. The top left panel of Figure 5.8 shows the impulse response of a low frequency filter in the filterbank along with its envelope, and the top right panel shows the impulse response of a high frequency filter. The time required for the low frequency filter's impulse response envelope to reach its peak is greater than the corresponding time required for the high frequency filter. In order to account for these differences, the output of each filter channel is time-shifted such that peaks in its impulse responses are time-aligned. This is done in two ways. Prior to temporal alignment, a desired delay for the synthesized signal

is chosen[5]; then the peak of each filter's impulse response envelope is compared to the desired delay. The method employed for time-alignment with the desired delay depends on this comparison.

1. In the first case, the envelope's peak is less than the desired delay. An example of this situation is depicted in the top right panel of Figure 5.8 where the desired delay is indicated as a vertical line. Here, the output of the filter is delayed so that the envelope's peak is aligned with the desired delay. Then the filter's output is multiplied by a complex value such that the real part of the filter's impulse response has a local maximum at the envelope peak. The result can be seen in the bottom right panel of Figure 5.8. This case generally occurs for filters with high center frequencies because their group delays are shorter.

2. In the second case, the envelope's peak is greater than the desired delay. An example of this situation is depicted in the top left panel of Figure 5.8. Here, the filter's impulse response envelope is not altered, but its temporal fine structure is multiplied with a complex constant so that it is maximized at the desired delay time, i.e. the real part of the impulse response is equal to its envelope at that time. The result can be seen in the bottom right panel of Figure 5.8. This case generally occurs for filters with low center frequencies because their group delays are longer.

To align the fine structure of the impulse response, the complex filter output is multiplied by a complex constant as described by Hohmann as

$$\tilde{y}'_k(n) = \tilde{b}_k \cdot \tilde{y}_k(n) \tag{5.17}$$

$$\text{with} \quad \tilde{b}_k = e^{-j2\pi \cdot f_{c,k} \cdot t_k} \tag{5.18}$$

where k is the channel index, f_c is the channel's center frequency, and t_k is the desired time at which the fine structure of the impulse response should be maximized. After temporal alignment, the real part of the filterbank's channels are linearly combined with a weighted sum,

$$y(n) = \sum_{k=1}^{52} g_k \cdot y'_k(n), \tag{5.19}$$

where the g_ks are the channel weights. The channels must be weighted to account for imperfect superposition of the channel's impulse response, which results in a non-flat

[5]In the example given by Hohmann, the desired delay is 4 msec. Shi *et al.* (2006) do not specify the temporal delay chosen, only that the filterbank of (Hohmann, 2002) was employed; therefore a temporal delay of 4 msec is chosen in the implementation used in this thesis.

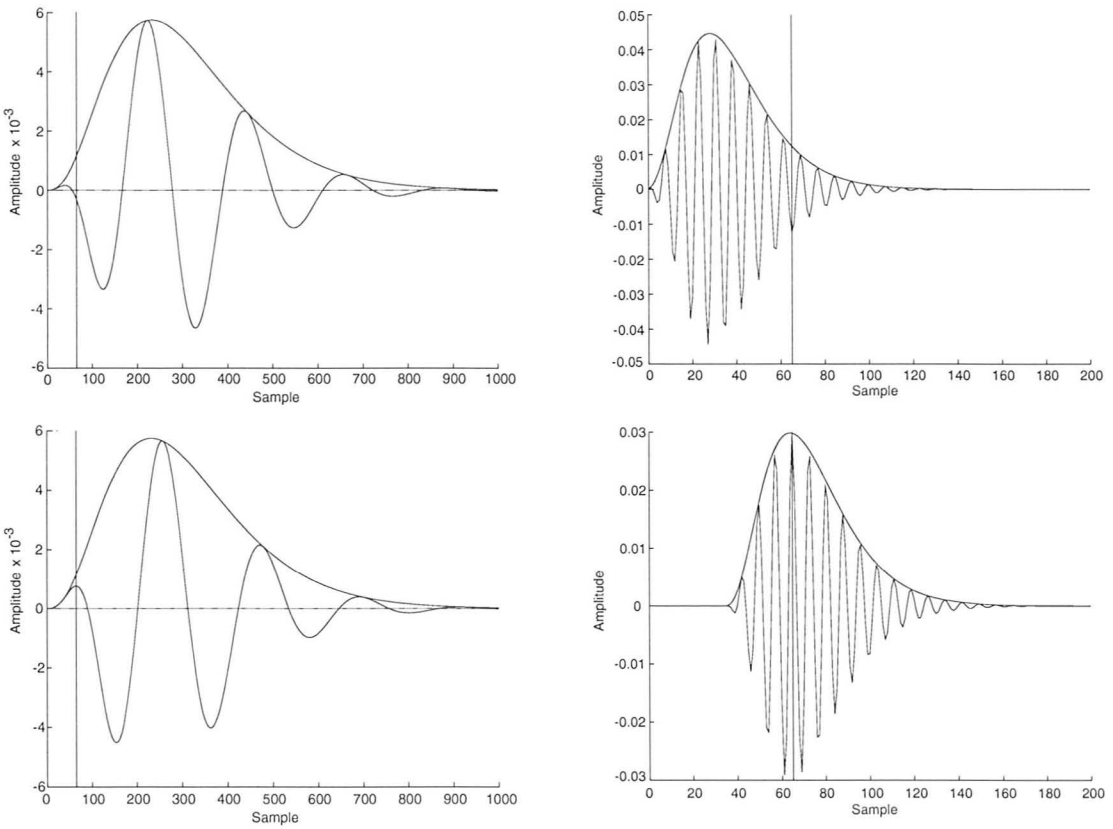

Figure 5.8: Procedure for aligning filterbanks at the desired delay (vertical line). *Top Left*: Impulse response of low frequency channel where the peak of the envelope is later than the desired delay. In this case, the phase of the impulse response's fine temporal structure is phase shifted to be maximized to the envelope at the desired delay point (shown in Bottom Left). *Top Right*: Impulse response of a high frequency channel whose envelope peak is earlier than the desired delay. In this case the impulse response is delayed so the envelope peak is aligned with the desired delay and phase shifted so its fine temporal structure is maximized at the peak (shown in Bottom Right). (Reprinted from Figures 4 and 5 of Hohmann, 2002)

frequency response of the system. Hohmann suggests that these weights be optimized numerically but does not report the optimization method. The implementation of the gammatone filterbank used in this thesis simply measures the gain of the filterbank's impulse response at the center frequencies of the constituent channels with no weighting applied, then sets the weights according to Equations 5.20 and 5.21:

$$g_k = 1/Y(f_c) \tag{5.20}$$

$$\text{with} \quad Y = |\mathscr{F}\{\sum_{k=1}^{52} 1 \cdot y'_k(n)\}| \tag{5.21}$$

where $\mathscr{F}$ represents the Fourier transform and the y_ks are impulse responses. The results of linear combination of filterbank channels are depicted in Figure 5.9, where the filterbank's gain-frequency response is shown before and after weighting is applied. Application of the weights has brought the gain-frequency response close to flat, but there are still some deviations from unity at low and high frequencies.

In addition to this, the insertion of delays by SPC has been found to confound the superposition of filterbank channels, resulting in a non-flat frequency response. This may require dynamic re-weighting of filterbank channels, as is discussed in Section 8.4.1.

Insertion of Delays

The time varying 'insertion delays' prescribed by the modelling path in the SPC are inserted into each channel in the filterbank using a pure delay FIR filter. This filter consists of a discrete impulse with unity gain whose position varies depending on the amount of delay that is required, the FIR filter is shown in Figure 5.10. The initial position of the impulse is determined by the desired group delay of the gammatone filterbank and time varying delays are implemented by shifting the position of the impulse. Only integer delays or advances can be implemented, so the delays prescribed by the modelling path are rounded to the nearest sample period. Delays and advances may be realized by shifting the impulse in the directions indicated by Figure 5.10. The delay FIR is updated for every sample period, $T_s = 1/F_s$. In this way, time-varying delays are introduced into each channel.

Latency in the Modelling Path and the Processing Path

It has been determined during this thesis, that insertion delays prescribed at a particular point in time in the modelling path of the SPC scheme must be applied to each channel in the processing path at the corresponding time point. In order to

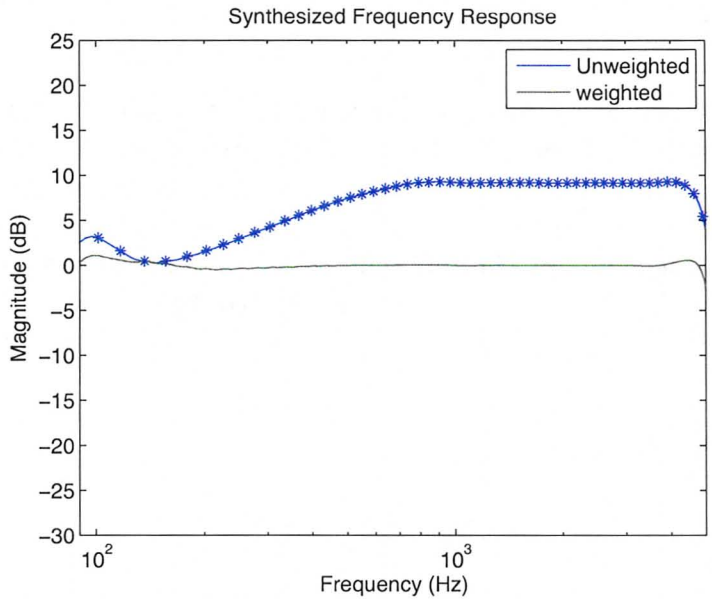

Figure 5.9: Gain-frequency response of the Hohmann analysis-synthesis gammatone filterbank. The blue trace shows the simple summation across aligned frequency channels, and the green trace shows the weighted sum as in Equation 5.21. Markers indicate the position of filterbank center frequencies.

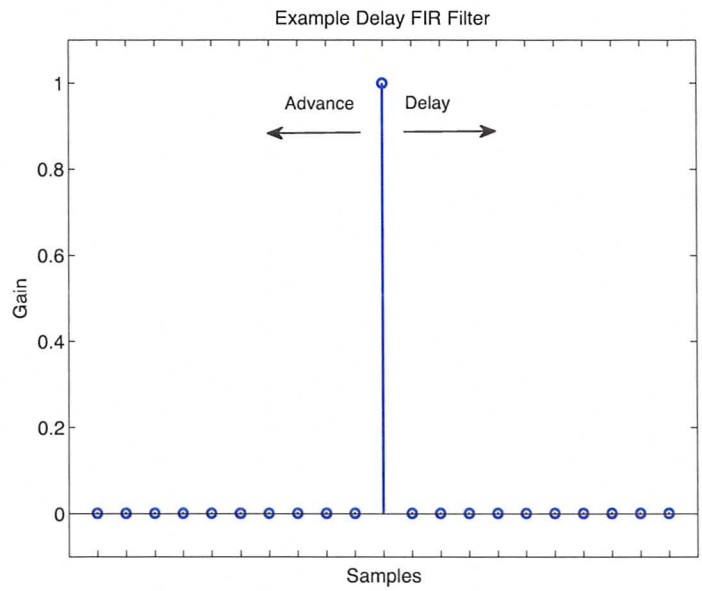

Figure 5.10: Simple delay FIR filter which can implement an advance or delay relative to the overall delay by shifting the position of the impulse.

ensure this, the latency through the middle ear filter in the modelling path, and the latency through the gammatone filter in the processing path is quantified. Figure 5.11 shows a block diagram of the modelling path for one auditory nerve fiber on top of the processing path for one gammatone filter in the filterbank. It is clear that these two paths are different, and therefore have different processing latencies associated with them. Shi *et al.* (2006) did not report on accounting for these differences in the original SPC scheme. Therefore, another version of the SPC scheme has been developed that does account for the different latencies through the modelling path and the processing path of SPC. This version is denoted 'SPC-1', and 'SPC-0' is used to denote the version of SPC that does not account for the different path latencies.

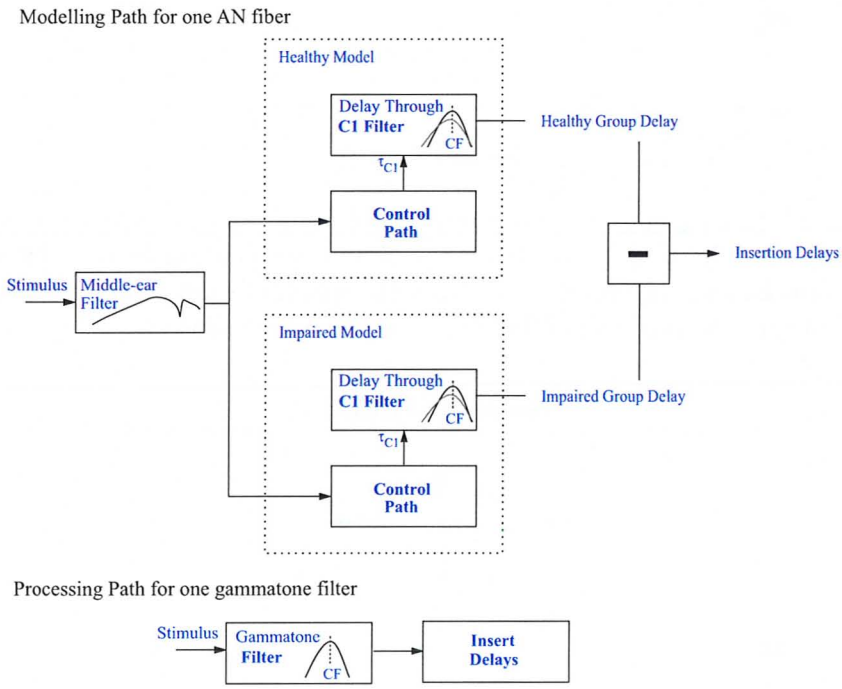

Figure 5.11: *Top*: Block diagram of the modelling pathway in the SPC scheme. SPC insertion delays are derived from the difference between the group delay through the C1 filter in the healthy model and the corresponding impaired model. *Bottom*: Block diagram of the processing path for a single filter in the SPC filterbank.

A filter's processing latencies can be described in two ways: one is with the filter's group delay, which can be thought of as the delay experienced by the low-frequency envelope of a complex signal as it passes through the filter; the other is with the filter's phase delay, which is the delay that a narrow-band signal will experience due

to the phase response of the transfer function associated with the filter. Both of these latencies are accounted for when applying the time-varying SPC insertion delays. Figure 5.12 displays the latencies that are associated with each path as a function of frequency: the top panel shows the phase delays and group delays through the middle ear filter in the processing path, and the bottom panel shows the phase and group delays through the gammatone filter in the processing path.

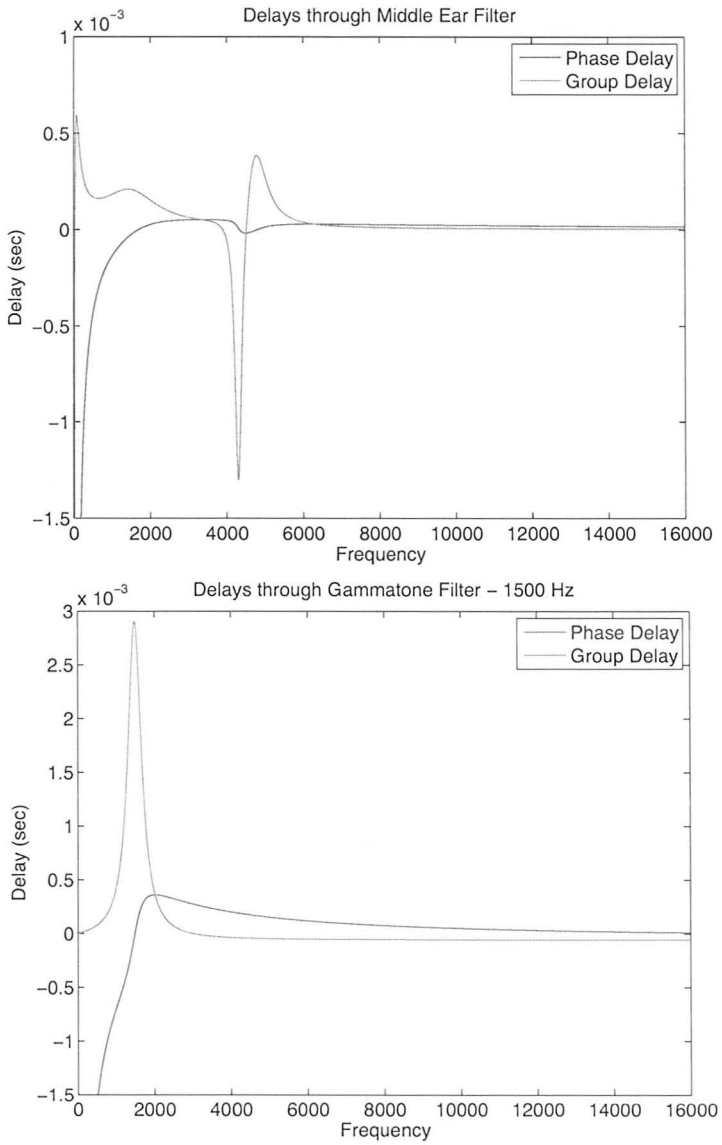

Figure 5.12: *Top Panel*: Processing latencies through the middle ear filter. *Bottom Panel*: Processing latencies through a gammatone filter with $f_c = 1500$ Hz.

The modelling path in Figure 5.11 demonstrates that the stimulus passes through the middle ear filter before presentation to the C1 filters in the healthy and impaired models. The prescribed SPC insertion delays are given as the difference between the group delays through these C1 filters and are given for every time instant. This describes the time-varying insertion delay, which is then applied to the processing path after the stimulus has passed through a gammatone filter. It easy to determine that the processing latencies through the middle ear filter and each gammatone filter in the processing path are different, and should be accounted for. It is somewhat harder to determine whether or not the latency through the control path filters in Figure 5.11 should be accounted for as well. The current study does not account for the latency through the model control path for three reasons. The first is because the latency through the control path is time-varying, which makes its value difficult to quantify. The second is because the latency through the control paths may be different for the healthy and impaired model. Since the SPC insertion delays come from the difference in group delays through the C1 filter in the healthy and impaired model, it is unclear whether to account for the latency through the healthy model or through the impaired model. The third reason is that the insertion delays at a given time-point are defined by the C1 filter group delays for the middle ear output as the corresponding time-point is passing through the C1 filter. This means that the information about the latencies through the control path may already be incorporated in the difference between the healthy and impaired C1 filter group delays.

The difference in path latencies are then accounted for by finding the *group delay* and *phase delay* through each of the gammatone filters in the SPC processing path and the middle ear filter of the modelling path. The phase delay gives the latency due to the phase shift in units of time, and is by definition:

$$\tau_\phi = -\frac{\phi(\omega)}{\omega}. \tag{5.22}$$

First, the delay required to account for the difference in *group delays* between the two paths is found at each auditory filter center frequency, or CF. Then the delay required to account for the remaining difference in *phase delay* between the two paths is found. This phase delay takes into account the phase shift that is applied to each channel in the gammatone filterbank in order to align their impulse responses (see Section 5.2.2).

The group delay, τ_{grd}, through each gammatone filter in the SPC filterbank at the channel's characteristic frequency, ω_{CF}, is found as follows:

$$\tau_{grd,\mathrm{CF}} = -4 \times \frac{(\lambda_{\mathrm{CF}}^2 - \lambda_{\mathrm{CF}} \cos(\omega_{\mathrm{CF}} - \beta_{\mathrm{CF}}))}{1 + \lambda_{\mathrm{CF}}^2 - 2\lambda_{\mathrm{CF}} \cos(\omega_{\mathrm{CF}} - \beta_{\mathrm{CF}})} \tag{5.23}$$

where $\tilde{a}_{\mathrm{CF}} = \lambda_{\mathrm{CF}} e^{j\beta_{\mathrm{CF}}}$ are the difference equation coefficients for the gammatone

filterbank found in Equation 5.10. Equation 5.23 is a simple application of the formula for group delay of a digital filter presented by Oppenheim and Schafer (1999, Equation 5.56), where the factor of four comes from the fact that each gammatone filter is fourth order.

The phase shift $\phi_{\gamma t}(\omega_{\mathrm{CF}})$ through the gammatone filters are found as follows:

$$\phi_{\gamma t}(\omega_{\mathrm{CF}}) \stackrel{\mathrm{def}}{=} \arg\{H_{\gamma t}(j\omega_{\mathrm{CF}})\} = -4 \times \angle(1 - \tilde{a}_{\mathrm{CF}} e^{(-j\omega_{\mathrm{CF}})}) \tag{5.24}$$

where $H_{\gamma t}(j\omega_{\mathrm{CF}})$ is the gammatone transfer function evaluated at the characteristic frequency and $\tilde{a}$ is the difference equation coefficient found in Equation 5.10. This phase shift is used to determine the phase delay of a narrowband signal through the gammatone filter.

The group delay and phase delay through the middle ear filter are then calculated in a similar manner, and evaluated at each CF in the SPC processing filterbank. The insertion delays at each filterbank channel are then time-shifted to account for the difference in group delays first, then the difference in phase delays, between the middle ear filter in the modelling path and the gammatone filters in the processing path. A schematic of this procedure is shown in Figure 5.13. This method attempts to ensure insertion delays prescribed in the modelling path are applied to the signal in such a way that the phase relationship between the signal and the insertion delays is maintained.

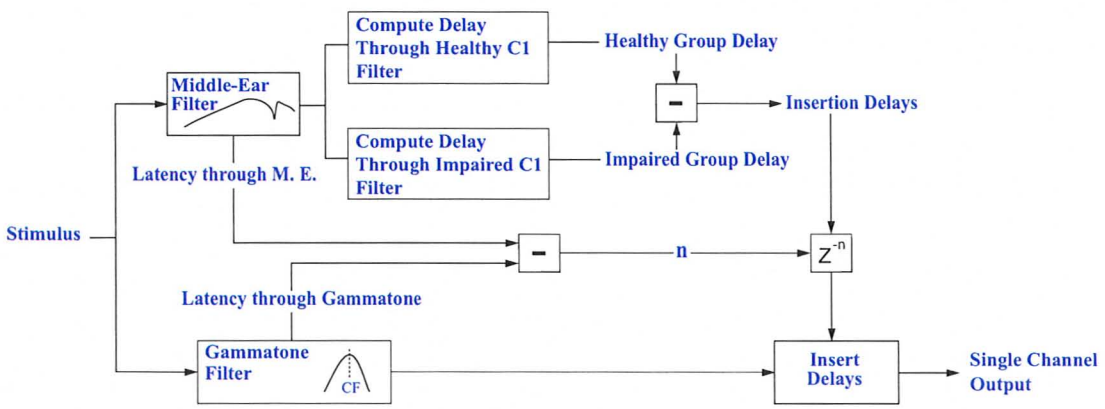

Figure 5.13: Block diagram of a single channel in the SPC-1 processing scheme. The modelling path is on the top and the processing path is on the bottom. This figure illustrates how the insertion delays are themselves delayed to ensure proper temporal alignment when they are applied to the signal path.

5.3 Low-Pass filtering of SPC Insertion Delays

The time-varying insertion delays prescribed by SPC are updated every sample, based on a sampling frequency of 32 kHz. This results in rapid temporal fluctuations in the SPC insertion delays; when applied to each channel this results in high-frequency modulations, which will be discussed in Section 8.4.4. These modulations produce noise in the processed signal that is audible to normal hearing listeners. In order to reduce this effect, a third version of SPC is evaluated that lowpass filters the SPC-1 prescribed insertion delays with a zero-phase, 500 Hz cutoff lowpass filter. This version is denoted SPC-2 and it is also evaluated in Chapter 7.

5.4 Combining MICEFS and SPC

The combination of MICEFS and SPC is a simple serial combination where the signal is first processed with MICEFS and then presented to the SPC processing block. It is important that the signal be processed with MICEFS first so that synchrony to the formants of voiced speech can be at least partially restored. This will ensure that the insertion delays prescribed by SPC will be based on an improved representation of synchrony in the impaired ear.

In Chapter 7, both versions of MICEFS (MICEFS-0 and MICEFS-1) are evaluated as standalone processors. As well, all three versions of SPC (SPC-0, SPC-1 and SPC-2) are evaluated by themselves. The processors with the best performance are then used to evaluate the combination of MICEFS with SPC. The methods of evaluation will be outlined next.

Chapter 6

Methods ii. Representing Auditory Nerve Response

This chapter describes several metrics used to qualify and quantify the differences between the auditory nerve response in a healthy ear and that in an impaired ear. These metrics can therefore be used to evaluate the benefit that a hearing aid may have in restoring the auditory nerve response in an impaired ear. First, a number of metrics that have been employed previously by other researchers are explained. These are useful metrics for representing auditory nerve response in general, and their relevance to this study is described. Then a novel error metric is developed. It uses auditory nerve fiber discharge cross-correlations to analyze their relative phase in a manner that does not, in principle, require specific knowledge about the stimulus.

6.1 Neurograms

The auditory model described in Chapter 3 is used to represent the neural response of a population of auditory nerve fibers across a range of characteristic frequencies. This allows for visualization of the spatiotemporal response pattern, or neurogram. The neurogram is similar to a spectrogram in that it shows activity across frequency as a function of time; it differs in its representation of frequency: where the spectrogram shows the energy across frequency bins in the stimulus, the neurogram shows the neural discharge rate across fibers that are ordered as a function of their characteristic frequency. The neurogram depicts the neural activity over time for a range of fibers, as such it can be used to depict either the model's synaptic output or the AN spike times. Example neurograms in response to a synthesized vowel are shown in Figure 6.1; the neurogram in the middle panel is based on the model's synaptic output and the neurogram in the lower panel is based on the spike times.

Neurograms can be used as a tool for visualizing the auditory nerve response to

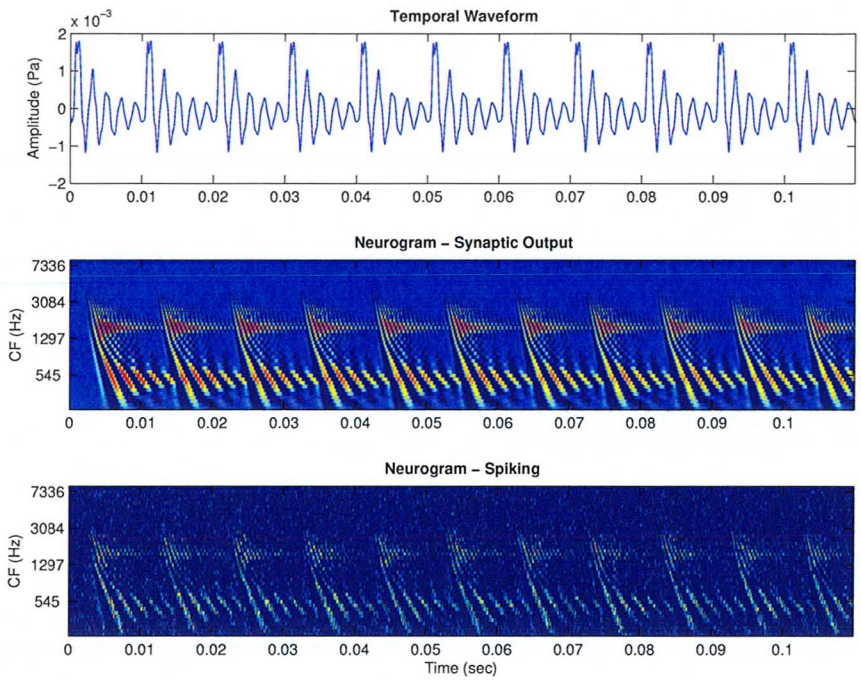

Figure 6.1: *Top*: Temporal waveform of a synthesized vowel. *Middle*: Healthy model neurogram based on the synaptic output of the Zilany and Bruce model. Probability of spikes in units of spikes/sec is shown where blue indicates a low spiking rate and red indicates a high spiking rate. *Bottom*: Healthy model neurogram based on the spike times.

a given stimulus, and as such, they can aid in qualifying the effects of impairment on neural activity and the effectiveness of a hearing aid in restoring neural activity. However, their usefulness is limited in the subjective nature of their analysis. It is hard to quantify the effectiveness of a signal processing scheme through visual comparison of a healthy neurogram and an impaired-aided neurogram[1]. For this reason it is important to identify and develop metrics that can quantify the error between two neurograms.

A simple comparison between a healthy and impaired auditory nerve response is an absolute difference between their neurograms:

$$\mathrm{E}_{\mathrm{abs}} = \sum |N_{\mathrm{Healthy}}(\mathrm{CF}, t) - N_{\mathrm{Impaired}}(\mathrm{CF}, t)|, \tag{6.1}$$

[1]i.e. an impaired neurogram generated in response to stimulus that has been 'aided' with a signal processing algorithm.

where N stands for the neurogram, CF is the fiber characteristic frequency and t is time. This provides a basic measure of the abnormal neural activity in an impaired ear.

It is important that the neural error metric be sensitive to the spike rate as well as the timing of spikes; the absolute difference given in Equation 6.1 is sensitive to both. However, this metric is susceptible to producing a large error if the two neurograms are not precisely time aligned, that is, if corresponding temporal fluctuations of the spiking rate in the two neurograms occur at slightly different times. Even in the absence of this problem, it is unlikely that a hearing aid will be able to restore the impaired neurogram completely, such that the absolute difference would go to zero. However, the hearing aid may be able to restore certain features of the neurogram more so than others. The absolute difference is not preferentially sensitive to any specific features of the spatiotemporal response pattern, so it does not indicate whether or not perceptual correlates of the auditory nerve response will be restored, which could be useful information.

6.2 Synchronized Rate

The synchronized rate is another useful tool for representing auditory nerve response. It gives an indication of the degree to which auditory nerve fibers are responding to various frequencies in the stimulus. The equation for synchronized rate is developed in a paper by Miller *et al.* (1997) and has since been used by a number of researchers to quantify synchronization in the auditory nerve (e.g., Wong *et al.*, 1998; Bruce, 2004). The synchronized rate is basically a Fourier transform of the instantaneous spiking rate as recorded, or modelled, in an auditory nerve fiber. This enables analysis of the frequency components present in the neural response. The formal mathematical description of the synchronized rate used in this thesis for a 26 msec time window $w(n)$ is:

$$R(kf_t) = \frac{\left|\sum_{n=0}^{N-1} w(n)s(n)e^{-j2\pi kn/N}\right|}{\sqrt{N\sum_{n=0}^{N-1} w(n)^2}}, \tag{6.2}$$

where $N = 416$, $s(n)$ is the model synaptic output in spikes/second resampled to 16 kHz, f_t is the frequency spacing of each frequency bin in the DFT, and k indexes the frequency bins.

This equation can be used for auditory nerve fibers across a range of characteristic frequencies and combined in a 'box-plot' to provide a representation of the synchrony in a population of fibers. The box-plot, seen in Figure 6.2, discretizes the synchronized rate calculated for the time window $w(n)$ in each frequency bin and plots a marker of corresponding size. Each auditory nerve fiber evaluated is organized with increasing

characteristic frequency along the abscissa (x-axis) of Figure 6.2 and the frequency bins included in the synchronized rate calculation are represented on the ordinate (y-axis). This representation of auditory nerve response is useful for characterization of phase-locking and identification of upward spread of synchrony. This particular figure is generated from the healthy model synaptic response to a synthesized vowel. A large amount of synchrony to the first three formant frequencies is noticeable because of the clusters of large markers located on the dashed lines, which indicate the formants.

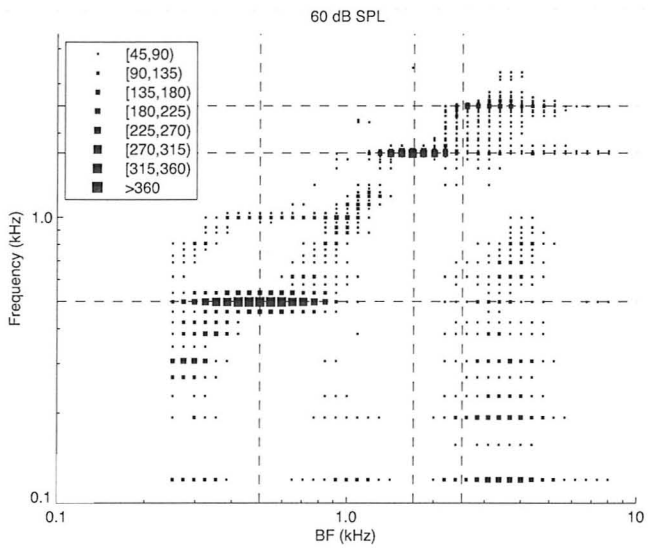

Figure 6.2: Box plot of the synchronized rates in a healthy model neural response to a synthesized vowel /ɛ/ presented at 60 dB SPL. Auditory nerve fibers are represented along the x-axis in order of increasing characteristic frequency. Stimulus frequency components are represented along the y-axis. The first three formants of the stimulus are shown as horizontal dashed lines; vertical dashed lines indicate the formant positions relative to AN fiber characteristic frequency. Marker size is dependent on the value of the synchronized rate (Equation 6.2) at each frequency, a large marker size indicates a high degree of response at that frequency.

A comparison between the box-plots of healthy and impaired auditory nerve response is useful for measuring the loss of synchrony or any upward spread to synchrony that may be present in the impaired response. This also applies for evaluation of a hearing aid's ability to restore synchrony in the impaired response. A simple difference between the two box-plots may assign a quantity to the error, but it cannot predict the degree of restoration of perceptual correlates like formant intelligibility because information regarding the location of fibers contributing to the error is not preserved. For assessment of perceptual restoration, it is likely that visual analysis of the box-plot is more useful.

6.3 Power Ratio

The synchronized rate can be manipulated to calculate a power ratio, which can assess the restoration of important features of the auditory nerve response to vowels. Recall that the relative frequency of vowel formants are thought to be unique to their vowel type (Peterson and Barney, 1952), and preservation of vowel formants is likely required for accurate vowel identification (e.g., Aaltonen, 1985; Kiefte *et al.*, 2010). This means the formants should be represented well in the auditory nerve response. The power ratio calculation, described by Miller *et al.* (1997), can analyze the degree of synchrony to a vowel formant, a phenomenon which could be important for accurate vowel identification.

The power ratio calculation for an arbitrary formant F_x describes the ratio of the power in a single fiber's synchronized rate associated with formant F_x to the total power in the synchronized rate. This allows for assessment of the degree of synchrony capture by a formant, where a power ratio of 1 indicates that the fiber is completely captured by F_x (i.e. synchronized to F_x), and a power ratio of 0 indicates that there is no synchrony to F_x. The power ratio calculation that is used in this thesis appears in Equation 6.3 in a similar manner to the implementation by Bruce (2004):

$$\mathrm{PR}(F_x) = \frac{\sum_{k_x} R^2(k_x)}{\sum_{k=2}^{208} R^2(k)} \tag{6.3}$$

where R is the synchronized rate and k is a vector indexing the frequency bins included in the DFT of the synchronized rate calculation. The variable k_x is a vector of indices corresponding to the frequency bins in the DFT near the formant frequency F_x. If the formant frequency F_x was exactly equal to the frequency of a bin in the DFT, then k_x included the index of that frequency component, and the index of its nearest neighbouring frequency components to each side. If the frequency of F_x was not exactly equal to the frequency of a bin in the DFT, then k_x included the indices of the two nearest frequency components to each side of F_x. In this way the vector k_x contained either three or four frequency components of the DFT, respectively.

The power ratios of a population of AN fibers to the formants of a synthesized vowel are shown in the left panel of Figure 6.3. This visualization of the power ratio is a useful tool for analyzing the amount of auditory nerve response to formant frequencies, F_x, in a vowel. However, it requires knowledge of the formant frequencies present in the stimulus, information that is not always readily available. Bruce (2004) analyzes the power ratio to the first three formants in running speech by making use of the formant tracker of Mustafa and Bruce (2006), which can be computationally expensive.

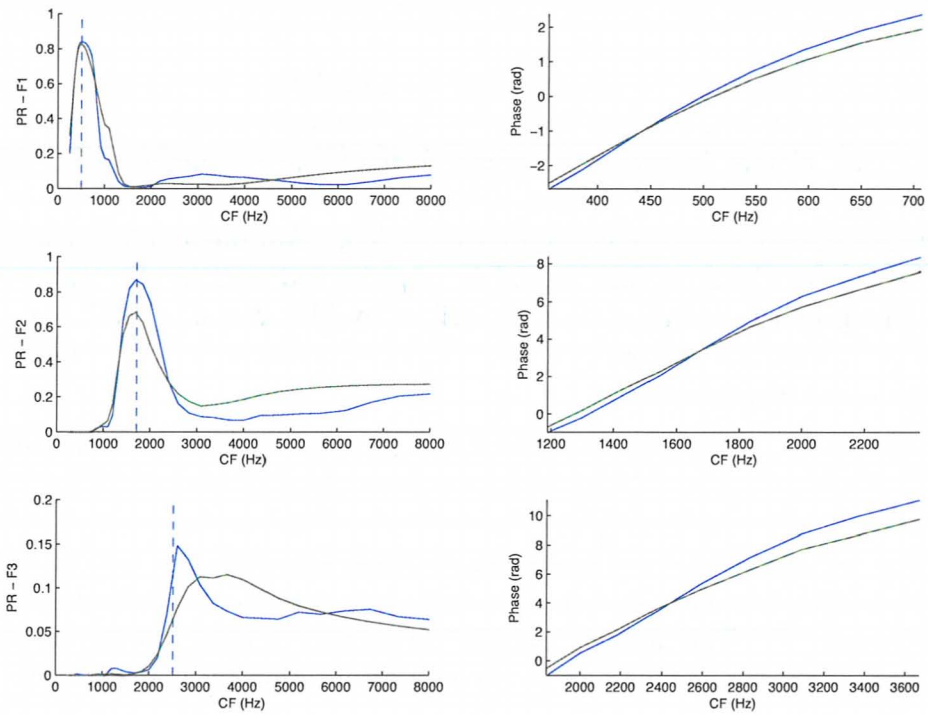

Figure 6.3: *Left Column*: Healthy and mildly impaired power ratios to the first three formants of a synthesized vowel, plotted as a function of fiber CF. Healthy response is in blue, and impaired response is in green. *Right Column*: Phase of the auditory nerve response evaluated at the formant frequencies shown.

6.4 Phase Response

In addition to analyzing the magnitude of auditory nerve fiber response as a function of frequency, it is useful to study its phase, or the timing of its synchronized response. In Section 2.6.3 the importance of the relative phase in the spatiotemporal response as it relates to loudness perception was outlined. Here methods for quantifying the relative phase of auditory nerve fiber response are investigated. One method that was used by Dinath (2008) calculated the relative phase of a population of fibers responding to formant frequencies of interest. The phase for an individual fiber at formant frequency F_x can simply be calculated as the angle of the synchronized rate at that frequency component. For a population of fibers across a range of characteristic frequencies, the angle can be unwrapped to replace phase jumps of π or more with

its 2π counterpart. The relative phase of the first three formants in a synthesized vowel for a population of healthy fibers is compared to the phase of a population of impaired fibers in the right panel of Figure 6.3.

6.5 Cross-Correlation Metric

Another method for representing the phase and magnitude of the spatiotemporal response that is developed here involves computing an estimate of the cross-correlation between AN fibers with neighbouring characteristic frequencies. The real-valued cross-correlation is defined as:

$$R_{xy}[m] = E\{x[n+m]y[n]\} \tag{6.4}$$

and can be estimated as

$$\hat{R}_{xy}[m] = \sum_{n=-\infty}^{\infty} x[n+m]y[n] \tag{6.5}$$

where x and y are the two AN discharge sequences, E is the expected value, R is the true cross-correlation and $\hat{R}$ is an unnormalized estimate of R. The cross-correlation gives a measure of the similarity between two neighbouring auditory nerve fibers and the lag can indicate the relative phase between two fibers. An example development of the cross-correlation between two AN fiber discharge responses is shown in Figure 6.4. Here the top two panels show the individual model fiber response to a synthesized vowel, and the bottom panel shows the positive half of the cross-correlation sequence. This can be computed for neighbouring fiber pairs across a population of fibers for a short time window of the spatiotemporal response. Features of this population cross-correlation can be used to characterize, and compare, the magnitude and phase of the healthy and impaired spatiotemporal response patterns.

The time-lag of the maximum value of the cross-correlation, or peak-lag[2], is related to the relative phase of the two fibers under analysis. If the two fibers are synchronizing to the same frequency component of the stimulus, and there is a large difference in phase between their responses, then the peak-lag will have a relatively large value. Conversely, if the two responses have similar phase, that is their spike times are more coincident, then the peak-lag of the cross-correlation will have a small value. In this way, the cross-correlation can give an indication of the relative phase between two fibers without prior knowledge of the frequency that they are synchronizing to. It is also sensitive to the temporal relationship between two fibers that

[2]The term peak-lag will be used to refer to the lag at which the maximum value of the cross-correlation of two fiber responses is found. Peak-lag can refer to the lag in terms of samples or in terms of time.

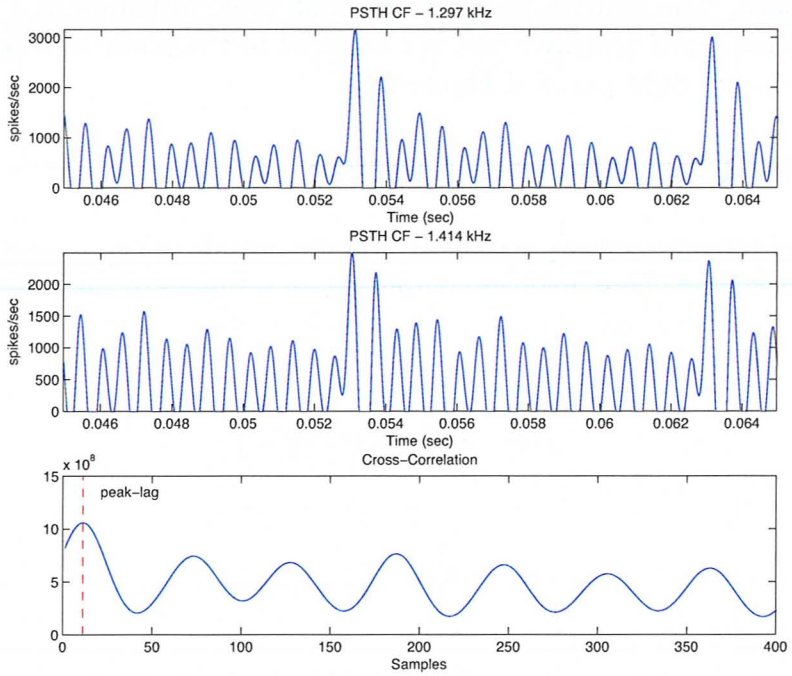

Figure 6.4: The model synaptic response to a synthesized vowel is shown in the top two panels at two AN fibers with similar characteristic frequencies. The positive half of cross-correlation between these two responses is shown in the bottom panel. The lag at which the cross-correlation has a maximum is indicated as the peak-lag.

may be synchronized to different frequency components of the stimulus.

The cross-correlation contains information about the discharge rate in the two fibers, as it is a sliding inner product of the two sequences, which can be useful when comparing healthy and impaired responses. This allows for simultaneous comparison of discharge rates and phase response. Another useful feature when using the cross-correlation for comparison of two spatiotemporal response patterns is that it is insensitive to overall delay mismatches between the two response patterns that may otherwise cause an increase in error. Refer to Section 6.1 to recall why this may a problem.

A number of more specific error metrics may be derived from the neighbouring fiber cross-correlation metric. These include:

1. *Peak-Lag Error.* This error metric is derived from analysis of the position of the neighbouring AN fiber cross-correlation peak-lags. The peak-lags are found for the cross correlation between the model neural response of fibers with adjacent

CFs; this is done across the whole population such that if there were 31 AN fibers represented in the model, there would be 30 calculations of peak lag. The absolute difference is taken between the peak-lags calculated from the healthy neural model and the peak-lags calculated from the impaired neural model. This error metric contains information about the relative phase response of model auditory nerve fibers. It will be discussed in more detail in the next section because it is used extensively in Chapter 7.

2. *Peak-Value Error.* This error metric is the absolute difference between the values of the neighbouring fiber cross-correlation peaks in the healthy model and the impaired model. It contains information about the average discharge rate of AN fiber pairs because it measures the magnitude of their cross-correlation.

3. *Population Cross-Correlation Error.* This error metric is simply an absolute difference between the neighbouring AN fiber cross-correlation calculation for the healthy and impaired neural response for the whole population of AN fibers represented by the model. This error metric contains information about both the average discharge rate of fiber pairs and their relative phase response. However, it may also contain information about the relationship between portions of neighbouring fiber responses that do not correspond to the same time-point in the stimulus. This is because the cross-correlation calculation computes the inner product at all time lags; with a large time lag, an early portion of the neural response of one fiber will be multiplied with a late portion of the other fiber's neural response. This could result in meaningless information because it is not likely that the mammalian auditory system performs cross-correlation on AN fiber discharges with large time lags.

6.5.1 Peak-Lag Error

The peak-lag error metric is described in more detail here as it concentrates on measuring the relative timing of AN fiber discharge patterns, which has been hypothesized to be a neural correlate to the representation of loudness. This metric is used in Chapter 7 to quantify the SPC processing scheme's effectiveness in restoring the relative timing of AN fiber discharge patterns.

Figure 6.5 demonstrates the development of the peak-lag error metric. A stimulus composed of a combination of three pure tones, called a 3-tone complex, is presented to the healthy auditory model and an impaired model with a mild, sloping hearing loss. A portion of the resultant healthy neurogram is shown in the top left panel of Figure 6.5, together with the healthy neighbouring fiber cross-correlation. It is easy to see from the neurogram that each frequency component in the tone complex is represented well by synchrony in the auditory nerve response. That is, there is a

local population of AN fibers that have a periodic response corresponding to each of the tones present in the stimulus. The peak-lag of the AN fiber cross-correlations are found for each adjacent fiber pair and depicted in Figure 6.5 with a blue line that is superimposed on the plot of AN fiber cross-correlations. The peak-lag is then found for both the healthy and impaired cross correlations (impaired cross-correlation not shown), these are compared in the top right panel of Figure 6.5. The absolute difference is found between the healthy and impaired peak-lag and summed across all frequencies, this quantity represents the error in peak-lag.

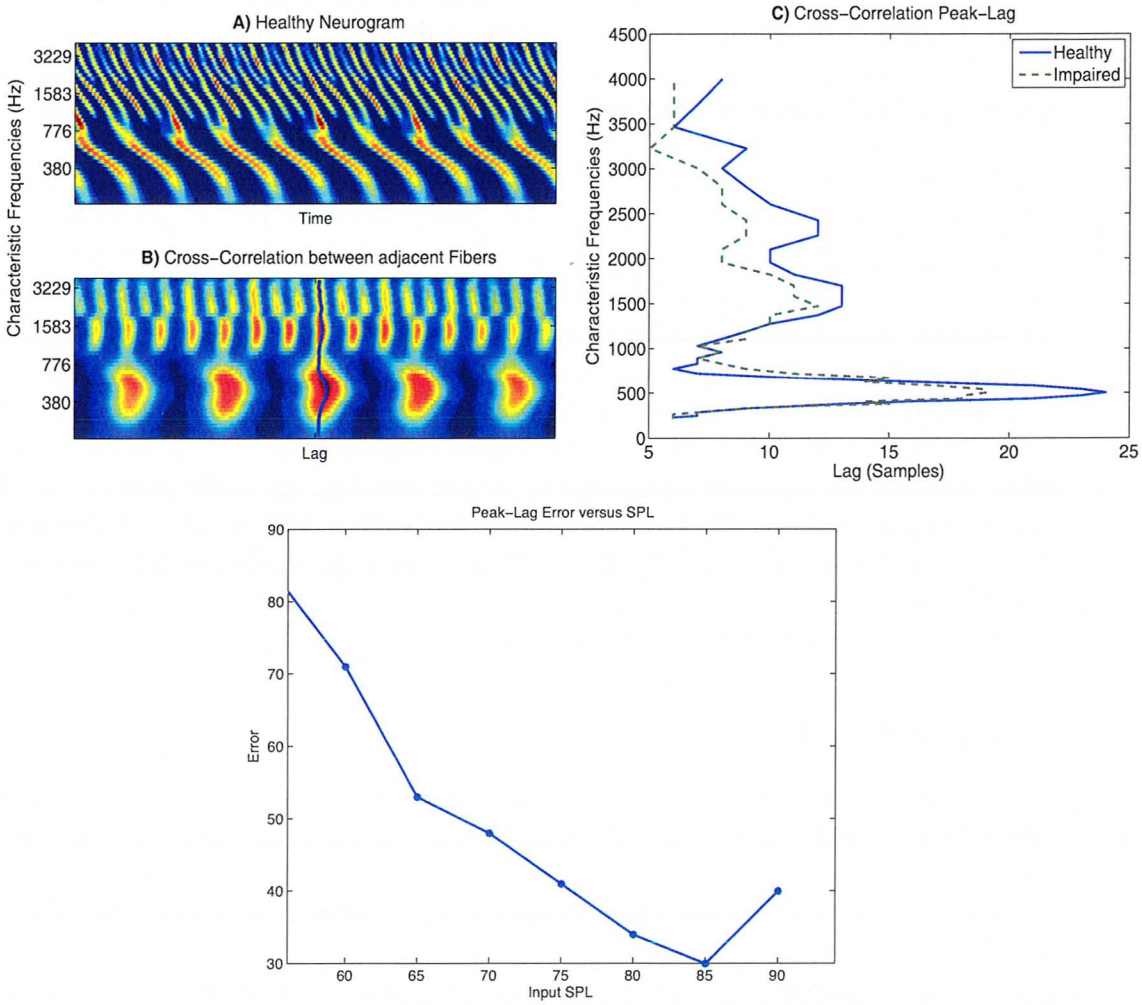

Figure 6.5: Development of peak-lag cross-correlation error metric. *Top Left*: Healthy neurogram and neighbouring fiber cross correlation in response to a 3-tone stimulus. *Top Right*: Peak-lags at each neighbouring fiber pair for the healthy and impaired response to the stimulus at one SPL. *Bottom*: Difference in peak-lag is summed across each neighbouring fiber pair, and plotted as a function of stimulus SPL.

Furthermore, this error can be found for a variety of input sound levels and plotted as a function of SPL; this is shown in the bottom panel of Figure 6.5. Note that the peak-lag error systematically decreases as a function of input sound level, until high SPLs are reached[3]. This systematic decrease can be explained by the flattening of healthy auditory filter phase response as sound level increases (recall Figure 2.16) and the already flattened phase response of the impaired auditory filters. As the sound level increases, the peak-lags of the healthy neural response decrease relatively quickly, approaching the value of peak-lag for the impaired neural response, which decreases slowly due to the already broadened impaired auditory filters.

Issues with peak-lag error

In certain situations, the position of the peak-lag between two AN fibers is significantly different from the peak-lag that would be expected based on the frequency content of their responses. In these cases, the peak-lag is greater than one period of the lowest dominant frequency present in the two responses. An example of this scenario is shown in the left panel of Figure 6.6, where the peak-lag for the highest frequency pair in the healthy response to a synthesized vowel is very different than the peak-lag at other frequencies. This occurs at the lowest stimulus SPL and is highlighted in the figure by a magenta ellipse. The right panel of Figure 6.6 demonstrates that the cause of this peak-lag difference is due to the interactions between the envelope of the neural response and its fine temporal structure. The top panel on the right side of this figure shows a short time course of the healthy model neural response to the stimulating synthesized vowel at the three highest frequencies modelled. The cross-correlations of these neighbouring fibers are calculated and shown in the bottom half of the right panel, along with vertical dashed lines indicating their peak-lags. The peak lag of the highest CF fiber pair, shown in green, is approximately one period of the dominant frequency larger than the second highest CF fiber pair (shown in blue). It is evident that the interactions of the signal envelope with the dominant frequency of the fine structure in the cross correlation calculation determine this difference.

In order to address this issue, fiber-pairs are only included in the peak-lag error calculation if the ratio of the peak-value of the cross-correlation to the total power exceeds a specified threshold. To determine this, the cross-correlation power ratio, $\mathrm{PR}_{\mathrm{xcor}}$, is found as follows:

$$\mathrm{PR}_{\mathrm{xcor}}(i) = \frac{\max(\tilde{R}_{\mathrm{AN}_i\mathrm{AN}_{i+1}})}{\sum(\tilde{R}_{\mathrm{AN}_i\mathrm{AN}_{i+1}})} \times T \tag{6.6}$$

[3]The increase in peak-lag error at high SPLs is likely due to a phase reversal that occurs at some frequencies in the impaired neurogram. An explanation for this could be the C1/C2 transition (refer Sec. 2.4.2 under the subheading 'Characersitics of AN Phase').

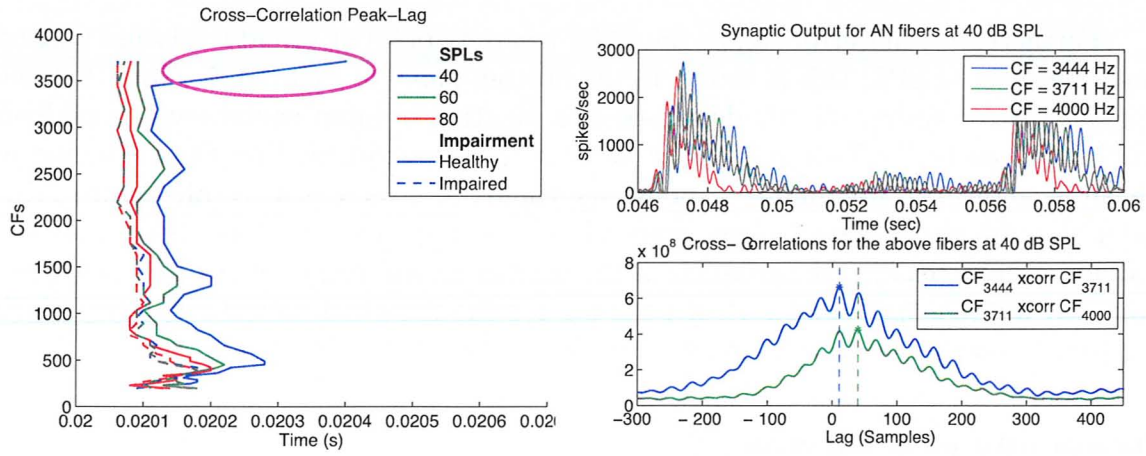

Figure 6.6: Description of potential problems with peak-lag error. *Left Panel*: Peak-lags shown for a healthy and impaired auditory model in response to a synthesized vowel presented at a range of sound pressure levels. Dashed lines indicate impairment and healthy response is shown with solid lines; colour indicates SPL. *Top Right*: AN response to the vowel for three fiber characteristic frequencies. *Bottom Right*: Cross-Correlation of the fiber pairs. Note the difference of approximately one cycle in the cross-correlation's temporal fine structure.

where AN_i is the model neural response at one characteristic frequency, $\tilde{R}$ is the cross-correlation estimate, and $T = 20$ msec is the duration of the neural response included in the cross correlation calculation. The value of $\mathrm{PR}_{\mathrm{xcor}}(i)$ must be greater than the value specified in Table 6.1, to be included in the peak-lag error calculation.

Table 6.1: Threshold requirements for peak-lag error inclusion

Parameter	Threshold value
$\mathrm{PR}_{\mathrm{xcor}}$	1.2×10^{-5}
$\mathrm{PR}_{\mathrm{fiber}}$	0.06

As well, there is a periodic correction done to the peak-lag error calculation if the initial difference between the two peak-lag positions being compared is greater than one period of the dominant frequency component in the healthy response. This means that there is no penalty for shifts in the peak-lag error by integer multiples of the dominant frequency's period, or $k \times 2\pi$. An example of this case appears in Figure 6.6. Here the absolute difference between the healthy and impaired peak-lags is taken to find the peak-lag error, then the modulo operator is implemented with

regards to the dominant frequency component. This ensures integer multiples of 2π are not penalized.

In order to implement the periodic correction, accurate estimation of the dominant frequency component in the model neural response is required. To ensure this, a second power ratio calculation, $\mathrm{PR}_{\mathrm{fiber}}$, is performed with regards to the dominant frequency in the model fiber's neural response. If the fiber's power ratio exceeds the value given in Table 6.1, then the periodic correction is performed, and the fiber is included in the peak-lag error. The left and right panels of Figure 6.7 demonstrate the peak-lag error as a function of characteristic frequency, before and after the corrections are implemented for the situation that has been presented in Figure 6.6.

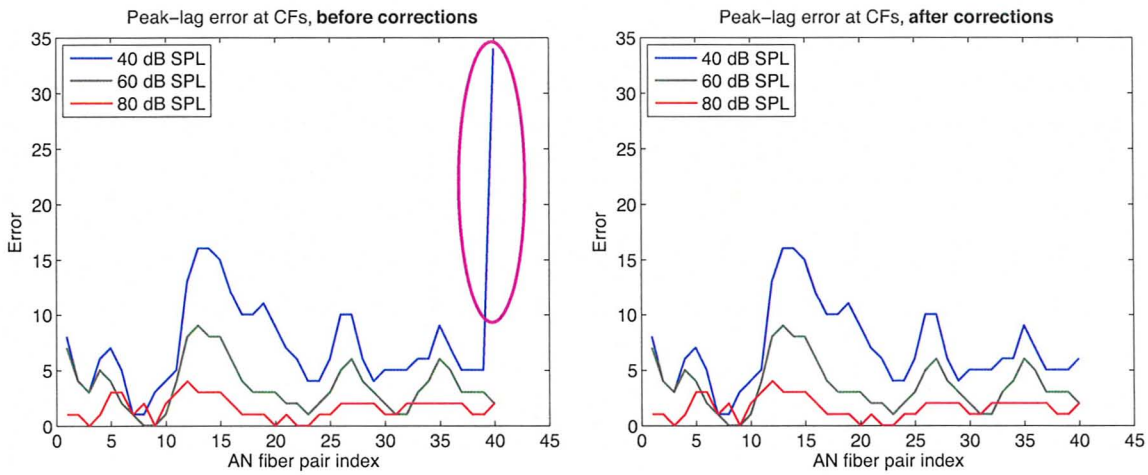

Figure 6.7: *Left*: Peak-lag error corresponding to Figure 6.6 prior to correction. *Right*: Peak-lag error after periodic corrections have been made.

Chapter 7

Evaluation

This chapter evaluates the original MICEFS and SPC processing schemes and compares the results with improvements made to these processing schemes in this thesis. The serial combination of the MICEFS and SPC processing schemes are then evaluated to determine if any benefit can be drawn from restoring AN synchrony before correcting for impaired group delays. The hearing loss types used in this analysis are described first, then each of these loss types are used to evaluate the MICEFS processor, the SPC processor, and the combined MICEFS/SPC processor.

7.1 Evaluation Configuration

7.1.1 Sound Stimuli

The processing schemes are evaluated in response to a synthesized vowel, /ɛ/, with known formants.[1] The first four formants are located at frequencies 500, 1700, 2500, and 3300 Hz, and can be identified as envelope peaks in the magnitude spectrum of the vowel shown in Figure 7.1. A vowel with known formants is used in order to evaluate the synchrony in the auditory nerve response to these formants.

7.1.2 Hearing Loss Types

Two hearing loss profiles, or audiograms, are represented by the auditory model for evaluation with the processing schemes developed in this thesis. One is a mild hearing loss, with threshold shifts ranging from 20 dB to 50 dB. The other is a moderate-to-severe hearing loss with threshold shifts ranging from 25 dB to 90 dB. Both audiograms, which are depicted in Figure 7.2, have a sloping high frequency loss,

[1]The vowel /ɛ/ is taken from the synthesized word 'besh', which has similar pronounciation to the word 'mesh'.

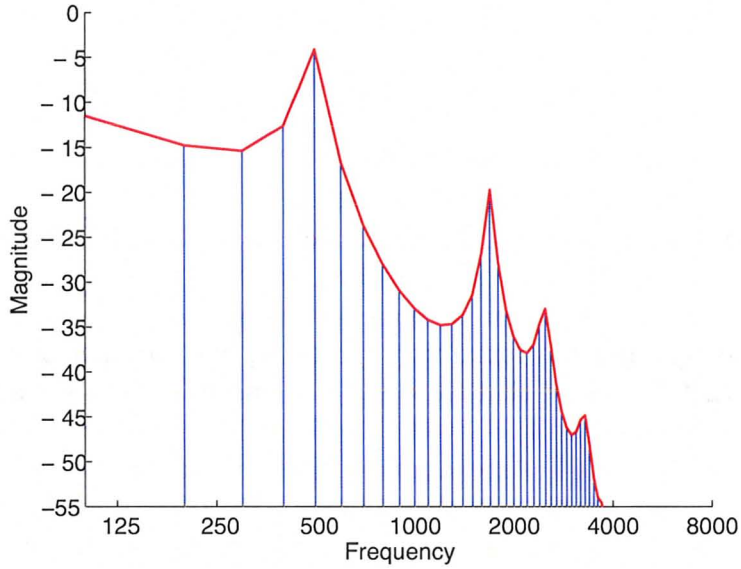

Figure 7.1: Magnitude spectrum of the synthesized vowel /ɛ/ shown in blue. The spectral envelope is plotted in red. Formant peaks are as follows: F1 = 500 Hz, F2 = 1700 Hz, F3 = 2500 Hz, F4 = 3300 Hz.

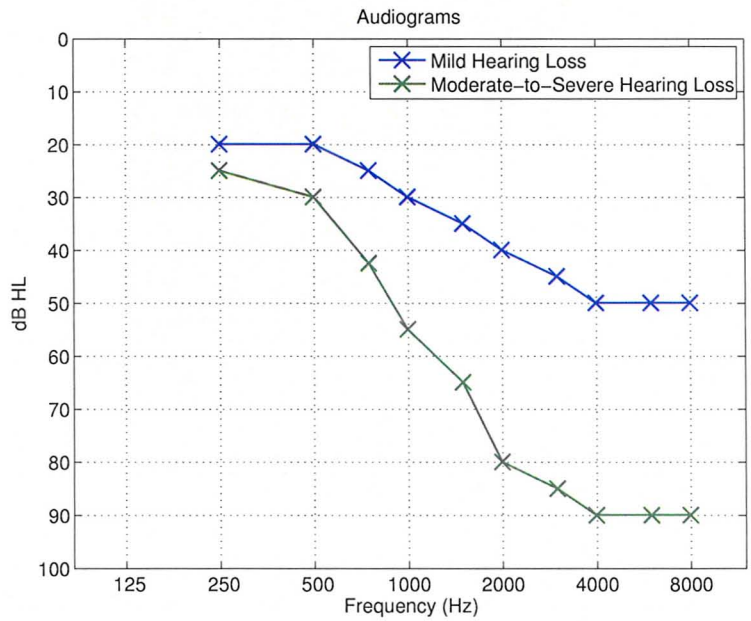

Figure 7.2: Two audiograms used in the evaluation. Loss is given in terms of a negative threshold shift in dB. These audiograms were used by Bruce *et al.* (2007)

meaning the hearing thresholds get worse as frequency increases. The audiograms in Figure 7.2 are plotted with a reversed y-axis, so hearing loss is shown below the x-intercept.

The auditory model of Zilany and Bruce (2006, 2007) represents impairment with model parameters corresponding to either inner or outer hair cell damage, or both. It is therefore a requirement that threshold losses that appear in the audiograms be translated into inner and outer hair cell damage.

Both audiograms are represented by the model with a mixed hair cell loss; that is, the threshold shifts are a result of both inner and outer hair cell damage. The specific mixed loss case evaluated below corresponds to 2/3 of the hearing loss being due to outer hair cell damage, and the remaining 1/3 being due to inner hair cell damage. Zilany and Bruce (2007) suggest that these proportions agree with impairment in cats due to acoustic trauma and with estimated typical OHC damage in humans (Plack *et al.*, 2004; Moore *et al.*, 1999). An example of these proportions are shown in Figure 7.3, where the audiogram of the mild hearing loss is divided into the proportion that results from outer hair cell damage and the proportion that results from inner hair cell damage.

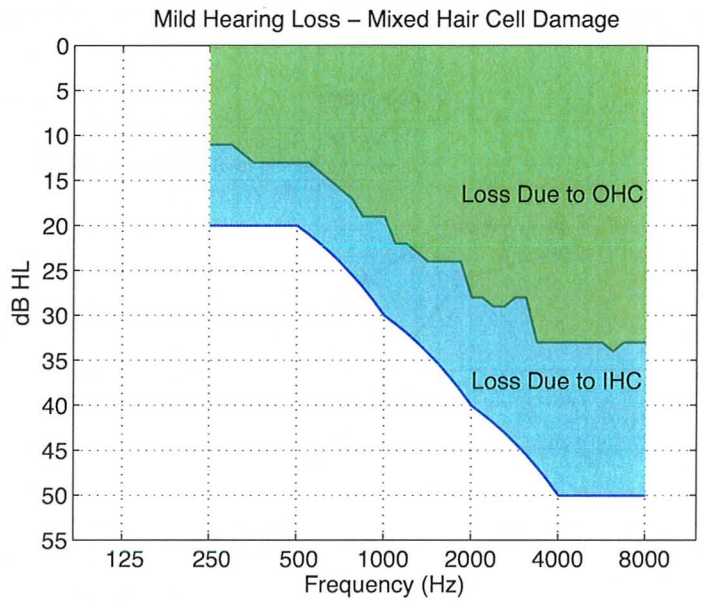

Figure 7.3: The contribution of OHC and IHC damage to the threshold shifts observed in the mild hearing loss (shown as a blue solid line). Threshold shift due to OHC damage is (shown in green) approximately 2/3 of the audiogram.

In addition to evaluating this specific mixed loss case for both the mild audiogram and the moderate-to-severe audiogram, the mild audiogram is also modelled with only outer hair cell damage. In this case, all of the hearing loss in dB is attributable to damage to the outer hair cells, which corresponds to an impairment of the compressive non-linearity. It is only the mild hearing loss that is evaluated with exclusively outer hair cell impairment, because the threshold shifts in the moderate-to-severe audiogram are too great to be attributable to outer hair cell loss alone. Moore *et al.* (1999) suggest that the threshold shift due to exclusively outer hair cell impairment cannot be greater than 65 dB based on physiological estimates of the amplification provided by the healthy compressive non-linearity.

In summary, the hearing loss types evaluated here are:

- Mild Hearing Loss, Mixed Hair Cell damage
- Mild Hearing Loss, Outer Hair Cell damage
- Moderate-to-Severe Hearing Loss, Mixed Hair Cell damage

7.2 Evaluation of MICEFS alone

Two versions of the MICEFS processing scheme are evaluated here, where MICEFS-0 refers to an implementation of the original MICEFS proposed by Ansari (2005) using Filter type I, and MICEFS-1 refers to an improved implementation developed in this thesis and presented in Chapter 5. Since the AN response is evaluated with a vowel with known formants, the outputs of the formant tracker are set directly to those formants to ensure MICEFS uses accurate formant estimates[2].

First the gain prescribed by each MICEFS algorithm in response to the synthesized /ɛ/ vowel is presented. The gain provided by MICEFS is independent of the type of hair cells that are damaged, so only two hearing loss cases are presented here: the mild hearing loss, and the moderate-to-severe hearing loss. The Input/Output level curves are presented in Figure 7.4 where the gain for the mild hearing loss appears in the subfigures in the left column, and the gain for the moderate-to-severe hearing loss appears in the right column. The top row of subfigures corresponds to MICEFS-0, and the bottom row corresponds to MICEFS-1. The dashed lines are provided as a reference, and represent the level for a 1-to-1 input-output, or unity gain, while the solid curves represent how the output level of MICEFS grows as a function of input level. The compression is notable in all cases as the slope of the solid line is shallower than that of the dashed line. The overall gain of MICEFS-0 is less than that of MICEFS-1 in this case.

[2]For description of the performance of MICEFS with noisy formant estimates see Ansari (2005).

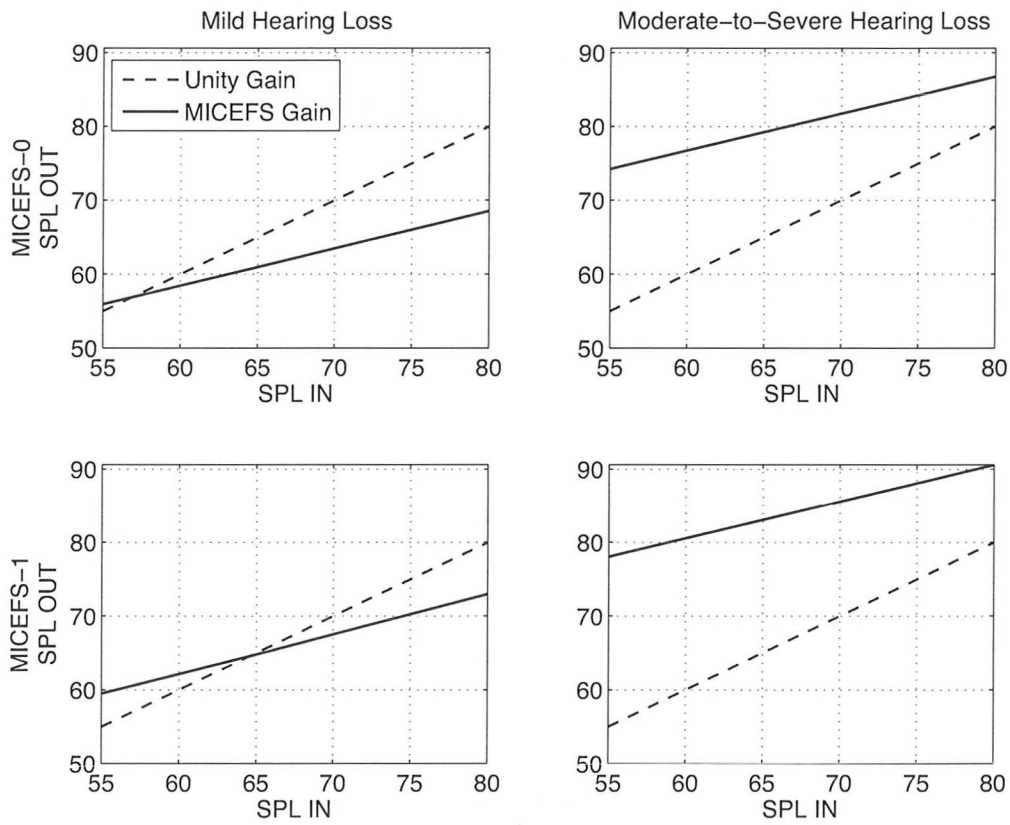

Figure 7.4: Input-Output level curves for two versions of MICEFS implemented in response to a synthesized vowel (solid lines). These are compared to a unity gain, shown with a dashed line. *Top Left Panel*: MICEFS-0 programmed for the mild hearing loss. *Top Right Panel*: MICEFS-0 programmed for the moderate-to-severe hearing loss. *Bottom Left Panel*: MICEFS-1 programmed for the mild hearing loss. *Bottom Right Panel*: MICEFS-1 programmed for the moderate-to-severe hearing loss

The gain versus frequency profile provided by both versions of MICEFS in response to the synthesized /ɛ/ vowel at three input levels are shown in Figure 7.5. Here they are compared to the gains provided by the NAL non-linear Version 1.1 (NAL-NL1) prescription scheme. Both the NAL-NL1 and the MICEFS schemes include multi-band compression, and therefore as the input level increases, the gain decreases. The original MICEFS processing scheme (MICEFS-0) displayed in the top two subfigures provides peaks of amplification at the second and third formant. MICEFS-1 provides some amplification at the first formant, in addition to the second and third formant.

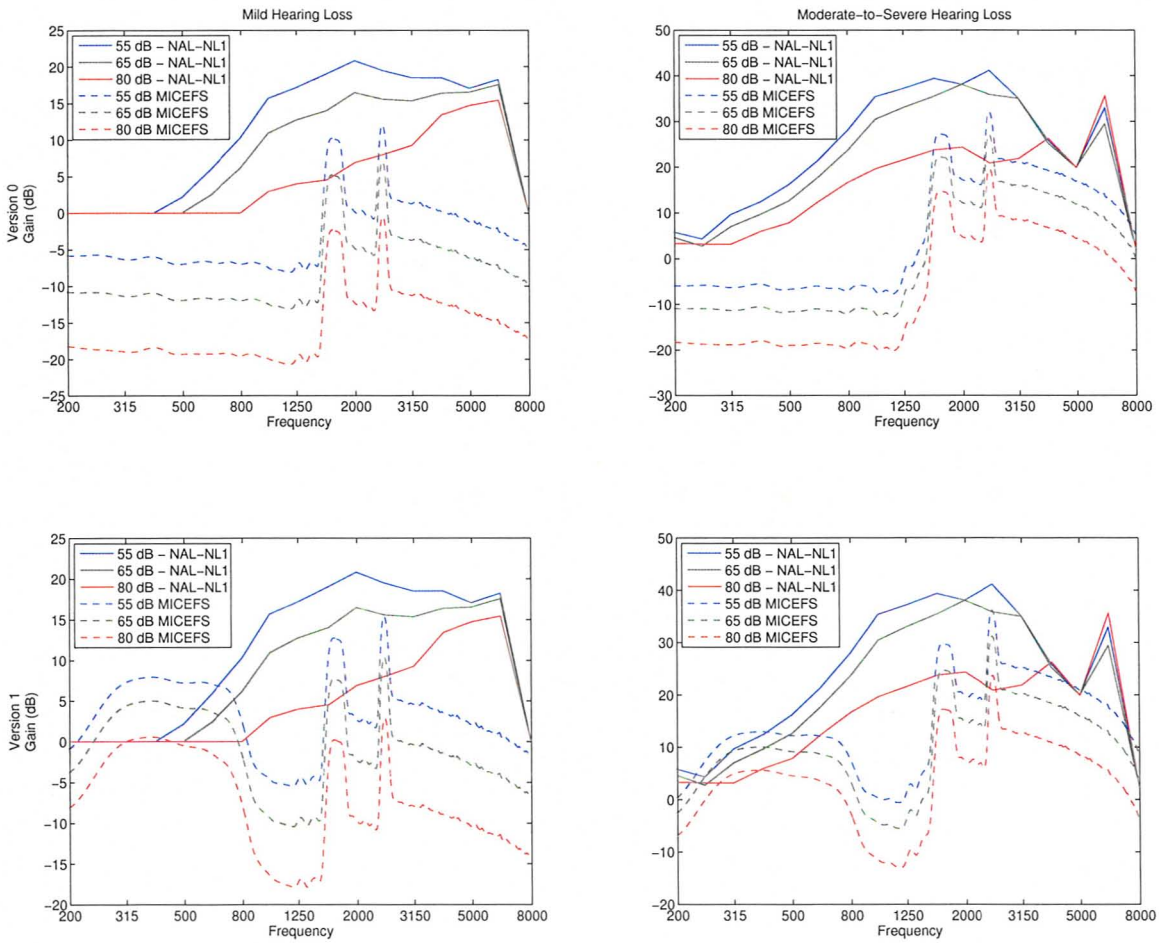

Figure 7.5: Gain versus frequency profiles of two versions of MICEFS (dashed lines) amplifying a synthesized vowel at various input levels. This is compared to the NAL-NL1 gain prescription (solid lines). Subfigure organization is the same as Figure 7.4.

7.2.1 Synchrony

The effectiveness of MICEFS processing of the synthesized vowel will be evaluated with the box-plots that were introduced in Section 6.2. Box-plots demonstrate the localization and degree of synchrony in the response of a population of model auditory nerve fibers. This qualitative metric is useful in determining the robustness of AN response to the vowel's formants. Figure 7.6 shows four box plots of the AN response in a healthy ear model to the synthesized vowel /ɛ/ presented at four sound pressure levels: 50, 60, 70, and 80 dB SPL. The auditory nerve response in the healthy ear demonstrates good synchrony to the first, second and third vowel formants, especially at low sound pressure levels. This is interpreted from the box-plots by the clusters of markers indicating synchrony at the intersection of the horizontal dashed lines and the vertical dashed lines. The horizontal dashed lines represent the formant frequencies in the stimulus and the vertical dashed lines represent the location of the formant frequencies in relation to the AN fiber characteristic frequencies. The box-plots indicate a range of AN fiber characteristic frequencies that are synchronizing to each formant. As the presentation level of the input increases, clusters of synchrony markers spread out horizontally at the stimulus' formant frequencies, indicating a spread of synchrony in the auditory nerve response. As well, at high SPLs, there is a slight upward spread of synchrony to F2 and F3, and the synchrony to F3 is somewhat diminished. This figure represents the synchrony behaviour in a healthy ear and will be used as a reference for evaluating the effectiveness of MICEFS ability to restore formant synchrony. Next, the neural synchrony of the three hearing loss profiles in response to MICEFS will be explored.

Mild Hearing Loss, Mixed Hair Cell Damage

Figure 7.7 shows four box-plots of the AN response of the impaired ear with mild hearing loss, and mixed hair cell damage. The organization of this figure is the same as that in Figure 7.6. It demonstrates that at low stimulus input levels, much of the formant synchrony is diminished for this impairment. As the sound level increases, the synchrony is dominated by F1 and synchrony to F2 is somewhat restored at high levels.

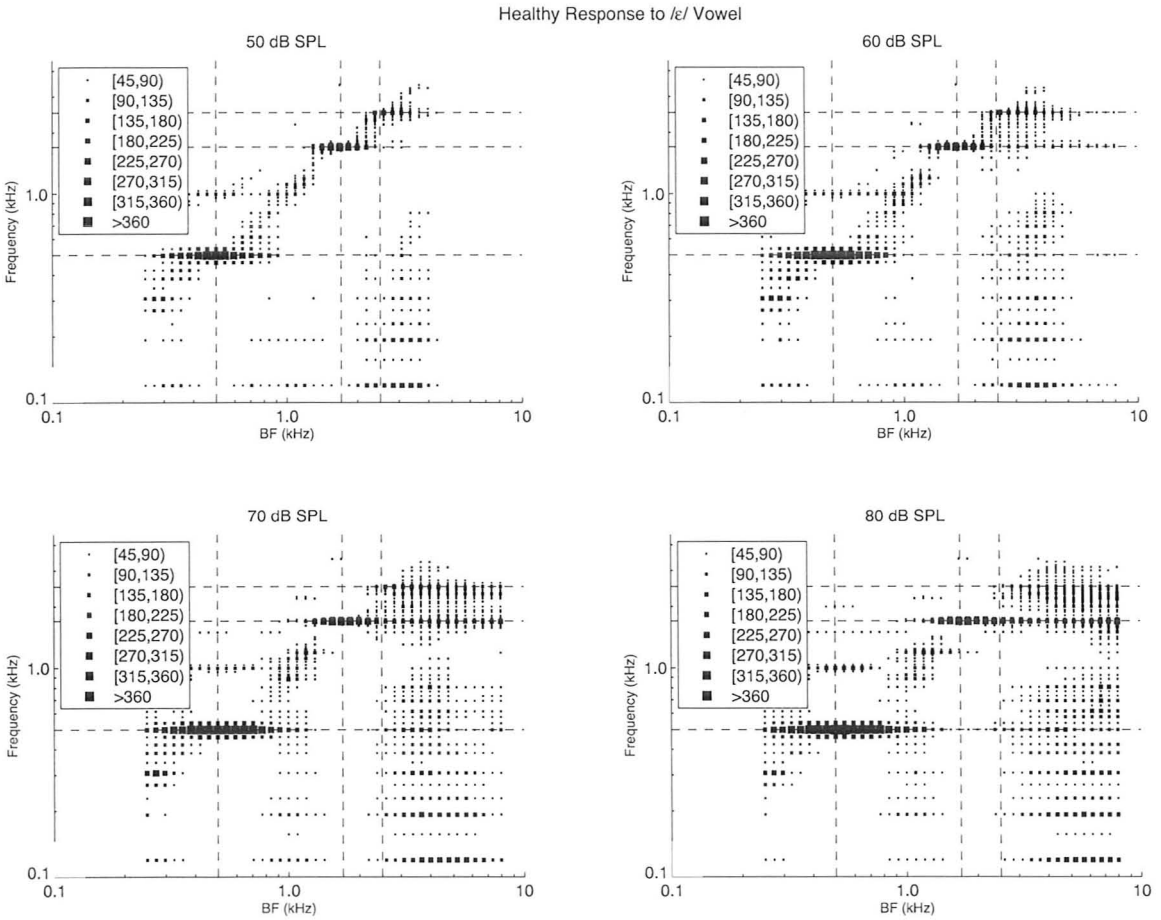

Figure 7.6: Box-plots of the synchronized response of a healthy model auditory nerve in response to the synthesized vowel /ɛ/. Each subfigure represents the response to the stimulus at a different sound pressure level as indicated in their titles. Marker size is dependent on the value of the synchronized rate (Equation 6.2) at each frequency and given as a probability of spikes in units of spikes/sec.

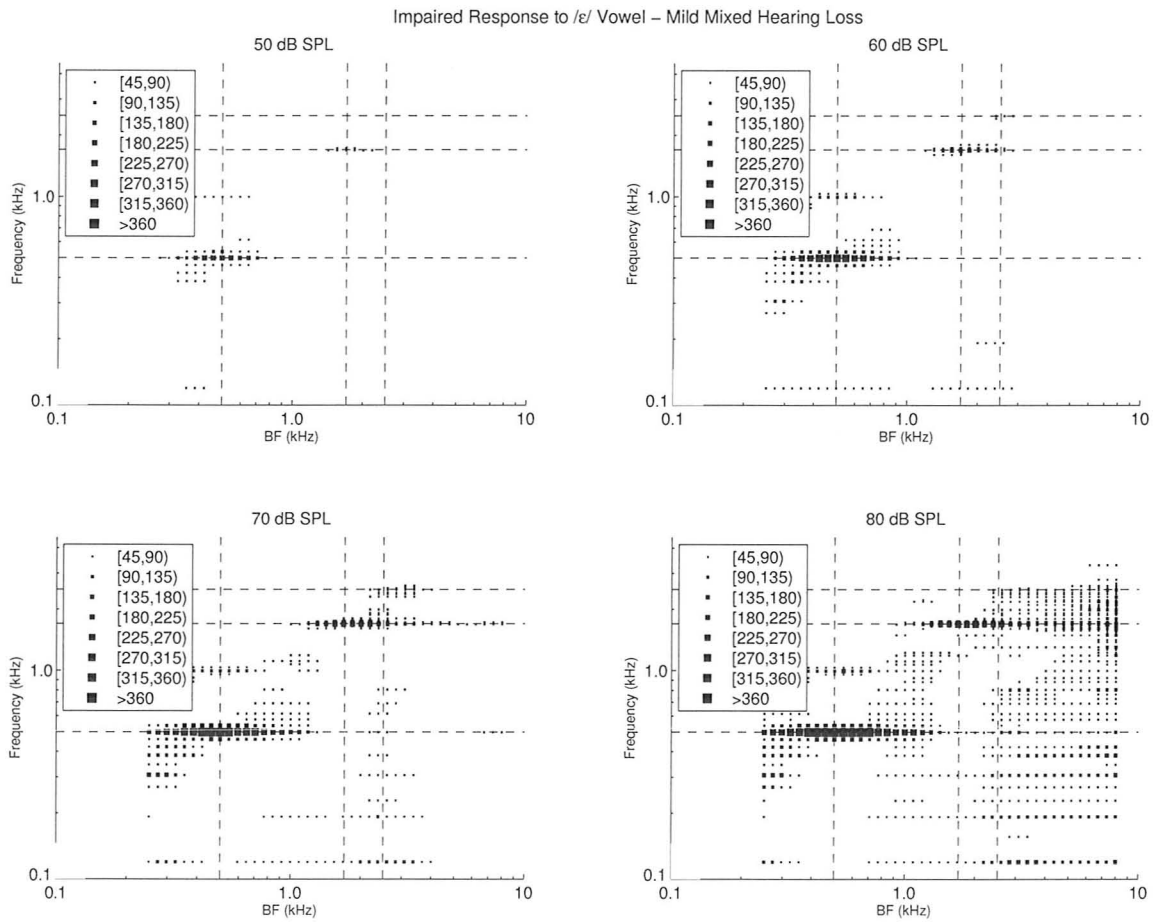

Figure 7.7: Box-plots of the synchronized rate for a mildly impaired AN model with mixed hair cell loss in response to the synthesized vowel /ɛ/.

Figure 7.8 demonstrates the synchrony of the mild, mixed hearing loss to the synthesized vowel after amplification with the MICEFS-0 processing scheme. Here, synchrony to the second formant is restored well for low SPLs, but has some degree of upward spread for high SPLs. Synchrony to the third formant is restored slightly, but is still minimal when compared to the healthy synchrony in Figure 7.6. MICEFS-0 has diminished the synchrony to the first formant when compared to the unaided impaired response of Figure 7.7. This is due to the low frequency cut seen in the MICEFS-0 gains in Figure 7.5, which were developed to target an upward spread of synchrony to F1 in the impaired ear.

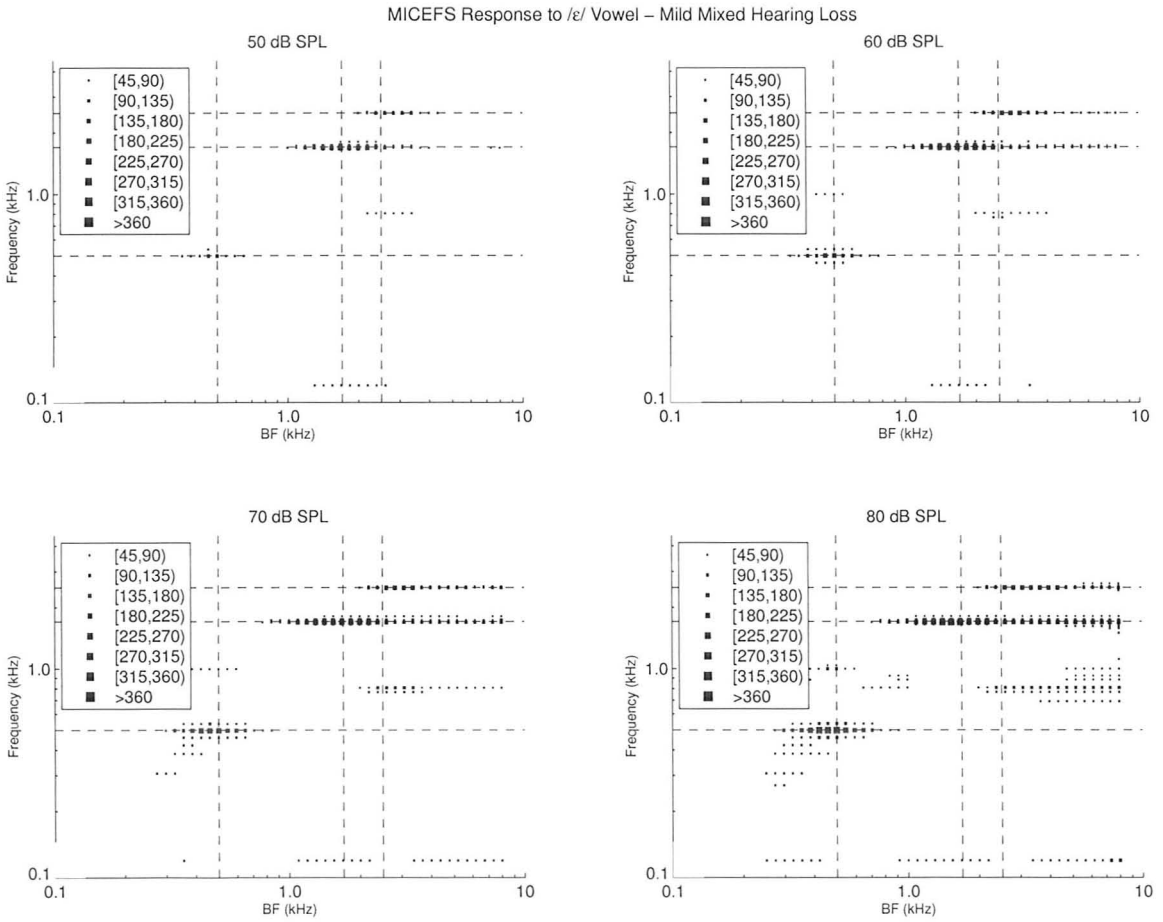

Figure 7.8: Box-plots of the synchronized rate for a mildly impaired AN model with mixed hair cell loss in response to the synthesized vowel /ɛ/, processed with MICEFS-0.

The decreased synchrony to F1 by MICEFS-0 motivates the inclusion of some amplification of F1 in the contrast enhancement scheme of MICEFS-1. This can be

noted in the insertion gains of Figure 7.5. The effects of MICEFS version 1 on the synchrony of the impaired auditory nerve response with mild, mixed hair cell damage can be seen in Figure 7.9. Here, synchrony to the second and third formant are restored just as well as is the case with MICEFS-0, and in addition, synchrony to the first formant is well restored.

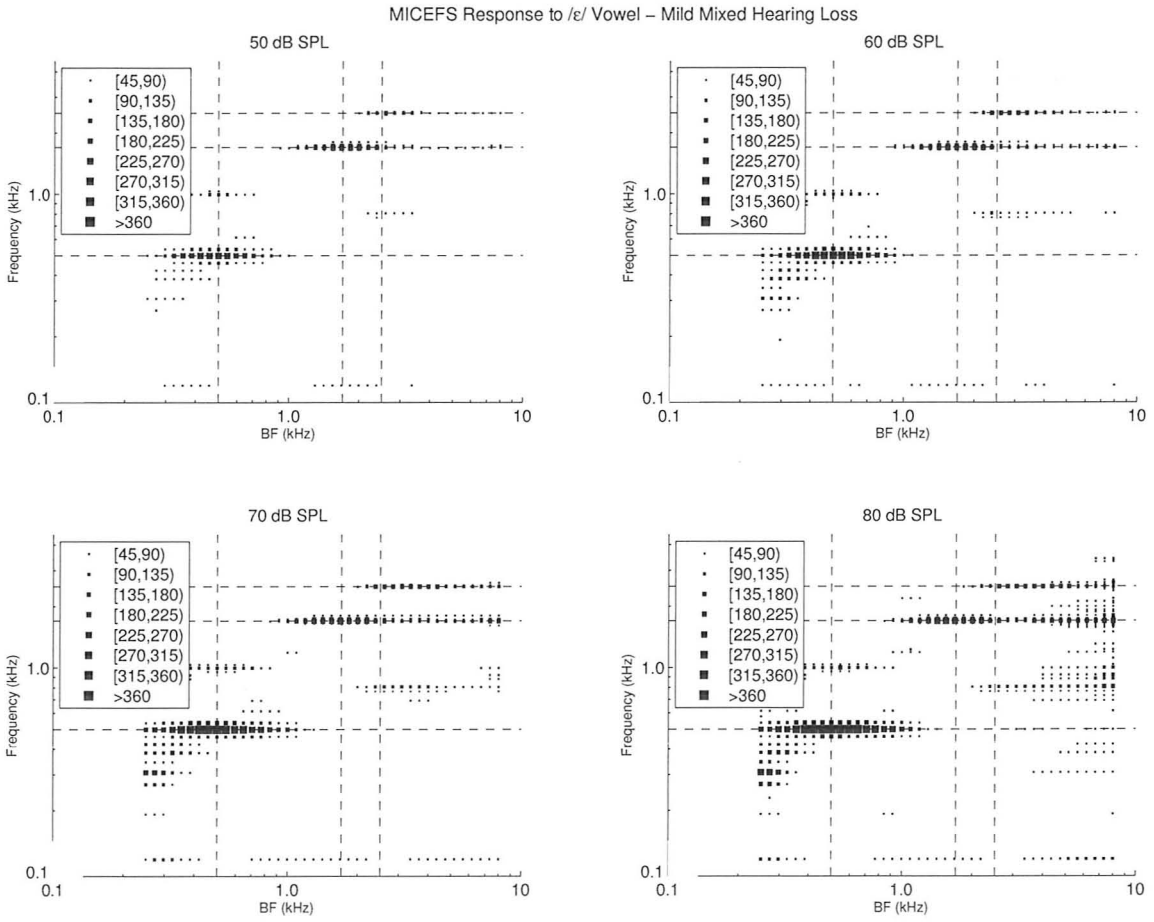

Figure 7.9: Box-plots of the synchronized rate for a mildly impaired AN model with mixed hair cell loss in response to the synthesized vowel /ɛ/, processed with MICEFS-1.

Mild Hearing Loss, OHC Damage

The effects of MICEFS on the synchrony of the impaired AN response in the case of mild hearing loss with exclusively outer hair cell damage can be seen in Figures 7.10 through 7.12. These figures are organized in a similar fashion to Figures 7.7 through

7.9, where the box-plots of the unaided, impaired response for this hearing loss type are followed by the MICEFS-0 aided response and the MICEFS-1 aided response.

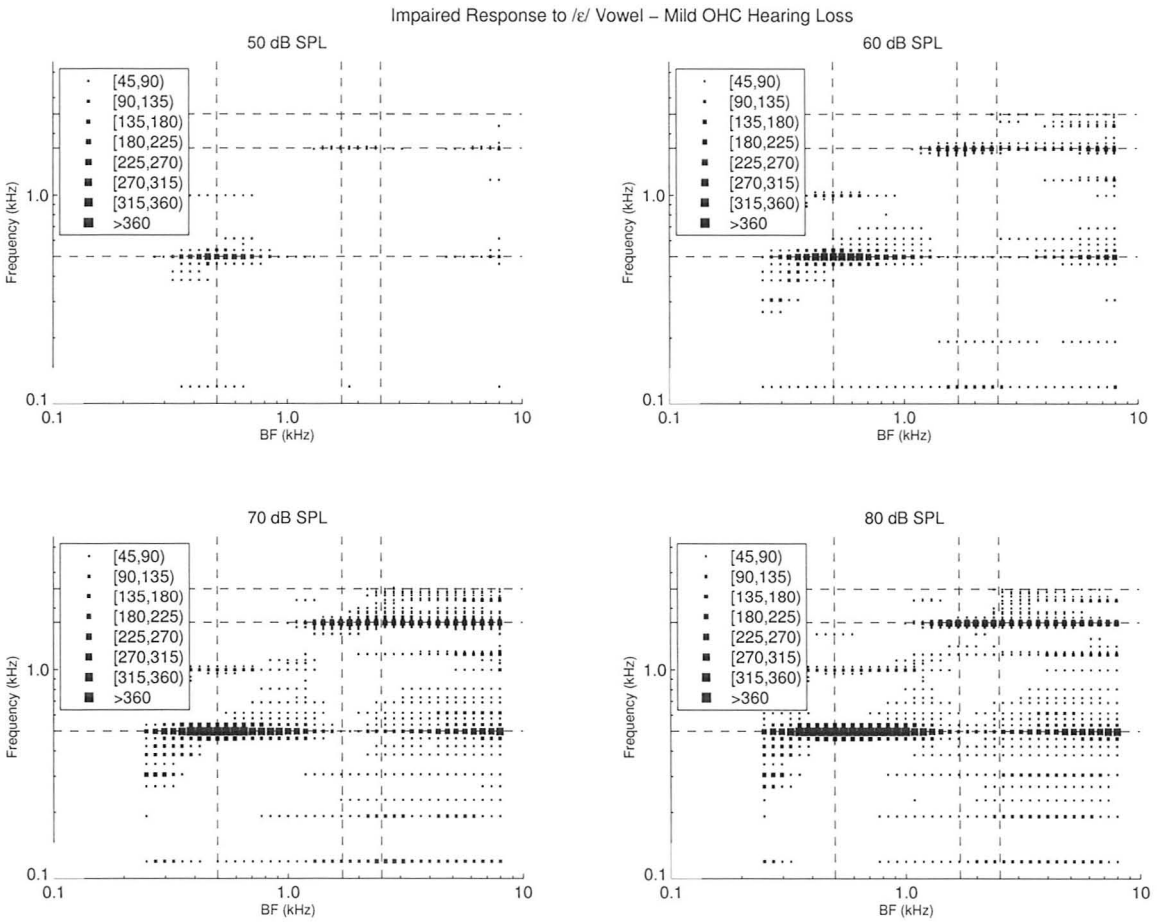

Figure 7.10: Box-plots of the synchronized rate for a mildly impaired AN model with exclusively OHC loss in response to the synthesized vowel /ɛ/.

The mildly impaired response with exclusively outer hair cell damage exhibits increased spread of synchrony to F1 and F2 relative to the mixed hair cell damage at most SPLs, excluding the lowest (50 dB SPL) where very little synchronized response is present. This increased spread of synchrony is likely due to the broadened bandwidths of impaired auditory filters that are associated with outer hair cell loss. The impaired response in Figure 7.10 shows almost no synchrony to the third formant of the vowel. This too, is likely a result of broadened auditory filter bandwidths that result in F2 capturing the response of fibers that might normally respond to F3.

The effect of MICEFS-0 in the outer hair cell damaged case, seen in Figure 7.11, is

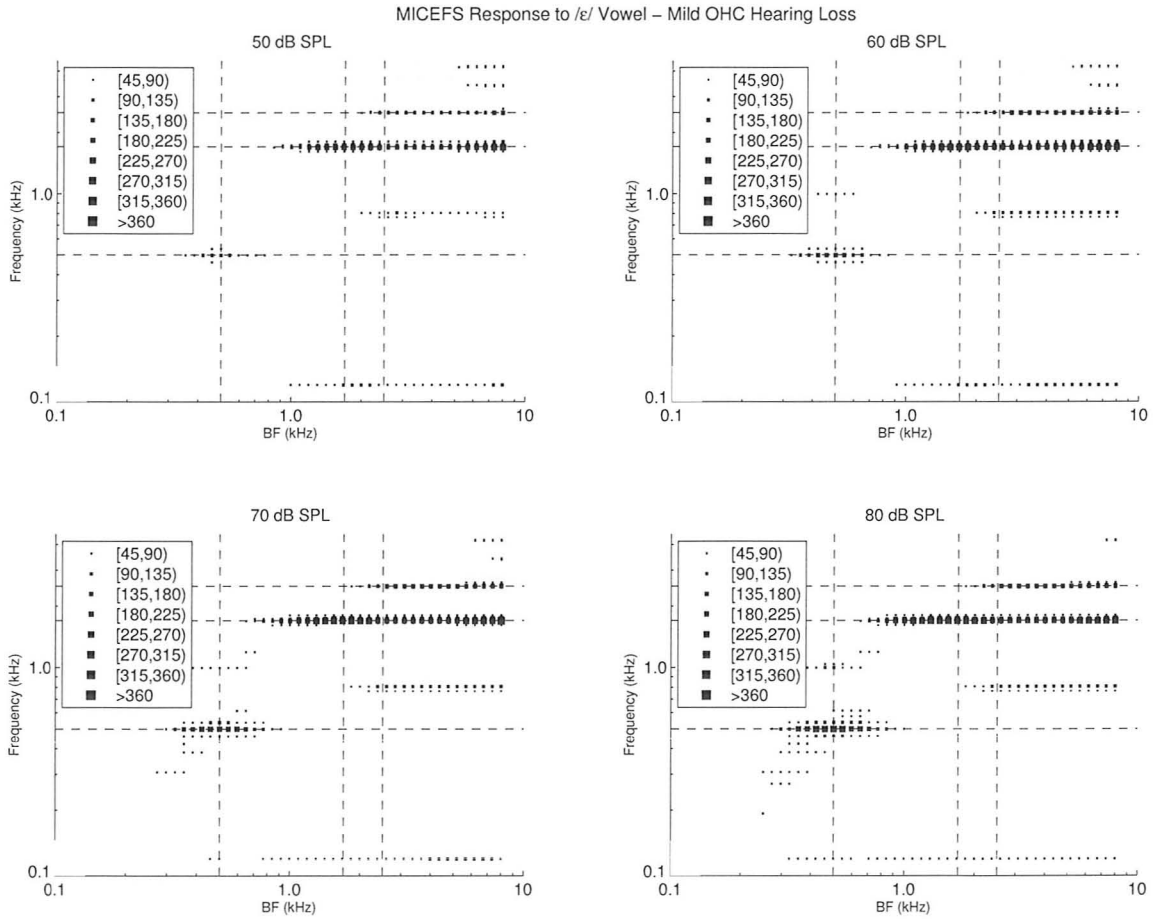

Figure 7.11: Box-plots of the synchronized rate for a mildly impaired AN model with exclusively OHC loss in response to the synthesized vowel /ɛ/, processed with MICEFS-0.

similar to the mixed hearing loss case except there is significant upward and downward spread of synchrony to F2 and F3. This again, is the result of broadened auditory filters associated with outer hair cell loss. The synchrony to F1 for the MICEFS-0 amplified vowel is minimal at all input levels, as it was for the mixed hair cell damage case.

The effect of MICEFS-1, shown in Figure 7.12, is that some synchrony to F1 is restored. At high sound pressure levels, there is a slight decrease in the downward spread of synchrony to F2 relative to the MICEFS-0, suggesting that some of the fibers in this location (approximately 1 kHz) are now responding preferentially to F1.

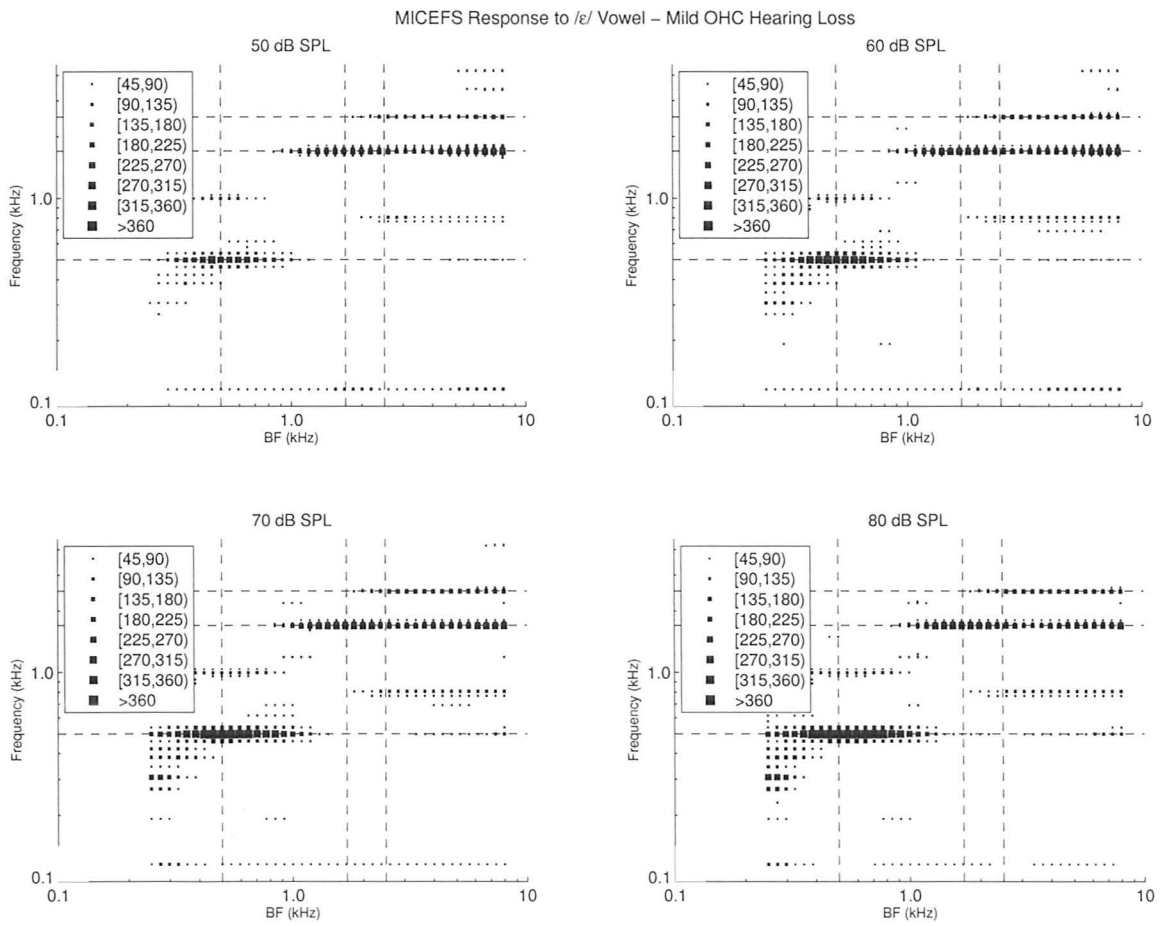

Figure 7.12: Box-plots of the synchronized rate for a mildly impaired AN model with exclusively OHC loss in response to the synthesized vowel /ɛ/, processed with MICEFS-1.

Moderate-to-Severe Hearing Loss, Mixed Hair Cell Damage

The effects of MICEFS on the Moderate-to-Severe hearing loss with mixed hair cell damage are similar to the case of mild hearing loss with mixed hair cell damage, as can be seen in Figures 7.13 through 7.15. Figure 7.13 shows box-plots of the synchronized rates for this hearing loss type. These demonstrate a loss of response to F2 and F3, which is expected from the high frequency loss that is present in the audiogram of Figure 7.2.

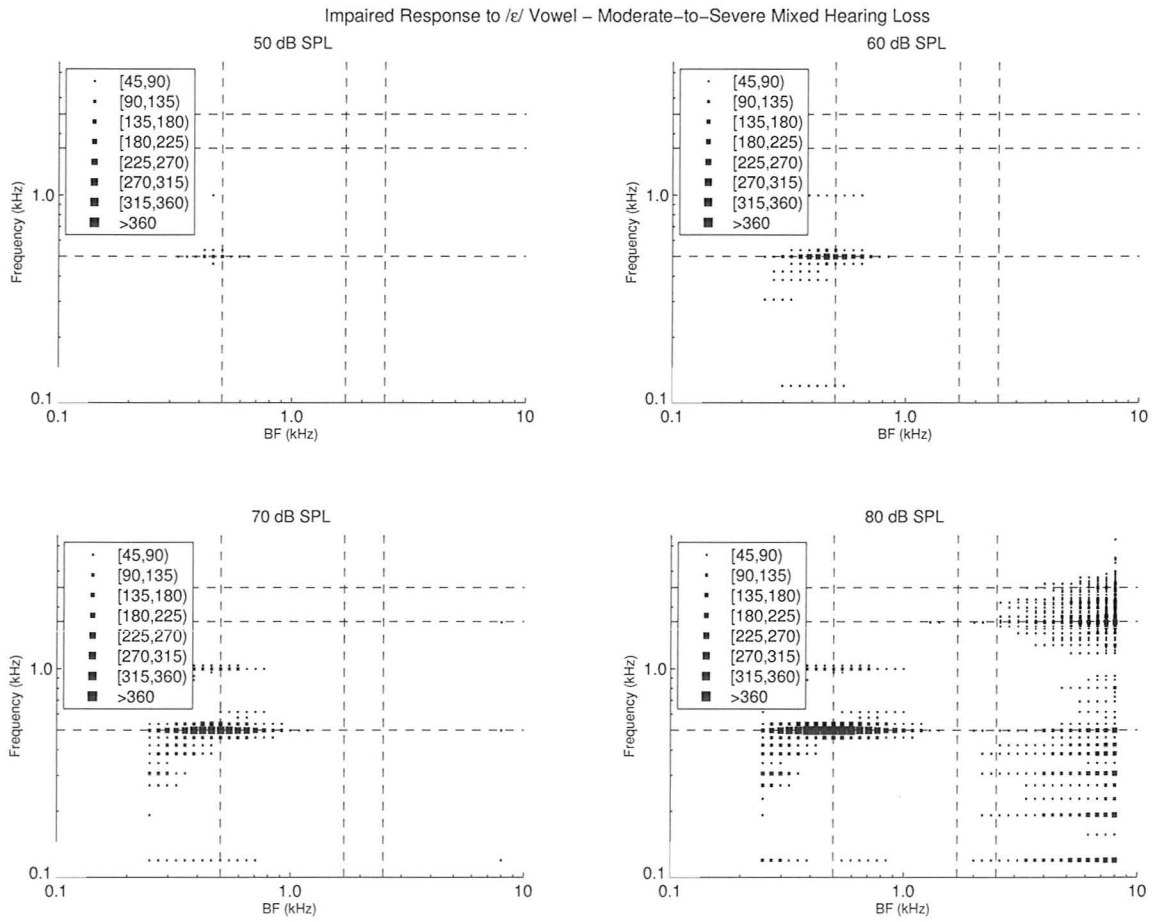

Figure 7.13: Box-plots of the synchronized rate for a moderate-to-severe hearing impaired AN model with mixed hair cell loss in response to the synthesized vowel /ɛ/.

Figure 7.14 shows the synchrony to the synchronized vowel after it has been processed with the MICEFS-0 processing scheme. Here a significant amount of synchrony to the second and third formants is restored. In fact, there is much more synchrony

to these formants here than is seen in the healthy response to the synthesized vowel in Figure 7.6. As well, the synchrony to F1 that was still present in the impaired case is absent, as was the case for MICEFS-0 applied to the other hearing loss types.

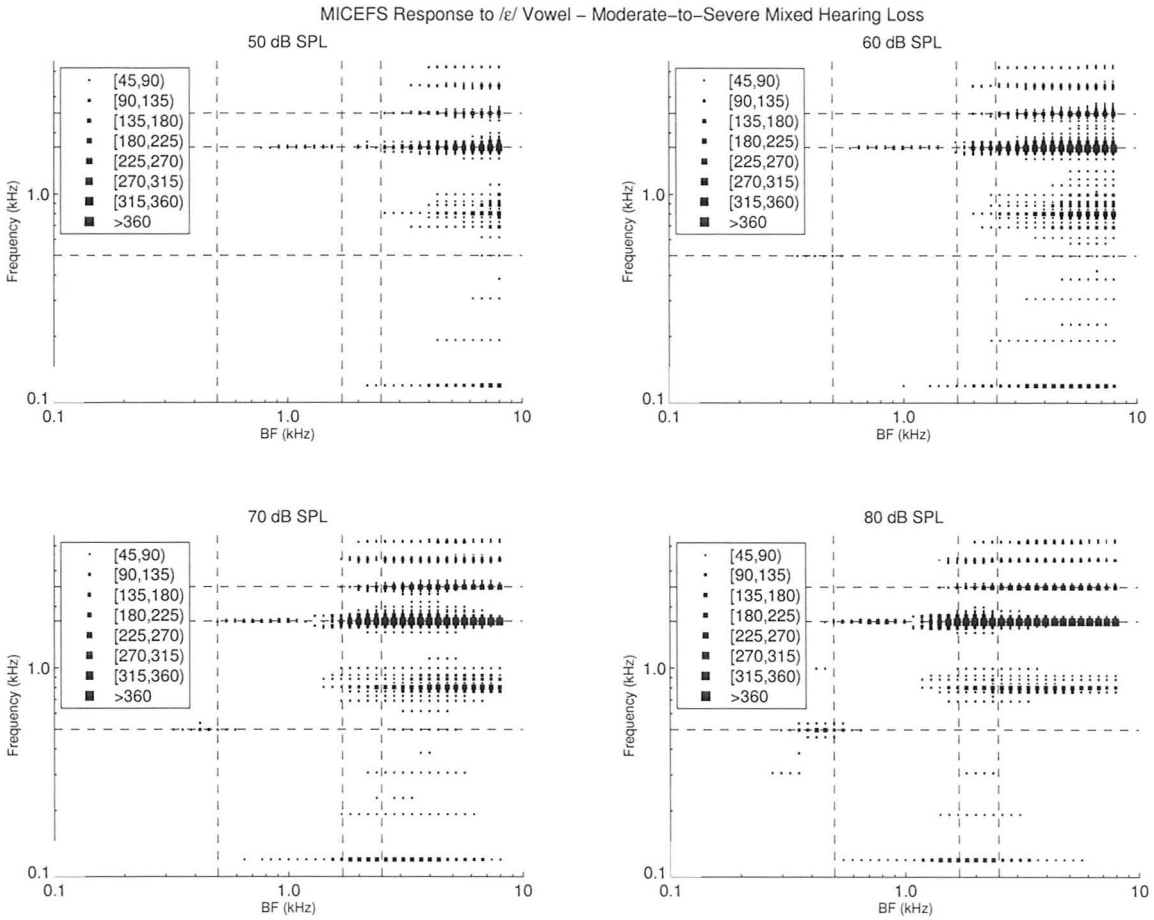

Figure 7.14: Box-plots of the synchronized rate for a moderate-to-severe hearing impaired AN model with mixed loss in response to the synthesized vowel /ɛ/, processed with MICEFS-0.

The response to the MICEFS-1 aided vowel programmed for the moderate-to-severe hearing loss type is shown in Figure 7.15. Here there is a restoration of some synchrony to F1, and not much change in the synchrony to F2 and F3 relative to the MICEFS-0 loss case.

For all hearing loss types evaluated, application of MICEFS seems to restore synchrony to F2 and F3. In some cases, the mildly impaired loss with outer hair cell damage, and the moderate-to-severe loss, there was an upward spread of synchrony

to these formants. MICEFS-0 attenuated the synchrony to F1 in all cases, but this was restored by application of MICEFS-1.

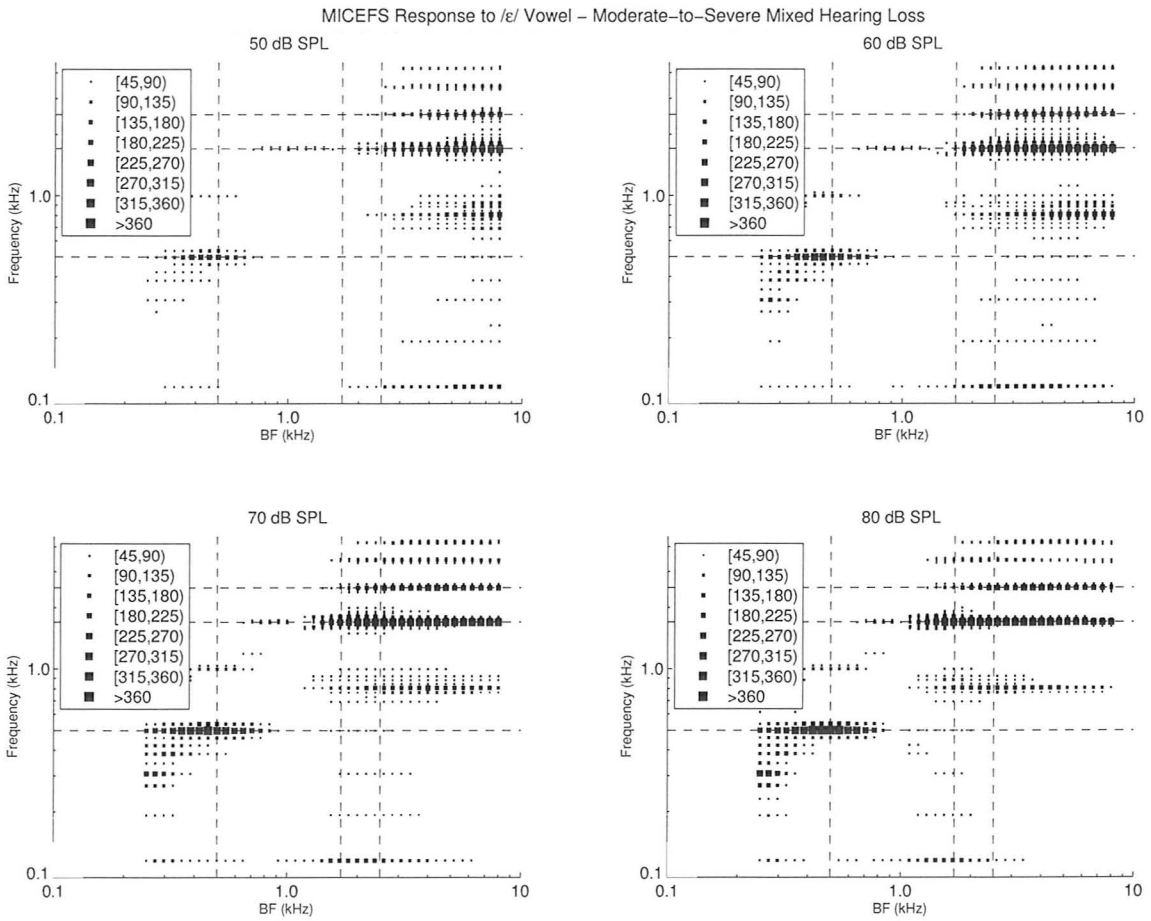

Figure 7.15: Box-plots of the synchronized rate for a moderate-to-severe hearing impaired AN model with mixed loss in response to the synthesized vowel /ɛ/, processed with MICEFS-1.

7.3 Evaluation of SPC alone

Three versions of the SPC scheme are evaluated here, where SPC-0 refers to the implementation of the original SPC that does not account for the different latencies through the modelling path and the processing path; SPC-1 refers to the version that does account for these delays, and SPC-2 refers to the version of the SPC that applies a low-pass filter to the insertion gains prescribed. Table 7.1 summarizes the difference in the versions of SPC.

Table 7.1: Processing components included in the different versions of SPC that are evaluated in this study.

SPC version	Path latencies accounted for?	Low-pass Insertion Delays?
SPC-0	No	No
SPC-1	Yes	No
SPC-2	Yes	Yes

7.3.1 Mild Hearing Loss, Mixed Hair Cell Damage

First, neurograms of the mildly impaired hearing loss with mixed hair cell damage in response to the synthesized /ɛ/ vowel at four presentation levels are presented in Figure 7.16. The neurograms of the healthy auditory model at these stimulus levels are included for comparison. The healthy neurograms appear in the left column, and the impaired neurograms appear in the right. Each subfigure is plotted with the same colour axis, so it is easy to see that there is reduced neural activity in the impaired auditory nerve, especially at high fiber characteristic frequencies and at low sound presentation levels. Careful inspection reveals a shallower relative phase response between fibers in the impaired case, resulting in slightly more coincident spike times for fibers responding to the same frequency component.

Next, the effects of processing the synthesized vowel with SPC-0 before presentation to the mildly impaired AN model are shown in the neurograms in the left hand column of Figure 7.17. These neurograms show that the overall level of neural activity is more or less the same as the impaired neural response in Figure 7.16. This behaviour is consistent with the SPC methodology whose filterbank attempts to have a flat, unity gain frequency response. What has changed in the impaired-aided neural response, is the relative timing of spiking activity between AN fibers with neighbouring CFs. However, it is apparent that this has not been achieved in the desired manner. The degree to which the relative phase of neighbouring fibers

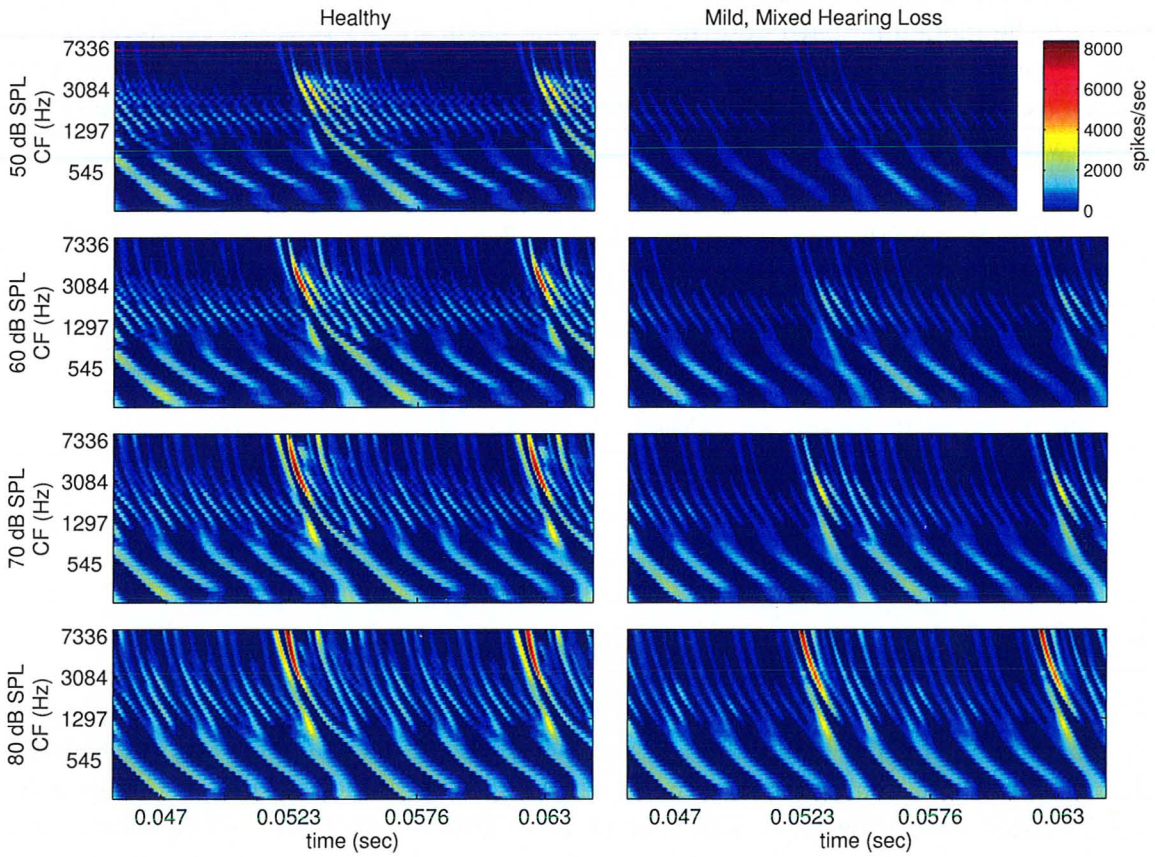

Figure 7.16: Healthy and impaired Neurograms for the mild hearing loss with mixed hair cell damage in response to the synthesized /ɛ/ vowel. Neurograms are derived from the model's 'synaptic' output which gives the probability of spiking in units of spikes/sec. Healthy neurograms are shown in the left column of subfigures, with each row showing the response at a different stimulus presentation level, as labelled. Impaired neurograms are shown in the right column of subfigures. Each subfigure is plotted using the same colour axis, on a colour spectrum scale from blue to red. Dark blue represents a very low spiking rate and red represents a high spiking rate in units of spikes/sec

has been adjusted is not consistent across AN populations that are synchronized to the same frequency; nor is it consistent for every cycle in the periodic response. As well, the frequency to which certain AN fibers appear to be responding has changed as well. These effects result in neurograms that appear to be significantly distorted from their healthy counterparts that are shown in the left hand column of Figure 7.16.

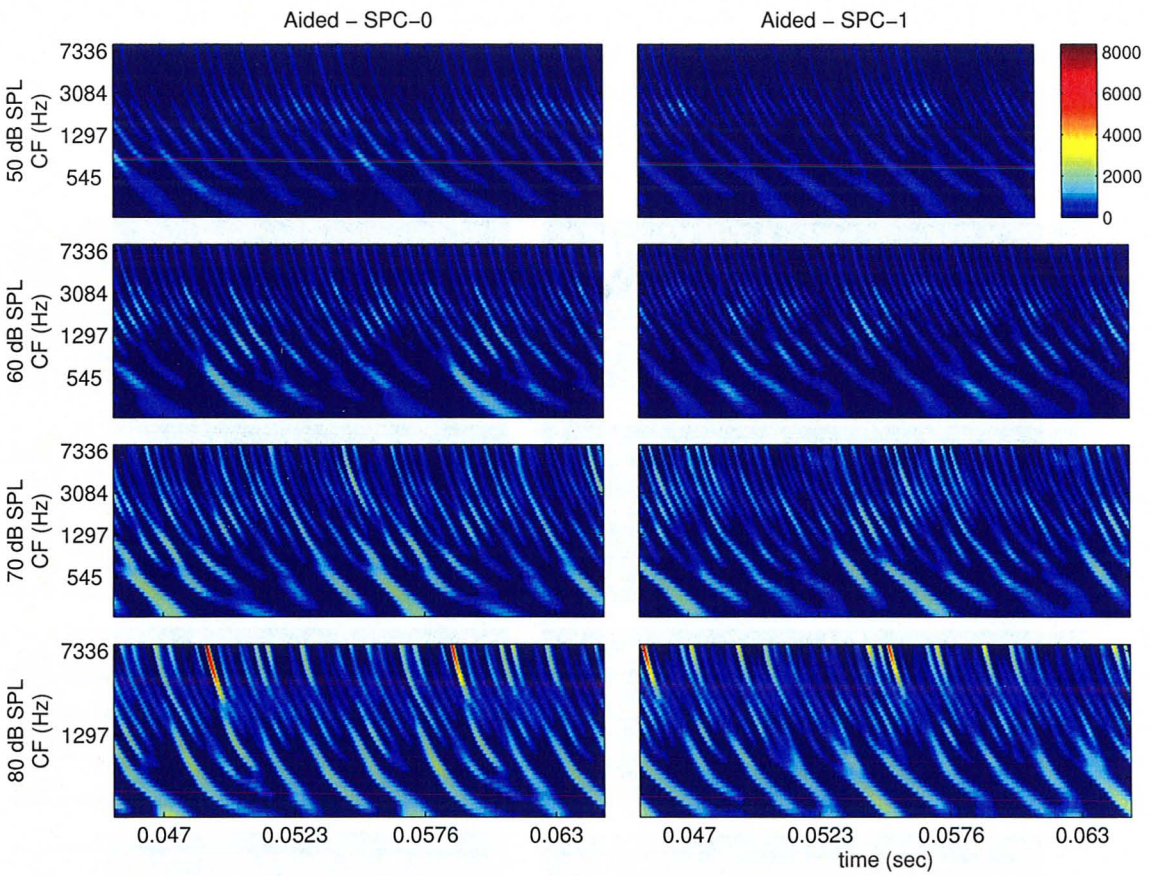

Figure 7.17: Impaired neurograms for the mild hearing loss with mixed hair cell damage in response to the synthesized /ɛ/ processed with both the SPC-0 and SPC-1 processing schemes. The neurograms that result after processing with SPC-0 are shown in the left column of subfigures while the neurograms that results after processing with SPC-1 are shown in the right column.

The impaired neurograms that result from processing the synthesized vowel with SPC-1 are shown in the right hand column of Figure 7.17. The effects of SPC-1 on the impaired neural response are similar to the effects of SPC-0 in that the relative timing

of spiking activity between neighbouring AN fibers is altered. However, there appears to be less distortion of the low frequency response, which still appears to synchronize to the first formant. This could be a result of more accurate timing in the application of SPC insertion delays by accounting for the difference in path latencies in SPC-1.

The neurograms that result from processing with SPC-2 are included in the right hand column of Figure 7.18. SPC-2, again, alters the relative timing of spiking between neighbouring AN fibers, however, the overall spatiotemporal response pattern appears to be slightly more similar to the impaired response to the unprocessed vowel in Figure 7.16. The quality of these neurograms can be compared with visual inspection, however it is hard to quantify their differences using this method, so it is useful to evaluate with additional error metrics.

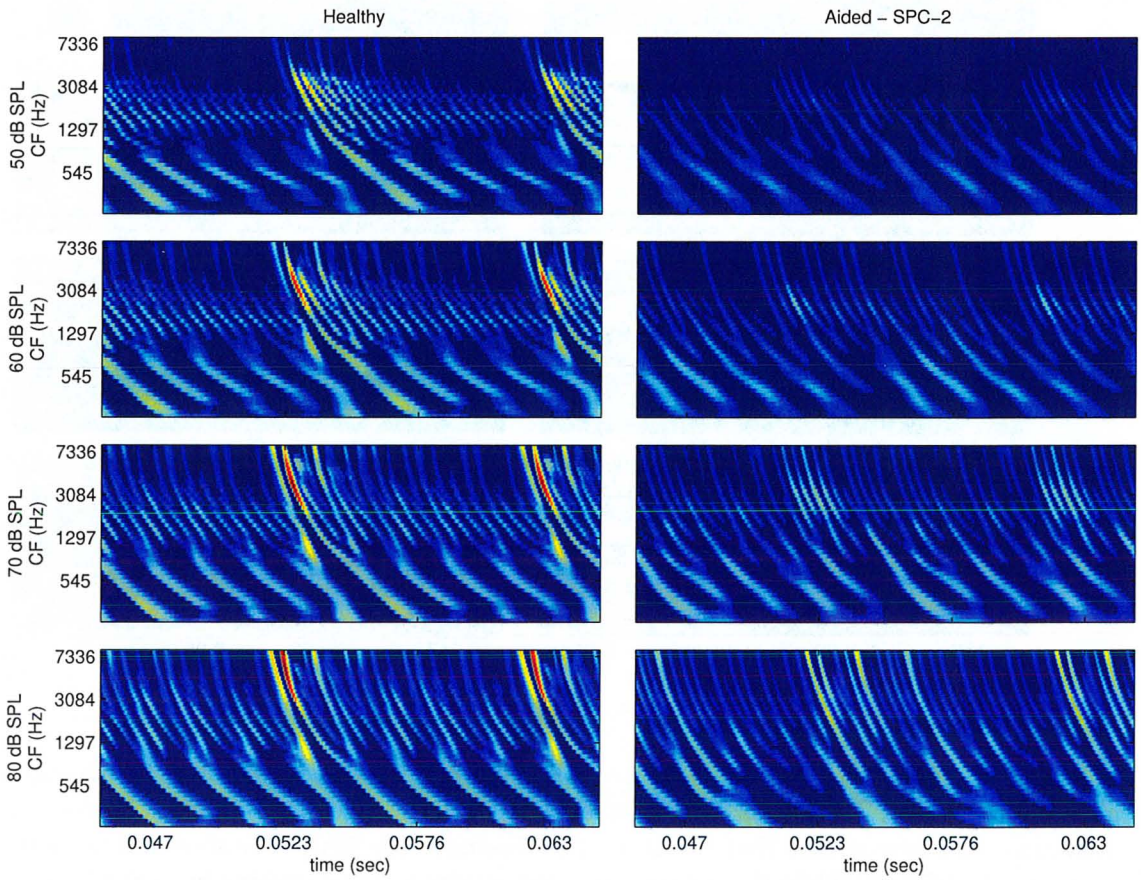

Figure 7.18: Healthy and Impaired neurograms for the mild hearing loss with mixed hair cell damage in response to the synthesized /ɛ/ processed with the SPC-2 processing scheme.

The synchronized rate shown in the box-plot of Figure 7.19 is useful for evaluating the AN fiber response to the formants in the vowel stimulus. The box-plots in this figure represent the neural synchrony of the mildly impaired, mixed hair cell loss in response to the synthesized vowel processed with SPC-1. They demonstrate that much of the synchrony to the second and third formants that remained in the impaired response in Figure 7.16 is now gone. Some of the synchrony to F1 remains, in particular at 70 dB SPL and at 80 dB SPL, but there also appears to be some synchrony to non-formant frequencies between F1 and F2. The effects of processing with SPC-0, and SPC-2 on the synchronized rate are similar, but are more pronounced. There appears to be more disruption of synchrony and more response to non-formant frequencies in response to SPC-0 and SPC-2; these can be seen in Figures F.11 and

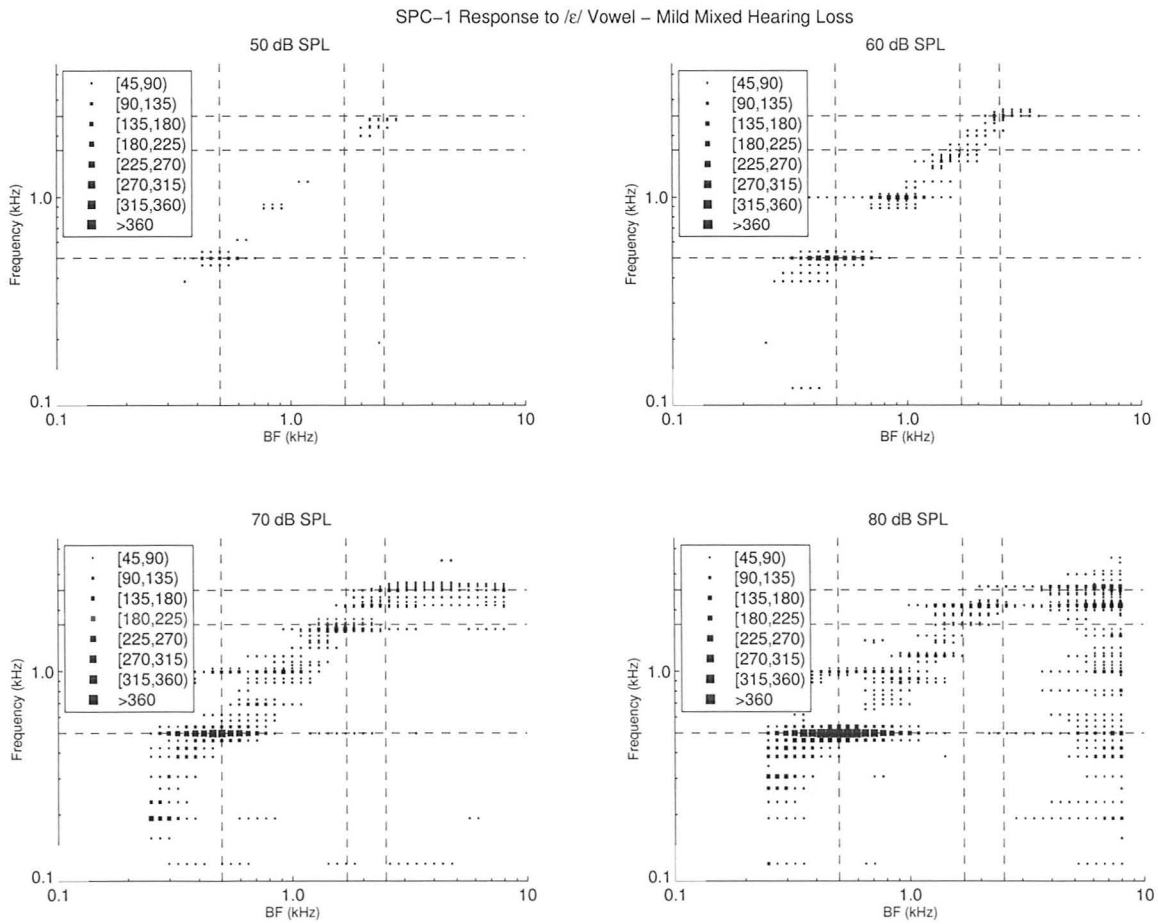

Figure 7.19: Box-plots of the synchronized rate for a mildly impaired AN model with mixed hair cell loss in response to the synthesized vowel /ɛ/, processed with SPC-1.

F.12, respectively, in Appendix F.

Analysis of the aided neurograms and the box-plot above indicate that all versions of the SPC scheme appear to distort the AN fiber response. However, the SPC-1 version, which accounted for path delays when prescribing insertion gains and did not low-pass filter these gains, seemed to distort the AN fiber response less than the others. As well, the peak-lag error, shown as a function of SPL in Figure 7.20, demonstrates that processing with SPC-1 and SPC-2 reduces the error in the impaired response for this metric at low SPLs. It also had a more consistent value as a function of SPL for these processors.

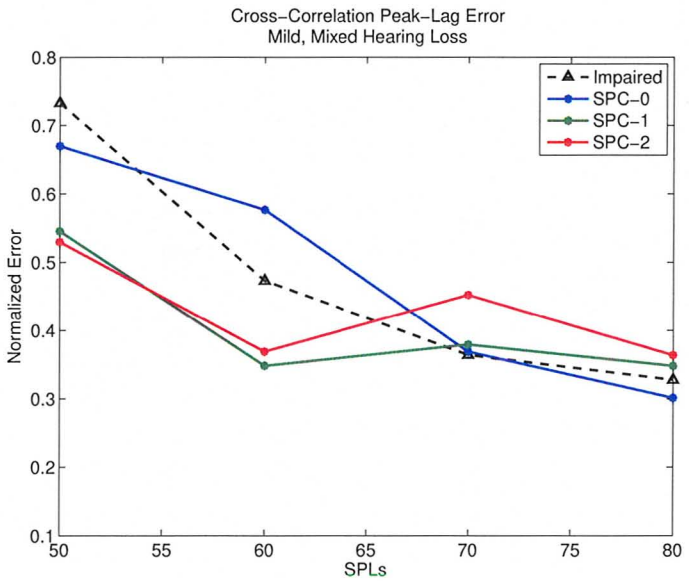

Figure 7.20: Peak-lag error for the synthesized vowel aided with three versions of SPC, shown as a function of stimulus sound pressure level. Error derived from the vowel response of an AN model with mild hearing loss and mixed hair cell damage.

7.3.2 Other Hearing Loss Types

The results in the previous section for the mild hearing loss with mixed hair cell damage are, for the most part, representative of the results seen for the mild hearing loss with exclusively outer hair cell damage. The neurograms of the unprocessed vowel demonstrate more wide-spread synchronization across characteristic frequencies as is expected, and the synchrony to the SPC-0, SPC-1 and SPC-2 processed vowel is disrupted. These figures appear in Section F.1.2, Appendix F.

As well, the neurograms for the moderate-to-severe hearing loss with mixed hair cell damage in response to the SPC aided vowel demonstrate a distorted spatiotemporal response pattern compared to the healthy ear. These figures can be seen in Section F.1.3, Appendix F.

The peak-lag error for the SPC-aided synthesized vowel is shown in Figure 7.21 for both the mild hearing loss with outer hair cell impairment (left panel) and the moderate-to-severe hearing loss (right panel). For the mild, OHC loss type, SPC-1 and SPC-2 showed an improvement at low SPLs. However, the resultant peak-lag error was not as consistent as a function of SPL as it was for the mild loss with mixed hair cell damage. For the moderate-to-severe hearing loss, there were relatively small improvements in the peak-lag error for the SPC-0 and SPC-1 processors.

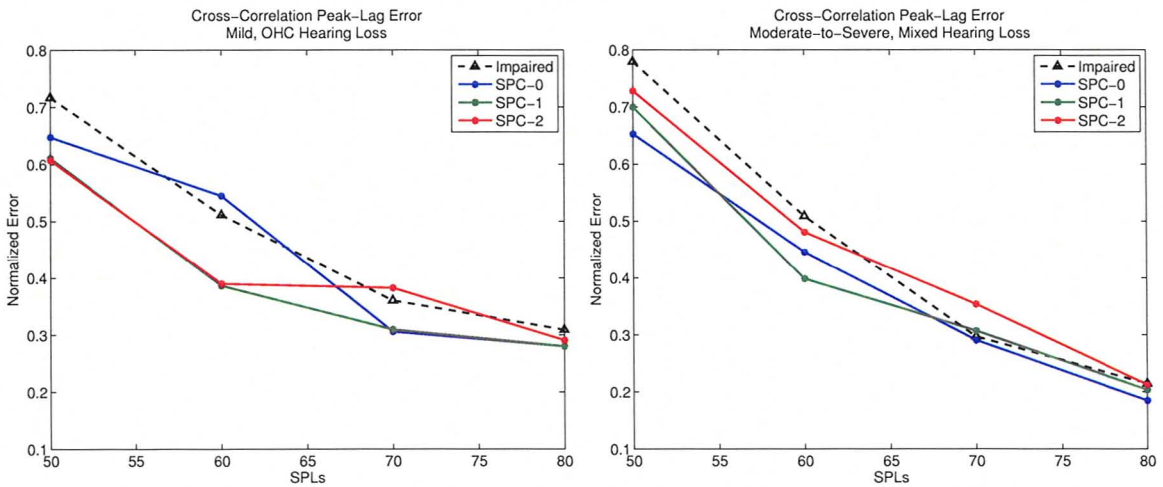

Figure 7.21: Peak-lag error for the synthesized vowel aided with three versions of SPC, shown as a function of stimulus sound pressure level. *Left Panel*: Error derived from the response of an AN model with mild hearing loss and exclusively outer hair cell damage. *Right Panel*: Error derived from the response of an AN model with moderate-to-severe hearing loss and mixed hair cell damage.

7.4 MICEFS and SPC

The sections above demonstrate that MICEFS-1 did a better job than MICEFS-0 at restoring the synchrony to the first three formants in the synthesized vowel for all hearing loss types. The performance of the different versions of SPC did not perform consistently across hearing loss types, however, the performance of SPC-0 was the worst for all error metrics analyzed. For these reasons, MICEFS-1 will be combined

only with the two versions of the SPC processing scheme that accounted for processing latencies through the SPC pathways: SPC-1 and SPC-2.

7.4.1 Mild Hearing Loss, Mixed Hair Cell Damage

Figure 7.22 shows the mildly impaired neurograms aided with the processor combinations MICEFS-1/SPC-1 and MICEFS-1/SPC-2. These neurograms are similar to the neurograms that result from SPC-1 and SPC-2 processing without MICEFS included; but here, there is more response to high frequencies. The overall spatiotemporal response pattern in response to the combined MICEFS/SPC processors is noticeably different from the healthy neurograms shown in Figure 7.16. However, the low frequency response in the MICEFS-1/SPC-1 aided neurograms is better than in the MICEFS-1/SPC-2 aided neurograms.

The synchronized rate of the mildly impaired AN response to the vowel aided with MICEFS-1/SPC-1 is shown in the box-plot of Figure 7.23. It is apparent that there is reduced synchrony to the second and third formants, which, when compared to the synchrony in response to MICEFS-1 (in Figure 7.9), suggests that the inclusion of SPC-1 causes this reduced synchrony. This behaviour occurs in the synchronized rates of the mildly impaired hearing loss with mixed hair cell damage for MICEFS-1/SPC-2 as well. This processor also produced significant synchrony to a non-formant frequency between F1 and F2 at high SPLs. This is shown in Figure 7.24.

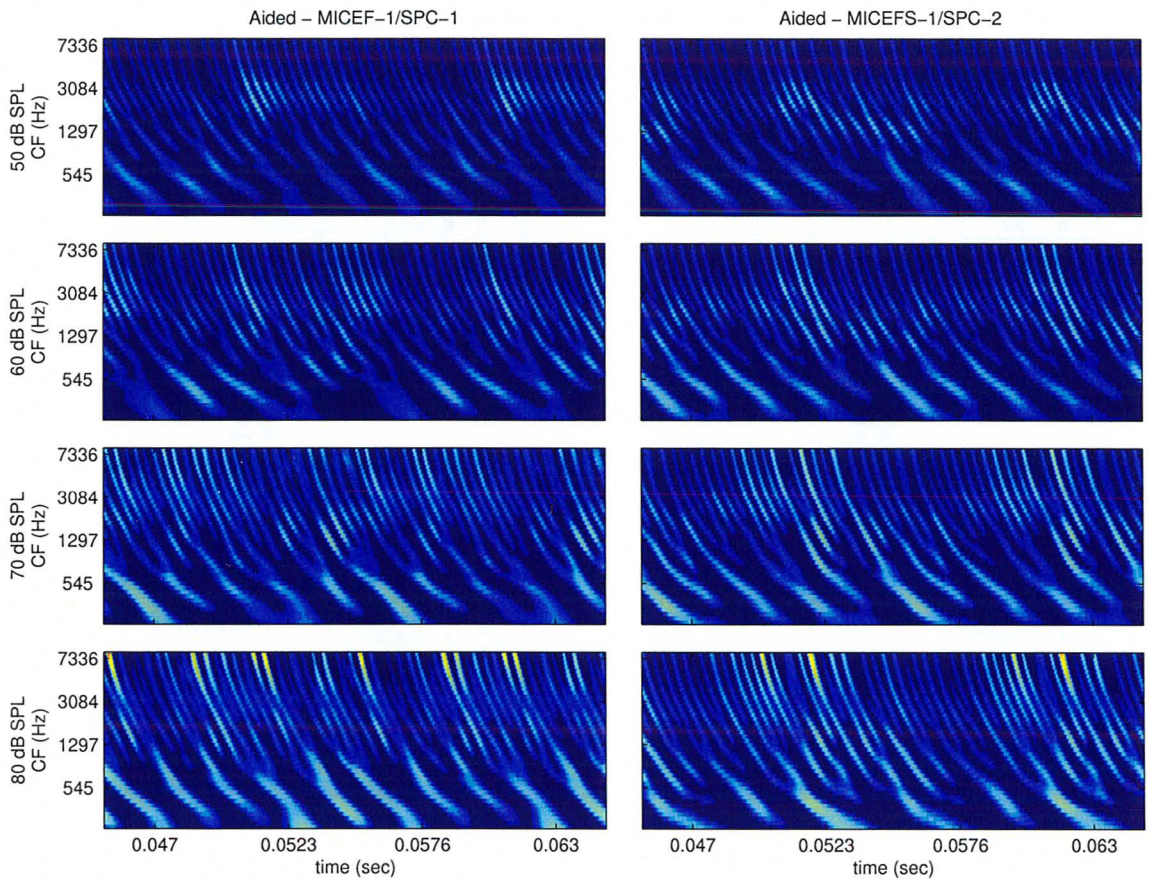

Figure 7.22: Impaired neurograms for the mild hearing loss with mixed hair cell damage in response to the synthesized /ɛ/ vowel processed with both the MICEFS-1/SPC-1 and MICEFS-1/SPC-2 processing schemes. The neurograms that result after processing with SPC-1 are shown in the left column of subfigures while the neurograms that results after processing with SPC-2 are shown in the right column.

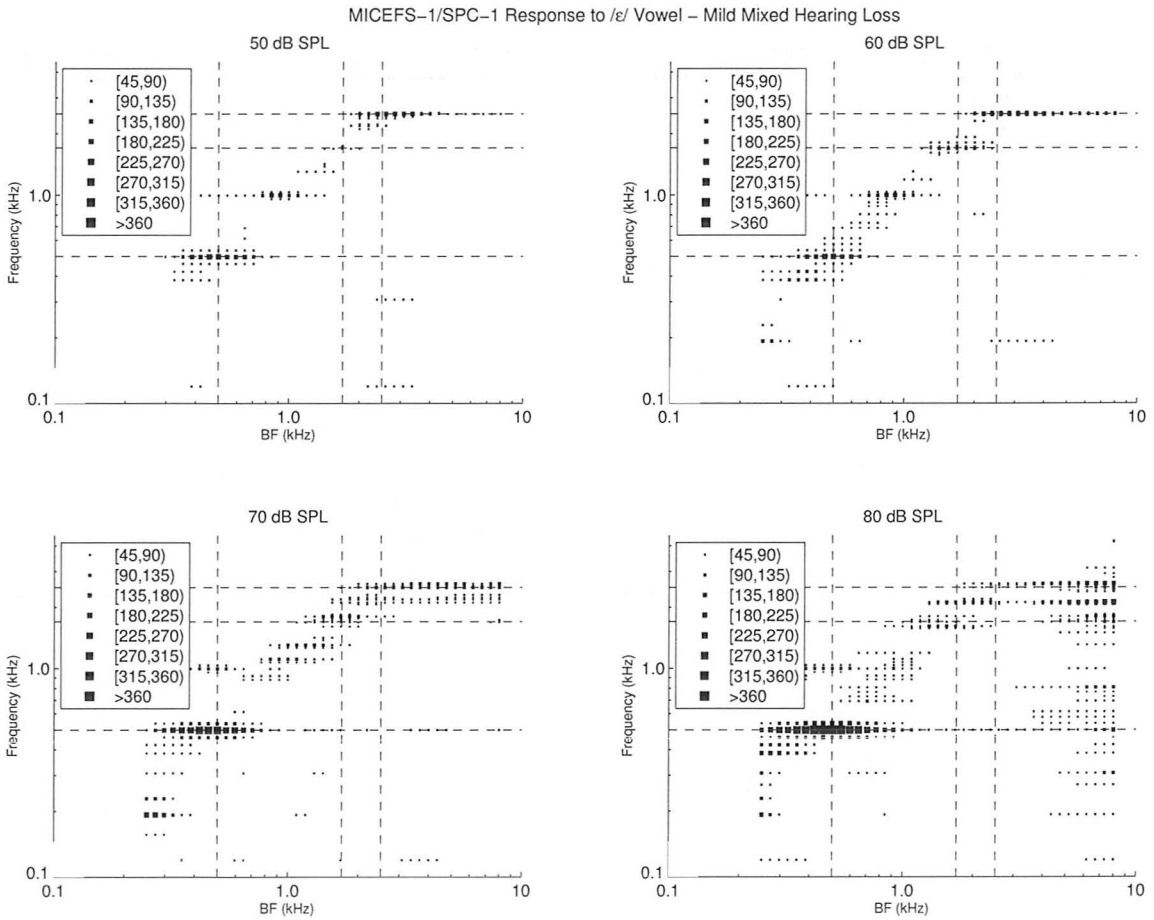

Figure 7.23: Box-plots of the synchronized rate for a mildly impaired AN model with mixed hair cell loss in response to the synthesized vowel /ɛ/, processed with MICEFS-1/SPC-1.

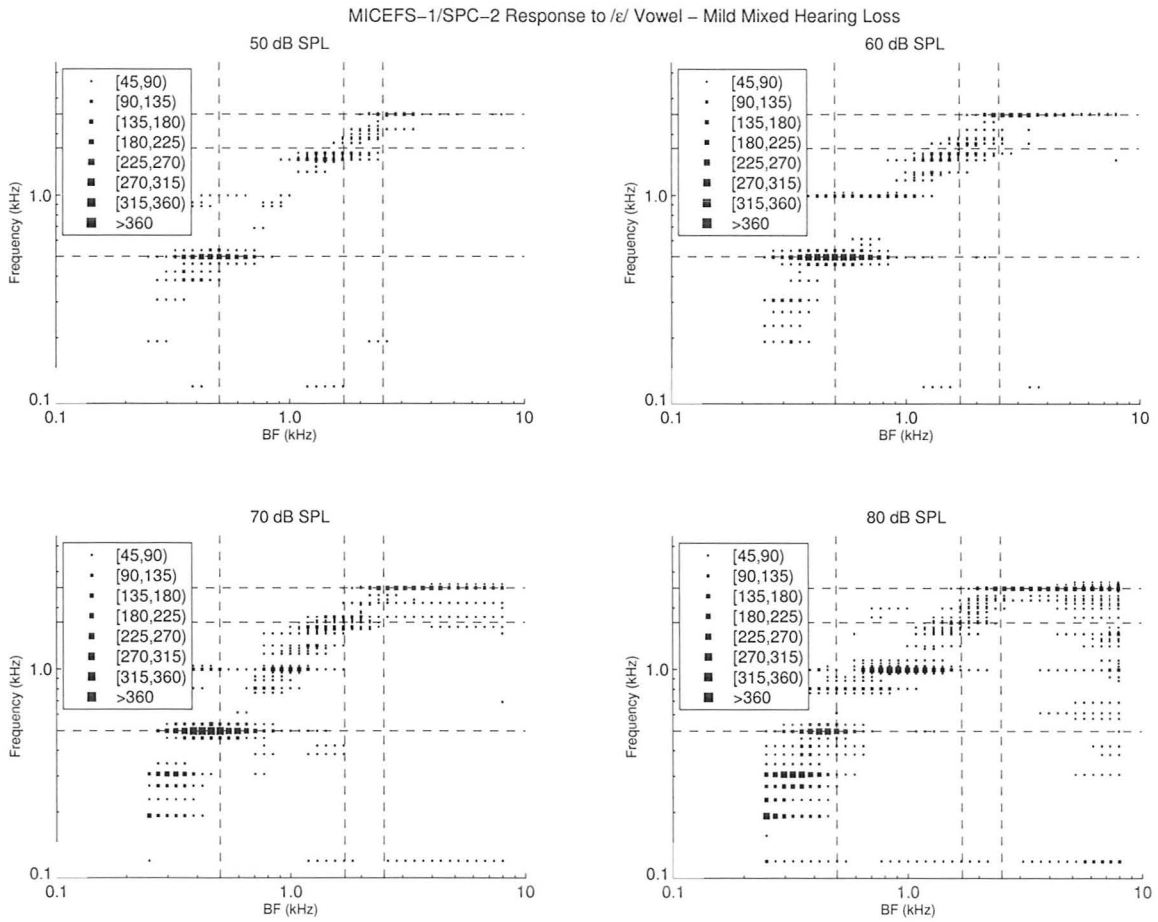

Figure 7.24: Box-plots of the synchronized rate for a mildly impaired AN model with mixed hair cell loss in response to the synthesized vowel /ɛ/, processed with MICEFS-1/SPC-2.

Figure 7.25 shows the peak-lag cross-correlation error as a function of SPL for the combination of MICEFS-1 with two versions of SPC evaluated. The figure demonstrates that MICEFS-1/SPC-1 performed better than MICEFS-1/SPC-2 for the mild hearing loss type with mixed hair cell damage. The MICEFS-1/SPC-2 processor had a larger peak lag error relative to the unaided ear for high SPLs, but the error was more consistent across sound pressure levels.

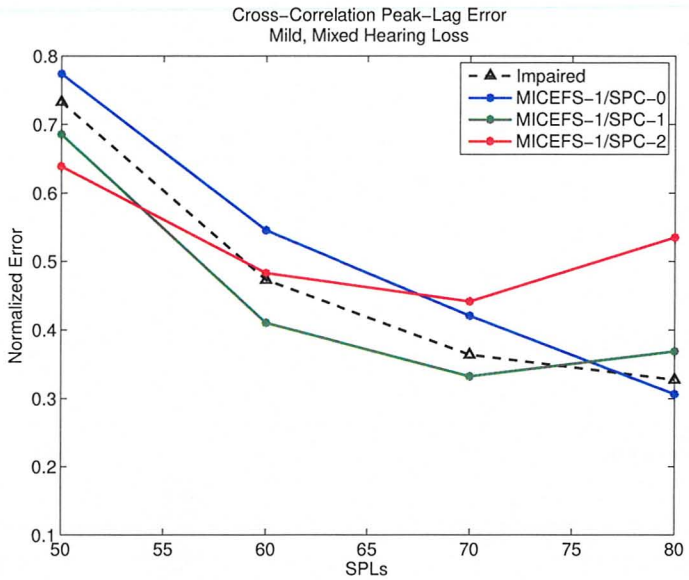

Figure 7.25: Peak-lag error for the synthesized vowel aided with MICEFS-1 combined with three versions of SPC, shown as a function of stimulus sound pressure level. Error derived from the vowel response of an AN model with mild hearing loss and mixed hair cell damage.

7.4.2 Mild Loss, OHC damage

The effects of the combination of MICEFS-1 with SPC on the neurograms, synchronized rate and peak-lag error mild hearing loss with outer hair cell damage are similar to the case with mixed hair cell damage reported in the previous section. For this reason, the figures for this hearing loss type are not shown here, but are included in Appendix F, Section F.2.1. In general, the neurograms appeared different from the healthy neurograms and the synchrony appeared to be diminished by MICEFS-1/SPC-x relative to MICEFS-1 alone.

7.4.3 Moderate-to-Severe Loss, Mixed Hair Cell damage

The previous sections suggest that inclusion of SPC with MICEFS disrupts the synchrony to F2 and F3 in vowel response. However, for the moderate-to-severe hearing loss with mixed hair cell damage, the attenuating effects of the inclusion of SPC on the synchrony to F2 and F3 is less pronounced. As an example of this, Figure 7.26 shows the synchronized rate for this hearing loss type in response to the synthesized vowel aided with MICEFS-1/SPC-1. It is clear that some synchrony to F2 and especially F3 is restored, although there is some minor attenuation of synchrony to F1.

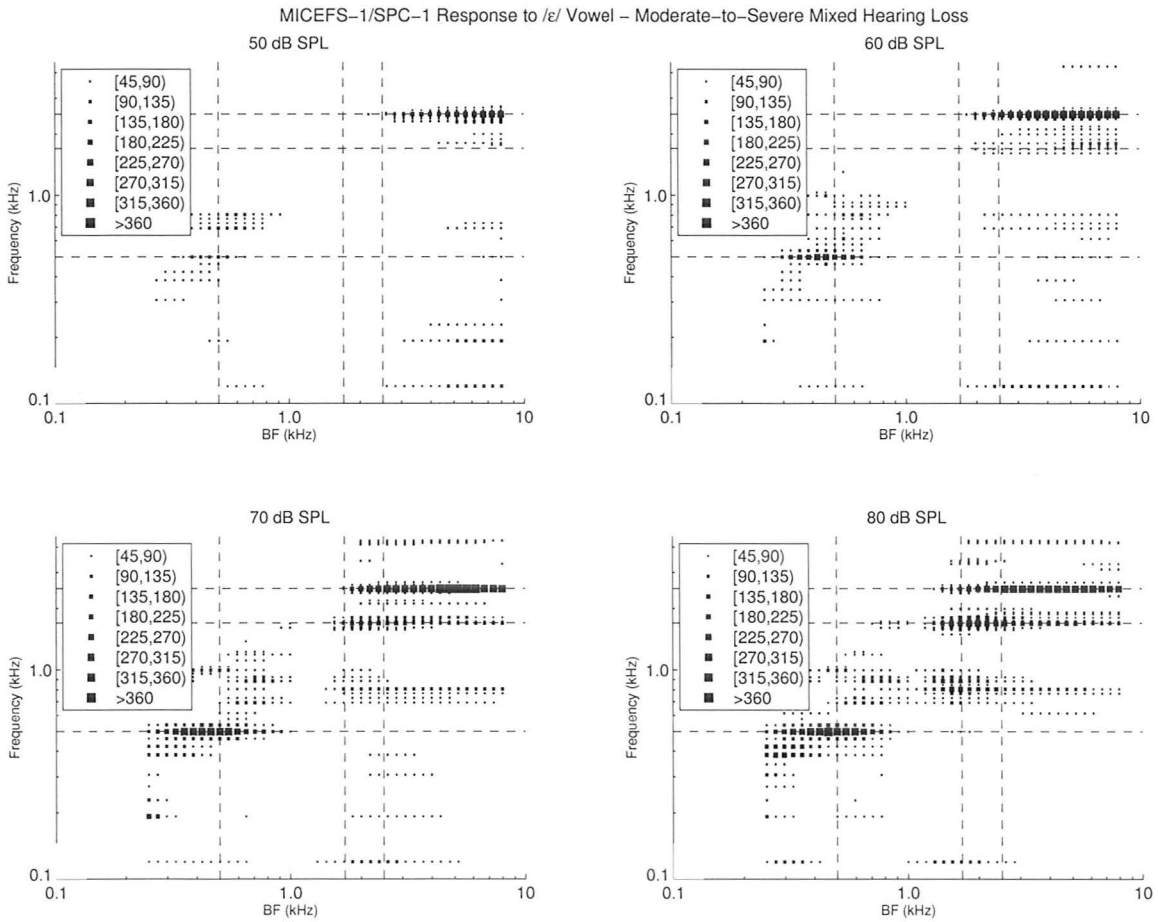

Figure 7.26: Box-plots of the synchronized rate for a moderate-to-severely impaired AN model with mixed hair cell loss in response to the synthesized vowel /ɛ/, processed with MICEFS-1/SPC-1.

The synchronized rate in response to the MICEFS-1/SPC-2 aided vowel is shown

in Figure 7.27. This figure demonstrates some restoration of synchrony to F3, and partial restoration of synchrony to F2, but synchrony to F1 is severely disrupted. The large amount of amplification at F2 and F3 may be a factor in the synchrony that is retained to these formants after SPC-2 processing.

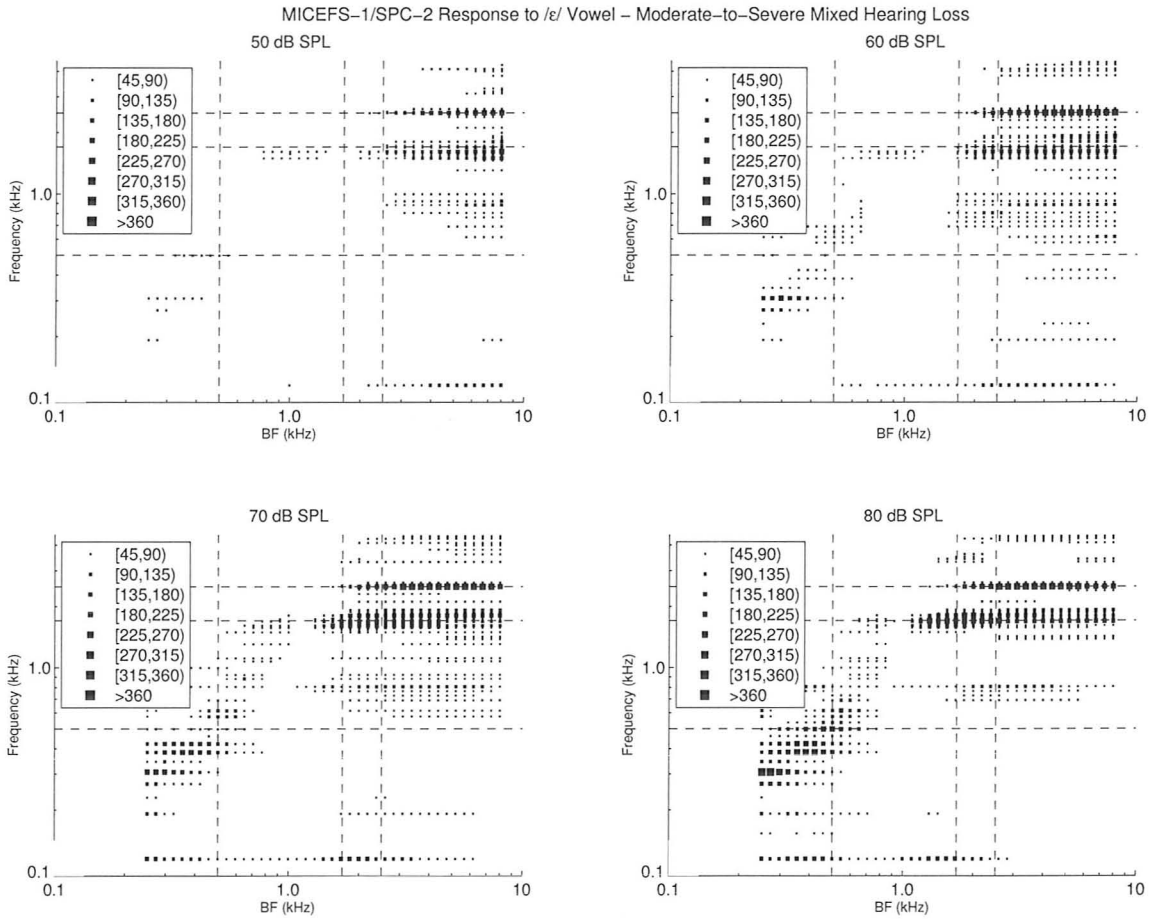

Figure 7.27: Box-plots of the synchronized rate for a moderate-to-severely impaired AN model with mixed hair cell loss in response to the synthesized vowel /ɛ/, processed with MICEFS-1/SPC-2.

The peak-lag cross-correlation error for this hearing loss in response to all combinations of MICEFS-1 and SPC are improved when compared to utilization of the SPC processor alone. In this case, the combination MICEFS-1/SPC-1 performed better than the combination MICEFS-1/SPC-2. These interpretations follow from the results shown Figure 7.28 which shows the peak-lag cross-correlation metric as

a function of SPL. This figure demonstrates that processing with MICEFS-1/SPC-1 significantly decreased the peak-lag error at low SPLs for this moderate-to-severe hearing loss.

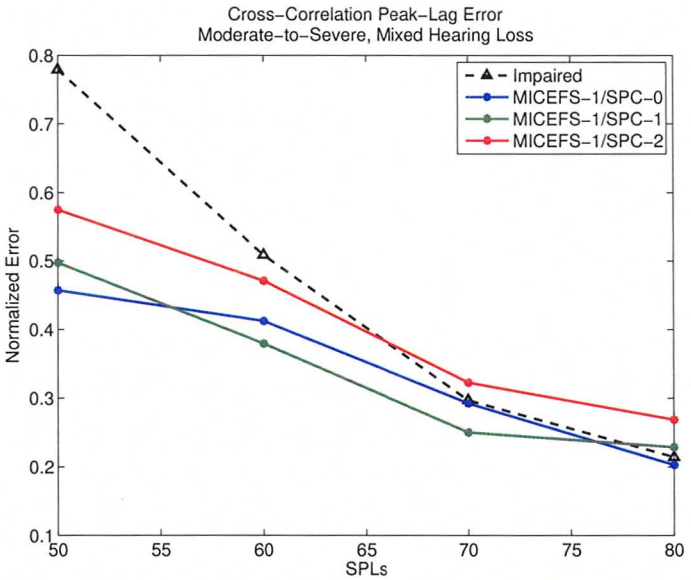

Figure 7.28: Peak-lag error for the synthesized vowel aided with MICEFS-1 combined with three versions of SPC, shown as a function of stimulus sound pressure level. Error derived from the vowel response of an AN model with moderate-to-severe hearing loss and mixed hair cell damage.

Chapter 8

Discussion

In the previous chapter, the results of evaluating several processing schemes with a computational model of the auditory periphery were presented. These results demonstrated: i) that MICEFS performed well in restoring normal auditory nerve synchrony to vowels; ii) that SPC alters the auditory nerve spatiotemporal response pattern significantly, but can improve the peak-lag error; and 3) that when combined, the benefits of MICEFS in restoring synchrony appear to be attenuated by the inclusion of SPC. MICEFS-1 and SPC-1 had the best performance of the versions tested, and it was determined that restoration of synchrony by pre-processing with MICEFS-1 improved the performance of SPC-1 for the moderate-to-severe hearing loss. In addition to these results, several insights regarding the sensitivity of AN synchrony, the limitations of SPC processing, and the potential of the peak-lag error as a loudness metric are discussed here.

8.1 Effects of MICEFS on Synchrony

The evaluation of MICEFS as a stand alone processor demonstrated that it was able to restore the synchrony to the second and third formants that is degraded in the unaided auditory nerve vowel response for each of the hearing loss types investigated. Section 2.6.1 explains that vowels are unique in their formant composition, and adequate synchrony to the formants may be an important neural correlate to vowel intelligibility. That is, proper AN synchrony to the vowel's formants may be related to a listener's ability to discriminate between two different vowels. The restoration of the synchrony to the second and third formants by MICEFS suggests that it may be able to improve the intelligibility of voiced speech for listeners with sensorineural hearing impairment.

The results show that both MICEFS-0 and MICEFS-1 restored synchrony to the second and third formants. However, for both these processors, there were cases of

significant upward spread of synchrony to F2 and F3. That is, AN fibers normally tuned to frequencies higher than F2 and F3 were responding synchronously to those formants as a result of MICEFS processing. This occurred especially in response to loud stimuli, for the moderate-to-severe hearing loss, and for the mild loss with exclusively outer hair cell damage. It is not clear how this over-representation of the second and third formants may effect intelligibility. Since synchrony to both F2 and F3 is spread to AN fibers with higher CFs, it does not appear that the synchrony to either formant suppresses the other. However, a larger ratio of F2 and F3 synchrony to F1 synchrony could result in an abnormal neural representation of relative formant amplitudes, which in some cases can affect vowel discrimination (Kiefte and Kluender, 2005).

Filtering the synthesized vowel with MICEFS-0 restored synchrony to F2 and F3, but it suppressed synchrony to F1. This is likely due to the fact that MICEFS-0 did not include any gain at F1 in its contrast enhancement scheme, and attenuates F1 when combined with multi-band compression. By including some gain to F1 in the contrast enhancement scheme of MICEFS-1, normal synchrony to F1 was restored. These results demonstrate that MICEFS-1 did a better job at restoring AN synchrony to vowel formants. They also suggest that changes in the relative gain at the formants in the contrast enhancement scheme can have significant influence on the synchronized response.

8.1.1 Sensitivity to Changes in Contrast Enhancement

The effect of variations in the gain applied to formants by the MICEFS contrast enhancement scheme is analyzed in Figure 8.1. The left side of this figure shows the gain versus frequency profile of the MICEFS-1 contrast enhancement scheme and the synchronized rate of the mildly hearing impaired model aided vowel response. In the right side of this figure, the MICEFS-1 contrast enhancement scheme is altered so that the passband gain at F2 is increased by 5 dB, and the synchrony to the correspondingly aided vowel is shown below. The effect of increasing the F2 passband gain by 5 dB is to increase the upward spread of synchrony to F2, highlighted by the magenta ellipses, and to suppress the synchrony to F3.

This figure demonstrates that formant synchrony in the impaired model is rather sensitive to changes in the contrast enhancement, and to the resultant amplitude of the vowel's formants. The gain applied to the vowel's formants after MICEFS processing is in part determined by the half-gain rule (recall Table 5.1), and for severe hearing impairments this results in formants with larger amplitudes than for mild hearing impairments. In the synchronized rate for the moderate-to-severe hearing impairment in Figure 7.15, there was stronger synchronized response to F2 and F3 than there was for the mild hearing loss types. This greater synchrony could be a

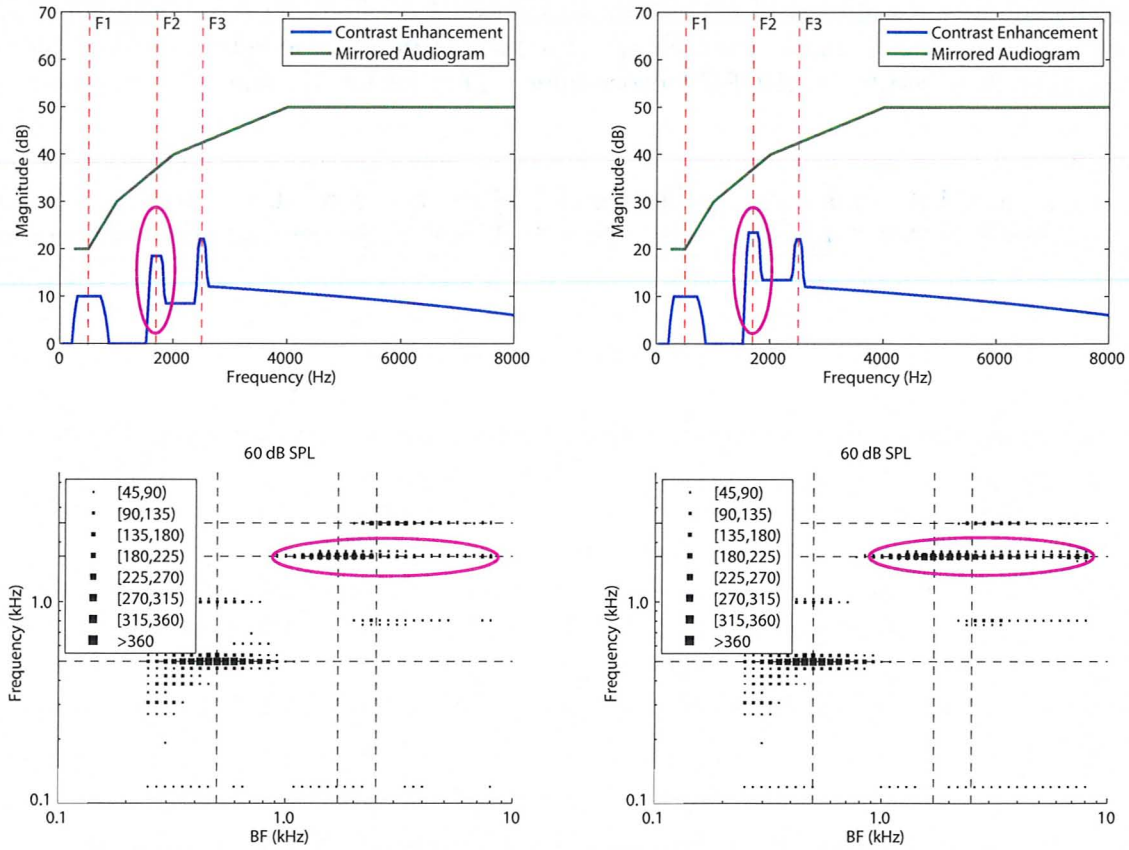

Figure 8.1: Demonstration of the effect of formant pass-band gain on the synchronized response. *Top Left Panel*: Mildly impaired audiogram and the MICEFS-1 prescribed contrast enhancement scheme. *Bottom Left Panel*: Synchronized rate of the mildly impaired AN model with mixed hair cell damage, in response to the MICEFS-1 aided vowel. *Top Right Panel*: As in the top left panel, but the gain at F2 is increased by 5 dB. *Bottom Right Panel*: As in the bottom left panel, but in response to MICEFS-1 with adjusted contrast enhancement.

result of formants with larger amplitudes. In order to reduce the overly synchronized formants in the moderate-to-severe hearing loss case, it could be necessary to individualize the contrast enhancement prescribed for each hearing loss, instead of using variations to the half-gain rule. This would require a non-linear prescription of the contrast enhancement gains as a function of hearing loss.

8.2 Effects of SPC Processing Alone

It is clear from the evaluation of the standalone SPC processor that it had a profound effect on the spatiotemporal response pattern in the model auditory nerve. The relative timing of spiking activity between neighbouring AN fibers was changed, but this was not consistent across fibers synchronized to the same frequency component, or for every cycle in the periodic response. This resulted in an apparently disrupted neurogram for each version of the SPC algorithm; however, SPC-1 preserved some formant synchrony. It is hard to judge perception of the vowel based on visual inspection of the neurograms, but the aided-impaired neurograms seemed to be more deviant from their healthy counterparts than the unaided-impaired neurograms were, suggesting at least some feature of the AN response was degraded.

The synchronized rate of the SPC aided vowel shown in the box-plots in the previous chapter indicate that synchrony to the formants was disrupted for each hearing loss type. As well, it appeared that fibers had some response at frequencies other than the vowel's formant frequencies. This response could be a result of frequency modulations that are present in the SPC aided signal, which is discussed in Section 8.4.4. The disrupted synchrony by SPC caused the aided box-plots to appear worse than the unaided box-plots, suggesting that SPC has a negative effect on vowel identification.

Despite these negative effects, the peak-lag error as a function of SPL was not affected negatively by using SPC alone. Using SPC-1 as a standalone processor improved the peak-lag error as a function of SPL for all hearing loss types evaluated. The most improvements were seen for the mild hearing loss with mixed hair cell damage, where the peak-lag error at low SPLs was significantly decreased and resulted in a relatively flat function of SPL. The improvements seen for the mild loss with only outer hair cell damage and for the moderate-to-severe loss were less profound, but still noticeable. Using SPC-2 improved the peak-lag error as well, and its performance was close to that of SPC-1 for all hearing loss types, although for the moderate-to-severe loss it did not provide much improvement over the unaided response.

Since the peak-lag contains information about the relative timing of neighbouring AN fiber response, a consistent peak-lag as a function of SPL suggests that the relative phase of fibers is more consistent as a function of SPL. Section 8.5 discusses whether this can give some insight into loudness perception and if flattening of the peak-lag error as a function of SPL can signify reduced recruitment of loudness in an impaired ear. Regardless of loudness perception, a reduction in peak-lag error signifies that a feature of the AN spatiotemporal response pattern is closer to normal.

The results of evaluating with SPC alone suggest that both SPC-1 and SPC-2 performed better than SPC-0. SPC-1 had noticeably less distortion of the low frequency response in the neurograms, and SPC-2 seemed to have less distortion at low SPLs. As well, both the SPC-1 and SPC-2 processors seemed to improve the

peak-lag error as a function of SPL while SPC-0 did not. These results suggest that it is important to ensure that insertion delays prescribed for a particular time-point in the modelling path of SPC are inserted at the corresponding time-point in the processing path.

8.3 Combined Effects of MICEFS and SPC

The results in Chapter 7 show that the neurograms in response to the synthesized vowel, aided with the combination of MICEFS and SPC, were altered from their healthy counterparts in much the same way as occurred by using SPC alone. However, there were higher overall rates of discharge for the higher frequencies, as is expected from the amplification of F2 and F3 by the inclusion of MICEFS. The spatiotemporal pattern for the MICEFS-1/SPC-2 aided response was altered from the unaided response in such a way that the relative timing of spiking probability between neighbouring fibers was inconsistent across a synchronized AN population. MICEFS-1/SPC-1 altered the spatiotemporal response pattern as well, but there seemed to be more consistency in the relative phase of fibers that were synchronizing to the same frequency component. As well, AN populations with synchronized activity were more readily identifiable for the MICEFS-1/SPC-1 case.

In addition, the combined processor MICEFS-1/SPC-1 was more effective than MICEFS-1/SPC-2 at reducing the peak-lag error and did not disrupt the formant synchrony as much as MICEFS-1/SPC-2. This suggests that the lowpass filtering of the insertion delays by SPC-2 results in worse performance than SPC-1 in terms of formant synchrony and reduction of peak-lag error.

For the moderate-to-severe hearing loss, pre-processing SPC-1 with MICEFS-1 resulted in an improved peak-lag error as a function of SPL. In this case, using MICEFS-1 to restore synchrony seemed to be beneficial for the operation of SPC-1, which is consistent with the motivation described in Chapter 4. However, for the mild hearing loss, using SPC-1 alone provided better reduction in peak-lag error than did the combination of MICEFS-1/SPC-1. It is not clear why MICEFS-1 was more beneficial for the moderate-to-severe case but it is evident that it resulted in stronger synchronization than for the mild hearing loss types. This may have meant that more synchrony was preserved when SPC-1 was applied, which could affect peak-lag position.

8.4 Limitations to SPC Processing

During the course of this thesis, several technical and conceptual limitations to the SPC processing scheme were discovered. These remain unresolved, but provide insight

into some of the negative effects of processing with SPC that likely resulted in poor performance during human testing by Shi *et al.* (2006). The limitations will be discussed here.

8.4.1 Non-Flat Frequency Response in SPC

The filterbank used by Shi *et al.* (2006) in their implementation of SPC was the one described by Hohmann (2002). In this implementation, upon completion of delay correction, the filterbank synthesis procedure combines individual frequency channels in a weighted sum described by Equation 5.19. The specific channel weights are given by Hohmann as a means of restoring a flat filterbank gain-frequency response, but the procedure for deriving these weights is not explained by Hohmann or Shi *et al.*. The channel weights used in this thesis are derived using Equations 5.20-5.21, which result in a relatively flat magnitude spectrum for the filterbank's impulse response (shown in Figure 5.9).

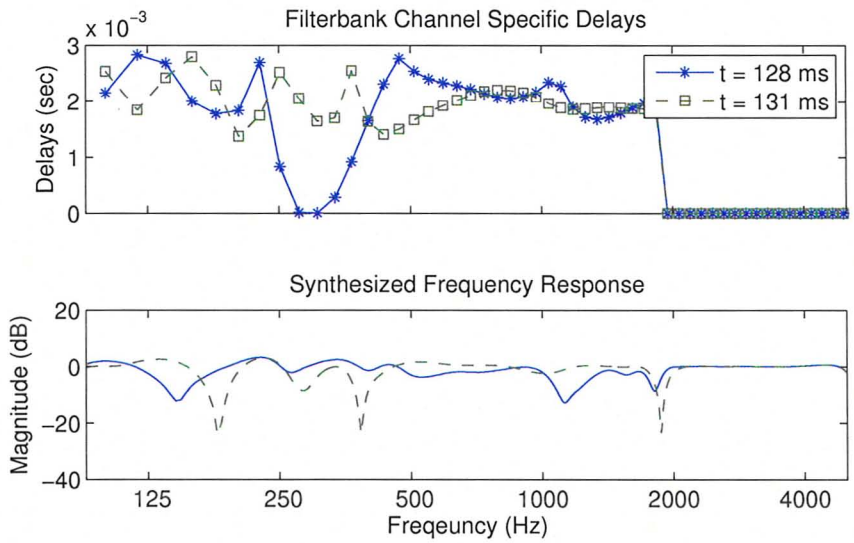

Figure 8.2: Top: The delays prescribed by SPC for each channel in the filterbank at two time-points. Channel center frequencies are indicated with markers. Bottom: Gain-frequency response of the filterbank at the time-points given in the top panel.

It has been determined that the filterbank's gain-frequency response becomes non-flat when channel specific delays are applied in addition to the overall specified delay of the Hohmann filterbank. Figure 8.2 demonstrates the behaviour of the filterbank's gain-frequency response when example insertion delays are applied at two time points. The channel delays shown in the figure's top panel come from the response to the

/ɛ/ vowel. As a result of these delays, there are clear valleys in the filterbank's gain-frequency response.

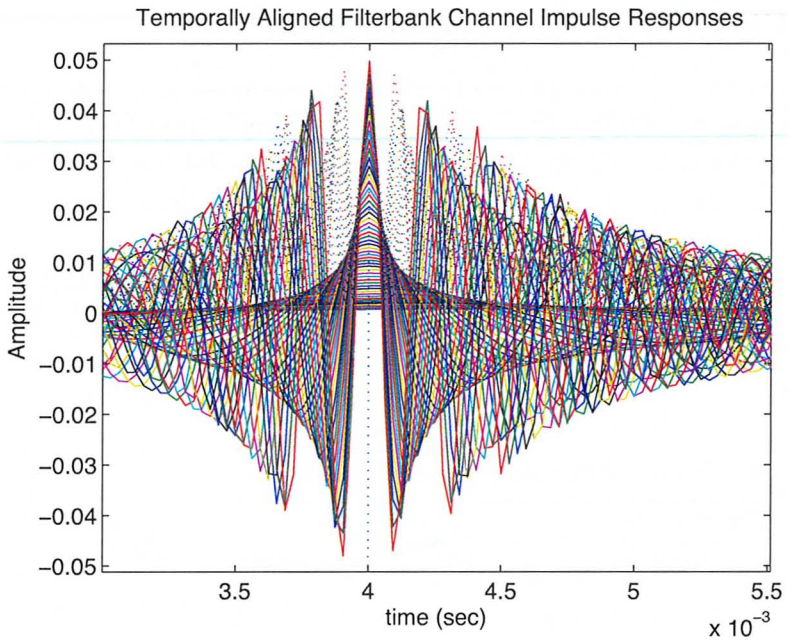

Figure 8.3: Impulse responses of individual frequency channels in the SPC filterbank aligned temporally at 4 msec.

The valleys that are apparent in the gain-frequency response in Figure 8.2 likely result from an alignment of filterbank channels that causes destructive interference in the summation of individual channel impulse responses. When there are no channel specific delays, individual filterbank channel impulse responses are aligned at 4 msec in a procedure that was illustrated in Figure 5.8. The aligned impulse responses of each filterbank channel that result from this procedure are shown in Figure 8.3. Using Figure 8.3 as a reference, it is easy to visualize that by applying channel specific delays, the impulse response alignment will be degraded and could result in destructive interference when filterbank channels are combined.

This is an inherent problem of the SPC algorithm which seeks to insert time-varying channel specific delays into the filterbank. To account for this problem, it seems that channels should be re-weighted continuously. However, the gain-frequency response will be unique at every instant, and it is likely impractical to measure this dynamically.

8.4.2 SPC Response to a Pure Tone

It has been determined that accounting for the difference in group delays between a healthy and impaired AN model cannot restore the normal spatiotemporal response to a pure tone. The normal spatiotemporal response to a pure tone can be visualized in the right panel of Figure 2.16, where the periodic peaks in spiking probability are dispersed in time across the AN population at low SPLs. The large relative phase between neighbouring AN fibers that are synchronized to the pure tone is consistent with a steep phase transfer function at low SPLs for individual AN fibers in the population. This can be noted from the auditory filter transfer function in the left panel of Figure 2.16.

In the impaired AN response, the individual AN fibers have a shallower phase transfer function at low SPLs, resulting in more coincident spiking across the synchronous population. One problem with using a filterbank that corrects for the impaired group delays through auditory filters using integer delays in order to restore the healthy response to a pure tone, is that applying a delay or advance to a pure tone will not change the relative phase response between fibers. It would simply shift the overall auditory nerve response in time. This would result in the same degree of coincident spiking.

Another problem with using the SPC filterbank in order to restore the response to a pure tone is that there is only one frequency component in the pure tone, while the filterbank corrects for group delay at many frequencies. Each AN fiber that is responding to the pure tone is generally synchronized to it, so it does not make sense to correct for shortened delays in each auditory filter evaluated at their characteristic frequency. This is because in most cases the AN fiber's characteristic frequency will not be equal to the frequency of its synchronized response.

8.4.3 Group Delay Based on Incorrect Frequencies

In Chapter 4, it was argued that impaired auditory nerve response had disrupted the synchrony to the formants of voiced speech, and the original calculation of group delay would therefore be based on AN response that lacked synchrony. In other words, the group delay calculations might be based on incorrect synchronization frequencies. In order to address this, MICEFS was employed first to restore synchronization in the impaired AN response before group delay was calculated for the SPC scheme. This reasoning remains sound, however it has been determined that even if perfect synchronization could be restored by MICEFS, group delay calculations in the SPC scheme could still be based on the wrong frequencies.

The group delay calculations used in the SPC processing scheme are evaluated at each model C1 filter's center frequency, or the AN fiber's characteristic frequency. However, the dominant frequency that is passing through the filter is not always at

this center frequency due to spread of synchrony that occurs in the healthy auditory nerve. This has been noted above for a pure tone, and in this case, the likelihood that an auditory filter is synchronized to its center frequency is small. Section 5.2.1 explained that it is difficult to calculate the frequency of synchronization in each auditory fiber continuously in time, and therefore the group delay is evaluated at the auditory fiber's characteristic frequency. This seems to be the best compromise, however it means that evaluation of group delay at incorrect frequencies must be tolerated.

Another confounding issue arises out of the mismatch in the frequency response of the individual filters in the modelling path and the processing path. The auditory filters in the modelling path are non-linear and broaden as stimulus level increases, causing spread of synchrony to a particular frequency component across a range of fiber CFs. Conversely, the gammatone filters in the processing path are of fixed bandwidth. This means that the group delay through a wide range of auditory filters can be driven by a single synchronization frequency, whereas the prescribed insertion delays that result are applied to a range of stimulus frequency components, instead of just the single synchronization frequency.

8.4.4 Frequency Modulations as a Result of Inserting Delays

Applying time-varying delays to the filtered signals in the SPC filterbank causes frequency modulation in the SPC-delayed channel. The top panel of Figure 8.4 shows the filtered synthesized /ɛ/ vowel in the 1344 Hz channel of SPC-1. In the figure's middle panel, the insertion delays prescribed for this channel are shown. The bottom panel shows the filtered vowel with the SPC delays applied to it. The frequency content in this SPC-delayed channel is noticeably different than the un-delayed channel. It is apparent that the insertion of delays causes this frequency modulation.

The SPC-delayed channel appears to be of a much poorer quality than the unfiltered channel; this is an unintended consequence of SPC processing. However, it is the SPC delays that cause these frequency modulations, meaning that they are likely an unavoidable consequence of attempting to correct for impaired auditory filter group delays in this way.

Abnormal frequency content remains in the synthesized SPC-1 processed signal, after all the channels have been combined. Figure 8.5 shows the unprocessed synthesized /ɛ/ vowel on top of the SPC-1 processed vowel. It is apparent that the time-domain waveform of the SPC-1 is quite different than the unprocessed waveform; this results in audible noise for listeners with normal hearing.

Applying a low-pass filter to the insertion delays, as prescribed by SPC-2, to reduce the noise also causes frequency modulation, but the modulations are much slower. Figure 8.6 shows the slower modulations that result from lowpass filtering the

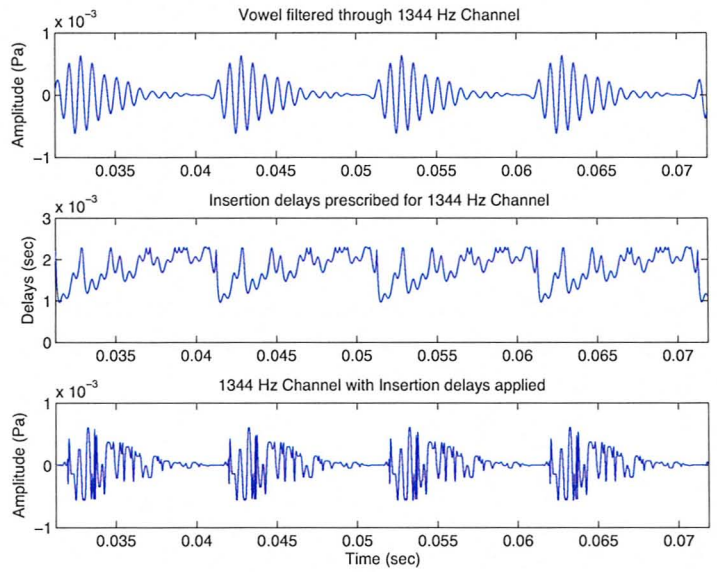

Figure 8.4: Demonstration of frequency modulations in individual channels by SPC-1. *Top Panel*: The synthesized /ɛ/ vowel filtered through the 1344 Hz SPC filterbank channel. *Middle Panel*: Insertion Delays prescribed by SPC-1 for this channel. *Bottom Panel*: Result of applying insertion delays to the filtered vowe.l

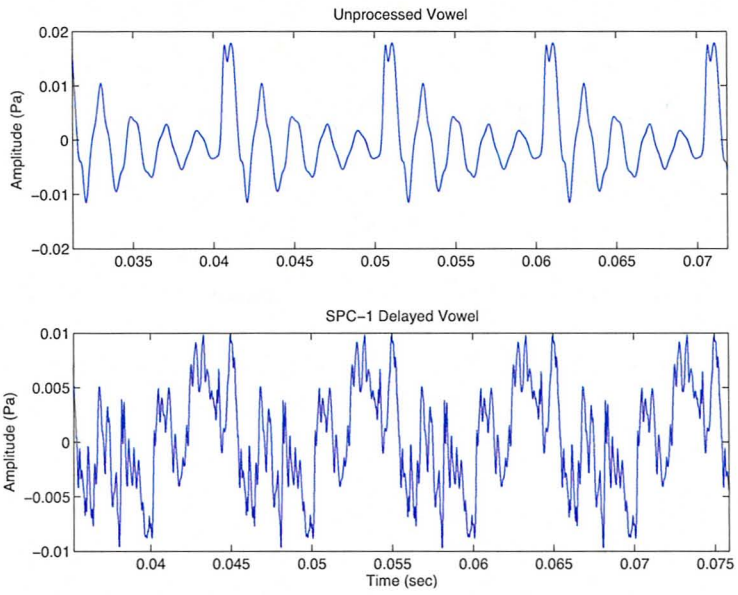

Figure 8.5: *Top Panel*: Unprocessed synthesized /ɛ/ vowel. *Bottom Panel*: SPC-1 processed vowel.

insertion gains prescribed for the 1344 Hz channel. Figure 8.7 shows the synthesized SPC-2 processed signal. There is clearly less high frequency power that is present in this signal compared to the synthesized SPC-1 processed signal in Figure 8.5.

This SPC-2 processed waveform results in less high frequency noise that is audible to listeners with normal hearing. But Chapter 7 demonstrates that this processor results in a deterioration of the synchrony to formants, and produces synchronized response to non-formant frequencies at high levels. The poor performance of SPC-2 in Chapter 7 follows from the fact that the insertion delays were not the actual delays prescribed by the model, but rather a lowpassed version. Any benefits that may have resulted in de-noising the processed signal have degraded the ability to restore the impaired spatiotemporal response to normal.

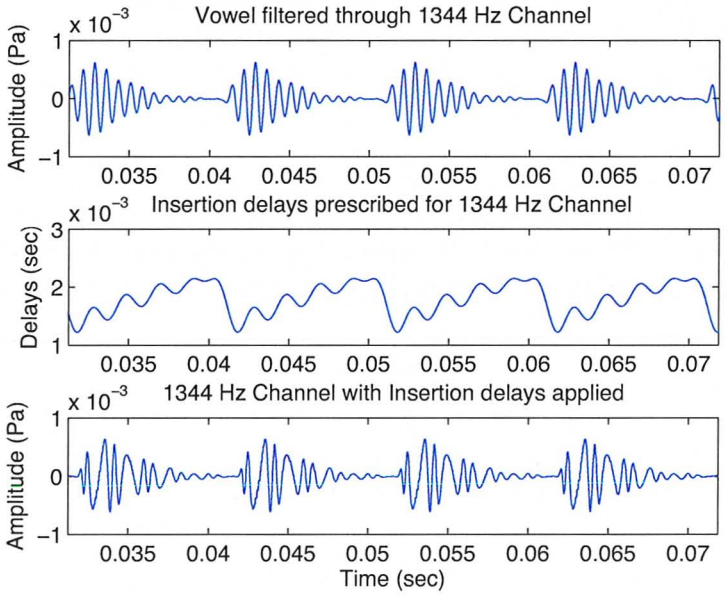

Figure 8.6: Demonstration of frequency modulations in individual channels by SPC-2. *Top Panel*: The synthesized /ɛ/ vowel filtered through the 1344 Hz SPC filterbank channel. *Middle Panel*: Insertion Delays prescribed by SPC-2 for this channel. *Bottom Panel*: Result of applying insertion delays to the filtered vowel.

As well, the SPC-1 processor performed better than the SPC-2 processor; a somewhat counterintuitive result given the appearance of the temporal waveforms of the vowel processed with the two SPC versions in Figures 8.5 and 8.7. This suggests that the noise associated with SPC-1 may be unavoidable in order to correct for impaired auditory filter group delays in this way. As well, it is hard to assess the consequences of a noisy processed vowel because the noise may not be audible to listeners with

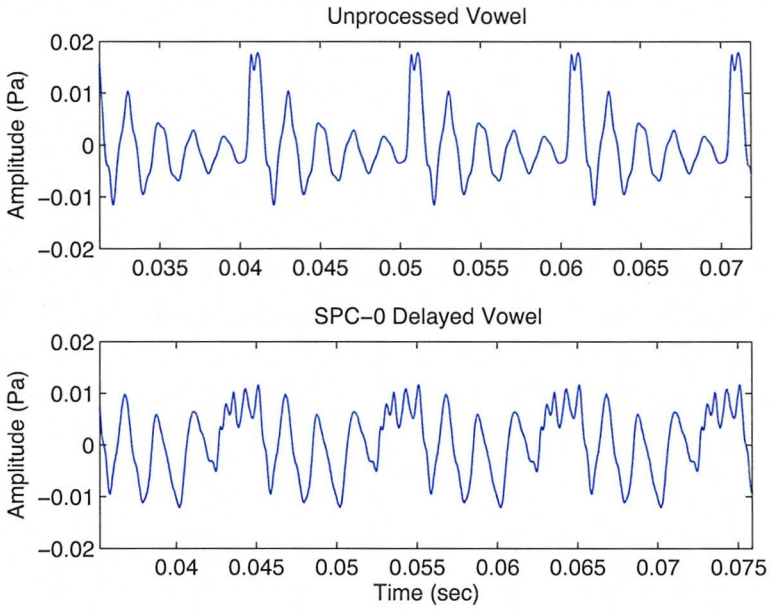

Figure 8.7: *Top Panel*: Unprocessed synthesized /ɛ/ vowel. *Bottom Panel*: SPC-2 processed vowel.

hearing impairment.

8.4.5 Response to Amplitude Modulated Stimuli

The insertion delays in the SPC processing scheme are not applied above 2 kHz because of reports of decreased benefit at high frequencies (Shi *et al.*, 2006). This is presumably due to a roll-off of phase locking to tonal stimuli that likely occurs around 2 kHz (Pickles, 2008). However, Joris and Yin (1992) have shown that AN fibers can phase lock to the low frequency envelope of amplitude modulated tones, which elicit responses similar to those generated with single formant speech (Kale and Heinz, 2010). As well, recent data shows that AN fibers in an ear with noise induced hearing loss demonstrate early and more salient onset responses to tone bursts (Scheidt *et al.*, 2010). Increased onset response in AN fibers is consistent with recently reported enhanced envelope coding following hearing loss, which may degrade speech perception by relatively suppressing the coding of temporal fine structure (Kale and Heinz, 2010). It is not clear whether the earliness of the onset response also increases the salience of the envelope relative to the temporal fine structure. But it may be worthwhile for SPC to address this issue by accounting for the group delay of amplitude modulated signals at high CFs.

8.5 Peak-Lag Error as it Relates to Loudness

Carney (1994) proposed that loudness could be encoded in the relative timing of the discharge pattern of a population of AN fibers and decoded by coincidence detecting neurons in the cochlear nucleus. The relative timing of the response to periodic stimuli of neighbouring AN fibers varies systematically as a function of level and it is clear from the model response that the peak-lag of their cross-correlation also varies systematically as a function of level. If the relative phase response between two fibers is small, as in the healthy AN response to high intensity stimuli, then the peak-lag between those fibers will be small. Conversely, if the relative phase response between two fibers is large, as in the healthy AN response to low intensity stimuli, then the peak-lag between those fibers will be large.

Section 6.5.1 described that the phase response of sensorineural impaired auditory filters is flattened, which results in a small value of peak-lag. This is similar to the peak-lag value that would be derived from a healthy AN response to high intensity stimuli. The peak-lag error used in the evaluations of Chapter 7 describes the difference in peak-lag between the healthy and impaired neural response; if their synchronized frequencies are similar, then this error too should be systematically dependent on the intensity of the stimulus. It should decrease as the intensity of the stimulus decreases, because as the healthy auditory filter's phase response flattens with level, it approaches the already flattened phase response of the impaired auditory filter. This was indeed the case in response to a three tone stimulus used in the development of the peak-lag error (refer to Figure 6.5, Section 6.5.1).

It will be interesting to note any relationships that the peak-lag error as a function of SPL has on perceived loudness in an impaired ear, and whether or not it is correlated to loudness recruitment. The peak-lag error function that was found in response to a three tone stimulus is compared in Figure 8.8 to some psychophysical data on perceived loudness in the impaired ear that was discussed in Section 2.6.3. Recall that Moore *et al.* (1985) asked listeners with unilateral hearing loss to match the loudness of tones of varying intensity presented to their healthy ear and their impaired ear. The results, reprinted in the top panel of Figure 8.8 for convenience, demonstrate the phenomenon of loudness recruitment in impaired ears, where loudness as interpreted from the impaired ear grows rapidly as a function of sound level. The psychophysical data comes from an average of five subjects chosen to have relatively flat hearing loss with thresholds in the range of 45-60 dB HL. The difference between the curve corresponding to a normal ear and the curve corresponding to an impaired ear in the top panel is normalized and plotted against the normalized peak-lag error versus SPL curve in the bottom panel of the figure. The peak-lag error in this case was found for an impaired ear model with a flat 60 dB hearing loss and in response to the three-tone stimulus. A pure tone was not used in this case due to unpredicted behaviour

at low CFs that has been noted in the healthy model response to pure tones[1]. There is noticable correspondence between the two curves which suggests that the peak-lag error could give some insight about a person's perception of loudness.

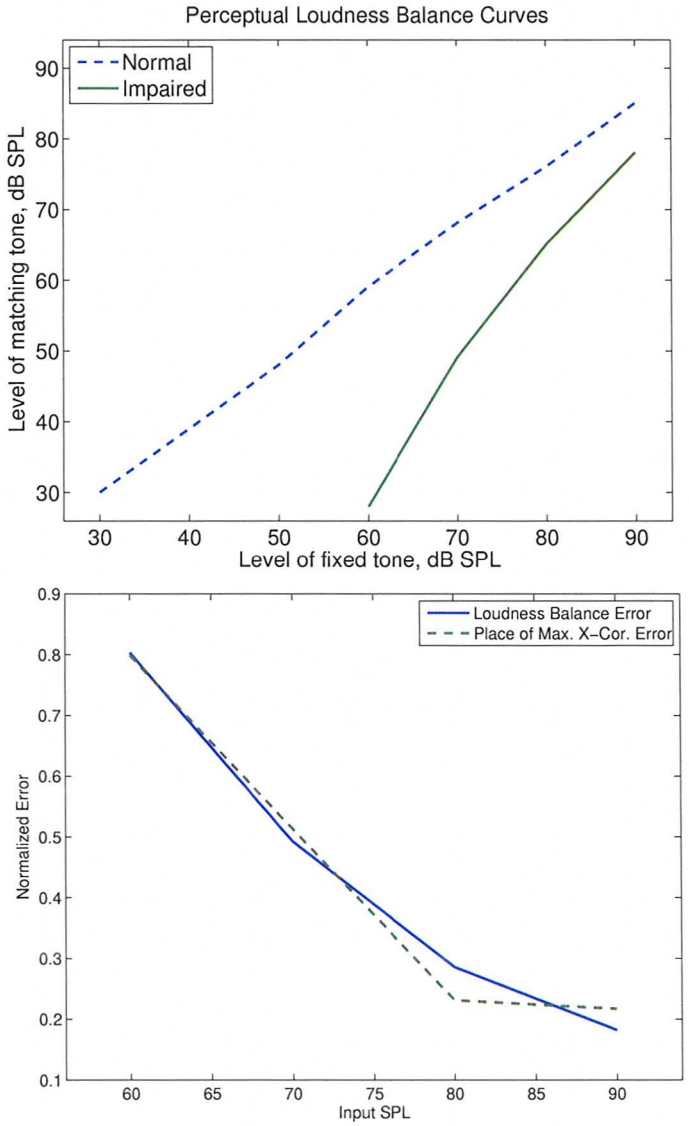

Figure 8.8: *Top Panel*: Average loudness balance curves for normal and impaired subjects. Interpreted from Moore (2003), Fig. 4.9. *Bottom Panel*: The green dashed line shows the error between the healthy and impaired place of maximum neighbouring fiber cross-correlation in response to a tone complex as a function of SPL. The blue line represents perceptual loudness balance error from the top panel.

[1]This unpredicted behaviour is detailed in Appendix B.

The analysis above demonstrates that peak-lag error could be related to loudness perception in impaired ears. However, more results are required to determine whether or not this metric may be used as predictor of loudness recruitment. Comparisons to psychophysical data need to be done, ideally using experiments where participants with unilateral hearing loss match the loudness of vowels, or tone complexes in their two ears. The auditory models must be individualized to the audiograms of the participants. In this way, it is possible that a relationship could be found between loudness recruitment and the peak-lag error.

Chapter 9

Conclusions

9.1 Summary

An improved version of the spatiotemporal pattern correction (SPC) scheme was developed and analyzed with a computational model of the auditory periphery to determine if correcting for impaired model group delays could improve the spatiotemporal response pattern in the impaired auditory nerve. Several technical and conceptual problems with the SPC scheme were discovered including: i) A non-flat gain frequency response through the analysis-synthesis filterbank when delays were applied; ii) frequency modulations that result from inserting time-varying delays; and iii) group delay calculations that are based on potentially incorrect frequencies due to a mismatch between the frequency response in the modelling path filters and the processing path filters, as well as evaluation of group delay at the characteristic frequency. These problems contributed to results demonstrating that SPC disrupts synchrony in the AN to formants and produces artifacts in the spatiotemporal response patterns. An improved version of the SPC scheme developed here (SPC-1), which accounts for a difference in SPC path latencies, significantly reduced these artifacts and demonstrated that the peak-lag error was improved at low SPLs. This meant that a feature of the spatiotemporal response was improved over the impaired case. However, using SPC-1 produced an unavoidably noisy signal, suggesting there is a tradeoff between correcting for group delays in this way and the frequency modulations that result.

The problem of group delays being based on incorrect frequencies was alleviated by pre-processing the signal with an improved version of the multiband and improved contrast-enhanced frequency shaping (MICEFS) algorithm. This seemed to improve the synchrony in the spatiotemporal response pattern relative to using SPC as a standalone processor for all hearing loss types. In addition the peak-lag error was improved at low SPLs for the moderate-to-severe hearing loss. However, the peak-lag error was degraded for the combined MICEFS/SPC processor relative to the SPC

standalone processor alone for the mild hearing losses, but was still better than the unaided peak-lags. This suggests that listeners with moderate-to-severe loss gain more benefit from pre-processing to restore synchrony.

The version of MICEFS developed in this thesis (MICEFS-1) demonstrated better synchrony to the first formant of voiced speech when applied to the audiograms used here. For the moderate-to-severe hearing loss case, MICEFS demonstrated an over-representation of synchrony to F2 and F3. This, combined with the results of the combined MICEFS/SPC processor, suggests that: i) the synchrony in the AN response is quite sensitive to the relative amplitude of formants, ii) the contrast enhancement prescribed by MICEFS might have to be a non-linear function of hearing loss for better individualization, and iii) over-representation of synchrony by pre-processing with MICEFS could result in less disrupted synchrony after SPC processing.

In addition to these findings, the cross-correlation peak-lag metric was developed to compare the relative timing of the AN response of neighbouring fibers. This error was found to systematically decrease as a function of stimulus level in response to a synthetic vowel for mild and moderate-to-severe hearing losses. This is consistent with the behaviour of healthy and impaired auditory filters as they change with intensity level. Preliminary comparisons to psychophysical data suggest this metric could be correlated with loudness perception. If further tests support this idea, SPC-1 and the combined processing of MICEFS-1/SPC-1 may be able to improve loudness perception at low levels despite the disruptions to the spatiotemporal response pattern that they cause.

9.2 Suggestions for Future Work

This work has demonstrated several technical limitations with the SPC processor that contributed to disruptions in the spatiotemporal response, and produced noise that is audible to normal hearing listeners. It is likely that these limitations also contributed to the generally poor results of human testing found by Shi *et al.* (2006) and Calandruccio *et al.* (2007). Future works could begin to address these technical limitations.

It may be possible to alleviate the problem of C1 filter group delays in the SPC modelling path being based on incorrect frequencies even further by estimating the frequency of each C1 filter's output. This would allow for more accurate group delay calculations, however there would still be issues if multiple frequency components passed through the C1 filter. If the frequency of each C1 filter's output was estimated, then it might also be possible to correct for the phase delay of that frequency component in relative filterbank channels in addition to the group delay. As well, the problem of a non-flat SPC frequency response might be alleviated by dynamically re-weighting the the filterbank's channels, however it would be difficult to update the

weights in real-time. These solutions all involve a significant computational cost, and it is unclear whether they could be implemented on a real hearing aid at present. However, if the auditory model demonstrates that these suggestions improve the spatiotemporal response significantly, it may be important to achieve them in future.

The auditory model of Zilany and Bruce (2006, 2007) was used in this implementation of the SPC scheme because it made several improvements over the model of Heinz *et al.* (2001) that was used in the original SPC implementation by Shi *et al.* (2006). However, this meant that the tuning properties of the BM were based on cat physiology. Recent work by Ibrahim and Bruce (2010) modifies the cat model to incorporate the sharper cochlear tuning that has been estimated in humans (Shera *et al.*, 2002). Future work could incorporate human tuning into the SPC scheme so that spread of synchrony is better represented.

Appendix A

Equivalent Rectangular Bandwidth

The ERB of an arbitrary bandpass filter is defined as the bandwidth of a rectangular filter whose amplitude is equal to the maximum transmission of the filter's passband, and the total power that would pass through the two filters in response to white noise is equal (Moore, 2003). This can be represented graphically by plotting an equivalent rectangular filter alongside the filter of interest, as in Figure A.1. Note that the ERB is defined for the filter of interest (*in blue*) and is equal to the bandwidth of the rectangular filter (*in green*).

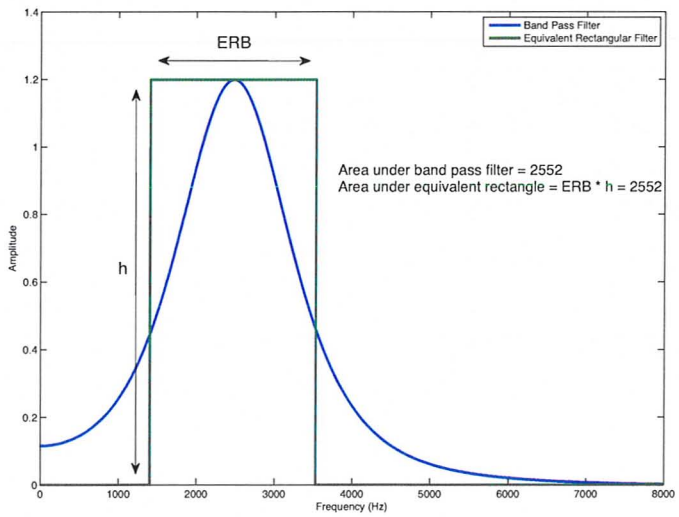

Figure A.1: Bandpass filter and its Equivalent Rectangular Bandwidth (ERB).

ERB of an auditory filter

Specific ERBs have been defined for auditory filters as a function of their center frequency and are given by Glasberg and Moore (1990) as

$$\mathrm{ERB}_{\mathrm{Aud}}(F) = 24.7(4.369 \times F + 1) \tag{A.1}$$

where F has units of kHz. This is equivalent to

$$\mathrm{ERB}_{\mathrm{Aud}}(f) = 24.7 + \frac{f}{9.265} \tag{A.2}$$

where f has units of Hz.

ERB scale

Frequency can also be scaled in terms of the number of auditory filter ERBs. The ERB scale is related to the linear frequency scale by [Hohmann (2002)]:

$$\mathrm{ERB}_{\mathrm{scale}}(f) = q \cdot \log(1 + \frac{f}{l \cdot q}) \tag{A.3}$$

$$f = (\exp(\frac{\mathrm{ERB}_{\mathrm{scale}}}{q}) - 1) \cdot l \cdot q \tag{A.4}$$

where $q = 9.265$, $l = 24.7$, and f denotes the linear frequency in units of Hz. Equal spacing along the ERB scale results in an approximately logarithmic spacing of frequencies in units of Hz.

Appendix B

Unpredicted behaviour of the Zilany and Bruce AN model in response to a tone

When stimulating the Zilany and Bruce (2006, 2007) cat auditory nerve model with a mid-frequency tone, some interesting behavior is observed in the response of healthy low-CF fibers. These low-CF fibers do not appear to synchronize completely to the stimulating tone, and sometimes have a higher synaptic spiking rate than fibers that are tuned to the stimulating tone. An example of this is seen in Figure B.2 which shows the instantaneous probabilities of spikes occurring across AN fibers in the left column, and the peri-stimulus time histogram in the right column in response to a 1700 Hz tone. Each row shows the response to increasing stimulus sound pressure levels and the area indicated in red shows the unsynchronized low CF behavior.

The outputs of the C1 filter and the C2 filter stages in the model were analyzed to determine whether the C1/C2 transition could be occurring at lower than expected SPLs (i.e. by 80 dB SPL). These signals are compared for a fiber tuned to 200 Hz in response to the 1700 Hz tone at 80 dB SPL in Figure B.3. As can be seen clearly from the figure, the C1 signal is dominant even at this relatively high signal level of 80 dB SPL. The output of the C2 filter also indicates that it is completely synchronized the the stimulating tone. After passing these signals through their separate C1 and C2 transduction functions, the C2 signal is negligible (see Figure B.4) and the unsynchronized behaviour is captured completely in the C1 signal. Figures B.3 and B.4 indicate that the C1/C2 transition has not yet occurred at a level of 80 dB SPL, and therefore the C1 signal pathway must be investigated to determine the cause of the unsynchronized low CF behaviour.

The time constant of the C1 filter is the output of the model's control path and it determines the bandwidth of the C1 signal-path filter. It also affects the bandwidth of

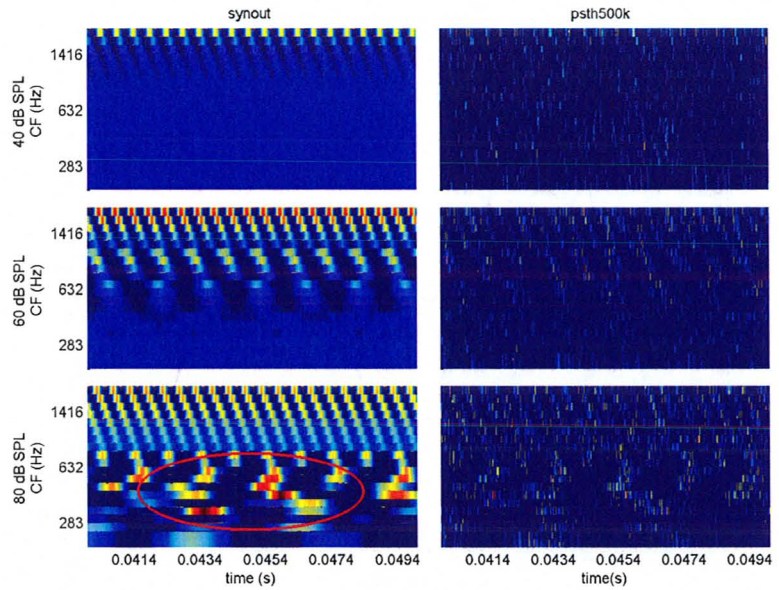

Figure B.2: *Left Column*: Healthy synaptic neurogram in response to a 1700 Hz tone. Descending rows correspond to tone presentation levels of 40, 60, 80 dB SPL. For all presentation levels, fibers with high characteristic frequencies (CF > 1200 Hz) are synchronized to the 1700 Hz tone. As presentation levels increase, low-CF fibers do not demonstrate synchrony capture by the 1700 Hz tone (see area indicated by the red ellipse). *Right Column*: Healthy PSTH across fiber best frequency in response to the same tones. Number of stimulus repetitions is 100. Fiber spontaneous rate is 50 spikes/sec.

the control path filter. Figure B.5 shows the behaviour of the C1 filter time constant, τ_{C1}, and the resultant C1 filter output in response to the 1700 Hz tone. This figure demonstrates that the unsynchronized low CF response is inherent in the value of the C1 filter time constant and is likely the cause of the unsynchronized synaptic spiking probabilities.

The time constant τ_{C1} can change on a cycle-by-cycle basis which means some cycles can be missed in the prescribed spiking probabilities. The reason for the fluctuation of τ_{C1} at low CF fibers could be because the stimulating tone falls on the upper edge of the the control filter's pass-band. Because the output of the control filter determines τ_{C1} and because of the feed-forward path, the control filter's pass-band can change on a cycle-by-cycle basis.

The behaviour of the control signal τ_{C1} seems to follow logically from the control path design. This raises the question of whether or not this behaviour is physiologically reasonable. And, if this behaviour is not supported by physiological measurements, whether any modifications to the model can be made to account for this.

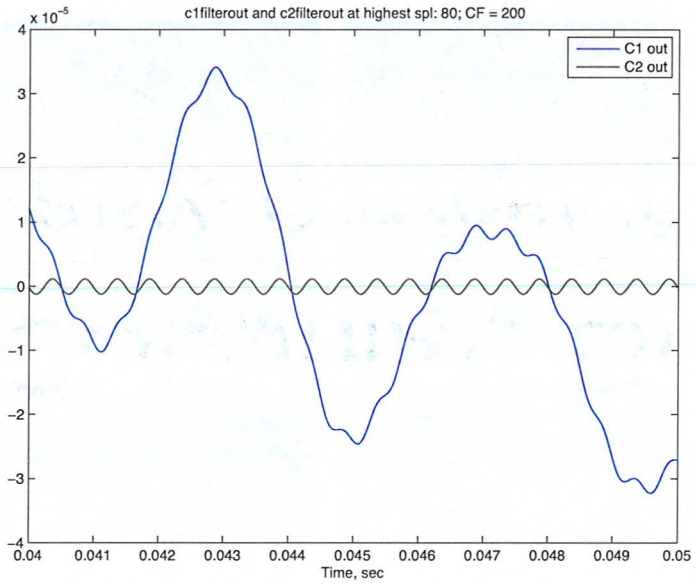

Figure B.3: The outputs of the C1 and C2 filters in the cat auditory nerve model Zilany and Bruce (2007) at a fiber with CF = 200 Hz, in response to a 1700 Hz tone at 80 dB SPL.

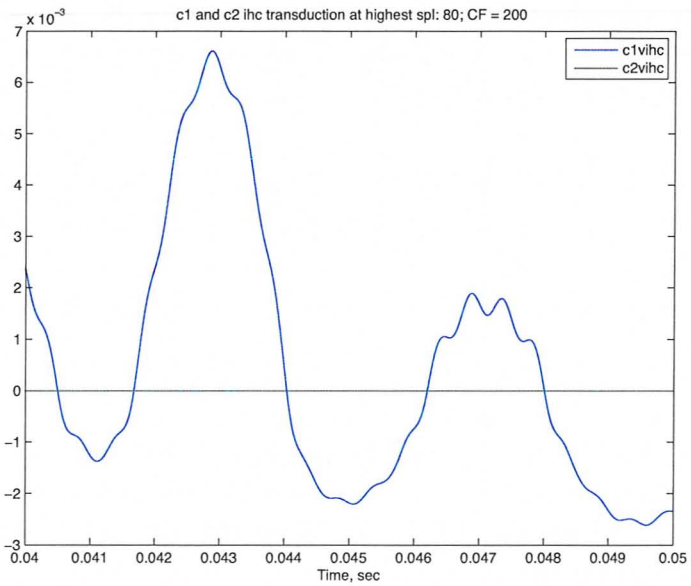

Figure B.4: The outputs of the C1 and C2 transduction functions in the cat auditory nerve model Zilany and Bruce (2007) at a fiber with CF = 200 Hz, in response to a 1700 Hz tone at 80 dB SPL. The C2 output is negligible.

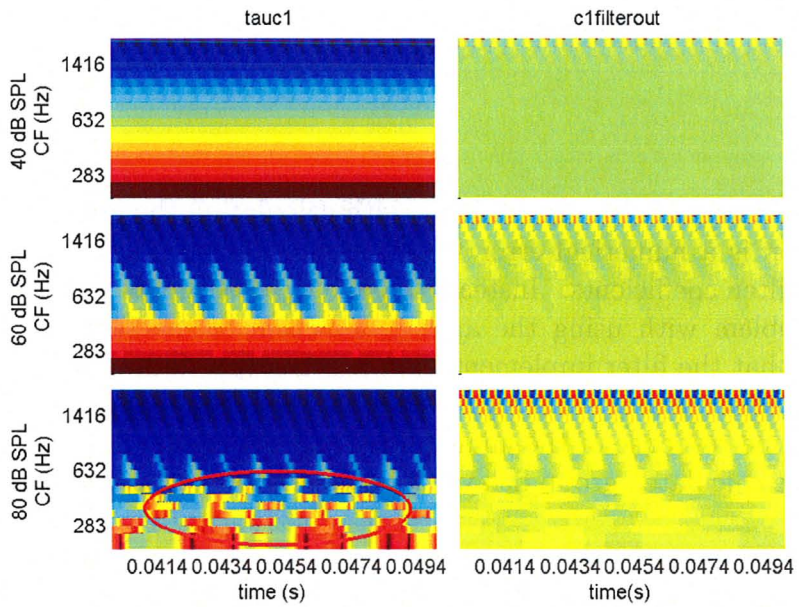

Figure B.5: Left Column: Healthy time constant τ_{C1} across AN fibers for 3 different levels of input: 40, 60 and 80 dB SPL. Right Column: C1 filter outputs in response to the same tones. Note that each color axis is independent but not shown in the interest of space.

Appendix C

Analog vs. Digital C1 Filter Group Delay

As described in Section 5.2.1, the problem with calculating the group delay from the digital definition is a numerical error which results from the large difference in the magnitude of filter coefficients. Instead the analog coefficients are used. However, a potential problem with using the analog definition of the filter for deriving the group delay is that the filter implementation is actually a digital one, found through the bilinear transformation. This means that as frequency increases, there will be a discrepancy in the response of the analog filter and the digital filter, and therefore a difference in the calculated group delays. This appendix demonstrates that this error is tolerable.

Figure C.6 shows the calculated group delay and the measured group delay for filters with center frequencies of 1 and 5 kHz. The group delay is calculated for the Zilany and Bruce C1 filter with a variety of methods. The methods based on the digital implementation of the C1 filter use the IIR coefficients, as labelled in the figure's legend, and result in a relatively flat function of group delay. This is a result of numerical error. The method based on the analog definition of the C1 filter, labelled 'Analog calcs (Pole/Zero)', results in a relatively accurate calculation of group delay. The actual group delay is then measured using the filter's impulse response and via the delay experienced by a number of amplitude modulated tones when they are filtered. The top panel of Figure C.6 shows that the analog calculations are the same as the measured group delay for a center frequency of 1 kHz. For a center frequency of 5 kHz, the analog calculation is slightly less than the measured group delay, however the error is very small. The largest error is approximately 2 samples based on a 100 kHz sampling rate, which corresponds to less than one sample for the 32 kHz sampling rate that is used in SPC. Since the SPC only inserts delays up to 2 kHz, the error due to using the analog definition of the C1 filter to find group delay

should be negligible.

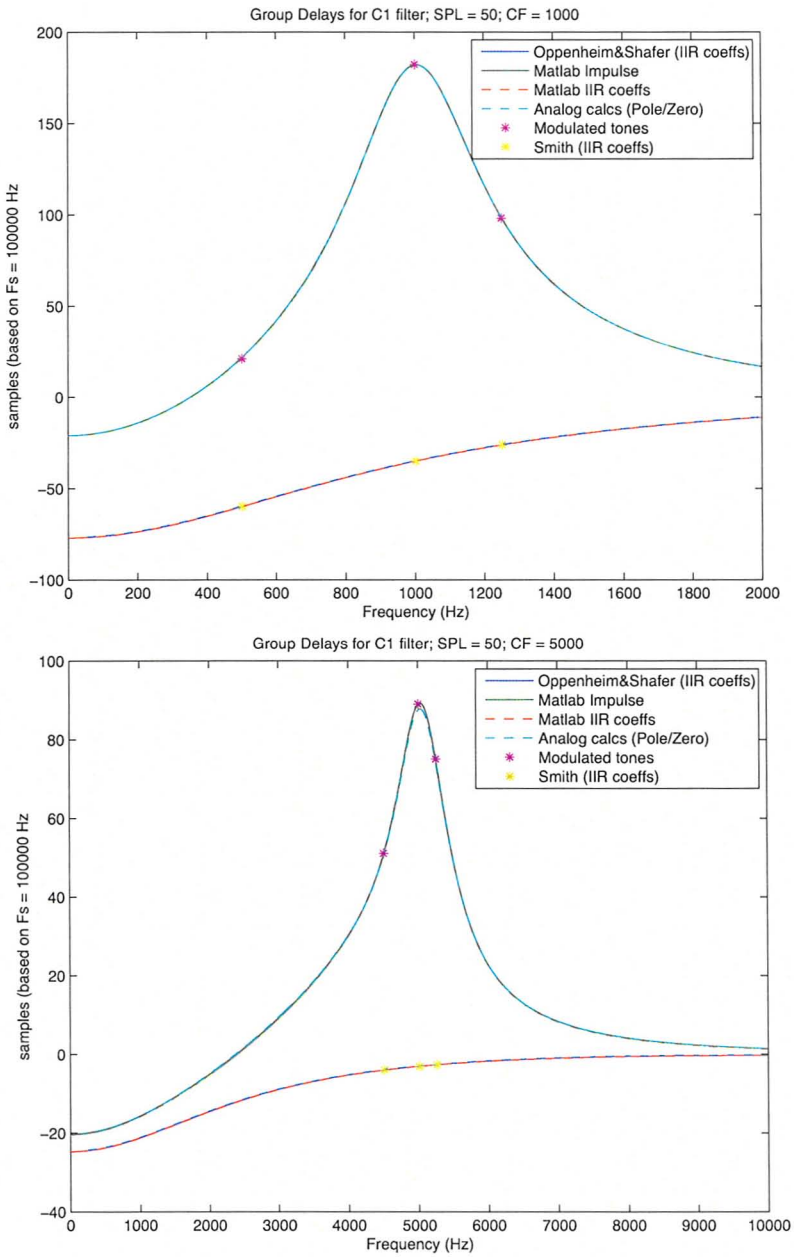

Figure C.6: *Top Panel*: Group delay versus frequency for a 1000 Hz tone found via a variety of methods. Note the error in using all the methods which use the digital IIR coefficients. *Bottom Panel*:Group delay versus frequency for a 5000 Hz tone

Appendix D

Implementation of MICEFS in Simulink

The MICEFS algorithm used in this thesis was first implemented in Matlab for development and later implemented using Simulink for future programming of a hearing-aid DSP (Digital Signal Processing) chip. Implementation on a DSP will allow for real time processing of natural sounds. Simulink is a development environment that facilitates the design of signal processing systems using pre-defined blocks that can be organized to simulate a hearing aid. For example, Simulink contains a pre-defined block that implements the STFT for conversion of the temporal waveform into the frequency domain and a block for the overlap-and-add (OLA) for signal reconstruction and conversion back to the time domain. These STFT blocks were employed in this implementation of MICEFS, and appear in Figure D.7, which is the parent Simulink model file for MICEFS-0. Each component of the MICEFS processor was required to be converted to Simulink, including multiband compression, contrast enhancement and the formant tracker.

Formant Tracker Implementation

Implementation of MICEFS in Simulink required that the formant tracker of Mustafa and Bruce (2006) be re-coded in Simulink and integrated into the MICEFS processor. The Simulink model block diagram of the formant tracker is shown in Figure D.8. In order to program systems for real-time use, Simulink requires the use of pre-defined blocks that are found in its library. As well, it requires that systems be causal; that is, systems containing feedback should be sensitive to the order in which certain blocks are processed. For these reasons, the implementation of a number of processing blocks within the formant tracker used in this thesis differs from its original Matlab implementation as described by Mustafa and Bruce (2006).

One such instance is the implementation of the highpass and lowpass filters in

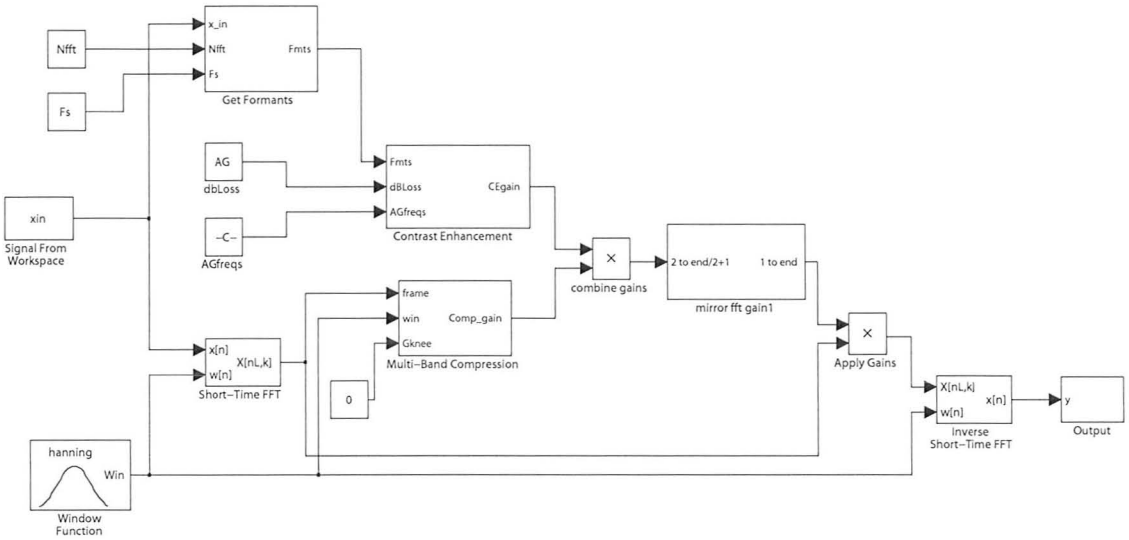

Figure D.7: Simulink model of the MICEFS-0 processing scheme.

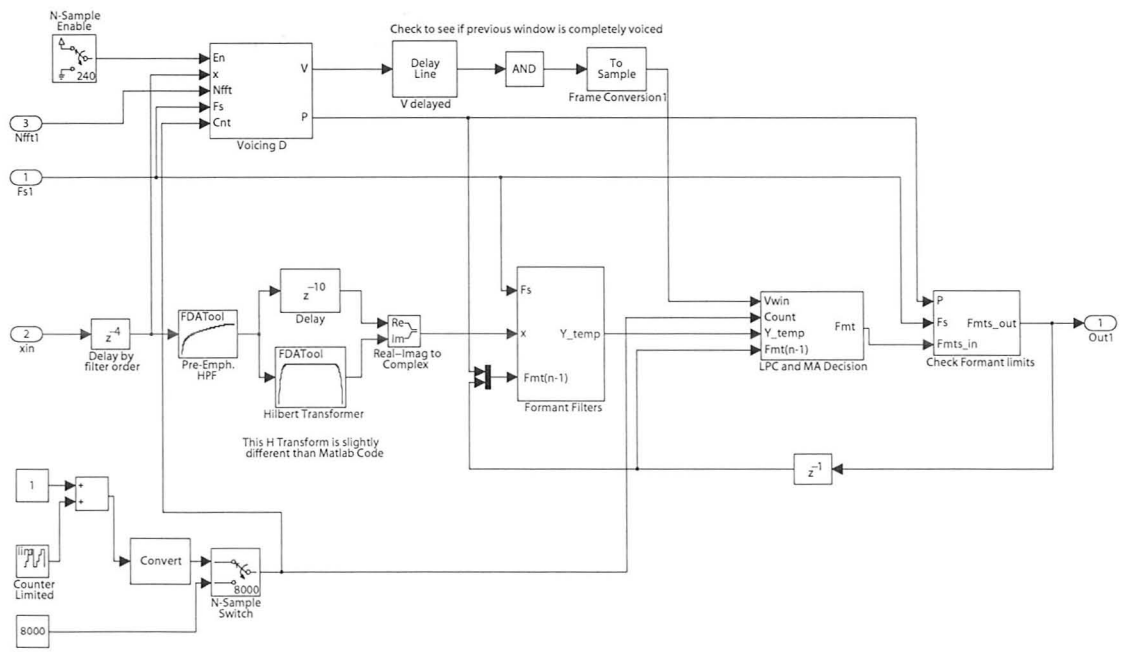

Figure D.8: Simulink model of the Mustafa and Bruce (2006) Formant Tracker.

the voicing detector as shown in Figure D.9. These filters are used to help determine whether or not the signal is voiced; the signal is more likely to be voiced if there is more energy through the lowpass filter compared to the highpass filter. These were originally implemented in the time domain, with a time-varying cutoff frequency based on the gender of the speaker. Because the filter cutoff frequencies are time varying, this requires a time varying definition of each filter's coefficients, which has not been possible using predefined processing blocks in Simulink. As a result, an approximation of these two filters has been implemented in the frequency domain and their energy ratio is determined by summing the low frequency bins in the STFT from 0 Hz up to the time-varying filter cutoff frequency, and comparing this to the sum of the high frequency bins.

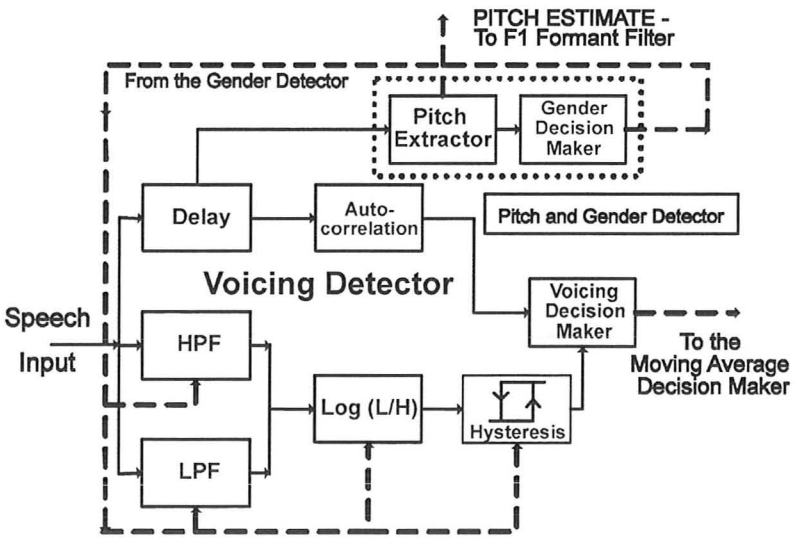

Figure D.9: Block diagram of the voicing detector used in the formant tracker of Mustafa and Bruce (2006). The result is a logical estimate of whether or not the speech is voiced.

These subtle changes in the implementation of the formant tracker have not changed its performance by any significant degree. Figure D.10 shows the formants estimated by both the Matlab code and the Simulink code for a portion of a synthesized sentence with known formants. The formants are tracked well for the voiced parts of the sentence.

Combining the Formant Tracker with MICEFS

The formant tracker processes the signal in the time domain with a sampling frequency of 8 kHz. This differs from the MICEFS processor which uses a sampling rate of 16

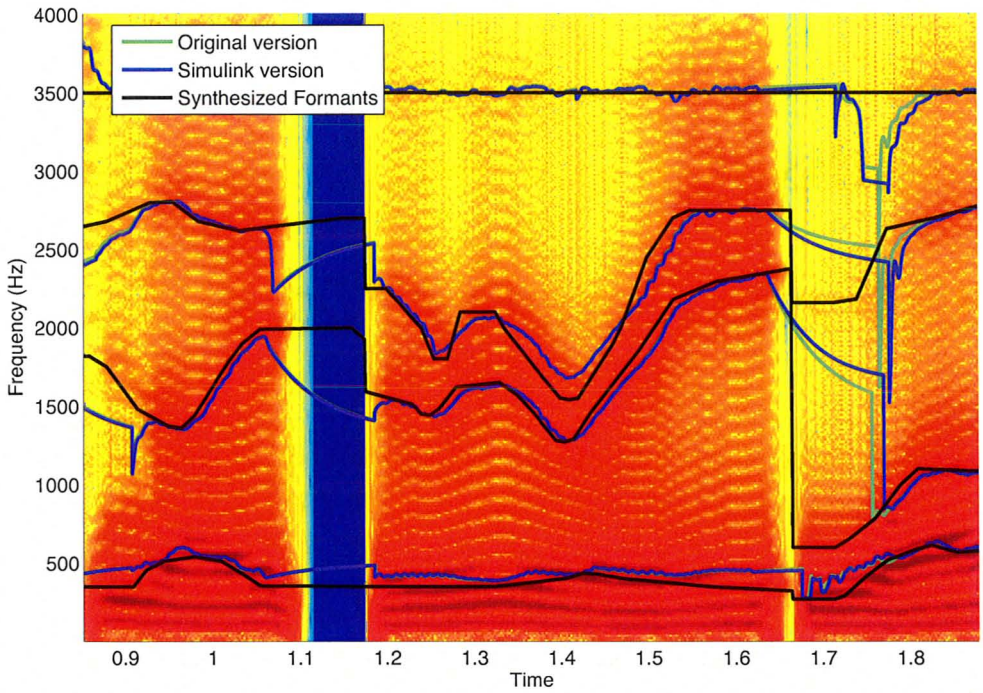

Figure D.10: Spectrogram of a portion of the synthesized sentence: 'Once upon a midnight dreary'. The first four known synthesized formants are shown in black, the formants tracked by the original Matlab implementation of the formant tracker are shown in green, and the formants tracked by the current Simulink implementation of the formant tracker are shown in blue.

kHz, and processes the signal in the frequency domain using the STFT. In order to allow for accurate formant estimates at each frame of the STFT used in MICEFS, the output of the formant tracker was buffered into blocks of 8 msec (64 samples/8 kHz sampling rate), overlapped by 50% and then averaged to be used in the corresponding 8 msec (128 samples/16 kHz) MICEFS frame.

Appendix E

Matlab and Simulink Code

Matlab and Simulink code for the work of this thesis is available upon request. Please email me at zeyltj@mcmaster.ca or timothyzeyl@gmail.com.

Appendix F

Additional Results

F.1 Evalutaion of SPC alone

F.1.1 Mild, Mixed Loss

SPC-0 Aided Box-Plot, Mild Mixed Loss

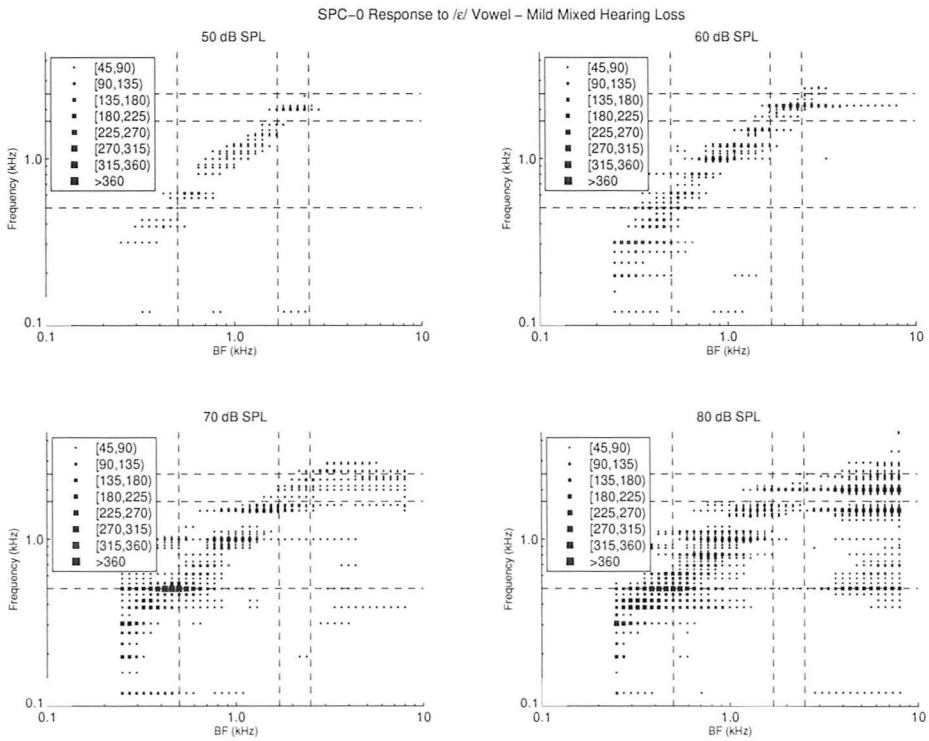

Figure F.11: Box-plots of the synchronized rate for a mildly impaired AN model with mixed hair cell loss in response to the synthesized vowel /ɛ/, processed with SPC-0.

SPC-2 Aided Box-Plot, Mild Mixed Loss

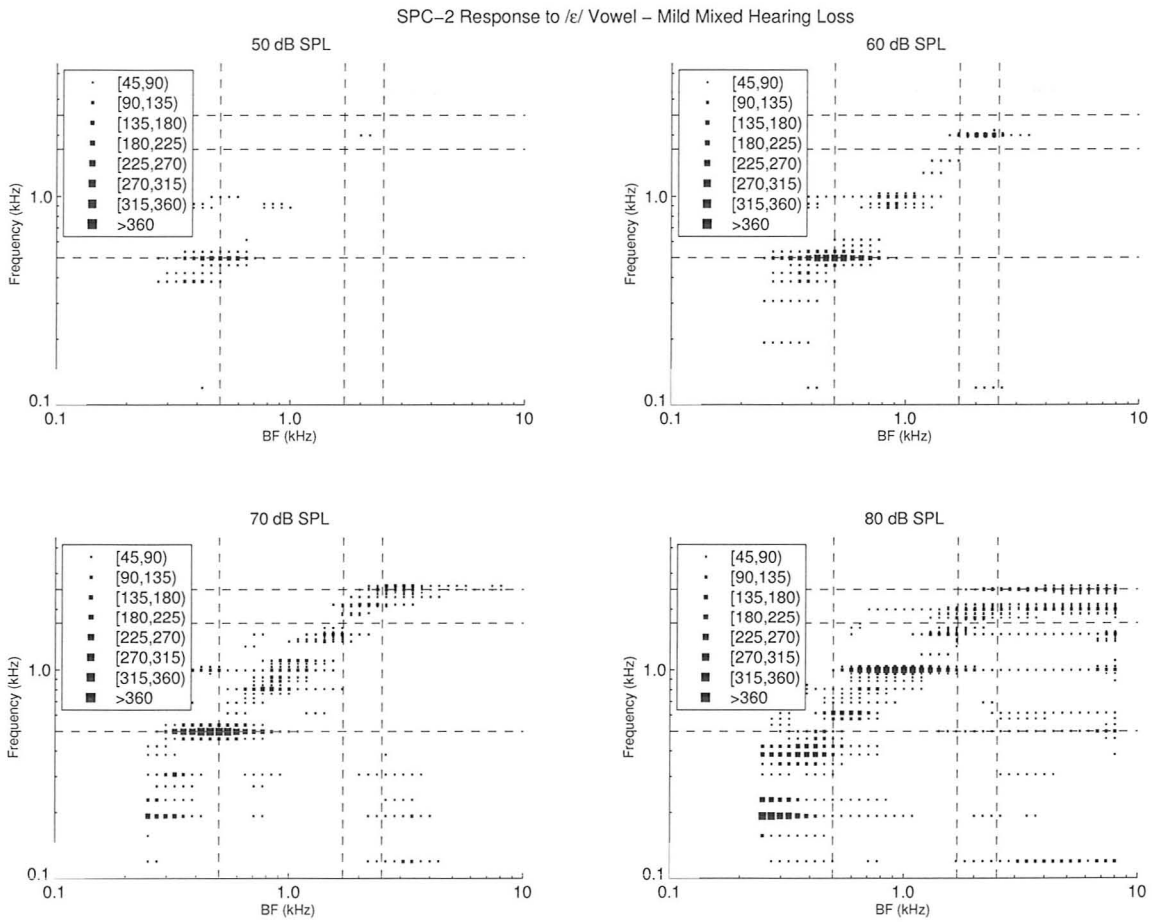

Figure F.12: Box-plots of the synchronized rate for a mildly impaired AN model with mixed hair cell loss in response to the synthesized vowel /ɛ/, processed with SPC-2.

F.1.2 Mild, OHC Loss

Healthy & Impaired Neurograms

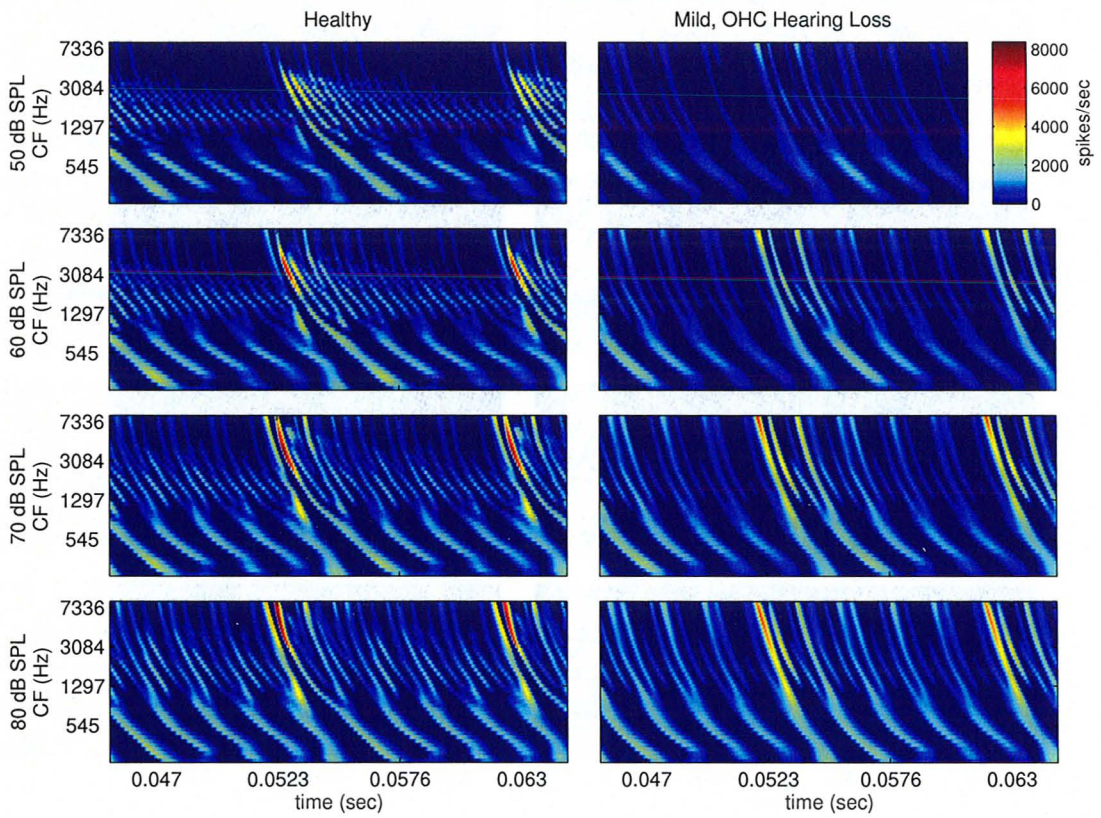

Figure F.13: Healthy and impaired Neurograms for the mild hearing loss with exclusively outer hair cell damage in response to the synthesized /ɛ/ vowel. Neurograms are derived from the model's 'synaptic' output which gives the probability of spiking in units of spikes/sec. Healthy neurograms are shown in the left column of subfigures, with each row showing the response at a different stimulus presentation level, as labelled. Impaired neurograms are shown in the right column of subfigures. Each subfigure is plotted using the same colour axis, on a colour spectrum scale from blue to red. Dark blue represents a very low spiking rate and red represents a high spiking rate in units of spikes/sec

Healthy & SPC-0 Aided Neurograms, Mild OHC Loss

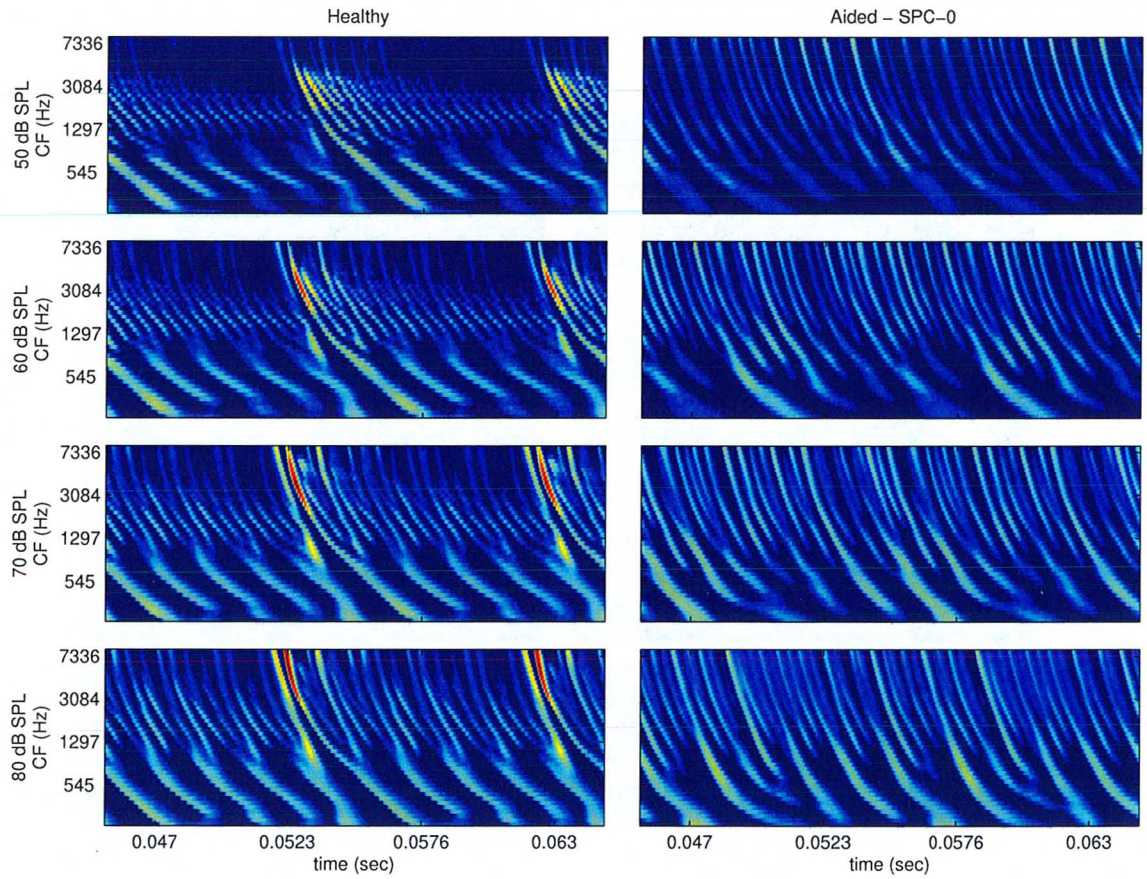

Figure F.14: Healthy and SPC-0 aided neurograms for the mild hearing loss with exclusively outer hair cell damage in response to the synthesized /ɛ/ vowel.

Healthy & SPC-1 Aided Neurograms, Mild OHC Loss

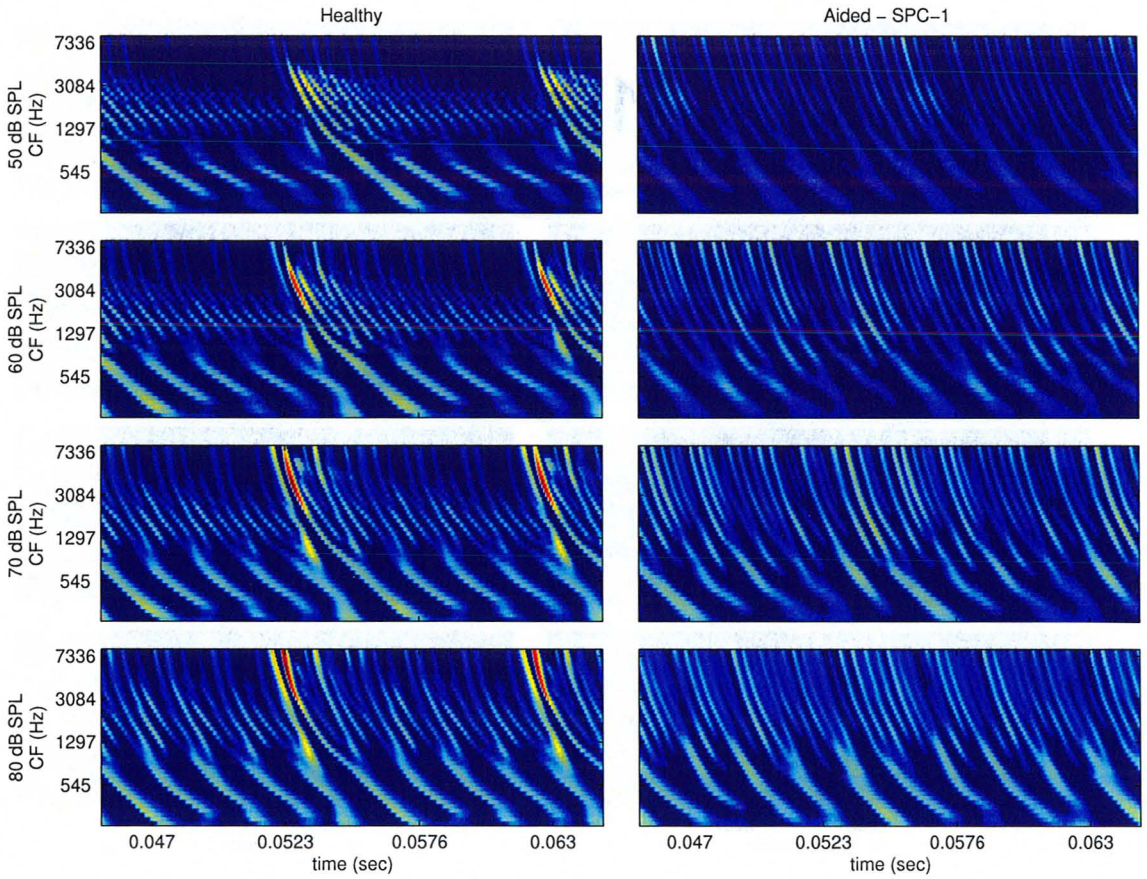

Figure F.15: Healthy and SPC-1 aided neurograms for the mild hearing loss with exclusively outer hair cell damage in response to the synthesized /ɛ/ vowel.

Healthy & SPC-2 Aided Neurograms, Mild OHC Loss

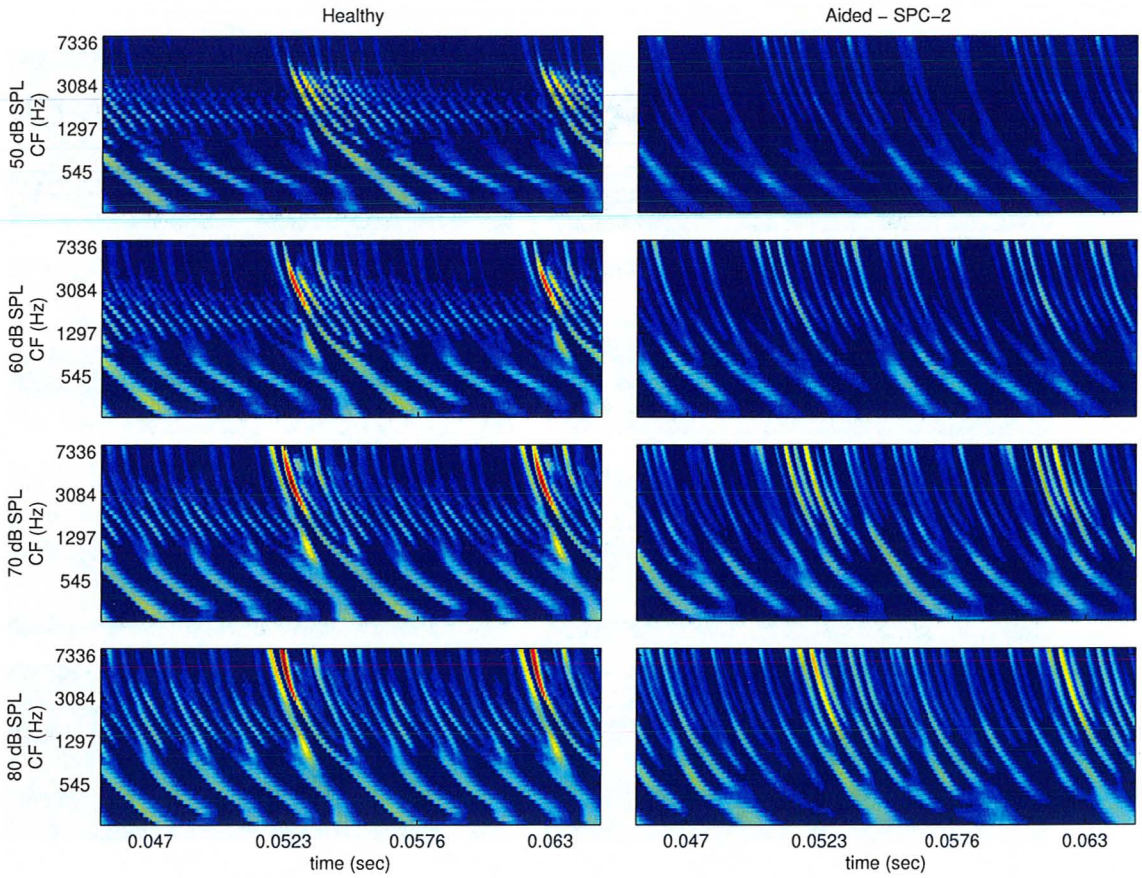

Figure F.16: Healthy and SPC-2 aided neurograms for the mild hearing loss with exclusively outer cell damage in response to the synthesized /ɛ/ vowel.

SPC-0 aided Box-Plot, Mild OHC Loss

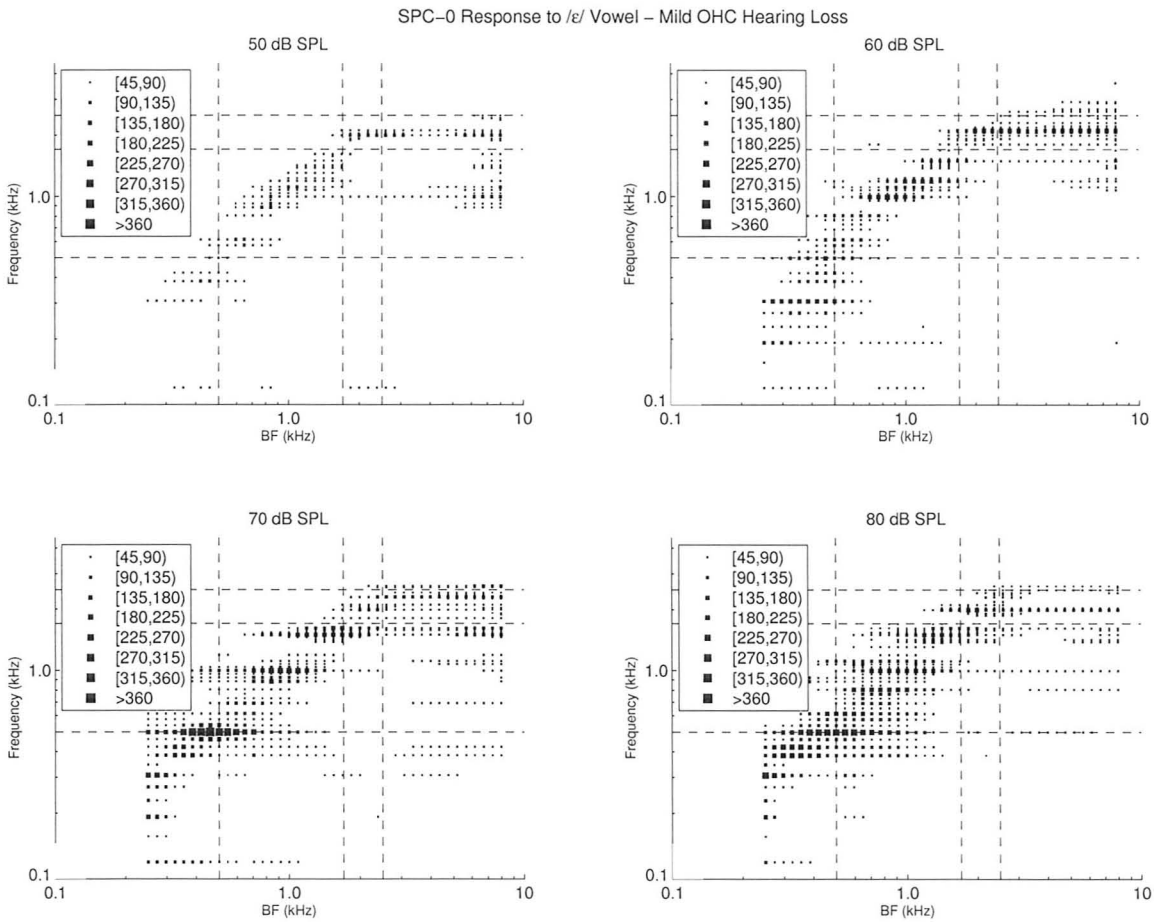

Figure F.17: Synchronized rate box-plots for a mildly impaired AN model with exclusively outer hair cell loss in response to the synthesized vowel /ɛ/, processed with SPC-0.

SPC-1 aided Box-Plot, Mild OHC Loss

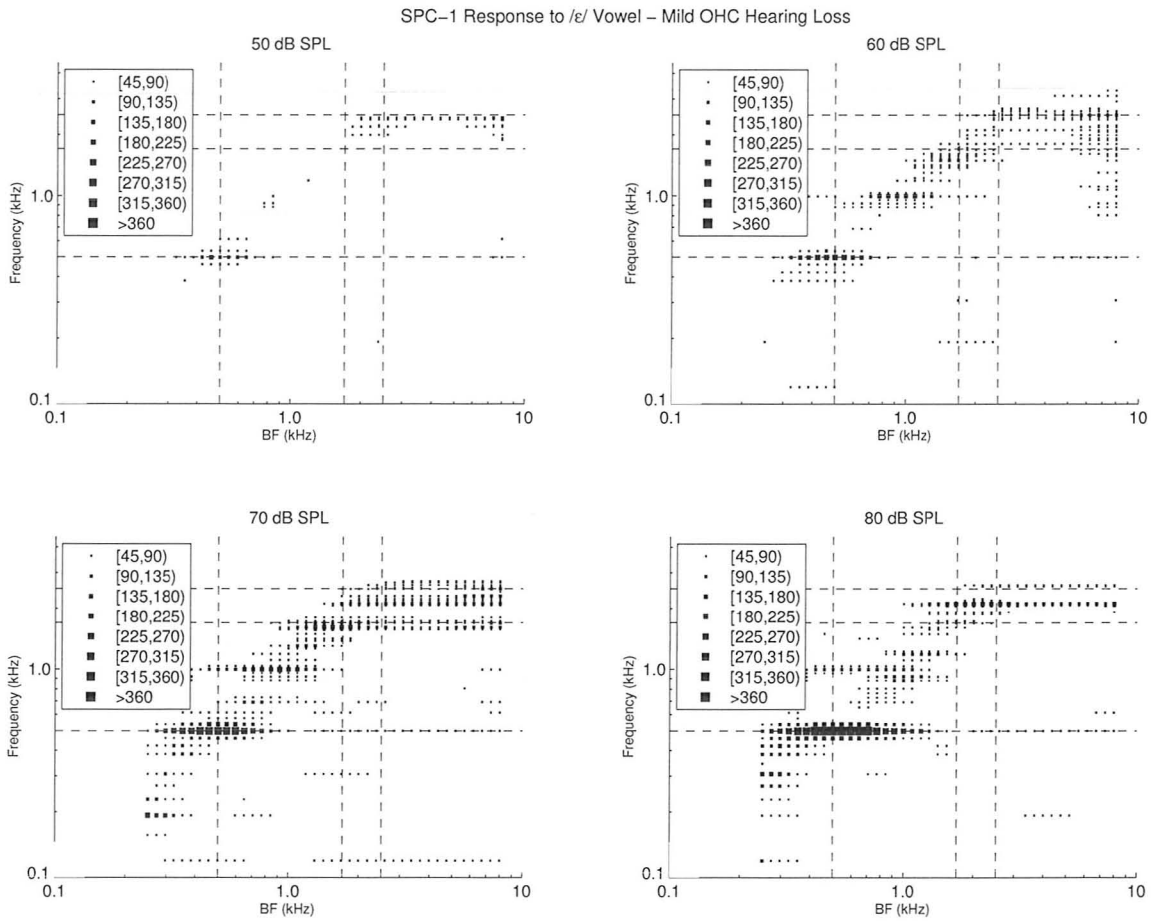

Figure F.18: Synchronized rate box-plots for a mildly impaired AN model with exclusively outer hair cell loss in response to the synthesized vowel /ɛ/, processed with SPC-1.

SPC-2 aided Box-Plot, Mild OHC Loss

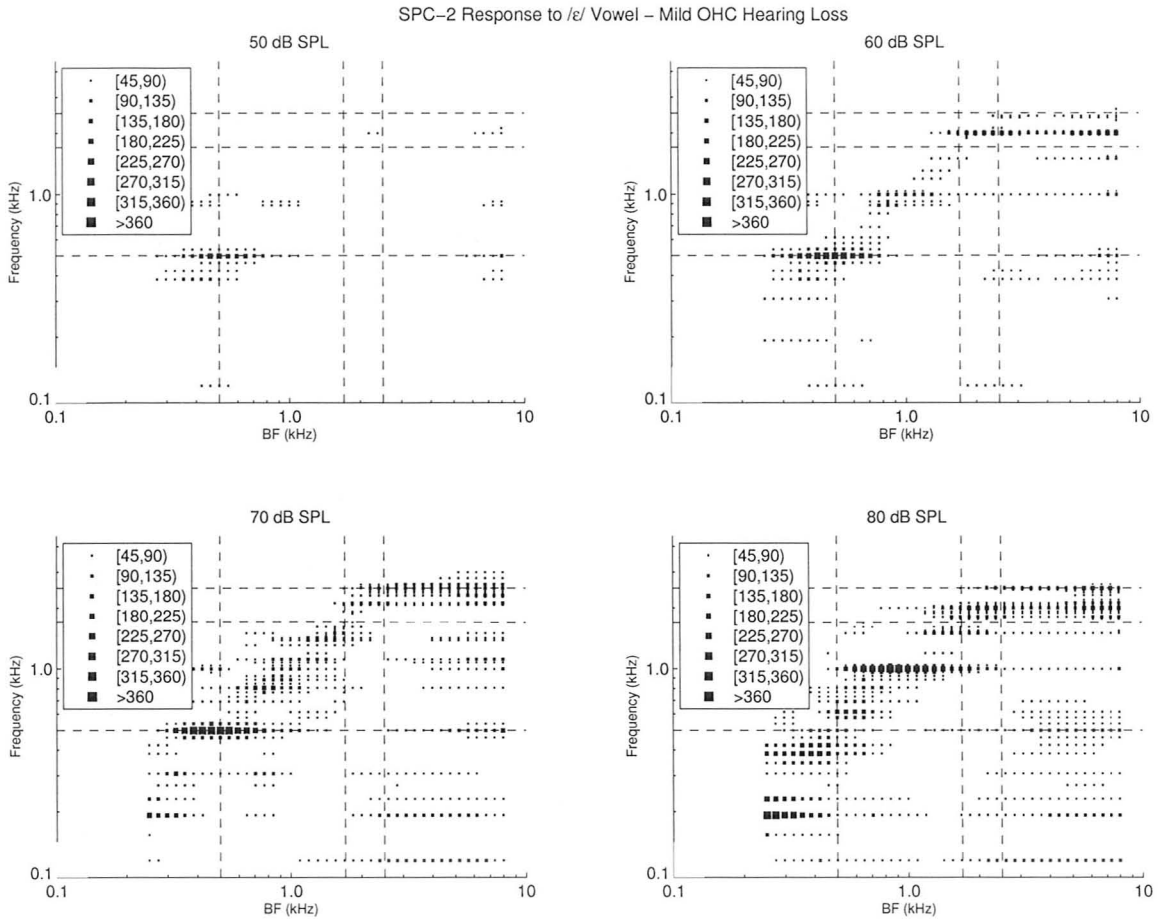

Figure F.19: Synchronized rate box-plots for a mildly impaired AN model with exclusively outer hair cell loss in response to the synthesized vowel /ε/, processed with SPC-2.

F.1.3 Moderate-to-Severe, Mixed Loss

Healthy & Impaired Neurograms

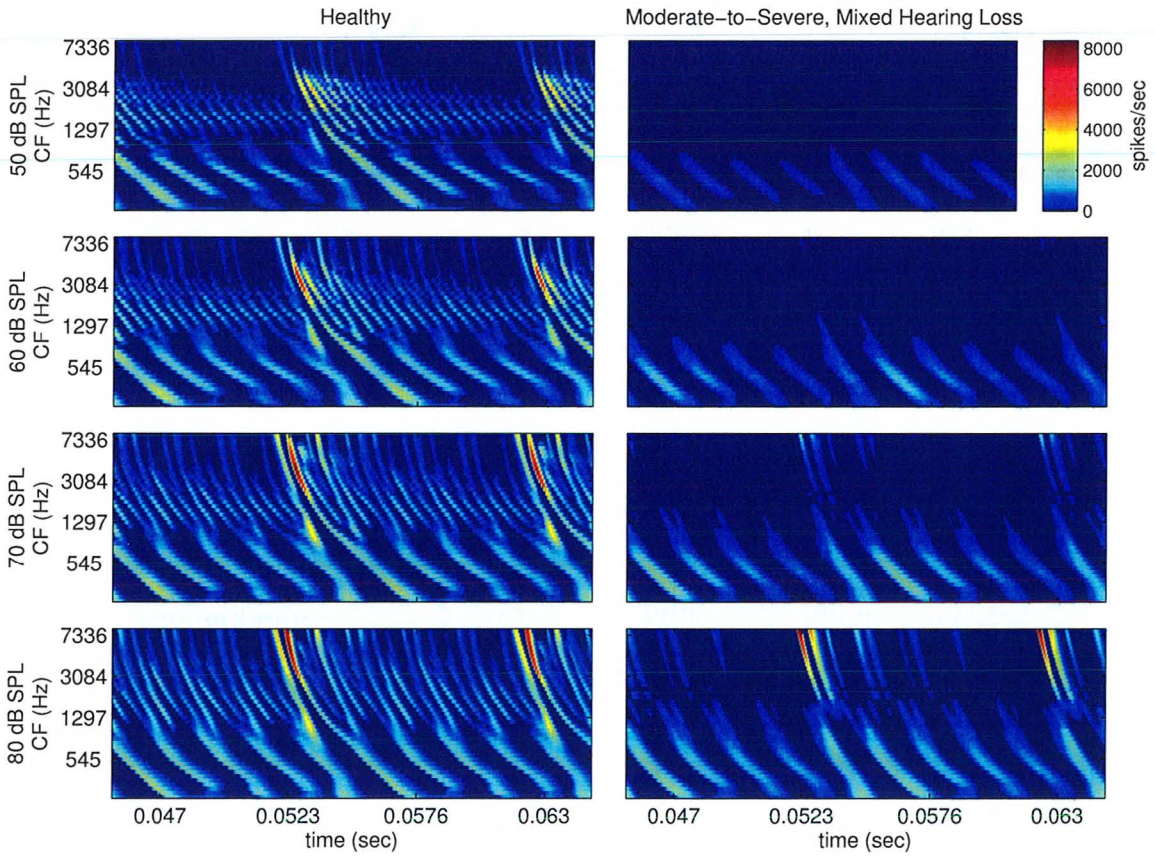

Figure F.20: Healthy and impaired Neurograms for the moderate-to-severe hearing loss with mixed hair cell damage in response to the synthesized /ɛ/ vowel. Healthy neurograms are shown in the left column of subfigures, with each row showing the response at a different stimulus presentation level, as labelled. Impaired neurograms are shown in the right column of subfigures.

Healthy & SPC-0 Aided Neurograms, Moderate-to-Severe Mixed Loss

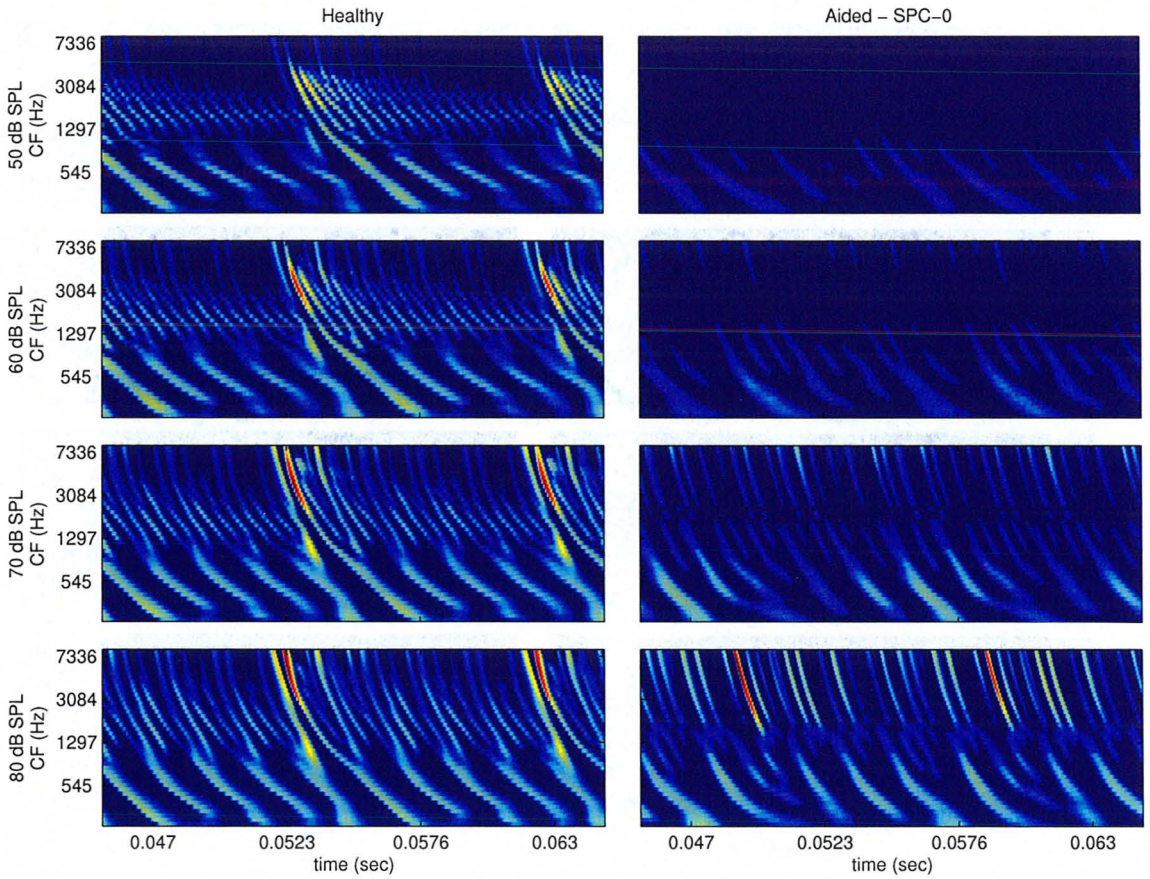

Figure F.21: Healthy and SPC-0 aided neurograms for the moderate-to-severe hearing loss with mixed hair cell damage in response to the synthesized /ɛ/ vowel.

Healthy & SPC-1 Aided Neurograms, Moderate-to-Severe Mixed Loss

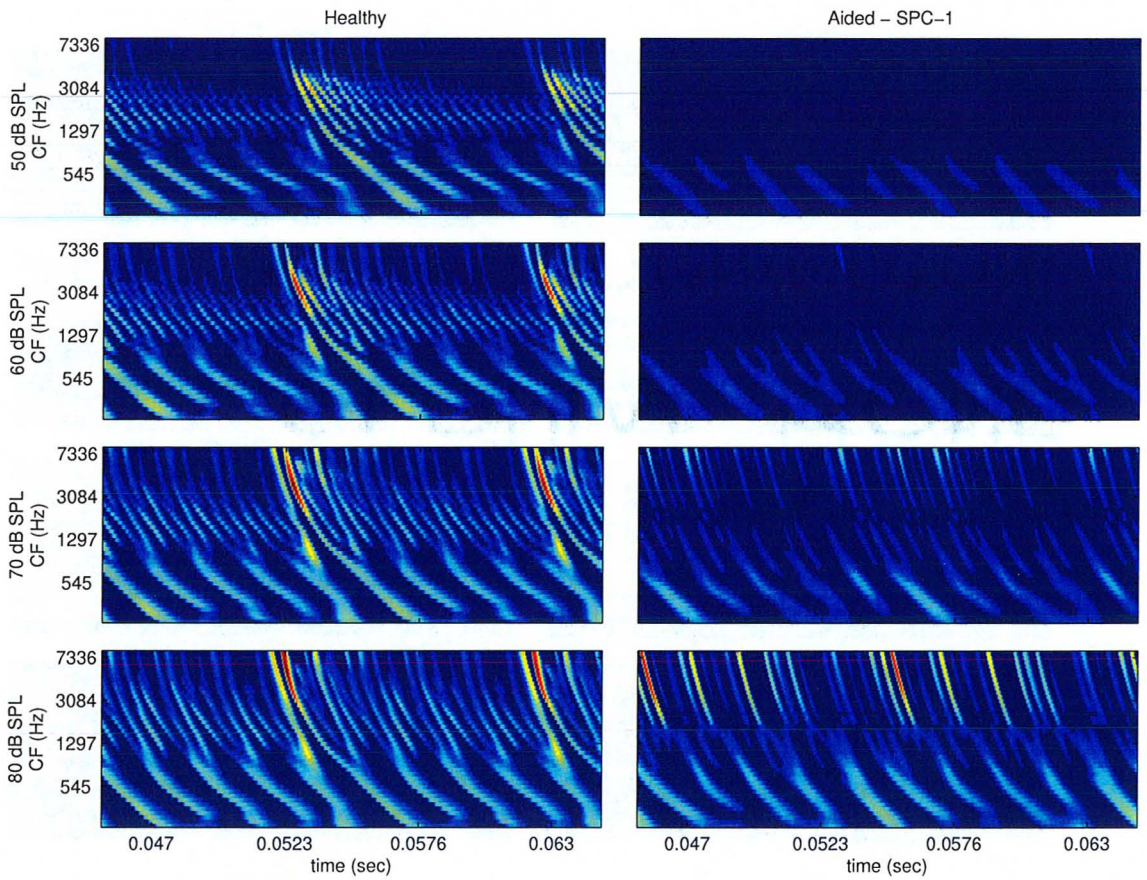

Figure F.22: Healthy and SPC-1 aided neurograms for the moderate-to-severe hearing loss with mixed hair cell damage in response to the synthesized /ɛ/ vowel.

Healthy & SPC-2 Aided Neurograms, Moderate-to-Severe Mixed Loss

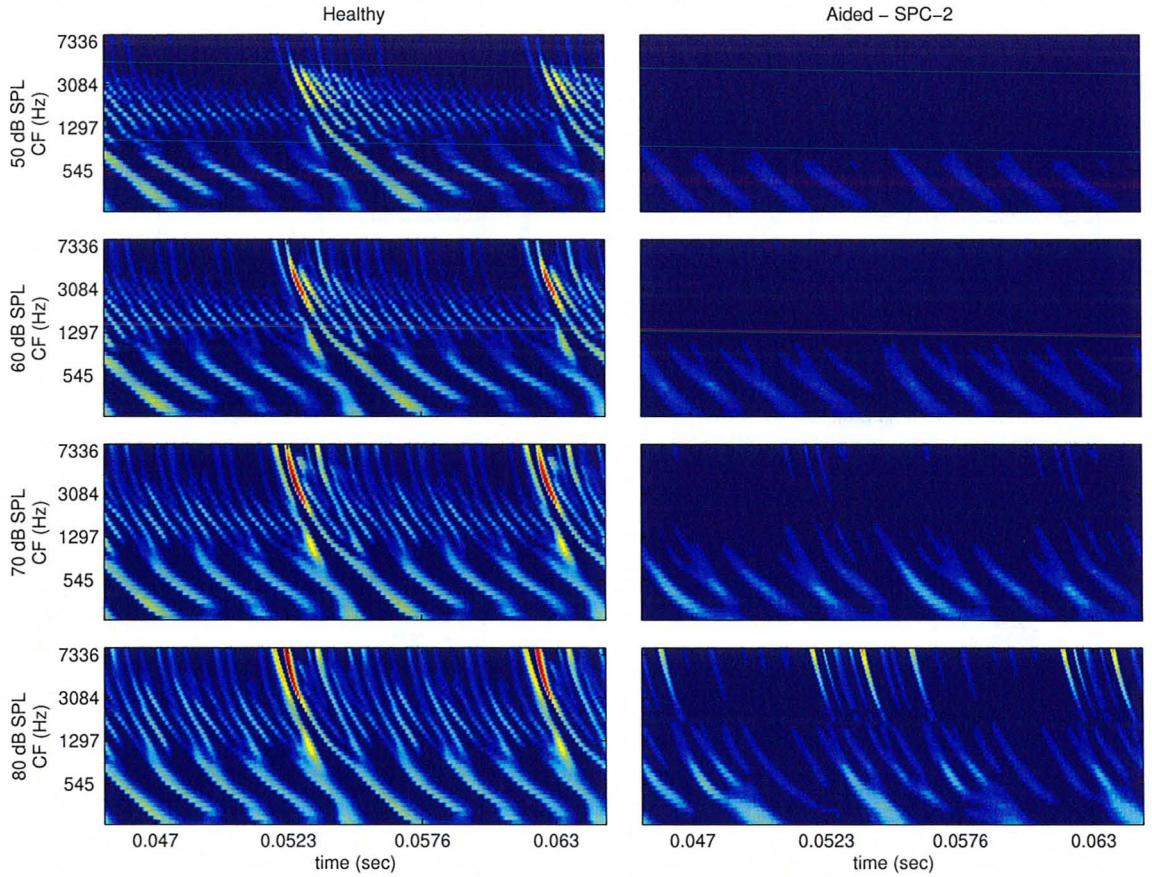

Figure F.23: Healthy and SPC-2 aided neurograms for the moderate-to-severe hearing loss with mixed hair cell damage in response to the synthesized /ɛ/ vowel.

SPC-0 aided Box-Plot, Moderate-to-Severe Mixed Loss

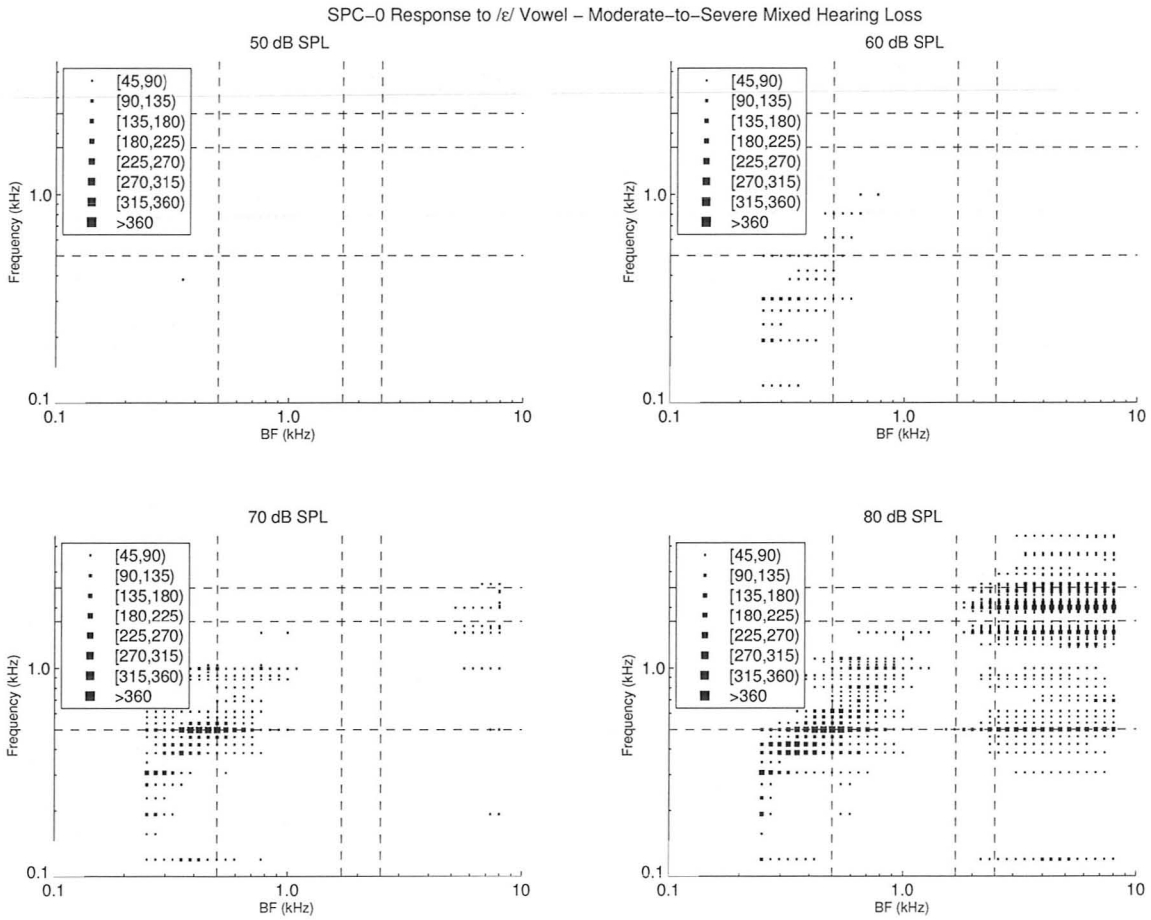

Figure F.24: Synchronized rate box-plots for a moderate-to-severely impaired AN model with mixed hair cell loss in response to the synthesized vowel /ɛ/, processed with SPC-0.

SPC-1 aided Box-Plot, Moderate-to-Severe Mixed Loss

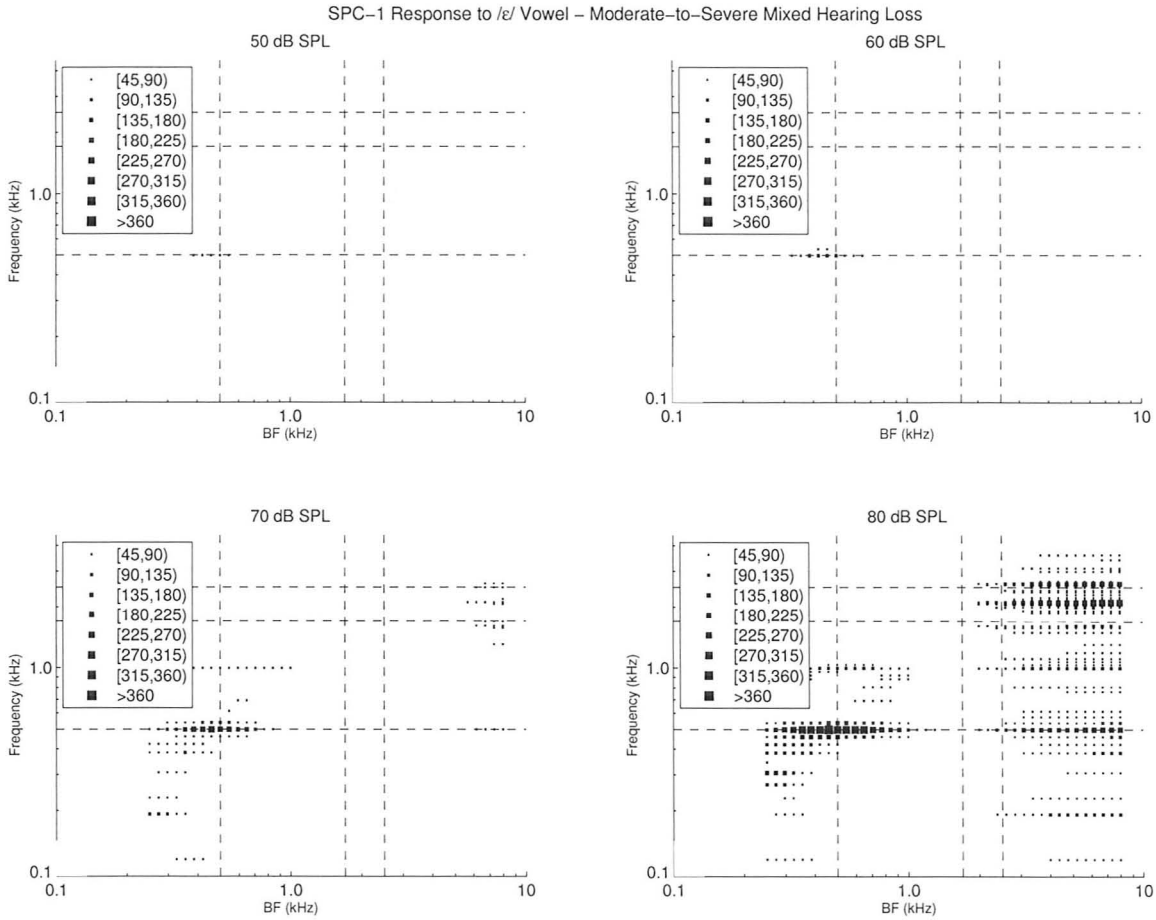

Figure F.25: Synchronized rate box-plots for a moderate-to-severely impaired AN model with mixed hair cell loss in response to the synthesized vowel /ɛ/, processed with SPC-1.

SPC-2 aided Box-Plot, Moderate-to-Severe Mixed Loss

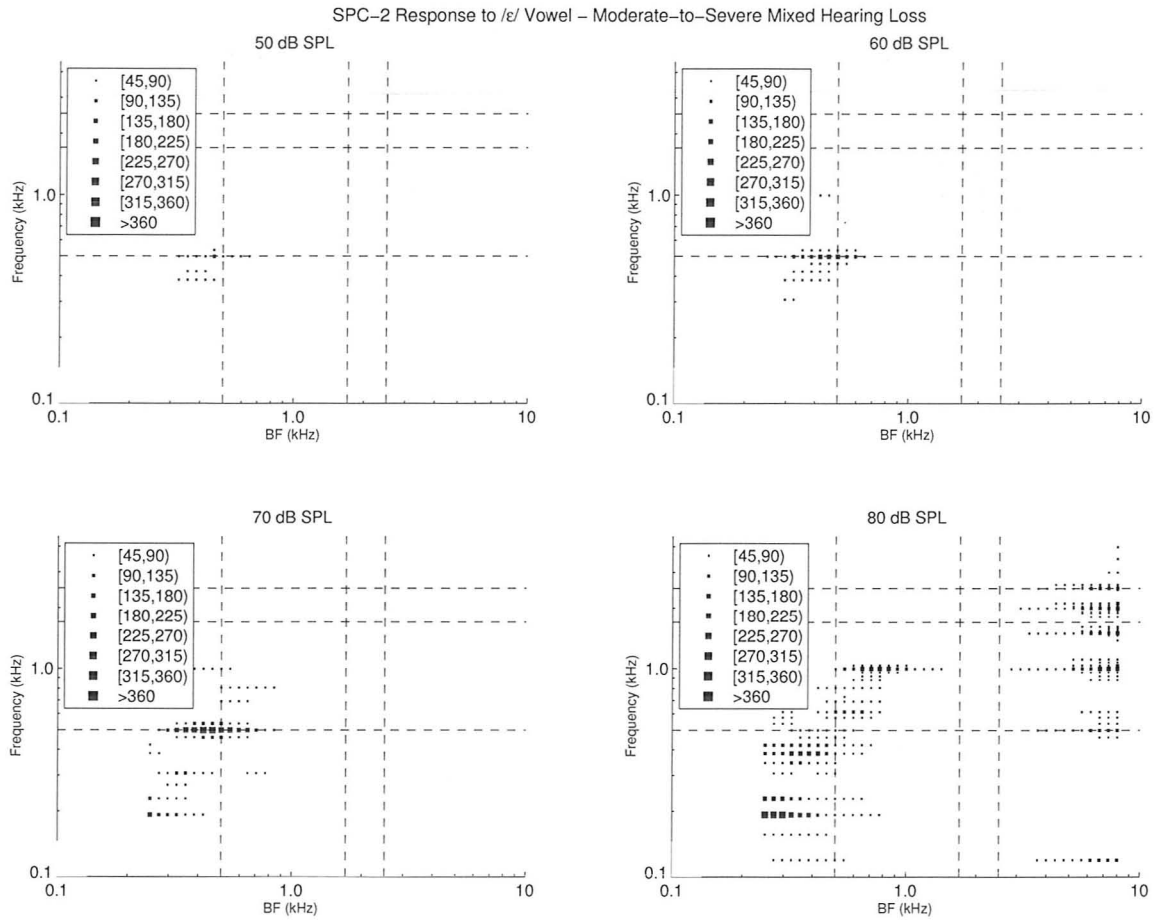

Figure F.26: Synchronized rate box-plots for a moderate-to-severely impaired AN model with mixed hair cell loss in response to the synthesized vowel /ɛ/, processed with SPC-2.

F.2 Evaluation of MICEFS and SPC

F.2.1 Mild, OHC Loss

MICEFS-1/SPC-1 and MICEFS-1/SPC-2 Aided Neurograms

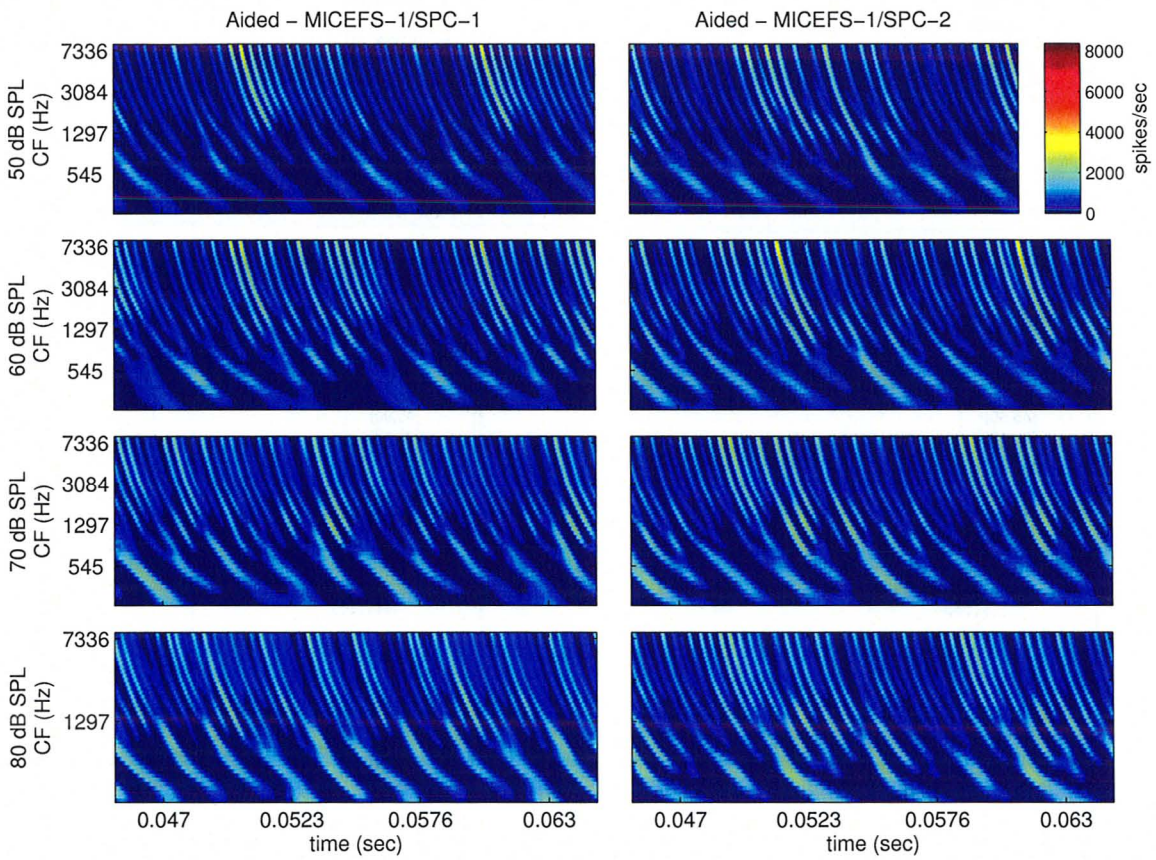

Figure F.27: Impaired neurograms for the mild hearing loss with exclusively outer hair cell damage in response to the synthesized /ε/ vowel processed with both the MICEFS-1/SPC-1 and MICEFS-1/SPC-2 processing schemes. The neurograms that result after processing with SPC-1 are shown in the left column of subfigures while the neurograms that results after processing with SPC-2 are shown in the right column.

MICEFS-1/SPC-1 aided Box-Plot, Mild OHC Loss

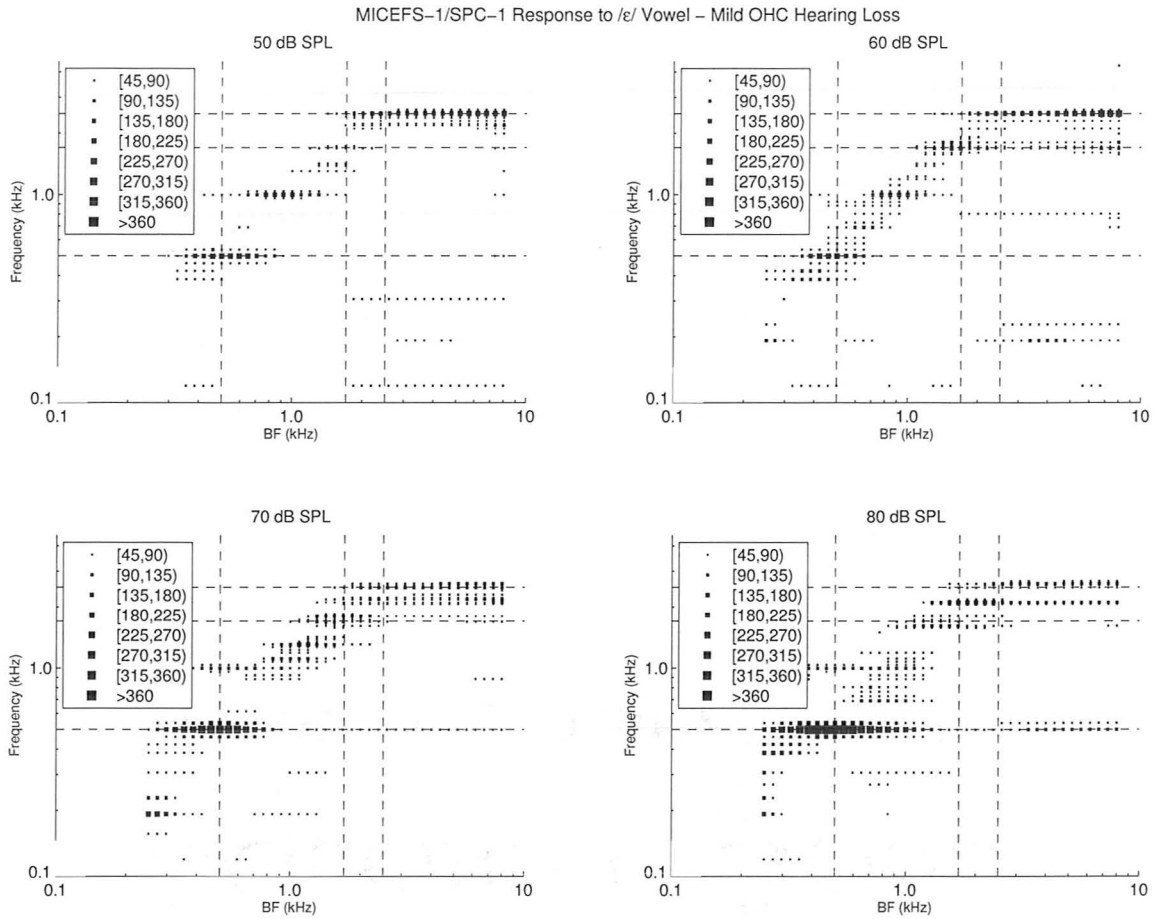

Figure F.28: Synchronized rate box-plots for a mildly impaired AN model with exclusively outer hair cell loss in response to the synthesized vowel /ɛ/, processed with MICEFS-1/SPC-1.

MICEFS-1/SPC-2 aided Box-Plot, Mild OHC Loss

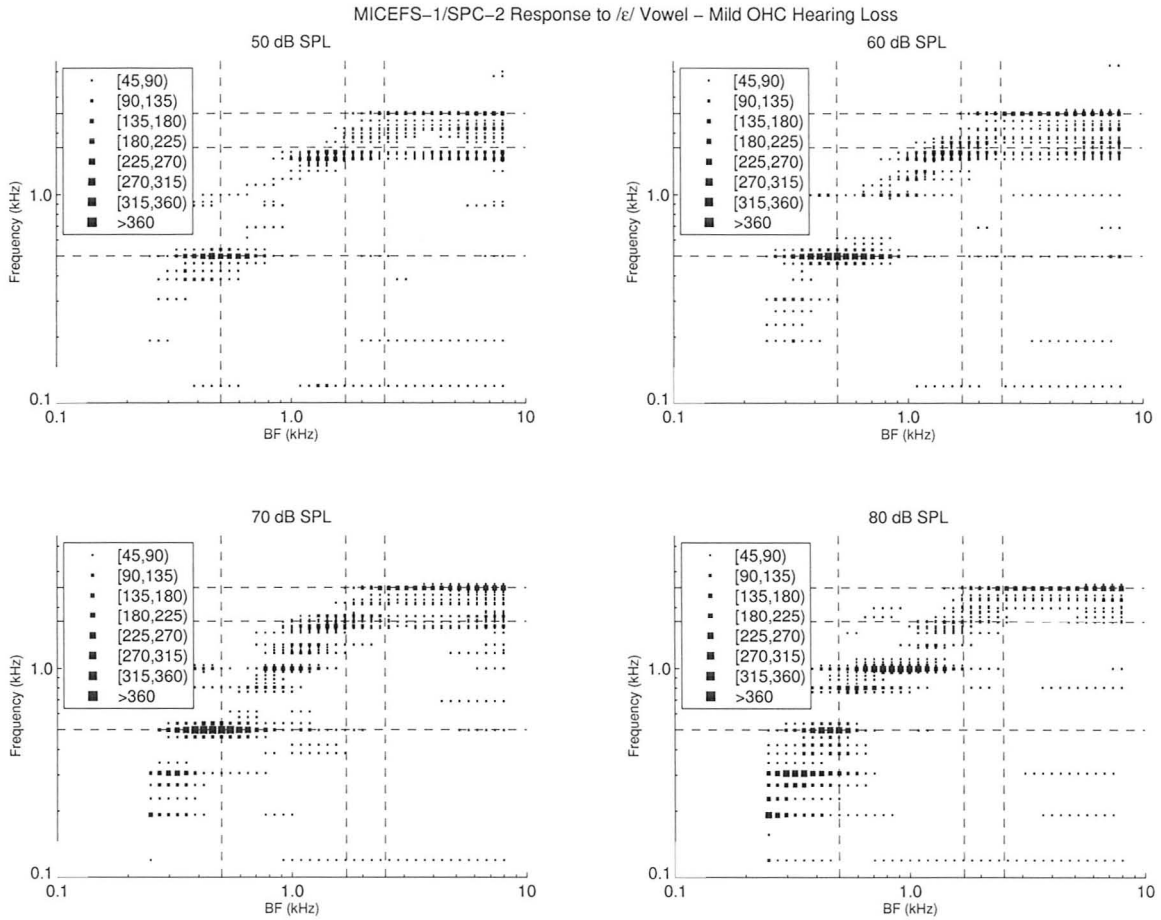

Figure F.29: Synchronized rate box-plots for a mildly impaired AN model with exclusively outer hair cell loss in response to the synthesized vowel /ɛ/, processed with MICEFS-1/SPC-2.

Peak-Lag Error, Mild OHC Loss

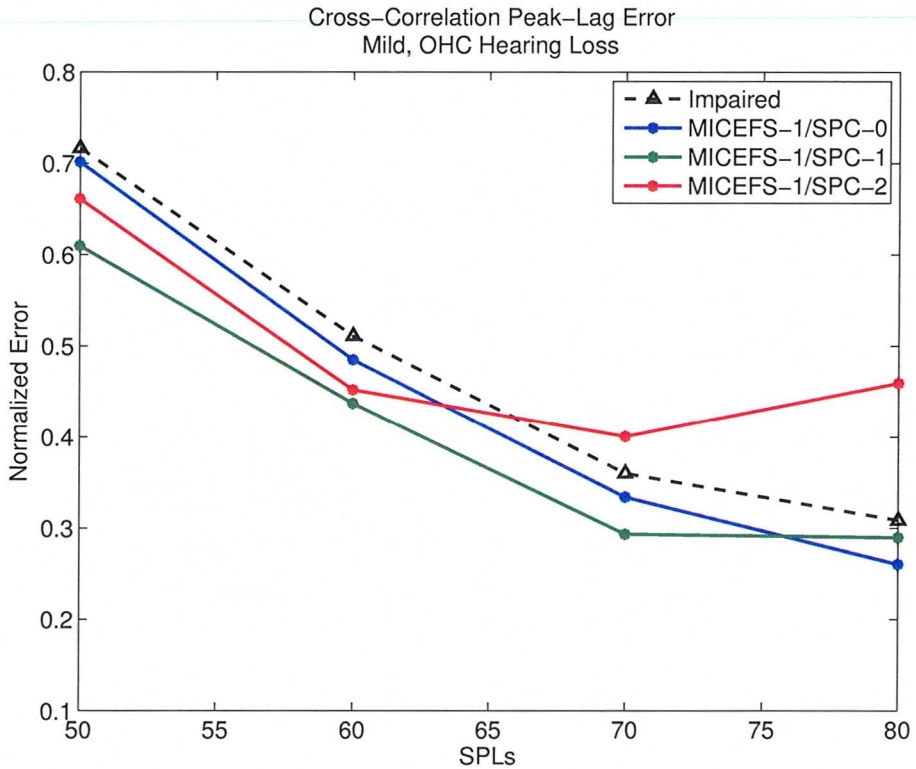

Figure F.30: Peak-lag error for the synthesized vowel aided with MICEFS-1 combined with three versions of SPC, shown as a function of stimulus sound pressure level. Error derived from the vowel response of an AN model with mild hearing loss and exclusively outer hair cell damage.

F.2.2 Moderate-to-Severe, Mixed Hair Loss

MICEFS-1/SPC-1 and MICEFS-1/SPC-2 Aided Neurograms

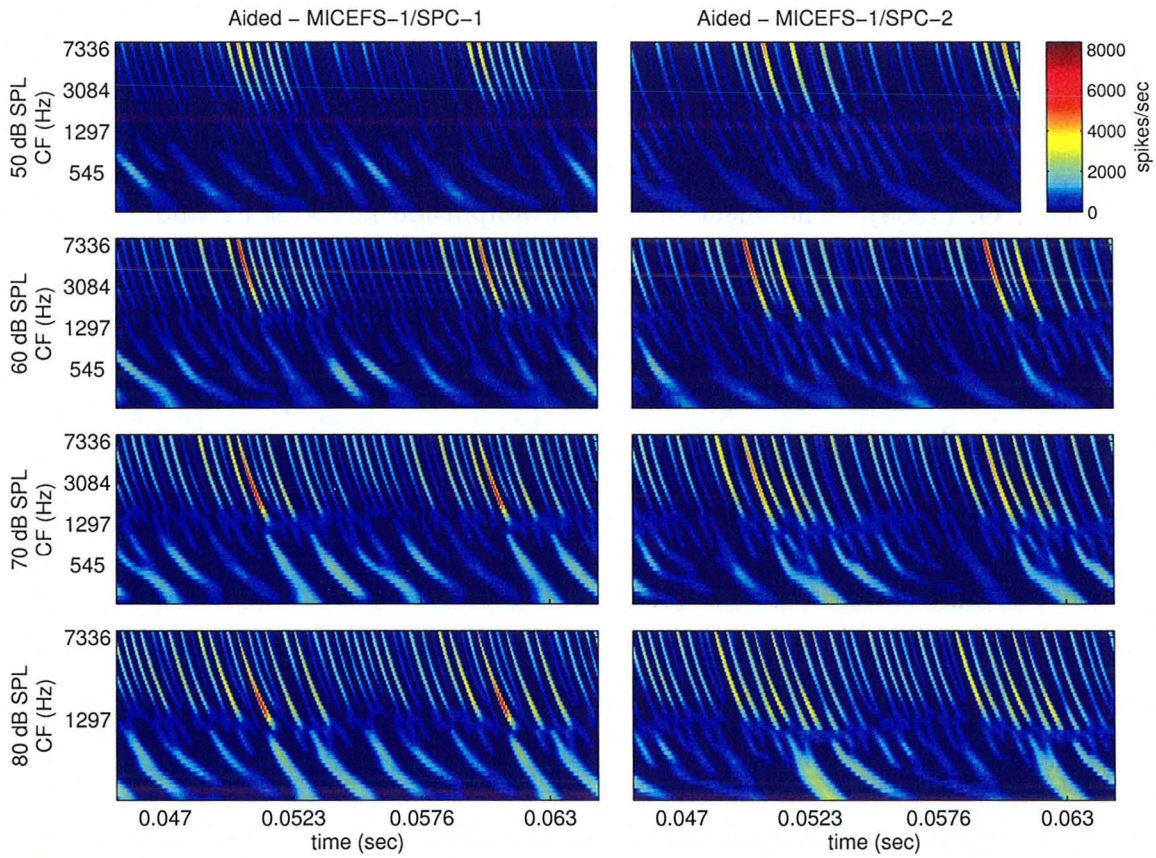

Figure F.31: Impaired neurograms for the moderate-to-severe hearing loss with mixed hair cell damage in response to the synthesized /ɛ/ vowel processed with both the MICEFS-1/SPC-1 and MICEFS-1/SPC-2 processing schemes. The neurograms that result after processing with SPC-1 are shown in the left column of subfigures while the neurograms that results after processing with SPC-2 are shown in the right column.

Bibliography

Aaltonen, O. (1985). The effect of relative amplitude levels of F2 and F3 on the categorization of synthetic vowels. *Journal of Phonetics*, **13**, 1–9.

Aibara, R., Welsh, J., Puria, S., and Goode, R. (2001). Human middle-ear sound transfer function and cochlear input impedance. *Hearing Research*, **152**(1-2), 100–109.

Ansari, S. (2005). *Efficiently Combining Multiband Compression and Improved Contrast-Enhancing Frequency Shaping In Hearing Aids.* Master's thesis, McMaster University.

Bloom, W. and Fawcett, D. W. (1975). *A Textbook of Histology.* WB Saunders, Philadelphia, 10 edition.

Bondy, J., Becker, S., Bruce, I., Trainor, L., and Haykin, S. (2004). A novel signal-processing strategy for hearing-aid design: neurocompensation. *Signal Processing*, **84**(7), 1239–1253.

Bruce, I. (2004). Physiological assessment of contrast-enhancing frequency shaping and multiband compression in hearing aids. *Physiological Measurement*, **25**(4), 945–956. Physiological Measurement.

Bruce, I., Sachs, M., and Young, E. (2003). An auditory-periphery model of the effects of acoustic trauma on auditory nerve responses. *Journal of the Acoustical Society of America*, **113**(1), 369–388.

Bruce, I. C. (2006). *Encyclopedia of Biomaterials and Biomedical Engineering*, chapter Hearing Aids. Marcel Dekker, New York, NY.

Bruce, I. C., Dinath, F., and Zeyl, T. J. (2007). Insights into optimal phonemic compression from a computational model of the auditory periphery. In T. Dau, J. Buchholz, J. M. Harte, and T. U. Christiansen, editors, *Auditory Signal Processing in Hearing-Impaired Listeners, Int. symposium on Audiological and Auditory Research (ISAAR)*, pages 73–81, Denmark. Danavox Jubilee Foundation.

Byrne, D. and Dillon, H. (1986). The national acoustic laboratories' (NAL) new procedure for selecting the gain and frequency response of a hearing aid. *Ear and Hearing*, **7**(4), 257–65.

Byrne, D., Parkinson, A., and Newall, P. (1990). Hearing aid gain and frequency response requirements for the severely/profoundly hearing impaired. *Ear and Hearing*, **11**(1), 40–9.

Cai, S., Ma, W.-L. D., and Young, E. D. (2009). Encoding intensity in ventral cochlear nucleus following acoustic trauma: Implications for loudness recruitment. *Jaro-Journal of the Association For Research In Otolaryngology*, **10**(1), 5–22.

Calandruccio, L., Doherty, K. A., Carney, L. H., and Kikkeri, H. N. (2007). Perception of temporally processed speech by listeners with hearing impairment. *Ear and Hearing*, **28**(4), 512–523.

Carney, L. H. (1993). A model for the responses of low-frequency auditory-nerve fibers in cat. *Journal of the Acoustical Society of America*, **93**(1), 401–417.

Carney, L. H. (1994). Spatiotemporal encoding of sound level: models for normal encoding and recruitment of loudness. *Hear Research*, **76**(1-2), 31–44.

Cheatham, M. and Dallos, P. (1999). Response phase: A view from the inner hair cell. *Journal of the Acoustical Society of America*, **105**(2), 799–810.

Chen, Z., Becker, S., Bondy, J., Bruce, I., and Haykin, S. (2005). A novel model-based hearing compensation design using a gradient-free optimization method. *Neural Computation*, **17**(12), 2648–2671.

Dallos, P. (1992). The active cochlea. *Journal of Neuroscience*, **12**(12), 4575–85.

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

Dinath, F. (2008). *Using Model Auditory-Nerve Responses to Optimize Compression Gains for Hearing Aid Amplification Schemes*. Master's thesis, McMaster University.

Dinath, F. and Bruce, I. (2008). Hearing aid gain prescriptions balance restoration of auditory nerve mean-rate and spike-timing representations of speech. In *Proceedings of the 30th International IEEE Engineering in Medicine and Biology Conference*, pages 1793–1796, Piscataway, NJ. IEEE.

Edwards, B. (2007). The future of hearing aid technology. *Trends in Amplification*, **11**(1), 31–45.

Fettiplace, R. and Hackney, C. M. (2006). The sensory and motor roles of auditory hair cells. *Nature Reviews Neuroscience*, **7**(1), 19–29.

Geisler, C. D. (1998). *From Sound to Synapse: Physiology of The Mammalian Ear.* Oxford University Pres.

Glasberg, B. R. and Moore, B. C. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, **47**(1-2), 103–38.

Harte, N., Ansari, S. U., and Bruce, I. (2006). Exploiting voicing cues for contrast enhanced frequency shaping of speech for impaired listeners. In *Proceedings of the 31st IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP2006)*, volume 5, pages V–137–V–140, Piscataway, NJ.

Heinz, M. and Young, E. (2004). Response growth with sound level in auditory-nerve fibers after noise-induced hearing loss. *Journal of Neurophysiology*, **91**(2), 784–795.

Heinz, M., Issa, J., and Young, E. (2005). Auditory-nerve rate responses are inconsistent with common hypotheses for the neural correlates of loudness recruitment. *JARO - Journal of the Association for Research in Otolaryngology*, **6**(2), 91–105.

Heinz, M. G., Zhang, X., Bruce, I. C., and Carney, L. H. (2001). Auditory nerve model for predicting performance limits of normal and impaired listeners. *Acoustics Research Letters Online*, **2**(3), 91–96.

Hohmann, V. (2002). Frequency analysis and synthesis using a Gammatone filterbank. *Acta Acustica united with Acustica*, **88**(3), 433–442.

Ibrahim, R. and Bruce, I. C. (2010). *The Neurophysiological Bases of Auditory Perception*, chapter Effects of Peripheral Tuning on the Auditory Nerve's Representation of Speech Envelope and Temporal Fine Structure Cues. Springer.

Johnson, D. (2009). Modeling the speech signal. *Connexions web site.*

Johnstone, B., Patuzzi, R., and Yates, G. (1986). Basilar membrane measurements and the travelling wave. *Hearing Research*, **22**(1-3), 147 – 153.

Joris, P. (2009). Recruitment of neurons and loudness. *JARO - Journal of the Association for Research in Otolaryngology*, **10**(1), 1–4.

Joris, P. X. and Yin, T. C. T. (1992). Responses to amplitude-modulated tones in the auditory-nerve of the cat. *Journal of the Acoustical Society of America*, **91**(1), 215–232.

Kale, S. and Heinz, M. G. (2010). Envelope coding in auditory nerve fibers following noise-induced hearing loss. *Journal of the Association for Research in Otolaryngology*.

Kessel, R. G. and Kardon, R. H. (1979). *Tissues and Organs*. W.H. Freeman and Company, San Francisco.

Kiang, N. Y. and Moxon, E. C. (1972). Physiological considerations in artificial stimulation of the inner ear. *Annals Otology, Rhinology, and Laryngology*, **81**(5), 714–30.

Kiang, N. Y., Liberman, M. C., Sewell, W. F., and Guinan, J. J. (1986). Single unit clues to cochlear mechanisms. *Hearing Research*, **22**, 171–82.

Kiefte, M. and Kluender, K. (2005). The relative importance of spectral tilt in monophthongs and diphthongs. *Journal of the Acoustical Society of America*, **117**(3), 1395–1404.

Kiefte, M., Enright, T., and Marshall, L. (2010). The role of formant amplitude in the perception of vertical bar i vertical bar and vertical bar u vertical bar. *Journal of the Acoustical Society of America*, **127**(4), 2611–2621.

Kochkin, S. (2005). Marketrak vii: Customer satisfaction with hearing instruments in the digital age. *The Hearing Journal*, **58**, 30–42.

Liberman, M. C. and Dodds, L. W. (1984). Single-neuron labeling and chronic cochlear pathology. iii. stereocilia damage and alterations of threshold tuning curves. *Hearing Research*, **16**(1), 55–74.

Miller, R., Schilling, J., Franck, K., and Young, E. (1997). Effects of acoustic trauma on the representation of the vowel /ɛ/ in cat auditory nerve fibers. *Journal of the Acoustical Society of America*, **101**(6), 3602–3616.

Miller, R., Calhoun, B., and Young, E. (1999). Contrast enhancement improves the representation of /ɛ/-like vowels in the hearing-impaired auditory nerve. *Journal of the Acoustical Society of America*, **106**(5), 2693–2708.

Moore, B., Glasberg, B., and Vickers, D. (1999). Further evaluation of a model of loudness perception applied to cochlear hearing loss. *Journal of the Acoustical Society of America*, **106**(2), 898–907.

Moore, B. C., Glasberg, B. R., Hess, R. F., and Birchall, J. P. (1985). Effects of flanking noise bands on the rate of growth of loudness of tones in normal and recruiting ears. *Journal of the Acoustical Society of America*, **77**(4), 1505–13.

Moore, B. C. J. (2003). *An Introduction to the Psychology of Hearing.* Academic Press, 5th edition.

Moore, B. C. J. (2008). The choice of compression speed in hearing aids: theoretical and practical considerations and the role of individual differences. *Trends in Amplification*, **12**(2), 103–12.

Mustafa, K. and Bruce, I. (2006). Robust formant tracking for continuous speech with speaker variability. *IEEE Transactions on Audio Speech and Language Processing*, **14**(2), 435–444.

Nadol, Jr, J. B. (1988). Comparative anatomy of the cochlea and auditory nerve in mammals. *Hearing Research*, **34**(3), 253–66.

Oppenheim, A. V. and Schafer, Ronald W. with Buck, J. R. (1999). *Discrete-Time Signal Processing.* Prentice Hall, 2 edition.

Oppenheim, A. V. and Schafer, R. W. (1989). *Discrete-Time Signal Processing.* Prentice Hall.

Ottersen, O., Takumi, Y., Matsubara, A., Landsend, A., Laake, J., and Usami, S. (1998). Molecular organization of a type of peripheral glutamate synapse: The afferent synapses of hair cells in the inner ear. *Progress In Neurobiology*, **54**(2), 127–148.

Oxenham, A. and Bacon, S. (2003). Cochlear compression: Perceptual measures and implications for normal and impaired hearing. *Ear and Hearing*, **24**(5), 352–366.

Parsons, T. D. (2006). Neurobiology - auditory fidelity. *Nature*, **444**(7122), 1013–1014.

Peterson, G. E. and Barney, H. L. (1952). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America*, **24**(2), 175–184.

Pickles, J. O. (2008). *An Introduction to the Physiology of Hearing.* Emerald Group Publishing Limited, 3rd edition.

Plack, C., Drga, V., and Lopez-Poveda, E. (2004). Inferred basilar-membrane response functions for listeners with mild to moderate sensorineural hearing loss. *Journal of the Acoustical Society of America*, **115**(4), 1684–1695.

Pressnitzer, D., Sayles, M., Micheyl, C., and Winter, I. M. (2008). Perceptual organization of sound begins in the auditory periphery. *Current Biology*, **18**(15), 1124–1128.

Recio, A., Rhode, W., Kiefte, M., and Kluender, K. (2002). Responses to cochlear normalized speech, stimuli in the auditory nerve of cat. *Journal of the Acoustical Society of America*, **111**(5), 2213–2218.

Relkin, E. and Doucet, J. (1997). Is loudness simply proportional to the auditory nerve spike count? *Journal of the Acoustical Society of America*, **101**(5), 2735–2740.

Robles, L. and Ruggero, M. (2001). Mechanics of the mammalian cochlea. *Physiological Reviews*, **81**(3), 1305–1352.

Rose, J. E., Hind, J. E., Anderson, D. J., and Brugge, J. F. (1971). Some effects of stimulus intensity on response of auditory nerve fibers in the squirrel monkey. *Journal of Neurophysiology*, **34**(4), 685–99.

Ruggero, M., Rich, N., Recio, A., Narayan, S., and Robles, L. (1997). Basilar-membrane responses to tones at the base of the chinchilla cochlea. *Journal of the Acoustical Society of America*, **101**(4), 2151–2163.

Ruggero, M., Narayan, S., Temchin, A., and Recio, A. (2000). Mechanical bases of frequency tuning and neural excitation at the base of the cochlea: Comparison of basilar-membrane vibrations and auditory-nerve-fiber responses in chinchilla. *Proceedings of the National Academy of Sciences of the United States of America*, **97**(22), 11744–11750.

Russell, I. J. and Sellick, P. M. (1978). Intracellular studies of hair cells in the mammalian cochlea. *The Journal of Physiology*, **284**(1), 261–290.

Sachs, M., Bruce, I., Miller, R., and Young, E. (2002). Biological basis of hearing-aid design. *Annals of Biomedical Engineering*, **30**(2), 157–168.

Scheidt, R. E., Kale, S., and Heinz, M. G. (2010). Noise-induced hearing loss alters the temporal dynamics of auditory-nerve responses. *Hearing Research*, **269**(1-2), 23–33.

Sewell, W. F. (1984). Furosemide selectively reduces one component in rate-level functions from auditory-nerve fibers. *Hearing Research*, **15**(1), 69–72.

Shera, C., Guinan, J., and Oxenham, A. (2002). Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements. *Proceedings of the National Academy of Sciences of the United States of America*, **99**(5), 3318–3323.

Shi, L.-F., Carney, L. H., and Doherty, K. A. (2006). Correction of the peripheral spatiotemporal response pattern: a potential new signal-processing strategy. *Journal of Speech, Language, and Hearing Research*, **49**(4), 848–55.

Shotwell, S. L., Jacobs, R., and Hudspeth, A. J. (1981). Directional sensitivity of individual vertebrate hair cells to controlled deflection of their hair bundles. *Annals of the New York Academy of Sciences*, **374**, 1–10.

Tortora, G. J. and Derrickson, B. (2006). *Principles of Anatomy and Physiology.* John Wiley and Sons, 11 edition.

von Békésy, G. (1947). The variation of phase along the basilar membrane with sinusoidal vibrations. *The Journal of the Acoustical Society of America*, **19**(3), 452–460.

Warren, R. M. (2008). *Auditory Perception.* Cambridge University Press.

Wiener, F. M. and Ross, D. A. (1946). The pressure distribution in the auditory canal in a progressive sound field. *The Journal of the Acoustical Society of America*, **18**(1), 248–248.

Wong, J., Miller, R., Calhoun, B., Sachs, M., and Young, E. (1998). Effects of high sound levels on responses to the vowel /ɛ/ in cat auditory nerve. *Hearing Research*, **123**(1-2), 61–77.

Yost, W. A. (2007). *Fundamentals of Hearing: An Introduction.* Elsevier, 5th edition.

Young, E. D. and Sachs, M. B. (1979). Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers. *The Journal of the Acoustical Society of America*, **66**(5), 1381–1403.

Zhang, X., Heinz, M., Bruce, I., and Carney, L. (2001). A phenomenological model for the responses of auditory-nerve fibers: I. Nonlinear tuning with compression and suppression. *Journal of the Acoustical Society of America*, **109**(2), 648–670.

Zilany, M. S. A. and Bruce, I. C. (2006). Modeling auditory-nerve responses for high sound pressure levels in the normal and impaired auditory periphery. *Journal of the Acoustical Society of America*, **120**(3), 1446–1466.

Zilany, M. S. A. and Bruce, I. C. (2007). Representation of the vowel /ɛ/ in normal and impaired auditory nerve fibers: Model predictions of responses in cats. *Journal of the Acoustical Society of America*, **122**(1), 402–417.

Zilany, M. S. A., Bruce, I. C., Nelson, P. C., and Carney, L. H. (2009). A phenomenological model of the synapse between the inner hair cell and auditory nerve: Long-term adaptation with power-law dynamics. *Journal of the Acoustical Society of America*, **126**(5), 2390–2412.