APPLYING MACHINE LEARNING TO SPEECH PROCESSING IN HEARING AIDS

APPLYING MACHINE LEARNING TO SPEECH PROCESSING IN HEARING AIDS

By

JEFF BONDY, B.A.Sc.

A Thesis
Submitted to the School of Graduate Studies
in Partial Fulfilment of the Requirements
for the Degree
Ph.D.

McMaster University

Ph.D. (2005)	McMaster University
(Electrical Engineering)	Hamilton, Ontario

TITLE:	**Applying Machine Learning to Speech Processing in Hearing Aids**
AUTHOR:	Jeff Bondy, B.A.Sc. (University of Waterloo)
SUPERVISOR:	Dr. Ian C. Bruce
NUMBER OF PAGES:	xvi, 226

Acknowledgements

First and foremost I have to thank my family. From my mother and father, to my grandmothers and grandfathers, to the aunts uncles and cousins. It's a great family to be a part of. A special thanks to Manny, who sadly departed before I could finish this thesis, and my grandfather who took the time to be my tour guide in Belgium.

I would be remiss if I did not thank each and every person at McMaster whom I have had the pleasure of meeting and working with over these five years. If I start at the beginning I must thank Dr. Simon Haykin for giving me the wide open question of how to make a hearing impaired person deal as well as a normal hearing person in a cocktail party environment. Lola Brooks helped me understand what needed to be done, and after moving labs I already miss discussing the little issues that surround working in post graduate studies. All the support staff in ECE have to be thanked profusely, Helen, Cheryl, Grace, and Fran are superstars.

My fellow students, whether it is kicking the ball around or having a beer this time hasn't been without its pleasantries. I especially have to thank Ramy for the Shisha time. Dr. Kiruba was very influential, not only on his teaching style and work habits, but also with his experience in seeing that things were not so bad, when I was floundering.

A special thanks must go to Dr. Ian Bruce, the supervisor with the most. Without Ian I'd have never finished this thesis, he was always fair and level headed, which was especially useful to offset my tendencies. Plus, without his paying for Melissa to PostDoc on his Tinnitus project, I might never have met my fiance.

And a huge thank you to Melissa, who never asked me when I am going to be done. Lastly, I'd like to thank the team, especially my brother and team captain, Jason; from the WestEnd to MNF or poker and games you were a cornerstone of sanity on which to build.

Abstract

The neurophysiological basis of sensorineural hearing loss is thought to be hair cell damage or stria vascularis atrophy in the cochlea. The normal cochlea is responsible for a very complex, dynamic, nonlinear analysis and coding of acoustic signals, which is distorted by cochlear impairment. To overcome hearing loss, a typical hearing aid provides linear gain or some simple form of dynamic compression. However, such simple processing cannot fully compensate for the effects of cochlear impairment. In this thesis, machine learning is used to investigate more optimal speech processing schemes for hearing aids.

A model of the auditory periphery is utilized to develop a set of neural predictors of human speech intelligibility. These are shown to have similar accuracy to acoustic predictors of intelligibility such as the articulation index. The neural predictors are then used as error metrics in a machine learning framework to train simple linear and compressive hearing aid algorithms. The results are consistent with empirically-derived prescriptions for hearing linear gain and compression.

It thus appears that to develop speech processing algorithms that provide greater benefits than those currently available in hearing aids, it is necessary understand more fully the distortions that are occurring in the cochlea due to hearing loss and to develop processing algorithms that specifically target compensation of these distortions. An analysis of the differences in compression, suppression and adaptation in the normal and impaired cochlea is performed using the model of the auditory periphery, and specific distortions are quantified. From this analysis, several speech processing algorithms are proposed that may more fully compensate for the effects of cochlear impairment on the neural representation of speech.

Contents

List of Figures

Chapter 1

Executive Summary

1.1 Goals

1. Design an adaptive hearing system that will make a difference to a hearing-impaired person and make it possible for him/her to engage in a conversation in a crowded room, doing so as comfortably as a normal-hearing person.

2. Restore near-normal firing patterns in the auditory nerve, in spite of hair cell damage. Central to this is the processing in the brain that is eminently capable of segregation, streaming, and decoding must still function.

3. Use neuro-physiologically based auditory models to develop predictive measures for offline evaluation and novel functional insights into the nonlinear operation of the cochlea.

4. Quantify how IHC and OHC loss affects the processing of the auditory system as well as how that processing affects perception.

1.2 Dissertation Introduction and Layout

There have been several advances in the understanding of the neurophysiological basis of hearing impairment. Hair cell damage in the cochlea alters the auditory system and has profound effects on the design of hearing-aid systems to combat this type of impairment. While conductive loss, arising from ossicle damage or an ear drum puncture, can largely be overcome with frequency-shaped linear amplification, the types of impairment associated with Inner Hair Cell (IHC) and Outer Hair Cell (OHC) damage directly affects the nonlinear signal processing in the cochlea. Up until the mid 1980's the mechanisms underlying the more prevalent type of impairment, hair cell loss or voltage loss from the atrophy of the stria vascularis, were not well understood. This led to a group of ad-hoc algorithms, largely based on the discerned symptoms (spectrally shaped sensitivity loss, identification in noise problems) as opposed to the mechanisms underlying the symptoms.

The processing of an acoustic signal by the peripheral auditory system can be summarized as follows. A sound signal is directed to the ear canal by the pinna (outer ear). The eardrum responds to the pressure wave by deflecting. This deflection causes the three small bones of the inner ear to move, producing a similar movement in the oval window of the cochlea. This vibration starts a travelling wave in the fluid of the cochlea. Up to this point, the system is well characterized by a linear transfer function, but beyond this point, the system is highly nonlinear and dynamic. The travelling wave produces a peak displacement at some point along the cochlea that is a function of frequency and OHC undamping. OHCs are motile members that precisely modulate the basilar membrane. IHCs transduce the mechanical displacement of the basilar membrane to auditory nerve (AN) firings. The OHCs undamping enhances the IHCs sensitivity and selectivity [Nobili et al., 1998].

The impairment or loss of these hair cells produces symptoms such as elevated thresholds, loss of frequency selectivity, and loss of temporal discrimination [Liberman & Dodds, 1984a]. The consequences of hair cell damage for auditory discrimination are far ranging, taking entire books to catalogue [Moore, 1995]. Chapter 2, provides a background on the auditory system and the associated psychophysics.

The normal hearing process can be described with the block diagram in Figure 1.1, where an input signal X is transformed by the auditory periphery, H, to produce a neural response Y.

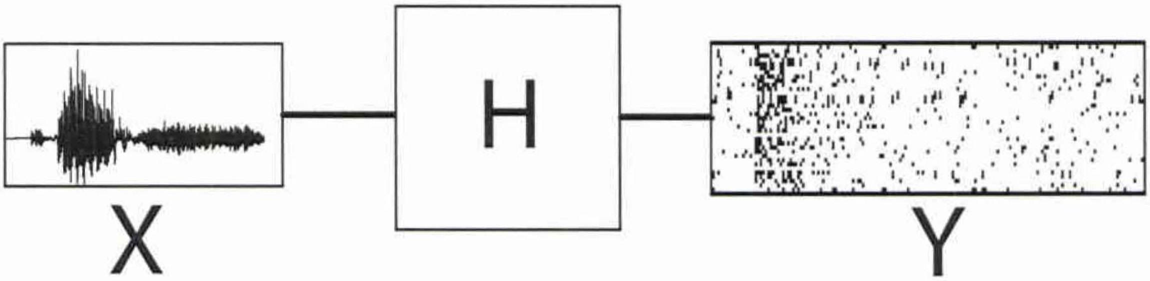


Figure 1.1: Block diagram representation of normal hearing system

The auditory periphery is treated as a black box in many signal-processing applications. In the hearing compensation application this approach has severe limitations. The success of the algorithm will be directly proportional to the amount of information about H that one embeds in the design. With the impairment of hair cells the functionality of H changes, resulting in a hearing impaired system as shown in Figure 1.2. That is, the same input signal produces a distorted neural signal, $\hat{Y}$, when processed by the damaged hearing system $\hat{H}$.

Can a hearing aid algorithm alter the input to the impaired ear in such a way as to return the firing pattern from figure 1.1 to the AN from figure 1.2? Figure 1.3

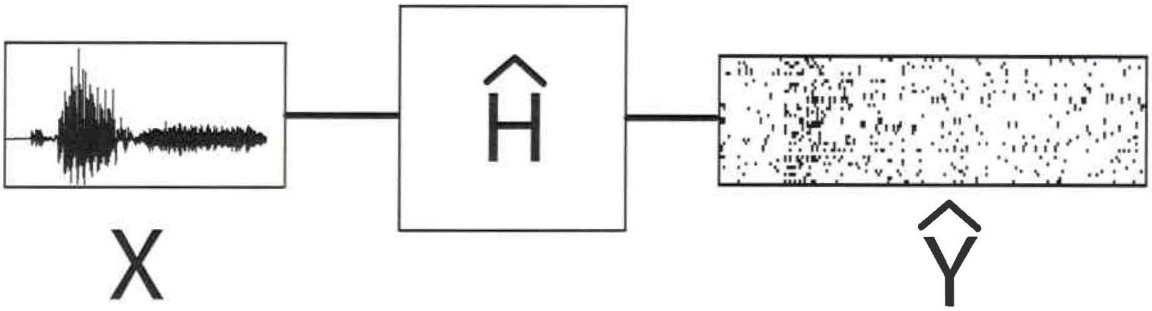


Figure 1.2: Block diagram representation of impaired hearing system

shows the hearing aid process, denoted N_c, before the impaired auditory periphery model, with the same spiking output as the normal ear.

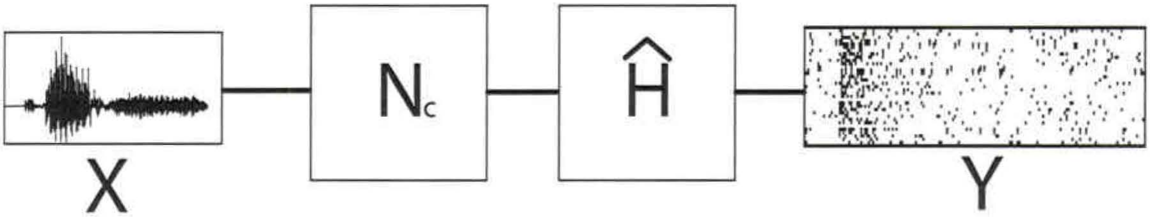


Figure 1.3: Block diagram representation of the compensated impaired representation

There is the real possibility that a perfect return to normal firing is not possible. There might be the loss of a nonlinear process, one to many mapping, or information bottleneck that is intrinsic to sensorineural hearing impairment. This framework then requires some way of evaluating the relative perceptual distortion between the spiking behaviour of two populations of neurons. Chapter 3 is the formation of a metric that predicts intelligibility from a perceptual relevant distance between the normal and hearing impaired AN representations. Chapter 3 details the development of a novel intelligibility metric that provides better results than the Steeneken [1992] Speech Transmission Index (STI) metric.

Chapter 4 then explores what exactly should make up the hearing aid processing block, N_c. The algorithm to alter the input signal is called the "Neurocompensator" [proposed by Becker & Bruce, 2002]; the N_c from figure 1.3. Initial attempts to prove out the machine learning framework by validating against empirical results are given in chapter 4, as well as expanding the framework into novel processing blocks.

If $\hat{H}$ was invertible the optimal hearing aid would be the cascade of the undamaged model and the inverse of the damaged system or $N_c = H\hat{H}^{-1}$, or $H = N_c\ \hat{H}$. This approach to hearing aid design has been explored by Anderson [1994], Chabries et al. [1995], and Anderson et al. [1995]. This algorithmic development belongs to the family of machine learning. Over the course of this dissertation an appreciation of why previous attempts have been unsuccessful because of the simplicity of their cochlear models is seen.

For example, the auditory system has very important nonlinearities [Julicher et al., 2001], time variances [Moore & Glasberg, 1986] and many to one mappings. The simple fact that a sound can be completely masked by the presence of a second sound is evidence that the auditory system discards information. This means a perfect inversion is not possible. However, even if H is non-invertible, one may still be able to capture its capabilities sufficiently to approach normal hearing. This requires embedding a detailed understanding of the auditory system, as well as the phenomenology of sensorineural impairment into the hearing aid signal processing.

One of the big advantages of the approach in this dissertation is the detailed, nonlinear, dynamic impaired and normal auditory models, taken from Bruce et al. [2003]. In the first modelling chapters the power of the model to accurately describe the dynamic nonlinearity has not been used. It became clear that the adaptive nonlinearity of the healthy cochlea was not being captured in the machine learning framework in chapters 3 and 4. Chapter 5 is an evolution of the theoretical basis of cochlear processes. It begins with summing up the differences between the normal and sensorineural impaired cochlea by stating that the impaired cochlea operates more linearly than the healthy cochlea. Modern hearing aid signal processing strategies do not re-introduce the important adaptive nonlinearities in the healthy cochlea that are compromised by sensorineural impairment. Chapter 5 discusses how important these nonlinearities are by polling the normal and impaired cochleas with a range of auditory coding metrics.

The discussion and future work chapter, chapter 6, discusses what could be done with the theoretical framework provided in Chapter 5. Novel algorithms aimed at addressing the core adaptive nonlinear problems of sensorineural impairment are introduced in chapter 6. These new processing strategies are key in mimicking the normal auditory periphery's dynamic nonlinearities. This dissertation concludes with a review of the novel concepts contributed in it, as well as a discussion of some future directions that may provide greater benefit to the sensorineurally impaired.

1.3 Contributions

1. A machine learning algorithm to design hearing aid .

2. New framework for producing fitting strategies that can adapt to particular pathologies offline.

3. Three new studies quantifying the effects of sensorineural impairment on temporal, spectral and level acuity.

4. Two novel processing strategies. The first, reduces the spread of masking associated with hearing loss that is governed by suppression. The second, replicates

the normal auditory systems rate-level growth curves versus a wide range of input stimuli.

1.4 Publications

This dissertation is the result of original research conducted by the author, except for contributions made by the thesis supervisor, Prof. Ian C. Bruce, and by co-authors of journal and conference papers arising from the research presented in this thesis. The publications resulting from each chapter and the contributions made by co-authors other than the thesis supervisors are as follows:

Chapter 3: Some of the results of this chapter were published in a paper presented at an international conference: Ian Bruce, Jeff Bondy, Simon Haykin and Sue Becker, "A Physiologically Based Predictor of Speech Intelligibility", International Hearing Aid Research Conference, Lake Tahoe, 2002. This work was extended in the refereed international conference paper: Jeff Bondy, Ian Bruce, Sue Becker and Simon Haykin. "Predicting speech intelligibility from a population of neurons", Advances in Neural Information Processing Systems 16, Sebastian Thrun, Lawrence Saul, Bernhard Schoelkopf (eds.), MIT Press, Cambridge, MA, 2004. Ian Bruce was pivotal in the inception and implementation of this chapter, his auditory model is central to this chapter and what follows. Sue Becker and Simon Haykin revised and suggested many changes.

Chapter 4: Some of the results of this chapter were published in a refereed journal paper: Jeff Bondy, Sue Becker, Ian Bruce, Laurel Trainor and Simon

Haykin. "A novel signal-processing strategy for hearing-aid design: Neurocompensation", Signal Processing 84(7), 2004, 1239-1253. As well, the nonlinear modelling portion of this chapter stems from the refereed paper presented at the international conference: Jeff Bondy, Ian Bruce, Rong Dong, Sue Becker and Simon Haykin. "Modeling intelligibility of hearing-aid compression circuits", Signals, Systems & Computers, 2003 The Thirty-Seventh Asilomar Conference on, Volume: 1 , Nov. 9-12, 2003, pp:720-724. Sue Becker, Laurel Trainor and Simon Haykin revised and suggest many changes. Rong Dong helped with initial programming.

Chapter 5: A portion of the results of this chapter were published in a paper presented at an international conference: Jeff Bondy, Ian Bruce, "Machine Learning and the Auditory Nerve", presented at: International Hearing Aid Research Conference (IHCON), Lake Tahoe, August 2004. As well as presented at the national conference: Jeff Bondy, Ian Bruce, "Degradation of acoustic coding in the auditory nerve by sensorineural impairment", presented at: Canadian Acoustical Association, Acoustics week in Canada, Ottawa, October 2004. Finally, the temporal modelling was originally presented at the international conference: Jeff Bondy, Ian Bruce, "From neurons to hearing aids", presented at: Toronto Auditory Temporal Processing Symposium, May 27-29, 2005.

Chapter 2

Introduction

The problem tackled in this dissertation is to derive novel signal processing strategies that can unleash the power of the recently introduced digital signal processing hearing aids to improve the quality of life of the sensorineural hearing impaired. There has been a historical difficulty with combining knowledge across the different disciplines involved with hearing aid research. The vast range of different knowledge from fields such as physiology, psychophysics, audiology, acoustics and engineering has not integrated into a processing scheme that improves speech intelligibility beyond that obtained with linear gain hearing-aids, with the possible exception of directional hearing aids.

This thesis attempts to be problem-centric, so the discussion starts with an introduction to the physiology of the auditory system (section 2.1), before moving on to how sensorineural impairment affects important psychophysical measures (section 2.2). Most of the models used to explain the different losses associated with hearing impairment hinge on alteration of one of three processes of the normal cochlea, namely compression, suppression and adaptation. Explaining the processing deficits in terms of changes to one of these dynamic nonlinearities drives much of the later chapters. After this, an introduction to how signal processing can be combined with pathology through the use of machine learning is introduced (section 2.3). Previous attempts with frameworks similar to the one proposed in this thesis are described in section 2.4.

2.1 Auditory System

The human auditory pathway is a complex structure that can be loosely categorized into the auditory periphery (including the pinna to the auditory nerve) and the central processing structures (the cochlear nucleus to the auditory cortex). Much more is known about the functional structure of the peripheral structure, including the main effects of sensorineural hearing impairment, then the auditory brain. The auditory periphery has an important impact on the understanding of speech, or quality of processing by the higher auditory brain centers. A detailed understanding of the auditory periphery is necessary to define sensorineural hearing loss clearly, while the hearing loss' impact on central processing structures is also key. Figure 2.1 shows a simplified representation of the auditory system.

The next section discusses the auditory system starting at the ear (section 2.1.1), and moving inward through the cochlea (section 2.1.2) to the auditory nerve (section 2.1.3). An attempt is made at providing functional and morphological descriptions of each subcomponent. Understanding the phenomenology of the auditory system is a modelling necessity, so modelling of the periphery is expanded upon in section 2.1.4. Quantifying the effect of auditory periphery impairment on the neural representation on the auditory nerve is a final open question, discussed in 2.1.5.

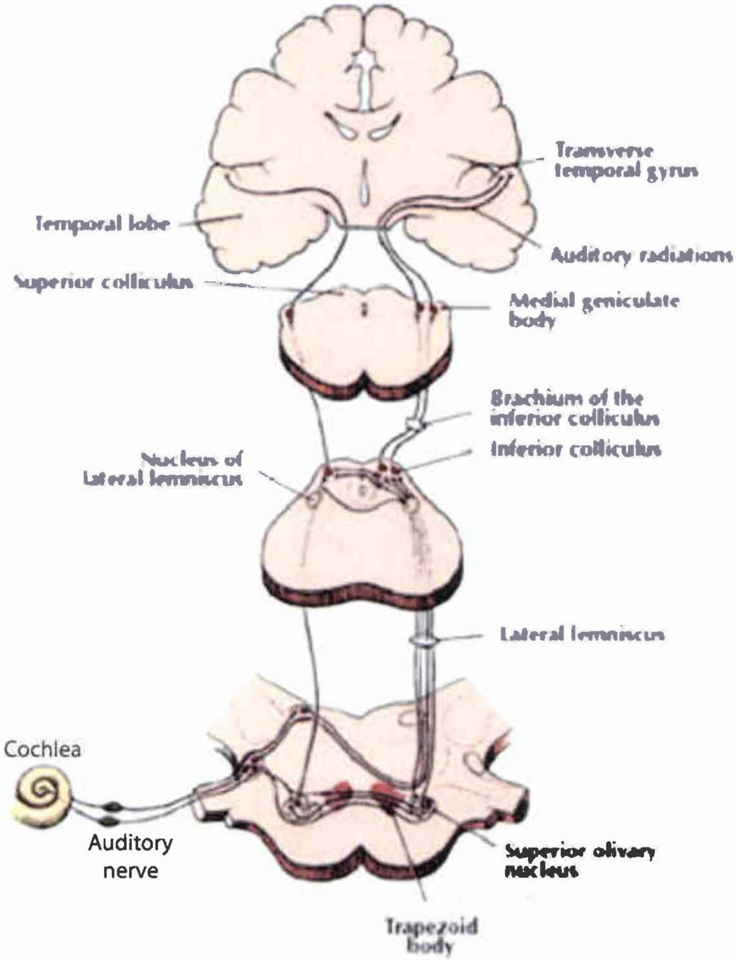


Figure 2.1: Pictorial representation of the major centers of the auditory system. Adapted from Yost & Nielsen [1977]

2.1.1 The Ear

The ear is often split into the outer, middle and inner ear sections. The representation in figure 2.2 processes the acoustic energy from the left, finishing with neural transduction on the right.

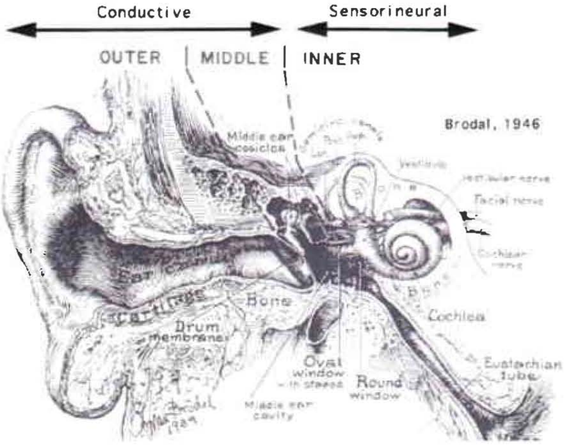


Figure 2.2: Cutaway representation of the ear. The outer ear is comprised of the Pinna and ear canal, the middle ear by the ossicles, and the inner ear by the cochlea. Adapted from Davis & Silverman [1970]

The pinna has evolved to gather and amplify acoustic energy. It has a frequency response that allows the auditory system to make accurate judgements on azimuth and elevation. The ear canal provides gain for important spectral components in speech.

The middle ear also has a frequency response, but it is usually thought of as an impedance transformer. The low acoustic impedance provided by the air must be transferred to the high acoustic impedance of the cochlear fluid. The directed acoustic signal hits the ear drum (or tympanic membrane) before the three small bones known as the ossicles. The small bones, malleus, incus, and stapes, efficiently couple the air vibrations by tapping on the cochlea's oval window.

2.1.2 The Cochlea

The cochlea is a coil, about 2.5 turns, for about 3.5 cm in human adults, often depicted as snail-like. The fluid filled cochlea transduces the acoustic wave into an electrical, neural response. When the stapes taps on the oval window at the base end of the cochlea, a travelling wave is sent along the basilar membrane that runs the length of the cochlea. This wave reaches a peak somewhere along the basilar membrane that is dependent upon the frequency content of the signal. The basilar membrane goes from narrow and stiff at the base of the cochlea to wide and flexible at the apex. This corresponds to a frequency specific deflection impedance which sets a tonotopic organization to the peak displacement of the basilar membrane. Figure 2.3 represents how high frequencies (20 kHz in humans) start at the basal end, and lower frequencies (100 Hz in humans) are at the apical end.

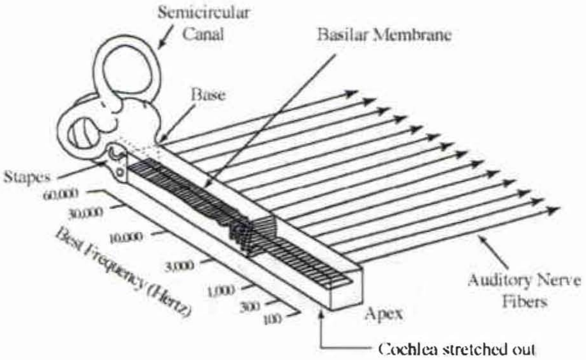


Figure 2.3: Representation of the unrolled cochlea of a cat. Adapted from von Bekesy [1960]

The cochlea has three ducts; the basilar membrane separates the scala tympani (perilymphatic) and scala media (endolymphatic); Reissner's membrane separates the scala vestibuli (perilymphatic) and scala media. On top of the basilar membrane, protruding into the scala media is the organ of Corti, which houses the mechanisms of mechanical to electrical transduction. Figure 2.4 is a cross-section of the healthy cochlea. A close up of the organ of corti is in figure 2.5.

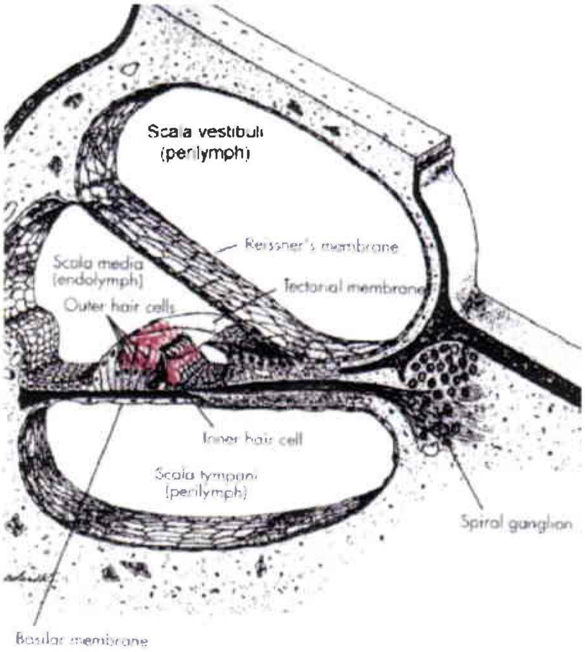


Figure 2.4: Cochlear cross-section with the three ducts. The inner, outer hair cells, basilar and tectorial membranes are all part of the organ of Corti. Taken from Nolte [1993]

The organ of corti is made up of various structural, supporting and sensory cells. Table 2.1 is an abbreviated list of the important cells in and around the organ of Corti.

The hair cells are arranged in rows running the length of the basilar membrane. There are three (or sometimes four) rows of outer hair cells; approximately 15,000 in all, and a single row of inner hair cells; about 4000. Hair cells have small stereocilia-like, "hair fibers", protruding towards the tectorial membrane. Figure 2.6 shows a representation of an outer hair cell. Inner hair cells are almost exactly the same, except they are not as vertical, they have a more arched body.

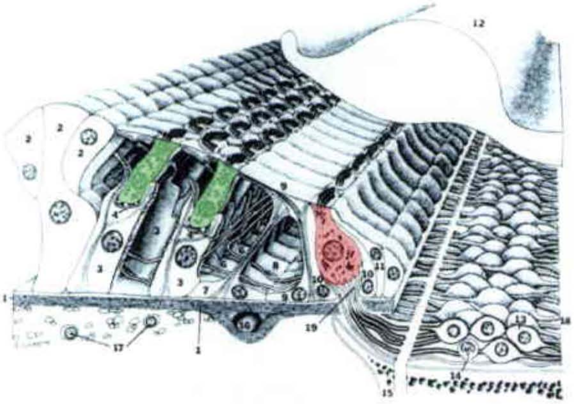


Figure 2.5: Detail of the Organ of Corti. Taken from Smith [1980]

Cell Name	Type	Description
Inner Hair	Sensory	Mechanical to electrical transducer
Outer Hair	Sensory	Electro-motile active element
Inner Tunnel Pillar	Structural	Support and forms part of tunnel of Corti
Outer Tunnel Pillar	Structural	Support and forms part of tunnel of Corti
Deiters	Structural	Under OHCs, effect active response
Hensen	Transport	Adjacent to Deiters, fluid or ion conduit
Claudius	Support	Extend from Deiters cells to spiral ligament

Table 2.1: Cell structures comprising the Organ of corti.

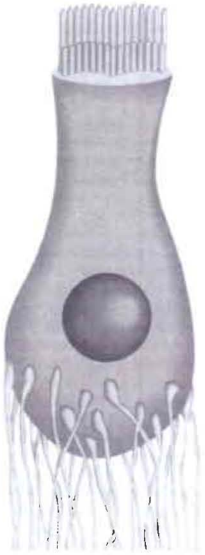


Figure 2.6: Detail of an Outer Hair Cell. Taken from Dallos et al. [1996]

The stereocilia have a height gradient, and the highest of them connect into the tectorial membrane for outer hair cells, but are thought not to for inner hair cells. The tectorial membrane is a gelatinous form running in parallel with the basilar membrane, and whose mechanical mass increases towards the apical end of the cochlea. This mass gradient produces a frequency peak displacement response analogous to the basilar membrane's elasticity derived tonotopic response. It is thought that the two membrane's are not exactly matched so a signal would produce a relative motion between the membranes, causing a shear on the stereocilia connection of the outer hair cells. Whether deflection by shear or deflection by the traveling wave, the motion of the stereocilia produces a potential change across the hair cell membrane. This potential change produces a motile response in outer hair cells, and leads to synaptic transmission in inner hair cells.

The motile response of the outer hair cells is essential to normal hearing. It is an active, nonlinear process that produces the frequency selectivity, sensitivity and adaptive, dynamic range compression which ultimately leads to the fantastic speech recognition ability of a normal hearing person. The inner hair cell is ultimately responsible for changing mechanical energy into the electrical energy encoded on the auditory nerve. Damage to OHCs and IHCs are the central issue in sensorineural hearing loss.

2.1.3 The Auditory Nerve

The AN, or nerve VIII, is the interface between the auditory periphery and the auditory brain. Acoustic information on the auditory nerve is carried in spikes and shows many of the dynamic nonlinear processes of the auditory periphery. The AN itself is a homogeneous bundle of approximately 30,000 mostly myelinated fibers in humans (50,000 in cats). This section deals with the normal operation of transduction of pressure waves into the AN representation.

The AN is comprised of axons from two types of spiral ganglion cells. 95 % are Type I neurons, with myelinated cell bodies which innervate inner hair cells. Each IHC has 15 to 20 of these synapsing to them. One IHC is connected to many neurons, conversely, single Type II neurons connect to many OHCs. The Type II neurons are unmyelinated and project to different areas of the cochlear nucleus from Type I neurons.

Not much is known about the Type II responses to acoustic stimulus, and Type I are hypothesized to carry the bulk of the stimulus information. The Type I neurons are typical spiking neurons, mediated by the electro-mechanical IHC process. Hyperpolarizing currents and small depolarizing currents produce small changes in membrane potentials, but large depolarizing currents exceeding a threshold produce an action potential (or spike) which travels along the axon without attenuation.

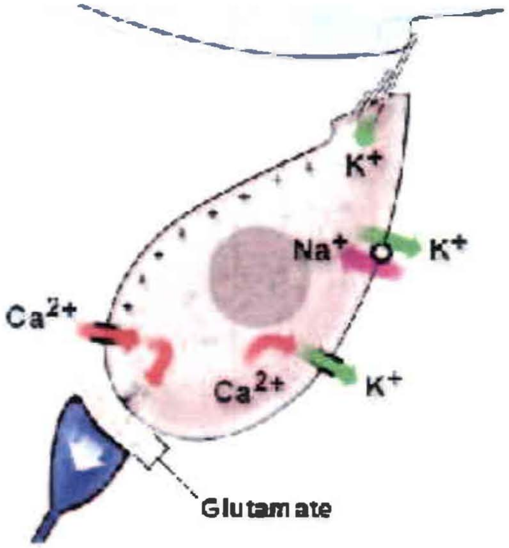


Figure 2.7: Schematic of the ion flow thought to describe IHC transduction. The light blue wave at the top represents a pressure gradient in the cochlear fluid that causes a large positive deflection of the IHC cilia. This opens gating channels in the cilia, and causes the cascade of depolarizing currents, and subsequent return to a quiescent state. Taken from Glowatzki [2004]

A positive pressure gradient in the scala media causes the IHC cilia tips to deflect in the positive direction. This stretches tip-link fibers between the cilia, which are attached to mechanical gates, and pulls these gates open. The gates are K^+ ionic channels, which depolarizes the IHC membrane potential, initially open Na^+ and K^+ ion channels on the IHC body, and Ca^{2+} channels at the base (these channels quickly close). The calcium channels induce neurotransmitter release. Vessicles filled with glutamate are pushed across the synaptic cleft, fusing there and causing excitatory postsynaptic potentials (EPSP), with enough frequency these EPSPs create action potentials.

A negative pressure gradient in the scala media causes the opposite effect, closing all the cilia gates. This is the cause of the half wave rectification on the AN. Even without a positive pressure gradient a portion of the IHC cilia K^+ gates are open, which can lead to spontaneous activity.

There are three groups that describe the spontaneous discharge rates (SR) of particular fibers. The most common, high-SR group (SR $>$ 18 spikes/s) provides 60% of the AN population. The medium-SR group (0.5 $<$ SR $<$ 18 sp/s) contains about 25 % of healthy fibers with the low-SR group (SR $<$ 0.5 sp/s) accounting for the rest [Liberman & Kiang, 1978]. Spontaneous discharge rate is inversely related to threshold at the CF, low-SR fibers have the highest thresholds, while high-SR fibers have the lowest thresholds.

By studying the AN response, one can infer cochlear processing. For example, in a single fiber, the pure tone response that produces activity levels statistically above spontaneous rates would be the neural equivalent of psychophysical tuning curves. Neural tuning curves are sharp for for low-CF ($<$ 2 kHz) fibers, and sharp, but with a broad tail extending to low frequencies for high-CF fibers. Low-SR fibers commonly exhibit larger Q10s than high-SR fibers, even though the same inner hair cell is innervated by fibers from all 3 groups [Miller et al., 1997].

Low-frequency (< 5 kHz) pure tones produce phase locking of spike discharges. Spikes tend to occur at a particular phase of a stimulus, but not every cycle. Phase locking is usually quantified using period histograms, which display the distribution of spikes within a stimulus cycle. With no phase locking the period histograms would be a uniform distribution, while perfect phase locking, would produce an impulse at a particular phase.

For pure tones above 1 kHz phase locking falls off. Above 5-6 kHz, the synchronization index reaches the noise floor. This fall-off loosely corresponds to the decrease in the AC component of the IHC receptor potential relative to its DC component due to the hair-cell membrane capacitance and resulting lowpass response. Additional stages of lowpass filtering must also play a role.

Phase locking seems a part of the half wave rectification process brought on by hair cell bundle displacement. The other cochlea nonlinearities are less obvious. Discharge rate growth versus input level functions of ANFs for tones at CF show both hard saturation and a "sloping" saturation. Low-threshold, high-SR fibers tend to have a hard saturation, while high-threshold, low-SR fibers a sloping saturation. This AN compressive response is modelled by cascading two nonlinearities; a peripheral soft compression (power-law type) is followed by a central, hard sigmoid nonlinearity whose operating point correlates with fiber threshold. The peripheral nonlinearity closely matches the 3:1 compression seen in basilar membrane motion after about 35 dB SPL. The central nonlinearity is thought to arise at the hair-cell auditory-nerve synapse. This two stage model predicts that the knee in the rate-level function should occur at the same SPL for all fibers innervating the same place, which is empirically correct [Sachs et al., 1989].

After compression, another nonlinear process that is shown in the AN response is adaptation. The onset of a tone-burst, produces discharge rates much greater than steady state rates. AN fibers quickly react to changes in a stimulus, then gradually decay to a steady level. The fastest, and largest adaptation decay is the fast adaptation that has a time constant of about 2 milliseconds. Other rates include processes from 40 milliseconds, to hundreds of milliseconds, with other adaptation responses extending past a full second [Fettiplace, Ricci & Hackney, Fettiplace et al.].

Adaptation is thought to arise at the synaptic cleft, a sort of filling and depleting of different storage mechanisms of neurotransmitter. Some research shows that there is no adaptation at the membrane potential. While postsynaptic potential data from the goldfish show adaptation consistent with a depletion of neurotransmitter on the presynaptic side [Westerman & Smith, 1988].

Adaptation changes the linear coding of speech in a very interesting way. The onsets of unvoiced phones, are captured by high CF fibers, while voiced phones show large adaptation spikes in low-CF fibers [Delgutte, 1980; Delgutte & Kiang, 1984]. This effect has brought some researchers to say that adaptation enhances the representation of rapid onset transients in speech. It also has the ability to demarcate when a stimulus is turned off. Fibers, who have strongly adapted will show poststimulatory depression after the stimulus is turned off. This poststimulatory depression lasts longer for large intensity adapting stimulus and is another way that the nonlinear, dynamic cochlea enhances contrast both spectrally and temporally.

Spectral contrast is also enhanced through the cochlear process of suppression. A tone can have its average discharge rate suppressed in response to another, excitatory tone at the CF. A tone at a certain CF can attenuate the firing rates of adjacent frequency bands [Sachs & Kiang, 1968]. The one tone versus another is called two-tone rate suppression. A fixed suppressor shifts the rate-level function for a CF tone towards higher intensities. Suppression increases with the intensity of the suppressor tone and strongly depends on suppressor frequency.

Suppression is asymmetrical. The growth of suppression for high side suppressors is less than a dB if the suppressor level is increased by 1dB. The growth of suppression for suppressors much lower in frequency than CF can exceed 2 dB/dB. This rapid growth of low-side suppression is responsible for "upward spread of masking". Or the psychophysical asymmetry, where it is easier to mask a tone with a low frequency masker than a high frequency masker.

2.1.4 Auditory Modeling: The Periphery

The auditory periphery model used throughout is from Bruce et al. [2003], following initial work by Bruce et al. [1999], Heinz et al. [2001] and Zhang et al. [2001]. The nonlinear, computational AN model includes gamma-tone filters with compressive magnitude responses, two-tone suppression, saturating rate-level curves, low-threshold high-spontaneous rate (HSR) fibers, rolloff in phase-locking, neural adaptation, and realistic onsets and offsets. The system is shown in Figure 2.8.

This model describes the function of the auditory system from the middle ear to auditory nerve. For outer ear functioning the head related transfer function from Wiener & Ross [1946] is used. The auditory model itself comprises several sections, each providing a phenomenological description of a different part of auditory periphery function.

The first section models middle ear filtering. The second section, labeled the "control path," captures the OHC's modulatory functions, and includes a wideband,

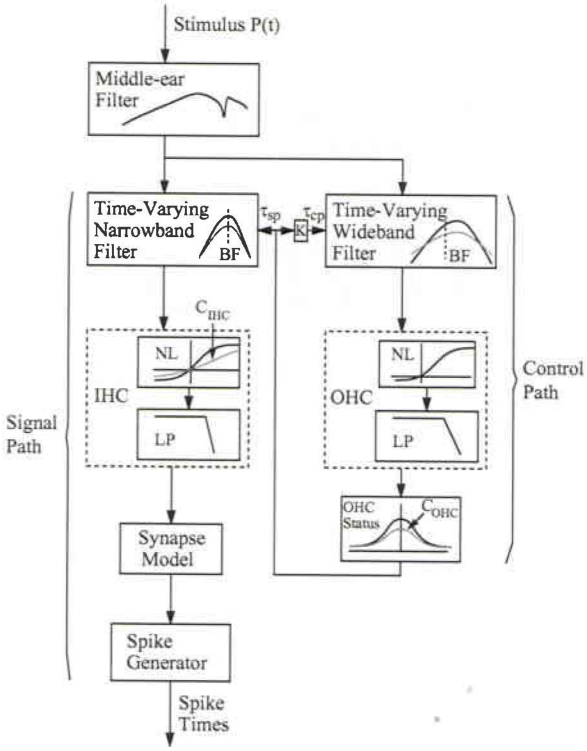


Figure 2.8: Block diagram of the computational model of the auditory periphery from the middle ear to the Auditory Nerve. Taken from Bruce et al. [2003].

nonlinear, time varying, band-pass filter followed by an OHC nonlinearity (NL) and low-pass (LP) filter. This section controls the time-varying, nonlinear behavior of the narrowband signal-path basilar membrane (BM) filter. The control-path filter has a wider bandwidth than the signal-path filter to account for wideband nonlinear phenomena such as two-tone rate suppression.

The third section of the model, labeled the "signal path", describes the filter properties and traveling wave delay of the BM (time-varying, narrowband filter), the nonlinear transduction and low-pass filtering of the inner hair cell (IHC NL and LP), spontaneous and driven activity and adaptation in synaptic transmission (synapse model), and spike generation and refractoriness in the auditory nerve (AN). In this model, C_{IHC} and C_{OHC} are scaling constants that control IHC and OHC status, respectively.

The gain functions of linear versions of the time-varying narrowband filter in the signal path, plotted as gain versus frequency deviation $\Delta f$from the filter's Best Frequency (BF) are given in Figure 2.9.

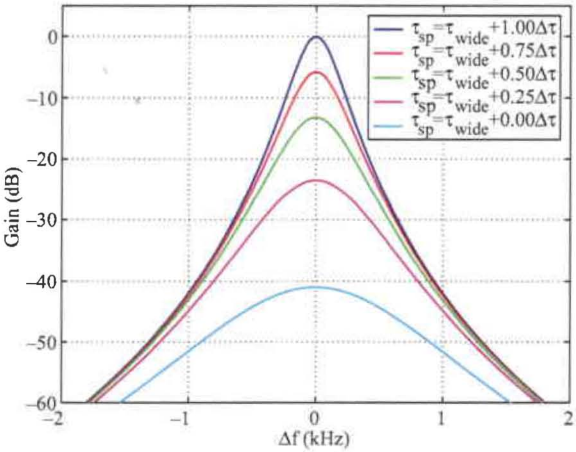


Figure 2.9: Filter shaping functions of the time-varying narrow-band filter in the signal path, plotted as gain versus frequency deviation (Δf) from BF. This example is at 1.7 kHz. Taken from Bruce et al. [2003].

The filter is fourth-order and is plotted for five different values of τ_{sp} between τ_{narrow} and τ_{wide}. τ_{sp} is the time-bandwidth control parameter, where larger values correspond to more frequency selectivity, and $\tau_{sp} \quad \varepsilon \ [\tau_{wide}, \tau_{narrow}]$. $\Delta\tau = \tau_{narrow} - \tau_{wide}$. τ_{narrow} was chosen to produce a 10 dB bandwidth of ~450 Hz, and τ_{wide} was chosen to produce a maximum gain change at BF of ~41 dB at 1.7 kHz. This plot can be interpreted as showing the nominal tuning of the filter with normal OHC function at five different sound pressure levels, or alternatively, as the nominal tuning of the filter for five different degrees of OHC impairment.

The success of the machine learning strategies presented in this dissertation depends upon the accuracy of the auditory model of the normal and damaged ear. The Bruce et al. [2003] model, while being based on cat physiology, is thought to correspond very closely with human physiology. This particular model has a long history of development and good fit to a wide range of empirical data. The auditory model can capture a range of phenomena due to hair cell dynamic nonlinearities, including loudness-dependent sensitivity and bandwidth modulation (as stimulus intensity increases the output response levels off and frequency-tuning becomes broader), and masking effects such as two-tone suppression. The model incorporates critical properties of the auditory nerve response including synchrony capture in the normal and damaged ear and replicates several fundamental phenomena observed in electrophysiological experiments in animal auditory systems subjected to noise-induced hearing loss. For example, with OHC damage, high frequency auditory nerve fibers' tuning curves become asymmetrically broadened toward the lower frequencies and tend to become synchrony locked to lower frequencies.

The Bruce et al. [2003] model is capable of simulating all known auditory nerve responses in both a normal and damaged human auditory system accurately. The damaged system model must be tuned to the parameters of a particular individual's hearing-impairment. This requires estimates of both inner and outer hair cell loss over a range of frequencies. The standard audiological assessment, the audiogram, simply measures the threshold for pure tones at each of a small set of frequencies. An elevation in pure tone threshold cannot differentiate between a reduction in OHC driven gain versus a loss of IHCs tuned to that frequency. In sensorineural hearing disorders, it is generally assumed that a moderate elevation in threshold primarily reflects OHC loss, while a severe elevation reflects an additional IHC loss. Although this pattern is typical in individuals with age-related and noise-induced hearing loss, the exact proportion of IHC to OHC loss may deviate from the typical pattern in some individuals, and also may not hold at all for individuals with less common types of sensorineural damage, e.g. drug-induced. Better methods for estimating, separately, the degree of inner and outer hair cell loss, such as using noise-masked tones (Moore et al. [2000]) are intrinsic to this strategy. Given accurate measurements, the model could be tailored to compensate for many individual patterns of deficits.

2.1.5 Neural Modelling: Differences in Neural Codes

Just as section 2.1.4 attempted to model the processes introduced in 2.1.2, this section deals with modelling the AN data introduced in 2.1.3. While the cochlea has a tremendous amount of data associated with trying to understand its workings, neural codes have probably been discussed an order of magnitude more. And while cochlear models have some agreement among researchers in the field the same cannot be said of the neural code. The neural code used by the AN is open to a multitude of interpretations.

The goal of this section is to introduce some neural coding hypotheses, and the tools to quantify the differences between neural codes. In chapter 5 quantifying the difference between neural codes with and without the imprints of the dynamic nonlinearities of the cochlea become supremely important.

As discussed in section 2.1.3 each fiber on the AN produces a set of action potential (AP) over time. There are a large number of theories on how information is encoded by APs.

1. **Binary Coding:** either AP is a signal, or it is silent

2. **Rate Coding:** the average firing frequency over a certain time period.

3. **Interspike Interval Coding:** Temporal sequence of spike times.

4. **Population Coding:** Output of a network is a pattern of activity across the population of neurons. There are different ways to view population coding, in essence saying that spikes are informative in relation to spiking behaviour of other fibers.

5. **Population, Local Coding:** each neuron represents a specific feature that the system distinguishes

6. **Population, Scalar Coding:** firing rate of each neuron encodes a feature. Redundancy and improved signal-to-noise ratio can be achieved by several neurons coding the same features

7. **Population, Vector Coding:** features are encoded in the firing rates of a subpopulation of neurons that have overlapping tuning curves in the feature space.

8. **Population, Volley Coding:** A single object in the real world may be encoded in the synchronous firing of neurons that code for each separate feature of the object. In this way multiple objects can be represented simultaneously and distinguished by the neurons representing one object firing out of phase with the neurons representing another object.

9. **Population, Synchronization/Oscillatory Coding:** There is evidence for an increase in synchrony between cortical neurons responding to a stimulus, without necessarily any change in their average firing rates. This synchrony is seen in nearby neurons, between neurons in different cortical areas and even across hemispheres. An example is in the auditory system, where in response to a sound stimulus, neurons do not change their average firing rate, but they do fire more in synchrony with each other - this synchrony may be detected by downstream neurons, thus recognizing the presence of the sound.

In reality, the brain probably uses a varied set of coding strategies that are optimal depending on neural resource, the input stimuli, and what the desired output format is. While a tremendous amount of research has gone into trying to provide a single general neural coding strategy, it has largely been less then entirely fruitful.

Shortly after Shannon's introduction of information theory Attneave [1954] and Miller [1956] discussed the possibility of treating biological sensory systems as communication channels. Uttley [1970] introduced *Informon*, that minimized mutual information between the input and output, and gave rise to discrimination functions. Linsker [1992] with *InfoMax*, provided a possible general rule for neural coding; that neural systems should maximize the mutual information between the input and output of the system. Becker [1996] put a novel twist on *InfoMax*, producing *Imax*, which maximized the mutual information between outputs of neighbouring neural networks, and extracted spatially coherent features for visual processing. Ukrainec & Haykin [1996] did the opposite, minimizing the mutual information, to produce spatial incoherent features that were effective at denoising radar images. These information maximization approaches have some usefulness.

Running in parallel with information maximization is redundancy reduction. While Attneave [1954] mentions this specifically, Barlow is most often quoted in describing the biological imperative for redundancy reduction [Barlow, 1961]. Here the goal for the neural code is to be as efficient as possible. Under certain conditions redundancy reduction is equivalent to maximization of input-output mutual information.

While fun, and sometimes useful, most attempts to derive real insights into a general neural code are little more than whistling in the dark. Yet, independent of the actual neural code, information theory does provide several tools to examine differences between AP sequences. Computing the mutual information between some small AP feature and the stimulus set can begin to form a foundation. This can then be extended into asking how do the small AP features code specific aspects of the stimulus? How do neurons interact to transmit information together? How well is a fiber encoding the acoustic stream if the cochlear preprocessing is damaged? While this is beyond the current dissertation, looking at small features of the AP is not.

In general, one neural fiber's spiking behaviour is compared to a similar fiber, but coming from a damaged cochlea. The small features of the spiking behaviour could be:

1. **Spike counts:** Counting the number of APs over a certain time period is used throughout section 5.1. For a homogenous Poisson process, it is a sufficient statistic. The AN response is not homogenous though, speech produces large

rate changes over time.

2. **Spike counts with inter-spike-intervals:** Spike counts can be extended for renewal processes. These are processes in which the distribution of inter-event intervals is independent of past events, but is not necessarily exponential as in a Poisson process. Renewal processes are often used for AN fibers because AN neurons reset after spiking. Biophysical processes in the cell such as membrane voltage and ion channel configuration in the soma and proximal dendrites return to a zero state, and with it the past spiking activity of a cell is decoupled from the future. This type of analysis is more complex and severely limited. In essence it takes the spike count and adds another statistic dealing with the temporal distribution of spike waiting times. It does not take into account complex patterns of spikes. Refined calculations can be made by adding on more complex connections between spikes. The inter-spike-interval can be viewed as the linear term in the Taylor series expansion of correlations between spikes. Using higher order statistics of inter-spike-intervals, say among groups can asymptotically increase the accuracy to rival the binary word method below.

3. **Spike counts with heterogenous firings:** Similar to the other spike counting paradigms. Heterogenous firing rates can be taken into account by discretizing time, and producing an analog probability of firing rate per period by summing across the AN responses. This type of calculation is used in chapters 3 and 4.

4. **First spike latency:** The timing of the first spike after stimulus onset carries a considerable amount of information about the stimuli. In an experiment from the visual domain, van Rullen & Thorpe [2002] show how a network that only processes the very first few spikes from a simulated retina can convey

almost all the information in a scene. An intuitive explanation is that neurons whose firing fits the homogeneous Poisson model, the inter spike interval is exponentially related to the firing rate. This allows an estimate of the rate in a very short time. While estimating spike counts requires averaging over a relatively long time window, the time interval between stimulus onset and the first spike conveys the same information but only requires the observation of a single spike. A method very similar to this is used in section 5.3

5. **Spike patterns as binary words:** The direct method captures potential information in temporal patterns of spikes. Each spike train is represented as a binary string. Time is discretized. This is exponentially more computationally complex than **Spike counts with heterogenous firings**. But with asymptotically infinite data and infinitesimal resolution, it captures the complete information that the spike trains convey regardless of their underlying distribution. With coarse resolution the binary word method extracts the coarse temporal structure of the responses, similar to the one captured by heterogenous rate models. This method was discarded during the initial phases of section 3.1 because of its computational complexity.

there are many other possibilities. In the best possible world, researchers would have an understanding of how the AN information is used. The best guess that is available right now, is that each researcher builds neural architectures mimicking the cochlear nucleus and applies Hebbian strategies to the incoming neural signals. This would show how the normal and impaired AN responses code information.

2.2 Psychophysics of Sensorineural Impairment

The reduction of sensitivity to low intensity acoustic pressure waves is probably the best understood symptom of sensorineural hearing impairment. Threshold shift, through the use of the audiogram has often been taken as the sole determining factor for fitting hearing aids. Threshold shift is often considered the largest factor creating hearing difficulty, leading to hearing aid algorithms hallmarked by attempting to returning audibility. But it is now a common understanding that audibility does not ensure intelligibility.

There is a large problem with thinking that audibility ensures intelligibility, because people with very similar audiograms behave very differently. People with moderate hearing loss have a wide range of deficits (2.5 to 5 dB signal to noise ratios (SNR)) on Speech Reception Thresholds (SRT) than normal hearing people [Glasberg & Moore, 1989]. Even with the same threshold shift, intelligibility can be quite different, so researchers have looked closer at the cause of hearing impairment.

Sensorineural impairment is thought to be caused by damage to stereocilia on the haircell bundles discussed previously. Different levels of damage to IHCs and OHCs can produce the same level of threshold shift while changing other symptoms. What follows is a discussion of the psychophysical consequences to sensorineural impairment or the symptoms. Each subsection has a short description before the detailed data. The consequences for the hearing impaired are given as a motivation to turn engineering into something that is socially beneficial. Finally, an attempt to describe the active processes that are damaged from sensorineural impairment and that give rise to the psychophysics are given in each subsection in an attempt to come to terms with the root causes.

The first several symptoms, threshold shift, reduced dynamic range and reduced frequency selectivity in sections 2.2.1, 2.2.2, and 2.2.3 are the most well known. These three symptoms of sensorineural hearing loss are quickly quantified in a clinic, and are often classified as spectral deficiencies. The next symptoms, changes, or lack thereof in tuning curves due to temporal masking, longer temporal integration and decreased temporal resolution in sections 2.2.4, 2.2.5 and 2.2.6 are less well known. They dovetail from the discussion on frequency selectivity, and highlight the microscopic aspects of temporal changes in the sensorineural impaired auditory system. The auditory system has been shown to integrate acoustic changes over time and frequency. Amplitude and frequency modulation discrimination are about the same in normal hearing and hearing impaired people, but there are some interesting differences, these are discussed in sections 2.2.7 and 2.2.8. The final three sections deal with hearing impaired people's difficulties in reverberation, source localization and competing speech. The sections 2.2.9, 2.2.10 and 2.2.11 deal with macroscopic definitions of sensorineural impairment. This is where obvious, large scale, intelligibility differences are seen between normal hearing and hearing impaired people (20 dB SNR deficits!), and because these sections are closer to real life use, why they are key for motivating the understanding of hearing impairment. Section 2.2.11 is the core problem encompassing the various symptoms. If one can understand the cochlear process in the competing speech regimes, one will have a clear representation of the tradeoffs necessary for optimal hearing aid processing.

2.2.1 Threshold Shift

2.2.1.1 Description

The loss of absolute threshold, or the minimum detectable level of a sound, without any other competing sounds accompanies sensorineural impairment. There are two basic ways of measuring the absolute threshold. The first places a probe microphone close to the subjects ear canal to record the sound pressure level (SPL) that elicits a response when the stimuli is presented through headphones. This is called the minimum audible pressure (MAP). The second way is to present the subject the sounds through loudspeakers in an anechoic chamber. This method's threshold is ordinarily called the minimum audible field (MAF). These two results differ because of every individual's differing pinna and canal response. The outer ear produces a gain for frequencies between 1 and 9 kHz, with up to 15 dB of gain at 3 kHz. Figure 2.10 shows the MAF for people generally considered to have normal hearing.

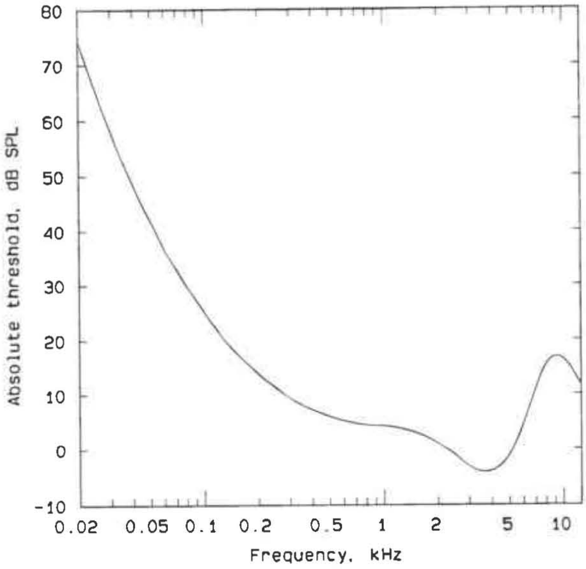


Figure 2.10: The Minimum Audible Field for young, normally hearing people. Taken from Robinson & Dadson [1956].

Thresholds have a basin between 1 kHz and 5 kHz, and then huge increases under 300 Hz and above 6 kHz. Sensorineural impairment can come in many different forms, but the most common is age related or presbycucis. Presbycucis is typified by a high frequency threshold shift. Classically, the audiogram has been made from recording the level shift in dB caused by hearing impairment versus the normal hearer's curve (Figure 2.10). The units are usually in dB HL (dB Hearing Level). Sounds can be characterized by the amount relative to hearing threshold at which they are presented, in dB SL (dB Sensation level), or in absolute pressure ratios versus 20 μ P, in dB SPL. For normal hearing versus a hearing impaired subject, the dB SL can be quite different given the same dB SPL.

Psychoacousticians typically plot thresholds as increasing upwards, as in Figure 2.10. Audiologists typically plot threshold degradations, or hearing losses, as plotted downwards, with the "normal" threshold as a horizontal line at the top of the plot. Thus, the degree of hearing loss is plotted below the normalization line.

2.2.1.2 Data

The most common way of quantifying hearing loss is in terms of the absolute threshold for sinusoids, in dB HL, averaged over the frequencies 500, 1000 and 2000 Hz. This is also known as the pure-tone average (PTA) hearing loss, and is used in many hearing aid fitting procedures, such as those from the National Acoustic Laboratory (NAL) of Australia. Goodman [1965] proposed the following classification:

At present, the level where mild hearing loss is typically demarcated is a PTA of 16 dB HL. The categories are an attempt to indicate broadly the difficulties in communicating in typical situations. That is mild or moderate to severe difficulties in normal interactions. A note of caution though: individuals with similar absolute

PTA dB HL	Description
-10 to 26	Normal limits
27 to 40	Mild hearing loss
41 to 55	Moderate hearing loss
56 to 70	Moderately severe hearing loss
71 to 90	Severe hearing loss
over 90	Profound hearing loss

Table 2.2: Categorization of hearing loss from the Pure Tone Audiogram. Taken from Goodman [1965]

thresholds can vary considerably in how their impairment affects their ability to hear in real situations.

2.2.1.3 Consequences

The most obvious consequence from threshold shift is the loss of audibility. Low level sounds, such as whispers simply cannot be heard. While the auditory system is eminently capable of filling in blanks, such as using a telephone with its 3.4 kHz bandwidth, unvoiced sounds may be under threshold, especially in low level conversations. Hearing aids have been largely defined by ensuring audibility. Chapter 4.1 details the seminal efforts in dealing with threshold shift and how it forms the basis for most hearing aid research.

2.2.1.4 Phenomenology

Liberman & Dodds [1984a] conducted studies in an attempt to relate hair cell damage to changes in AN responses. By using noise exposure and/or ototoxic drugs they were able to produce a mixed set of OHC and IHC damages in cats. This produced some animals with various ranges of IHC and/or OHC damage. Both IHC and OHC damage resulted in elevated thresholds. Complete OHC loss without corresponding IHC damage can result in a threshold shift from 30 dB for low frequency to an almost 60 dB shift at higher frequencies. Complete IHC loss without corresponding OHC damage can result in complete loss at a frequency, or a hole in hearing.

Recently, some other suggestions have come to light, such as loss of voltage from stria vascularis atrophy [Gratton et al., 1996]. It is a research goal to try to determine the connection between hair cell damage and psychophysical threshold detection. An example of this line of research is in Slepecky et al. [1982].

2.2.2 Dynamic Range

2.2.2.1 Description

A higher than normal absolute threshold does not correspond with an increase in the highest intensity a sound may be before it is deemed uncomfortable. This means that the growth of loudness with increasing sound level is greater than in normal auditory systems. For example, someone with 50 dB threshold shift equates a 50 dB sound with a normal hearers 0 dB sound, while both normal and impaired ears hear 120 dB sounds the same. The range of sounds that the impaired ear can comfortably listen to, or the auditory system's dynamic range is greatly reduced. The audiology term for this loss of dynamic range is loudness recruitment.

2.2.2.2 Data

Moore et al. [1985] studied loudness recruitment in subjects with unilateral hearing loss. By having the subject match the loudness of tone burst between the normal and impaired ear they came up with the mean results in Figure 2.11

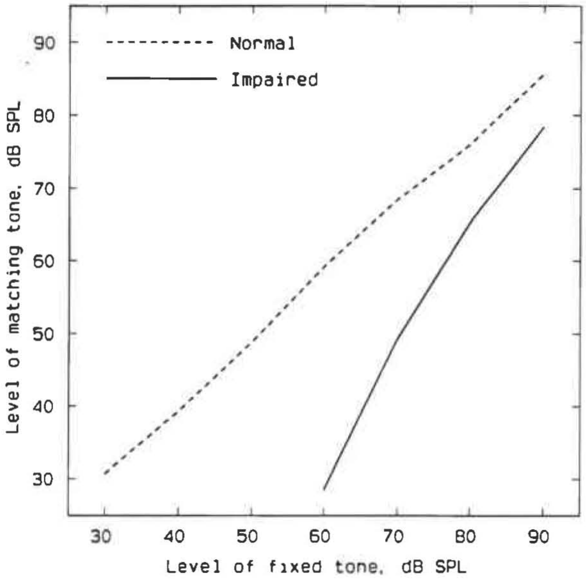


Figure 2.11: The solid line shows the intensity level matching for tone burst for people with asymmetric hearing loss. The dotted line is loudness matching for normal hearing people. Taken from Moore [1995].

From Figure 2.11 the slope for normal ears is close to one (there is a bias for the matching tone to be slightly less then the test tone, especially at high levels since subjects tend to avoid high sound levels). The curve for ears with unilateral impairment, have a slope greater than one. At low levels there is a greater discrepancy between the normal and impaired ears (33 dB), than at high frequency (12 dB).

Loudness recruitment can also be demonstrated in people with cochlear damage in both ears. One method of doing this involves the use of a categorical loudness scaling procedure. The subject is presented with a test sound, and is asked to judge its loudness by using one of several verbal categories [Pascoe, 1978].

For people with bilateral hearing loss, qualitative loudness tests can be used to judge recruitment. Allen [1990] provides a methodology where half-octave wide bands of noise centered at 500, 1000, 2000 and 4000 Hz were used as test stimuli. The first component of judging loudness has a subject respond with 'cannot hear', 'very soft', 'soft', 'comfortable', 'loud', 'very loud', or 'too loud' to a set of 3 noise bursts. The second component collects data by randomizing the center frequency and the level for all the stimuli except the 'too loud' and 'cannot hear' stimuli. From here, functions relating perceived loudness to level at each centre frequency are made; the steeper the slope, the more recruitment.

Although there can be considerable individual recruitment differences for subjects with similar audiograms, on average the steepness of the loudness growth curves increases with increasing absolute threshold [Hellman & Meiselman, 1993].

2.2.2.3 Consequences

Loudness is a funny thing. While most of the above discussion on dynamic range revolved around a reduced set of intensities that some one with hearing loss can judge from, does it really correspond to real life? Matching noise or tone bursts does not give the same results as matching speech. Loudness is a complex function of the stimuli, even the smallest details, such as phase [Gockel et al., 2003] are important factors for loudness and are not very well described by the loudness matching tests above.

Most hearing aid research discounts this, and has produced nonlinear algorithms meant to restore the compressed dynamic range of the hearing impaired. The largest set of nonlinear algorithms used in hearing aids are compression circuits. In general, these circuits are meant to make low level sounds audible, make normal level sounds intelligible, and to bring high intensity sounds into the comfortable range. A compressive hearing aid maps the impaired intensity/loudness function from figure 2.11 to the normal intensity/loudness function.

2.2.2.4 Phenomenology

Most researchers believe that for simple stimuli, loudness recruitment as well as threshold elevation are both due to the loss of the active mechanism in the cochlea. Loudness is then some function of the total evoked neural activity. Accompanying threshold loss is also a loss in the compressive nonlinearity of the BM, thought to be mediated by the electromotility of OHCs. With a steeper (less compressive) BM input/output function than normal the amount of total neural firings increases. At high SPLs the normal ear begins to operate as a linear system, and thus the impaired ear begins to match the normal I/O function. This is an explanation of why the loudness in an impaired ear usually mimics a normal ear at sound levels greater then 90 dB SPL.

Along with changes to the BM level function, there is a broadening of the cochlear filters (see section 2.2.3) with sensorineural impairment. This leads to a broader AN excitation pattern, ie., more neural activity, and hence from the above phenomenology, a louder sensation.

2.2.3 Frequency Selectivity

2.2.3.1 Description

Frequency selectivity often goes unmeasured by an audiologist, yet is the second most obvious difference between normal hearers and the sensorineurally impaired. Frequency selectivity is the ability of the peripheral auditory system to resolve a complex sound into spectrally distinct components.

Its most often quantified by studying masking effects. Masking is the effect of how one particular sound renders another undetectable, even though that impoverished sound is suprathreshold. If a sound at one frequency is masked by another sound at another frequency, then the auditory system has failed to resolve the two sounds, and this may lead to a loss in quality or intelligibility. By measuring when a sound is masked by another, it is possible to characterize how the frequency analysis capabilities of the auditory system are diminished with sensorineural impairment.

Frequency selectivity depends to a large extent on the filtering that takes place in the cochlea. A complex sound, undergoes a spectral analysis in the cochlea where the sinusoidal components of the sound are separated if their frequency separation is large enough. In the normal auditory periphery there are several active mechanisms that ensure this and that the acoustic landscape is coded independently in the auditory nerve.

The perception of a sound as a coherent whole depends upon an accurate representation of the individual components as well as accurate interpretation of their interconnection at some later stage in the auditory system. Damage to the cochlea leads to reduced sharpness of tuning on the basilar membrane and in single neurons of the auditory nerve, as measured in critical bands, or the auditory filter shape. Frequency selectivity is often worse than normal in people with sensorineural impairment.

2.2.3.2 Data

Much of the explanation of masking effects derive from Fletcher [1940] and early work in acoustics at Bell Labs. The concept of the critical band and power spectrum model stem from this body of work. Fletcher [1940] measured the threshold for detecting a sinusoidal signal when masked by a bandpass noise masker. The noise was centered at the same frequency as the sinusoid, and the noise power density was held constant while the bandwidth was altered. The total noise power increased as the bandwidth increased. Moore et al. [1993] prepared a modified version of this test, with normal and hearing impaired people. Their results are given in figure 2.12.

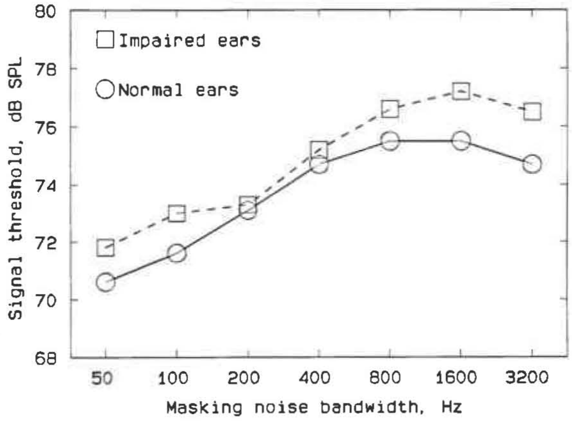


Figure 2.12: Averaged threshold for normal and impaired subjects for a 2 kHz sinusoidal signal plotted as a function of bandwidth of a masker centered at 2 kHz. Taken from Moore et al. [1993],

Qualitatively, the pattern of results is similar for the normal and impaired subjects, both show a leveling off at between 400 and 800 Hz. Quantitatively, the overall performance is worse for the hearing impaired, with thresholds increasing for a larger bandwidth of about 800 Hz. Fletcher [1940] suggested phenomenology was that the peripheral auditory system behaves as if it contained a bank of overlapping bandpass filters. This gave rise to the term "auditory filters". His explanation was that the signal was detected by using the output of the auditory filter centered on the signal frequency. Thus, increases in noise bandwidth result in more noise passing through that filter, and a lower signal to noise ratio at the output. That is while the noise bandwidth is less than the filter bandwidth. When the noise bandwidth exceeds the auditory filter bandwidth, an increase in noise bandwidth does not decrease the SNR the auditory brain works with. Fletcher termed this frequency dependent bandwidth, for which the signal threshold ceased to increase the "critical bandwidth" (CB). This auditory analysis explanation and critical bandwidths are key elements in modern audiology. Fletcher's phenomenological explanation led to the power spectrum model of auditory functioning. It is based on the following assumptions:

1. The peripheral auditory system can be represented as an array of overlapping linear bandpass filters.

2. For detecting a signal in a noise background, the listener makes use of the filter with a centre frequency close to that of the signal, or this filter that has the highest signal-to-masker ratio at its output.

3. Components in the noise outside of the auditory filter in use have no effect on masking the signal.

4. Detection is determined at a specific signal-to-masker ratio. That is, the stimuli are fully represented by their long-term power spectra; phases and the short-term fluctuations in the masker are not important.

None of these assumptions is strictly correct; auditory filters are level-dependent, listeners combine information from many auditory filters to enhance signal detection, noise outside of the auditory filter centered at the signal frequency can strongly affect detection and fluctuations and phase coherence in the masker plays a strong role. However, these simplifying assumptions are a valuable starting point in dealing with the complexities of the auditory periphery.

The psychophysical measurement of these auditory filters results in psychophysical tuning curves (PTCs); there is a counterpart procedure for determining the neural tuning curve [Small, 1959]. To measure a PTC, a test signal is input at a very low sensation level along with a sinusoid or narrow band noise masker. For several masker centre frequencies the level needed to mask the signal is determined. From the aforementioned power spectrum model assumptions this creates a PTC indicative of the masker level required to produce a fixed output from the auditory filter as a function of frequency.

From the linear filter assumption the PTC can be determined with a signal fixed in level and varying in frequency or by determining the input level required to produce the same output level. A typical normal and impaired PTC are given in Figure 2.13.

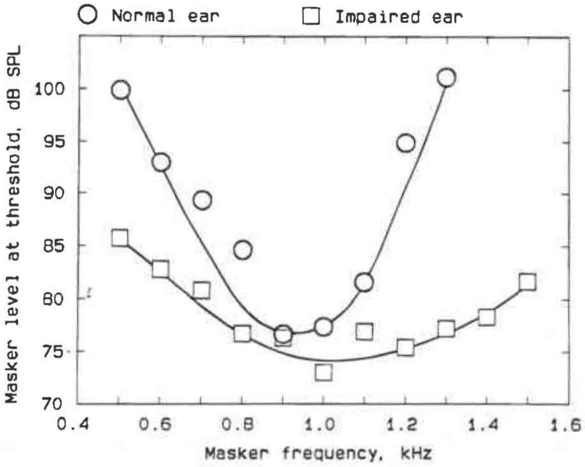


Figure 2.13: Simultaneous masking PTCs for a normal and hearing impaired ear on a unilaterally impaired subject at 1 kHz CF. Taken from Moore & Glasberg [1986].

The circles are for the normal ear, the squares from the impaired ear of the same subject. The PTC is broader for the impaired ear, a clear indication of wider tuning and loss of frequency selectivity.

In this example there may be off-frequency listening, or a break with assumption 3 of the power spectrum model. With the masker frequency above the signal frequency, the most propitious signal-to-masker ratio occurs for a filter centered *below* the signal frequency; when the masker frequency is below the signal frequency, the best signal-to-masker ratio occurs in a filter centered *above* the signal frequency. So the reality of off-frequency listening produces a higher determined masker level then if off-frequency listening did not occur. The overall effect is that the PTC has a sharper tip than would be obtained if only one auditory filter were involved [O'Loughlin & Moore, 1981].

A way to limit the amount of off-frequency listening is to use notched-noise, whose effect limits the frequency shift of the auditory filter, as well as severely deteriorating the signal-to-masker ratio in adjacent frequency bands. The auditory filter shapes for sensorineurally impaired subjects have been estimated many times using notched-noise maskers (for example Glasberg & Moore [1986]; Leeuw & Dreschler [1994]). Typically, the auditory filters are wider than normal in hearing-impaired subjects. Also, the amount of broadening loosely correlates with increasing threshold shift. Glasberg & Moore [1986] measured auditory filter shapes in subjects with unilateral sensorineural impairments, so that the differences between the normal and impaired ears cannot be due to age or attention. Results are given in figure 2.14

The auditory filters for normal ears are very sharp indicating good frequency selectivity, while the auditory filter shapes for impaired ears are much less sharp. Also there is a wide variance in shapes and asymmetries in the hearing impaired ears, and especially the much lower masking thresholds at lower frequencies. This is a hallmark of the hearing impaired person's problem with masking from low-frequency sounds, such as car noise, HVAC, cows mooing.

An additional problem with the power spectrum model and PTCs is the assumption about the auditory filter being linear. In reality the actual shape depends highly on the input SPL. By determining the PTC with a changing input the underlying

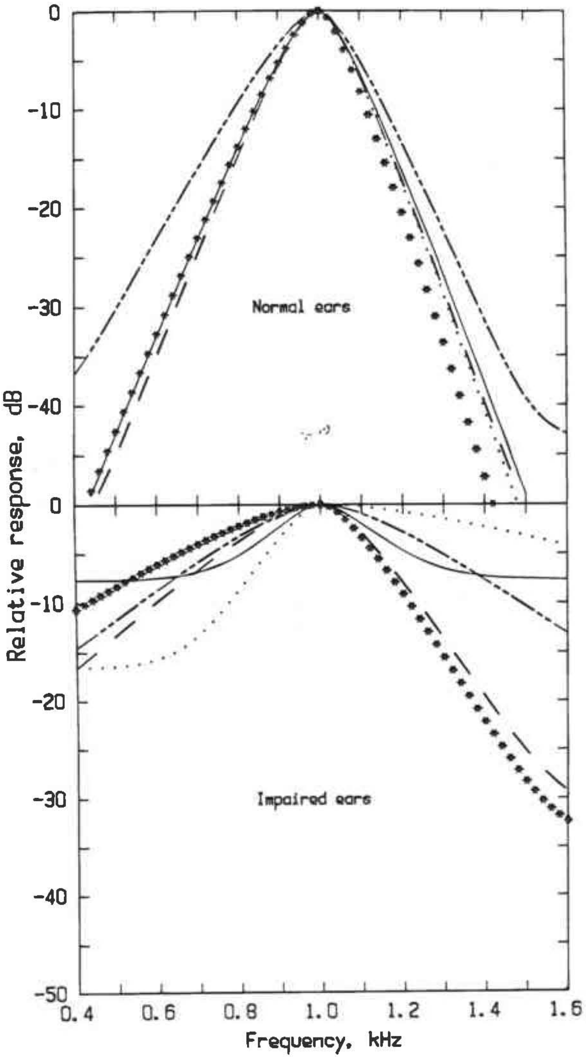


Figure 2.14: Auditory filter shapes for a CF of 1 kHz for six unilaterally impaired subjects. Taken from Glasberg & Moore [1986].

filter shape changes as the masker frequency and level is varied. Because of this, the slope of the low frequency skirt of the PTC is underestimated, along with a converse overestimation of the higher frequency skirt slope [Verschuure, 1981].

2.2.3.3 Consequences

With reduced frequency selectivity, people with sensorineural impairment are victims to increased masking, and wholescale changes to perception. Some of these aspects, such as frequency discrimination, pitch and timbre perception effect the quality and quantity of the impaired persons acoustic experience. The most direct consequence to the reduction in frequency selectivity is the increased susceptibility to masking. As delineated above, it is actually how one quantifies the frequency selectivity of the auditory system. There are many different ways that one can define a masking signal, the most obvious is the signal and masker overlapping in spectrum. In this case, the masked thresholds for hearing impaired people are usually only a little more for hearing impaired people then for normal hearing people. If the signal and masker are spectrally disjoint masking could be considerably greater in the hearing impaired, depending on their exact auditory filter layout. This is supremely important in competing speech, where the attended speaker has a much different short term spectrum then the other conversations in a room. That is, because of reduced frequency selectivity, a particular speaker's individual sounds are more likely to be masked.

2.2.3.4 Phenomenology

Studies into PTCs give a general agreement into the broader tuning in the auditory filter shapes for hearing impaired people. There is also some correlation between the broadened filter and desensitizing thresholds, but there is also wide variance between these two findings.

In particular, PTCs have been found that have two tips, or colloquially, that are W-shaped rather than the typical V-shaped [Hoekstra & Ritsma, 1977]. This result has also been found in neural tuning curves. Sensorineural impairment also can shift the CF of the auditory tuning curve, sometimes well away from the normal place. This is thought to stem from an almost complete loss of inner hair cells in low-frequency regions of the cochlea, making higher frequency neurons take low frequency input.

This is quite rare, the more prevalent case is damage primarily to the OHCs with the IHCs left more intact. This is thought to diminish the electromechanical active mechanism of the outer hair cells, leaving a relatively normal mechanism for transducing basilar-membrane movement into neural responses, but with a loss of the suppression, and nonlinear selectivity. If this is indeed the case, the reduction in frequency selectivity should correlate with elevation in threshold. Another possibility is deterioration of both OHCs and IHCs. Both the active nonlinearity and the transducer mechanism are damaged. There should be considerable reduction in both sensitivity and selectivity. Lastly, there is a rare case of damage solely to the IHCs. One expects high thresholds but tuning to remain sharp. For a realistic hearing loss the cochlear degradation is some combination of the above loss mechanisms, there is considerable variability in threshold shifts and selectivity patterns. So much so that the above phenomenology is somewhat debatable.

If OHC stereocilia damage does reduce tuning of the basilar membrane [Glasberg & Moore, 1986], it is not a simple function of threshold shift [Leek & Summers, 1993]. The increase in auditory filter equivalent rectangular bandwidth (ERB) can be anywhere from 20 to 500 percent, depending on the loss. Phenomenologically, this broadened tuning increases the correlation between different frequency bands, and therefore reduces the number of independent channels that the hearing impaired person has access to.

This, coupled with a lower SNR, because of larger channel bandwidths leads to the idea that an information bottleneck reduces speech intelligibility in the hearing-impaired cochlea. Attempts at enhancing the spectral sharpness of an incoming signal to restore the auditory response have led to minimal or mixed results, see for example Bunnell [1990], Bruce [2004]. The lack of positive results for spectral enhancement is thought to be connected to interaction with the compressive nonlinearity in the basilar membrane response.

2.2.4 Temporal Aspects of Frequency Selectivity

2.2.4.1 Description

Frequency selectivity has been quantified through the use of a simultaneous masking experiment, but there is also the possibility of probing the auditory filter through the use of a masker preceeding the signal, and deriving a PTC through this, forward masking paradigm. The fact that the two PTCs differ in shape gives rise to the idea that they are mediated by different mechanisms. The mechanism that governs simultaneous masking is known as suppression. In general, it is a dynamic mechanism of the healthy cochlea that attenuates (or suppresses) particular frequency components in response to other frequency components.

2.2.4.2 Data

Suppression effects in normal hearing and hearing impaired people are derived by studying the simultaneous and forward masking. The signal level is not varied, to reduce the issues intrinsic to the nonlinear growth and decay of the auditory periphery response. By changing the masker, forward masking studies attempt to determine the masking thresholds based on a fixed output of the auditory filter in the cochlea, not based on a fixed input as previously discussed.

Wightman et al. [1977] studied the PTCs derived from both simultaneous and forward masking paradigms for hearing impaired subjects. Their results indicate a much sharper forward masking PTC when the signal and the masker were in regions of normal sensitivity. Conversely, when the signal was applied to a frequency region with elevated threshold, the differences between the two PTCs were minimal. The explanation for this is that some type of unmasking was available to the healthy cochlea, but not when the ear was impaired. Wightman et al. [1977] conjectured that suppression was responsible for the differences between PTCs in simultaneous and forward masking in the healthy ear.

Moore & Glasberg [1986] updated this research by looking at the simultaneous and forward masking PTCs in subjects with unilateral losses. A notched noise, off-frequency listening limiting method was followed. The PTCs for the healthy ears were sharper in forward masking than in the simultaneous masking experiment. The PTCs for the damaged ears had negligible differences between the two paradigms. An example of this is shown in 2.15

The loss of suppression reduces the ability to temporally separate frequency components in people with sensorineural hearing impairment. Festen & Plomp [1983] also studied simultaneous and forward masking PTCs, but their experiment used a

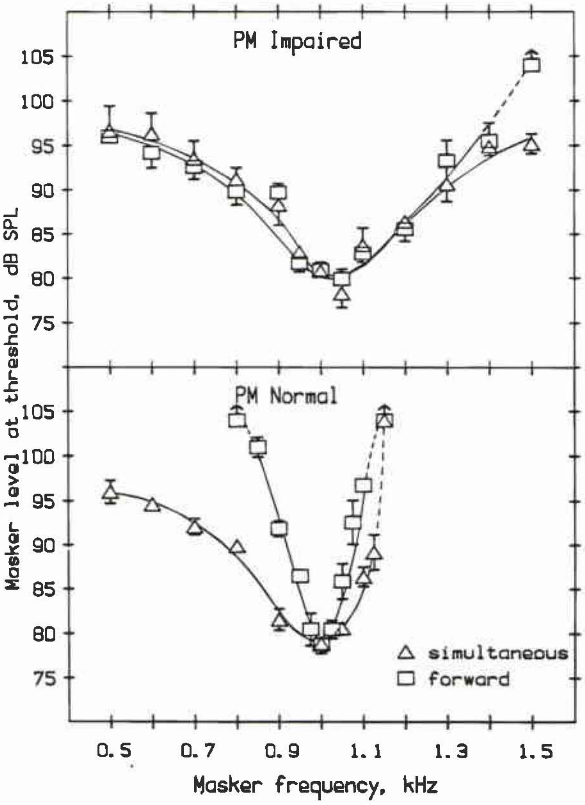


Figure 2.15: PTCs for simultaneous (triangles) and forward (squares) masking experiments from unilaterally impaired subjects. The top plot is determined from the damaged ear, the bottom from the normal hearing ear. Taken from Moore & Glasberg [1986].

rippled-noise masker. This diminished the differences between simultaneous and forward masking PTCs for impaired hearing subjects, corroborating that suppression effectiveness was greatly reduced in the hearing impaired.

In general, two-tone unmasking by preceding frequency components is diminished or entirely absent with sensorineural impairment, and the differences in frequency selectivity of complex stimuli, mediated by the suppression mechanism for simultaneous and forward masking are greatly diminished or absent in hearing impaired people when compared to normal hearing people.

2.2.4.3 Consequences

This loss of suppression has similar consequences to increasing frequency selectivity. On top of the problems already detailed, there is the exacerbating issue that normally hearing people have an even easier time resolving frequency components that have temporal characteristics. Suppression makes it easier to separate acoustic phenomena that happen serially in time, while frequency selectivity can be thought of as separating parallel components. This is key for dealing with competing speech, speech streams have ever changing temporal patterns that the healthy ear can use to unmask the attended talker, while the sensorineural impaired ear cannot unmask with temporal fluctuations.

Separating acoustic streams is of obvious importance, but so is processing a single, complex, acoustic stimulus. Complex stimuli created with the vocal tract or a musical instrument can be defined by their timbre. Timbre has the silly definition of: "that attribute of auditory sensation in terms of which a listener can judge that two sounds having the same loudness and pitch are dissimilar". A saxophone or piano playing middle C can be readily identified because of their different timbre. There are both spectral and temporal characteristics of timbre that are directly affected by the frequency selectivity and suppression mechanism of the ear. With the reduced frequency resolution brought on by cochlear damage the auditory system has less detail to determine the source instrument. Practically, this also effects speech recognition, as voiced sounds have information coded in its timbre, while the unvoiced sound is easily masked by large, preceding frequency components.

2.2.4.4 Phenomenology

The psychophysical data strongly support the idea that lateral suppression is reduced or absent in people with cochlear hearing loss, but the cochlear mechanism is not well described. It is theorized that a different time constant filter in tandem with the natural BM time constant can mediate the OHC electromechanical response. This secondary filtering interaction is damaged with OHC loss. The most common suggestion for this effect is the OHC stereocilia tip connection to the tectorial membrane acts as a feedback mechanism tuned to similar spectro-temporal components as the BM. The tectorial membrane is much more gelatinous then the elastic basilar membrane, giving a broader resonant frequency. The tip connection of the OHCs would then be modulated by different frequency components than the corresponding IHC gating filaments at a particular place. While this is just one conjecture, what is not in dispute is that the loss of OHC function results in a loss of suppression.

2.2.5 Temporal Integration

2.2.5.1 Description

Cochlear damage engenders a change in threshold intensity over a signal duration that is often smaller than for normally hearing people. Normal hearing people have a -3 dB threshold advantage for every doubling of stimuli duration, while the slope for hearing impaired people can be much shallower. That is, the hearing impaired have a reduced ability to sum frequency components over time. Also, changes in absolute threshold correlate to reduced temporal integration or shallower slopes (dB/Δ t). This shallower slope also translates into the sensorineurally impaired needing higher signal intensities to detect brief sounds.

2.2.5.2 Data

One explanation for the reduction in temporal integration brought on from sensorineural impairment is in terms of the detection of spectral splatter in finite duration signals [Hall & Fernandez, 1983]. For this explanation, the absolute threshold has to be determined by detection of frequency components in the splatter and not at the stimulus frequency. The experimental evidence suggests that this is not what is happening in the sensorineural impaired ear. People with flat losses have reduced temporal integration and normally hearing subjects tested with noise that mimics a hearing impairment still have temporal integration advantages [Florentine et al., 1988].

The spectral splatter hypothesis cannot be correct. The explanation that is most likely is that a reduction or complete loss of the compressive nonlinearity of the BM reduces the temporal integration. Zwisklocki [1960] and Penner [1972] both give models showing that a steeper I/O functions from the BM or steeper discharge rate-versus-intensity level functions in the auditory nerve lead to reduced temporal integration. Their models are based on a sliding temporal window (mathematically equivalent to a low-pass filter) followed by the auditory periphery's nonlinearity. The nonlinearity is less compressive in the impaired cochlea. For stimuli with slowly changing intensity, such as narrow bands of noise, this can lead to poorer temporal resolution, since the inherent fluctuations can be confused with the temporal feature to be detected.

2.2.5.3 Consequences

Since most sounds in everyday life are characterized by seemingly random fluctuations in intensity, hearing impaired people will have greater difficulty in following the temporal structure. It is often said that reduced temporal integration from cochlear damage is less severe for weak sounds than for long sounds. For example, a sound with a saturating intensity-loudness duration of 400 ms has less relative loss then a burst of 10 ms. The normal hearing person requires 4 dB SPL and 20 dB SPL, respectively for detection, but a typical hearing impaired person may have a 54 dB SPL and 60 dB SPL respectively. Thus the relative detection threshold shift is more profound for the longer duration signal then the short.

2.2.5.4 Phenomenology

Steeper discharge rate-versus-intensity level functions lead to reduced temporal integration. The threshold for detecting a sound when looking at a neurogram is defined by having a number of spikes above the noise condition. The discharge rate necessary at absolute threshold for a long duration sound in a normal auditory system is N_1 spikes per second. When the duration of the sound is halved the stimulus has to be amplified to restore the total spike count, $N_2 = 2$ x N_1. But the amplification necessary to get to N_2 spikes, or the same total spikes is higher for the compressive, normal functioning ear, than the impaired ear because of the assumed steeper rate-versus-level function in the impaired auditory system. That is, because of the steeper discharge rate-to-stimulus intensity curve for an impaired cochlea less integration is seen.

2.2.6 Temporal Resolution

2.2.6.1 Description

Apart from these fairly well understood symptoms of sensorineural hearing impairment, the temporal effects of hair-cell damage are not well understood. Temporal integration can be thought of as how the auditory system deals with long-term acoustic stimuli, while the very shortest duration signals are characterized by the auditory system's temporal resolution. A subject's temporal resolution determines how they deal with fast transients and quick gaps. Since a lot of the information in speech is coded by consonants that may last for only a few milliseconds, understanding sensorineural impairment's effect on temporal resolution is key.

2.2.6.2 Data

Temporal resolution, such as gap detection, of deterministic signals can be shorter for hearing-impaired people [Moore et al., 1989], but gap resolution is approximately 30 percent longer for hearing impaired people in octave band noise [Fitzgibbons & Wightman, 1982]. The conjectured mechanism is that the sensorineural impaired auditory system cannot reliably separate the random fluctuations in noise with the fluctuations of the signal. This effect may be from higher order auditory processes, or from the loss of the compressive nonlinearity in the auditory system.

2.2.6.3 Consequences

Temporal resolution can be diminished because the sounds are at low SLs or because the audible bandwidth is affected with cochlear damage. These temporal factors can lead to problems in understanding speech or environmental sounds, especially in noise.

2.2.6.4 Phenomenology

There is a hodge-podge of reasons used to try to explain the lack of temporal resolution in people with sensorineural hearing loss. The first point is the effect of sensation level. Normal hearing people show a deterioration in resolution at low sensation levels. For the detection of bandlimited noise and forward masking recovery, hearing impaired people operate near normal hearing people at equivalent SL, but at a detriment for the same SPL [Fitzgibbons & Wightman, 1982]. Another possibility for diminished temporal resolution is the loss in audible bandwidth. This is discussed in detail in section 2.2.7 in relation to amplitude modulations.

The loss of the compressive nonlinearity, specifically by how it influences loudness resolution, is often suggested for the reduction in temporal resolution. This stems from the diminished capacity to deal with gaps in noise versus sinusoids. The inherent fluctuations in noise produce a large swing in perceived loudness and these fluctuations may be confused with the actual gaps.

Glasberg & Moore [1992] coducted an experiment that compressed and expanded the envelope of narrowband noise by raising it to some power N. If N is greater than one, then this magnifies the fluctuations, mimicking impairment. N less than one represents a typical hearing aid compression circuit. Their results show that impaired gap detection can be mimicked with envelope manipulations greater than one.

2.2.7 Amplitude Modulation

2.2.7.1 Description

The Short Increment Sensitivity Index (SISI) revolves around the idea that loudness recruitment can be used to probe the level of sensorineural impairment of a subject. The SISI test involves detection of intensity levels of a continuous sound. This task is comparable to detection of a slow amplitude modulation. Buus et al. [1982] delineates subjects with cochlear impairment as having higher SISI scores, and conversely lower difference limens (DLs) for amplitude modulation detection when tested at equal SL. In contrast, at equal SPL, subjects with cochlear damage perform either equivalently or worse then normal hearing people.

2.2.7.2 Data

Glasberg & Moore [1989] tested nine subjects with unilateral cochlear impairments with relatively 'flat' moderate losses. The test consisted of two successive tone pulses; one was sinusoidally modulated with 4 Hz sinusoid, the other was unmodulated. Each tone was 1020 ms, that includes a 10-ms raised-cosine onset and offset ramp. Subjects had to determine which tone was modulated. The unmodulated tone was presented at 80 dB SPL to the impaired ears, while the normal ears were tested at the equivalent SL and the same SPL. The detection threshold was measured using an adaptive procedure that determined the peak-to-valley ratio for 71 % correct. Trials were conducted in quiet and in an one-octave wide noise designed to mask the high frequency excitation pattern of the test tone. The low frequency cutoff was twice the signal frequency with an overall level of 77 dB SPL. For the two normal ear conditions the results are quite different. The results when the two ears are at the same SPL have thresholds that are sometimes larger and sometimes smaller. While at equal SL, the AM detection thresholds are consistently smaller for the impaired ear in quiet, and mostly also in noise. The highpass noise increased AM difference limens (AMDLs) in both ears, highlighting that high frequency information is important for amplitude modulation detection. This is in contrast to the difference limens for intensity (DLI), which is a measure of how small a difference in intensity can be detected between two tones.

2.2.7.3 Consequences

Drullman et al. [1994] shows how slow amplitude modulations are dramatically important for speech intelligibility. The inability to follow these slow modulations in hearing impaired people may result in a large loss in the amount of information received.

2.2.7.4 Phenomenology

Depireux et al. [2001] shows how ferrets have specific spectro-temporal response fields (STRFs) dedicated to amplitude modulations. It is thought that most auditory brain centers, including humans, have a similar structure. So the basic idea is that higher order brain centers have an impaired representation because of what happens in the cochlea.

2.2.8 Frequency Modulation

2.2.8.1 Description

Along with short amplitude transients of some unvoiced sounds, and the slow amplitude modulations so important for speech intelligibility, there is a huge amount of information contained in frequency modulation. Speech is not characterized by a standard spectro-temporal template for each sound. Instead a speaker connects each individual phone into a slur of sliding formants and continuous, smooth vocal tract changes. In particular, some consonants in real speech are only characterized by their transition characteristics. These transitions are usually characterized by slides in frequency as the vocal tract resonator modulates itself for the next phone. FM detectors are found throughout the higher auditory brain centers.

2.2.8.2 Data

Zurek & Formby [1981] calculated the frequency modulation difference limens (FMDLs) for ten subjects with cochlear damage. An FMDL is the amount of modulation required for a person to tell the difference between two differently modulated tones. They used a 3-Hz glide rate and for test frequencies from 125 and 4000 Hz at a SL of 25 dB (the level found to produce results independent of level). FMDLs increased with increasing hearing loss, and for a specific threshold shift the performance was worse at low frequencies than at high frequencies.

This follows a different pattern than difference limens for frequencies (DLFs). DLFs are the point where a person can judge detect that there is a difference in frequencies between two tones of slightly different frequencies. While both FMDLs and DLFs are raised with hearing impairment, the DLFs for a hearing impairment do not improve with increasing frequency, a trend that is seen in normal hearing subjects.

Moore & Glasberg [1986] tested for FMDLs using a 4-Hz modulation rate and compared it to the AMDLs described above. The theory was that if the same cochlear damage mediated resolution of AM and FM then the ratio FMDL/DLF should equal the ratio AMDL/DLI. Moore and Glasberg found these two metrics were uncorrelated ($r = 0.06$), concluding that the excitation-pattern model was inconsistent in accounting for both DLFs and FMDLs. While the FMDLs can be well modeled with an excitation pattern paradigm [Sek & Moore, 1995], the AMDLs may be responsible for a different, possibly cognitive mechanism.

2.2.8.3 Consequences

Grant [1987] measured FMDLs with a stimulus amplitude modulated by a 3 Hz cutoff lowpass noise. It was expected that the random amplitude fluctuations would impair the use of cues for FM detection by using place changes in excitation level. The excitation-pattern model predicts that the random amplitude modulation would lead to increased FMDLs. This was seen in Grant's results, but the increase was much greater for the impaired than for the normally hearing subjects. The suggested explanation for this large difference is that at low modulation rates, normal hearing subjects can extract information about frequency modulation both from changes in excitation level and from phase locking [Sek & Moore, 1995] while the impaired person has difficulty accessing the phase locking information. So the random AM diminishes the information available from changes in excitation level, but does not alter the use of phase locking cues.

2.2.8.4 Phenomenology

The first suggested mechanism to explain FMDL variations between low and high frequency are the different strategies used in coding frequency. A temporal mechanism is predominant at low frequencies and a place mechanism is used for higher frequencies. That is, the low frequency mechanism is more disrupted with hearing loss than the place mechanism. Another, highly likely, widely supported through empirical testing, yet not much discussed, is the very real possibility that absolute thresholds, especially at low frequencies do not provide an accurate indicator of the extent of cochlear damage.

FMDLs for hearing-impaired people can be modelled by excitation-pattern models that make use of the reduced frequency selectivity of the damaged ear. Of course, that doesn't really explain the loss of phase sensitivity. It is one of the more specious arguments in the cochlear impairment field.

2.2.9 Reverberation

2.2.9.1 Description

When people speak, sound waves propagate away from the mouth until they hit some object, or wall, where some of the energy is absorbed and some reflected back. This occurs for all surfaces and sets up a complex situation where 3D pressure waves bounce about a room. The listener then hears sound arriving in two distinct parts. The first part is sound that travels directly from the speaker, or the direct sound field. After this, reflections from other objects begin to arrive. Considering the complexity of furniture layout, any room can have a very complex set of reflections and delays before the sound arrives at the listener. The indirect sound energy is known as reverberation and is often a large problem for hearing impaired people.

2.2.9.2 Data

Irwin & McAuley [1987] tested eight normal and sensorineural impaired subjects to determine the minimum detectable gap for a 71% correct score in a 2IFC tone bursts paradigm. They added distortions to the test stimulus, including two noise levels and two reverberation conditions. Hearing-impaired listeners needed significantly longer gaps for detection, similar to gap detection experiments for noise stimuli. On speech intelligibility tests, Irwin & McAuley [1987] showed longer reverberation times producing significantly higher thresholds than the shorter times for the hearing impaired. In all, the time constant was significantly correlated with the speech threshold measures (r = -0.58 to -0.74) and speech thresholds were correlated to hearing threshold (r = 0.53 to 0.95). The correlation between time constants and speech thresholds in real reverberation were of similar importance to those for hearing loss and simulated reverberation.

Payton et al. [1994] studied speech intelligibility for normal-hearing and hearing-impaired listeners in noisy, reverberant, and reverberant noisy environments. They studied clear and conversational speech. Clear speech is more intelligible across all noise and reverberation conditions, but is a special benefit to hearing impaired people; they experience a much larger increase in intelligibility that normal hearing people. On the subject of reverberation, hearing impaired people suffered greater degradation due to reverberation when compared to normal hearing people.

Harris & Swenson [1990] studied speech intelligibility in quiet and noise in three levels of reverberation (anechoic, R60 = 0.54 s and R60 1.55 s) for subjects with hearing impairment. Diminishing intelligibility due to noise and reverberation both correlated to absolute threshold shift. Plus, there is a correlation between noise and reverberation, showing a compounding of both is a special problem for the hearing impaired.

2.2.9.3 Consequences

Hearing impaired people have problems dealing with reverberation, and real environments are sometimes terribly reverberant. For example, classrooms or auditoria have a lot of focused research on them. Since it is harder to learn with an increase on listening stress many, even moderately hearing impaired children, can suffer a learning disability simply from the acoustics.

2.2.9.4 Phenomenology

Roberts et al. [2003] determined the effects of reverberation and noise on the precedence effect in listeners with hearing loss. They measured lag burst thresholds (LBTs) for 4-ms noise bursts for normal hearing and hearing impaired subjects in reverberant and anechoic environments in quiet and noise. In quiet, LBTs increased with SL in reverberant environments and decreased with SL in the anechoic environment, while threshold loss did not correlate with the LBTs. When noise was added it had a greater deleterious effect on the performance of listeners with impaired hearing. Their findings indicated that the ability to fuse direct sounds and early reflections is degraded with sensorineural impairment.

The initial wavefront information which a normal hearing person uses as a key to deconvolve the environmental effects are not as accurate in the hearing impaired auditory system. Also, the hearing impaired person has an inability to connect, or stream, the different acoustic markers that make up a single stimulus. It is theorized that the first window of stimulus takes precedence over later reflections, and while later segments of the stimulus can improve the estimation, the hearing impaired person is not as able to integrate those later segments. This is largely thought to be a cognitive effect.

2.2.10 Spatial Hearing

2.2.10.1 Description

One of the largest and most obvious consequences of sensorineural impairment is difficulty in localizing sound sources or using spatial separation to unmask a target stimulus. In some severe cases of cochlear damage the person cannot make use of pinna spectral cues. Without pinna cues, it is difficult to make elevation decisions and resolve the front-to-back ambiguity. Also, hearing aids dramatically alter the spectral patterns at the eardrum and reduce the important high frequency cues. Even when sounds are presented at an adequate SL into a non-occluded sensorineural impaired ear there is difficulty in determining the location.

Most people with sensorineural impairment have more difficulty in determining Interaural Timing Differences (ITD) and Interaural Intensity Differences (IID) than normal hearing people. This reduction in cues used to locate a sound reduces the ability to stream a stimulus based on spatial position.

Some intelligibility studies have used speech and noise coming from the same loudspeaker, thus making conservative estimates on the deficits the hearing impaired person faces versus the normal hearer, who can have a 5-7 dB advantage because of spatial unmasking. Normal hearing people have an SRT advantage of 10 dB for spatially separated speech and noise versus coincident presentation. People with moderate cochlear damage only see an advantage of 3-5 dB between these two cases.

2.2.10.2 Data

The two quantifiable performance metrics in sound localization are one's ability to correctly judge where a sound is coming from and how well one can ascertain small changes in the stimulus. The first metric is widely variegated in people with sensorineural impairment because of the loss of pinna and front-to-back cues distorted with a hearing aid. The other resolution metric is determined by finding the smallest detectable change in azimuth, or the minimum audible angle (MAA). The MAA is smallest for sounds coming from in front of the subject. A shift of only about 1^o can be detected for frequencies below 1000 Hz, with diminishing performance above 1500 Hz. This is explained with the duplex theory. At 1500 Hz, phase differences (ITD) between the two ears become ambiguous and interaural intensity differences (IID) are small. For normal hearing people the MAA increases as the test location is moved away from directly ahead. When hearing sounds from the sides, it is almost impossible to determine sounds above 1500 Hz with any accuracy.

For studying the cochlear and auditory brain responses directly, it is possible to move the stimulus out of free field and present it via headphones. With headphones one can adjust the signal to specifically probe ITD and IIDs. Detection of changes in ITD are smallest when mimicking the front, or centre of the head, or an ITD of zero [Yost, 1974]. ITD changes of 10ms can be detected around 900 Hz. This loosely corresponds to a free field shift of 1^o at speaking distances. For frequencies below 900 Hz, the ITD resolution diminishes slightly, plateauing at about 3^o, while above 900 Hz, the ITD resolution rapidly deteriorates. Above 1500 Hz, ITD ceases to be an accurate determiner of location. Similar to free field results, ITD resolution increases at all frequencies for the nonzero reference ITDs.

Similar to ITD, IID resolution is best at the zero reference IID, with about a 1 dB detectable difference. Unlike ITD, IID does not have a substantial drop-off at higher frequencies. In practice, large IID values are most likely at higher frequencies as the human head is a poor low frequency acoustic baffle.

The localization of the binaural auditory system is best for determining sounds that come from directly ahead a person (0 azimuth). Localization is dependent upon two cues, where low frequency resolution is mediated by ITD cues, while IID can be used over a larger frequency range, but for people without enormously large heads, is mostly used at high frequencies.

Nordlund [1964] determined localization of free field sinusoids at 500, 2000 and 4000 Hz for normal and hearing impaired people. Subjects with cochlear losses in both ears, typically showed normal results, while highly asymmetric loss subjects had a propensity for larger deviation from normal hearers.

For looking at localization cues specifically, Hall et al. [1984] measured ITDs resolution differences at 500 Hz and 70 dB SPL. The impaired subjects ITDs were on average 176 μs versus normal hearing subjects 65 μs. There was also some correlation between threshold shift and ITD resolution, but this ratio had large variability.

Smoski & Trahiotis [1986] show IID resolution differences for 500 Hz sinusoids at 80 dB SPL. The hearing impaired people, again, had higher ITD thresholds than normal hearers, but at equal SL of 25 dB, IID resolution thresholds were not substantially different. In general, the data on sinusoids suggest that symmetric cochlear damage does not impede localization. However, asymmetrical damage may lead to localization difficulties, yet this is not correlated to a loss in absolute thresholds.

2.2.10.3 Consequences

The data for spatial localization above stems from sinusoidal stimuli. Smoski & Trahiotis [1986] also measured ITD resolution thresholds using narrowband noises at 500 and 4000 Hz, presented over headphones at 80 dB SPL. The ITD resolution differences were small for the 500 Hz stimulus but large for the 4000 Hz stimuli, sometimes by an order of magnitude. At equal SL (25 dB), this result did not hold, as ITD resolution thresholds grew to 200-600 μs for normal and hearing impaired people.

Kinkel et al. [1991] measured ITD and IID resolution thresholds for narrowband noises at 500 and 4000 Hz, presented over headphones, at 75 dB SPL. This was a larger study using 15 normal hearing and 49 subjects, presumably with sensorineural hearing loss. The average ITD resolution thresholds were much larger for the hearing impaired than for the normally hearing subjects; 210 μs versus 38 μs for the 500 Hz stimulus; 530 μs versus 81 μs for the 4000 Hz stimulus. The average IID resolution thresholds were also larger for the hearing impaired, at 500 Hz (4.7 dB versus 2.6 dB) and at 4000 Hz (5.1 dB versus 2.2 dB). For both the ITD and IID some of the hearing impaired people had normal resolution thresholds, more so for the IID.

People with very similar audiograms based on similar causes can behave very differently on ITD and IID tests. There is enormous variability between subjects, making the corresponding pathology difficult to pinpoint. In general, people with asymmetric losses more often show larger resolution thresholds for detecting changes in ITD and IID than people with normal or bilateral loss.

2.2.10.4 Phenomenology

Several pathologies have been suggested for the difficulty in discriminating interaural arrival timing differences. The most hopeful suggestion deals solely with the absolute threshold shift and the corresponding low SL of the stimuli. Normal hearing people have deteriorating ITD discrimination below about 20 dB SL [Hausler et al., 1983]. Another explanation comes from the possibility of disruptions to the BM traveling wave. Here, hair cell damage produces discontinuities and an irregular phase response, reducing the information of spike initiation shared between the two ears [Ruggero & Rich, 1987]. This is more or less the specious phase locking argument. Poor IID discrimination also shares the SL explanation.

2.2.11 Competing Speech

2.2.11.1 Description

Competing speech combines all the preceding psychophysics. This is an important environment to understand because dealing with each symptom individually has produced no hearing aid algorithms that improve intelligibility in a cocktail party. A possible exception is directional microphone hearing aids.

The core problem is modelling how the compressive non-linearity of the cochlear amplifier, disturbed by sensorineural hearing loss, can be restored by signal processing in a hearing-aid. There is a complicated set of signal processing that is taking place in the cochlea that ultimately affects intelligibility. Little is constructively known about why there is such a large discrepancy between the hearing impaired and normal hearing person's ability to unmask competing speech. Understanding this disparity is key to building better speech processing algorithms.

2.2.11.2 Data

Carhart & Tillman [1970] show a SNR advantage between 12-15 dB for normal hearing people over hearing impaired people in identifying syllables in competing speech. Over time, testing methodologies have been refined, but results still show an enormous discrepancy between normal hearing and hearing impaired people's ability to understand speech against contending speech. Table 2.3 gives an overview of normal hearing versus hearing impaired peoples ability to recognize target speech with a masking speaker.

Study	**Description**	**SRT Normal/ Impaired**
Duquesnoy [1983]	20 elderly subjects with ski-slope high frequency loss; freefield; Competing @ 55 dBA	-17.6/-5.3
Festen & Plomp [1990]	20 mixed age and losses; monaural earphones; Competing @ 80 dBA	-11.4/-1.1
Hygge et al. [1992]	24 mixed age; freefield, binaural; Competing Speech.	-9.2/7.0 * SNR
Peters et al. [1998]	10 elderly subjects with ski- slope high frequency loss; monaural earphones; Competing @ 65 dBA	-11.9/0.8

Table 2.3: Intelligibility in speech and speech-like noise

To underscore Table 2.3, in noise with a long term average speech spectrum (LTASS) the difference in SRTs between normal and impaired hearing individuals is only 2-5 dB [Glasberg & Moore, 1989].

2.2.11.3 Consequences

It seems that to allow a sensorineural impaired person the ability to operate in the classical cocktail party in a way that approaches a normal hearing person, the auditory impairment must be understood in the competing speech regime, because of the complexity of competing needs.

2.2.11.4 Phenomenology

There is no real phenomenology to competing speech. Everything researchers know about sensorineural hearing loss probably comes into play given the cornucopia of stimuli and environments that competing speech comes up with. One thing is certain, the small pieces of the puzzle do not all add up to being able to build a hearing aid algorithm that improves the hearing impaired person's ability to operate like a normal hearing person at a cocktail party. As it stands, hearing aids reduce the ability to operate in a cocktail party.

2.3 Machine Learning

Machine learning is often introduced as the analogue to Hebbian learning, a theory of how humans learn. Hebbian learning is an unsupervised learning paradigm, but has a simple generalization to correlative learning and is the biophysical motivation for supervised learning strategies. In essence, supervised learning adapts a system to produce outputs that approach desired outputs in response to known inputs. These known inputs and outputs are referred to as the training set. Thus, supervised learning attempts to derive an algorithm, function or mapping between the input and output. That is, for a set of training data with pairs of input patterns, x, and corresponding desired outputs or targets, y, the goal is produce a function $f(x) \rightarrow y$. If f does closely match the functional relationship mapping the inputs to the target outputs then input taken from outside the training set, x', when applied to f, should produce proper results, $f(x') \rightarrow y'$. This is the generalization problem. While it may be simple to fit a function for the training data, it is often not known how well the training set approximates or encompasses all the real situations. Figure 2.16 gives the typical supervised learning framework.

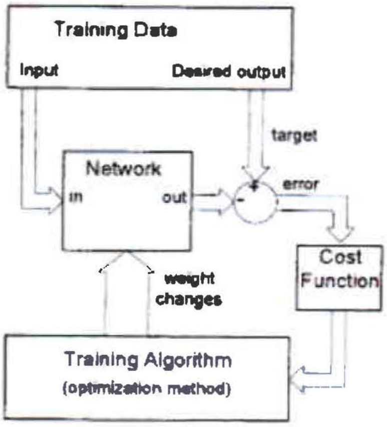


Figure 2.16: Block diagram of supervised learning. Taken from Reed & Marks [1998]

For this dissertation, this framework has two major implications. The first has already been touched on, if the network in Figure 2.16 is the auditory system, then what are the implications on a training set whose input is distorted by sensorineural impairment. The second is, if the network is a hearing aid processor, how can one transform the input acoustic signal so that the output when fed into a damaged ear has the greatest intelligibility.

For the second interpretation, the network is a function with a set of weights that have to be determined, or the network is an artificial neural network (ANN). At the beginning of training, each input pattern is fed through the network, in simplest terms the training optimization tries to correlate the output of the network with the desired output. This is the correlative learning, or Hebbian learning connection. For real data, the network is almost never perfectly trained; there is some assumption about the importance of each deviation of the output to training set.

For this framework several questions must be answered:

1. **Training Set**. What data is input and how does one define the desired output? In the case of hearing aid processor training, the input is an acoustic waveform, and the desired output shall be the AN response of a normal auditory system.

2. **Network**. What type of function should be approximated by the network, or what should the hearing aid processor be able to do? In the following chapters there is an evolution from simple processing blocks, replicating linear hearing aid fittings, to network structures which are based on adaptive, nonlinear processing that is lost with sensorineural impairment. This really set the viability of implementing a solution on silicon. Most networks without feedback, such as the multi-layer perceptron, or feedforward frameworks are computationally inexpensive, and produce little delay, other than windowing for frequency implementations. While other networks are not very conducive to hearing aid implementation, notably the hopfield or elman networks which are computationally more expensive, with larger temporal requirements.

3. **Cost Function**. The cost, or error function is the statistical measure of quality. As with the network, this evolved over the development of this dissertation. The main goal for a hearing aid processor is to return intelligibility, so the cost function grew out of predicting intelligibility by looking at the ANF discharge

rates.

4. **Training Algorithm**. How are the weights changed in the network in response to the error function? This is probably the least important aspect of the learning for hearing aids with present-day computing power.

5. **Initial Conditions**. What type of prior knowledge can be used to speed up convergence, or limit solutions that are physically not possible? In the end, the hearing aid processors output is listened to, but it is impossible to code the quality of a sound into the training algorithm. Heuristics such as limiting high frequency gain have shown some utility to keeping machine learning running smoothly in the auditory domain.

6. **Generalization**. How well will the end network work in a real environment? Does training capture the important statistical structure of the acoustic environment? The training set may be too small, the training algorithm may only replicate the desired outputs but produce unusual results for other inputs. There are a huge number of reasons that the network once trained is incapable of dealing with the enormous corpus of acoustic environments.

The cost function will be seen to be key to producing good hearing aid algorithms. The cost function guides the search for the solution, so it has a fundamental effect on the outcome. The most common error cost function is the mean squared error (MSE)

$$E_{MSE} = \frac{1}{PN} \sum_{p=<P>} \sum_{i=<N>} (t_{pi} - y_{pi})^2 \tag{2.1}$$

p indexes the training set of P vectors, i indexes the number of output nodes, N, and t_{pi} and y_{pi} are, respectively, the target and actual network output. An expansion of the power explicitly states the correlative term

$$E_{MSE} = \frac{1}{PN} \sum_{p=<P>} \sum_{i=<N>} \left(t_{pi}^2 + y_{pi}^2 - 2t_{pi}y_{pi}\right) \tag{2.2}$$

The cross term, relates to the correlation between the desired output and the networks output. The two squared terms are a normalization, making the error approach zero when the outputs approach the target. For an error signal that has zero mean and is Gaussian distributed, this leads to maximizing the correlation between the output and target.

2.4 Prior Art

In the following sections, previous attempts along these lines are expounded. Section 2.4.1 was one of the original attempts to use auditory modelling to derive new hearing aid processing strategies. While this is a good starting point, this previous work did not address the essential nonlinearities or dynamics of hearing impairment. Section 2.4.2 also went with a simple linear fitting strategy. This section will expound on why intelligibility prediction, the first research chapter of this dissertation (Chapter 3), is so important. Lastly, section 2.4.3 ends with some insights into the dynamic properties lost with sensorineural hearing impairment. This last section, while unsuccessful at developing hearing aid strategies, is a solid foundation for the processing strategies developed in this dissertation.

2.4.1 Anderson (1994)

Anderson used a model of the human cochlea, encompassing inner and outer hair cell function. The outer hair cells were modelled with active mechanical feedback elements. The OHC model was able to account for the compressive nonlinearity and tuning sharpness. The compression was approximately logarithmic, mimicking the auditory system's loudness just noticeable differences (JNDs). IHC functions were modeled as hyperbolic tangent transducers of mechanical to spike rate firing. Anderson [1994] fit his model with loudness data from Stevens' power law for loudness growth. The full model is shown in figure 2.17

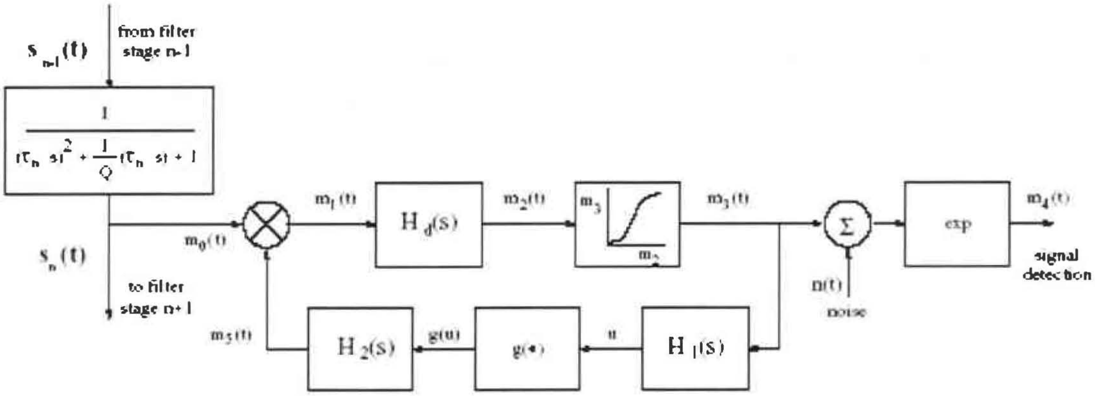


Figure 2.17: The cochlear model for both Normal and Impaired hearing people, taken from and used by Anderson [1994]

For figure 2.17

$$H_d(s) = \frac{\tau_n s}{\tau_n s + 1} \tag{2.3}$$

$$H_1(s) = \frac{A}{\tau s + 1} \tag{2.4}$$

$$H_2(s) = \frac{A}{\kappa s + 1} \tag{2.5}$$

$$g(u) = \frac{\alpha u}{e^{\left(u/\beta\right)}} \tag{2.6}$$

Then this cochlear model was altered to mimic hearing impairment. That is, it had the same functional form, but the values controlling filter shape and level changed. The form of this hearing model is very important because it allows for the development of an inverse model. Anderson's inverse model is in 2.18

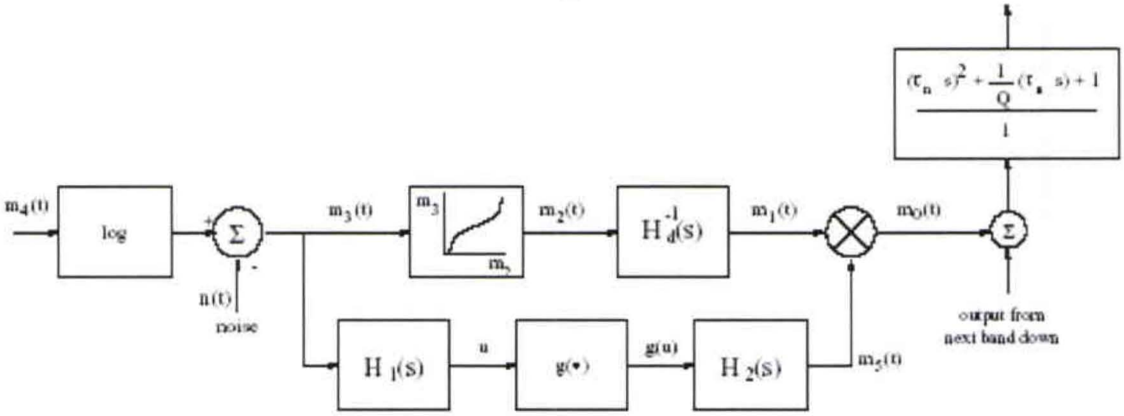


Figure 2.18: The INVERSE cochlear model for both Normal and Impaired hearing people, taken from and used by Anderson [1994]

The basic premise is that it is because of its linearity that the optimal hearing aid circuit is then the cascade of the inverse hearing impaired model and the normal model. After tinkering with the format to reduce complexity, Anderson [1994] ended up with the hearing aid processor in 2.19

Anderson et al. [1995] tested the inverse hearing aid algorithm implemented on an HP series computer. Testing was performed on eight hearing impaired subjects. All eight subjects showed improvement in their speech discrimination scores over control

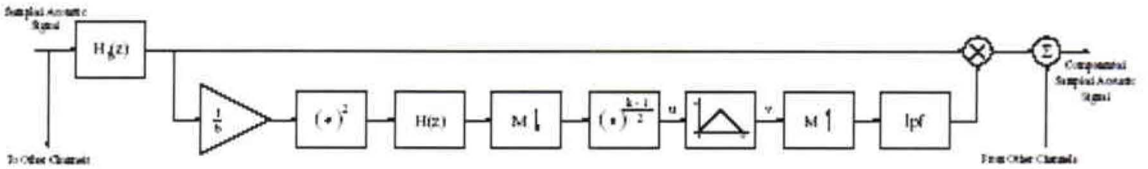


Figure 2.19: The hearing aid algorithm as implemented in Anderson [1994]. H_b is a one third octave bandpass filter, H(z) is a 16 Hz lowpass filter, the M decimator has an output of 60 Hz and b is an input intensity normalization factor

hearing aids. Results are summarized in Table 2.4

Study	**SL/SNR [dB]**	**Unaided**	**Aid A**	**Aid B**	**New Aid**
Quiet	10 dB	0	4	0	14
Quiet	20 dB	0	17	9	71
Quiet	30 dB	12	58	32	95
Quiet	40 dB	33	90	66	97
Babble	10 dB	0	39	17	65
Babble	20 dB	19	71	64	86
Babble	30 dB	41	90	82	96
Babble	40 dB	77	96	93	94
Babble	10 dB	0	39	17	65
LTASS	10 dB	7	9	12	17
LTASS	20 dB	18	45	48	61
LTASS	30 dB	43	73	69	75
LTASS	40 dB	72	91	95	93

Table 2.4: Intelligibility in Quiet, multitalker babble and long term average speech shape noise

Table 2.4 shows a clear advantage of the modelled aid over the other hearing aids. The problem with the data reported is that the control hearing aids weren't actually the ones that the subjects used in everyday life, so the results, in reality, are not as obvious.

There are some obvious problems with Anderson's work. The assumption of linearity, and invertibility is not truly held. The reasons behind the exquisite frequency selectivity and ability to deal with complex stimuli are more incumbent on the adaptivity and nonlinearities in the normal auditory system, than the averaged, linear operation. Anderson [1994] is in essence a complicated AGC, but with the growing data showing the difficulties in optimizing nonlinear parameters [Smoorenburg, 2004], this may be a great help in understanding the phenomenology of the empirical data.

2.4.2 Rankovic (1991)

Rankovic [1991] attempted to derive a hearing aid processing strategy by maximizing the articulation index (AI). Referring to figure 2.16, her approach would have the network implement a linear fitting strategy and the cost function would be the AI. The ensuing amplification scheme was evaluated on 12 hearing impaired subjects versus such standard amplification schemes as NAL (Byrne and Dillon, 1986) and POGO (McCandless and Lyregaard, 1983).

The empirical data showed a relationship between the intelligibility of nonsense syllable and AIs calculated on the NAL and POGO outputs. Subjects with severe sloping high-frequency hearing losses demonstrated nonmonotonicity on the AI maximization condition. This fitting prescribed much more gain at high than at low frequencies. This study essentially showed that the AI prescription provided no benefit over the empirically derived strategies in maximizing word-intelligibility scores.

The AI has significant corpus and environment effects. Rankovic [1991] discussed improving this process by altering the AI formula used, or adding kludge factors based on things such as entropy of material or reverberation.

2.4.3 Kates (1993)

Kates [1993] improved on Rankovic's approach by trying to find the optimal frequency gain response that would produce the minimum mean square error between the outputs of a normal and impaired auditory model. He used real speech stimuli to come up with the optimal gain across frequencies. The idea was similar to the basic hypothesis of this dissertation: An ideal hearing aid for a peripheral hearing loss would process the incoming signal in order to give a perfect match between the cochlear outputs of the impaired ear and a reference normal ear.

Kates [1993], similar to Anderson [1994], develops a normal and impaired peripheral auditory system. Both researchers auditory models include the compression and the neural transduction process, but Kates [1993] adds in suppression effects.

The machine learning framework is followed from Figure 2.16. The target signal is the analog discharge rate, and the network is a linear filter. The mean square difference between the target and the output of the damaged auditory model with the linear filter on the input drives the training of the filter.

One of the exciting new ideas added by Kates was the need to change the linear filter with different input stimuli. In general Kates [1993] produced a three-channel, adaptive compression system. This system is shown in Figure 2.20

This system stemmed from the analysis of several short term stimuli. Figures 2.21, 2.22 and 2.23 give the input spectra, the auditory analysis representation and the derived optimal gains

As the stimulus changes it is necessary for the optimal hearing aid to follow those changes. Kates [1993] found that the optimization process is not stable, that the derived filter shapes were different depending on starting conditions, and that the training algorithms used did not necessarily converge. He also stated that the largest

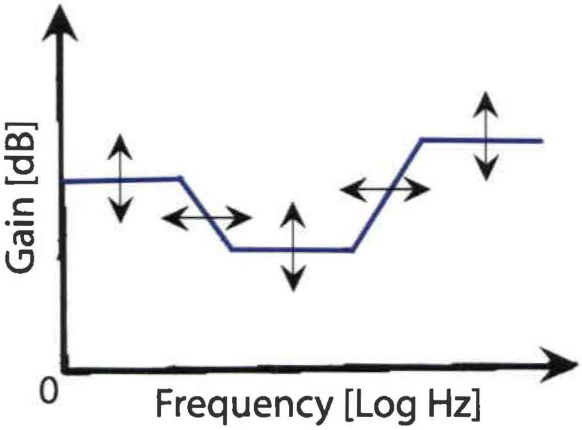


Figure 2.20: The simplified optimal hearing aid, taken from Kates [1993]. The arrows indicate an adaptive gain and bandwidth function.

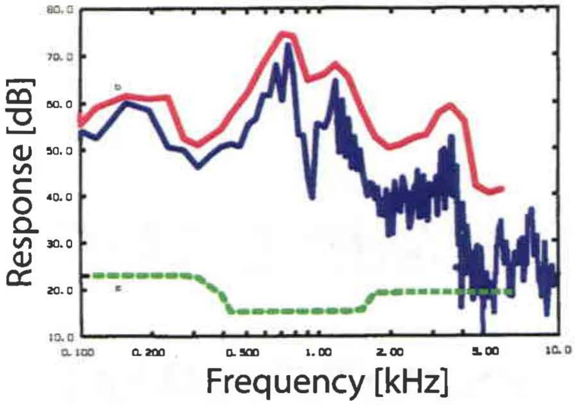


Figure 2.21: The input spectrum (blue), auditory analysis spectrum (red) and optimal hearing-aid gain (green) for a simulated, flat 60 dB hearing loss for the /a/ in "ka".

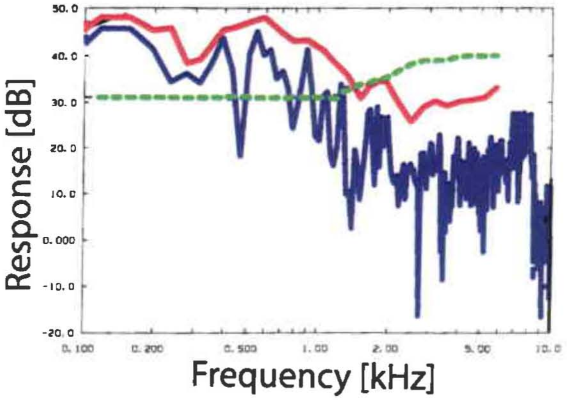


Figure 2.22: The input spectrum (blue), auditory analysis spectrum (red) and optimal hearing-aid gain (green) for a simulated, flat 60 dB hearing loss for the /p/ in "pa".

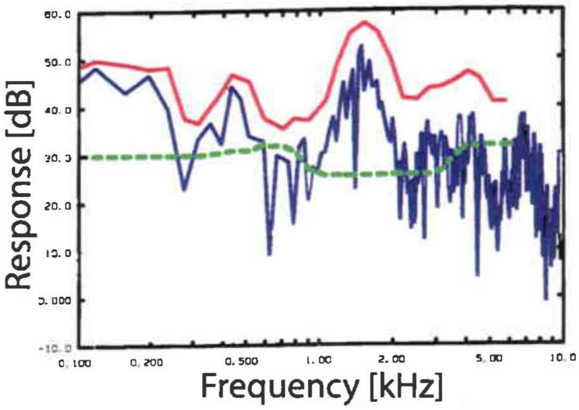


Figure 2.23: The input spectrum (blue), auditory analysis spectrum (red) and optimal hearing-aid gain (green) for a simulated, flat 60 dB hearing loss for the /k/ in "ka".

problem with this framework was adapting the filter shapes. From Figure 2.21 and 2.23 it is obvious that the filter must change significantly, yet these two sounds are adjacent to one another, the timing of the change makes it impossible to retain optimality without introducing processing artifacts. One of the main goals for this thesis was to derive the optimum dynamic response for a hearing aid.

Chapter 3

An AN Error Metric: The NAI

For machine learning to be effective at producing hearing aid algorithms, a good error signal is pivotal, and following from Rankovic [1991] it should revolve around the idea of intelligibility. This chapter deals with the development of a novel intelligibility predictor; one that encompasses the auditory periphery, and hence the differences between normal hearing and hearing with cochlear damage. The first, semi-successful attempt at an error signal/intelligibility predictor was the Information Theoretic Intelligibility Metric or ITIM. The development of ITIM is in section 3.1. ITIM was further refined by using Articulation Index (AI) theory, and the Speech Transmission Index (STI). This gave rise to a good predictor of intelligibility, the Neural Articulation Index, or NAI, which is detailed in section 3.2. This novel intelligibility predictor is used to successfully design linear hearing aid algorithms in chapter 4. Later chapters evolve from the idea that maximizing intelligibility over an ensemble is not as important as optimizing intelligibility for each token in that ensemble. In later chapters this is the cornerstone for building on the differences between the normal and impaired cochlea and deriving novel signal processing strategies.

3.1 Information Theoretic Intelligibility Metric

The initial attempt at a error signal/intelligibility predictor was the Information Theoretic Intelligibility Metric, or ITIM[1]. The basic premise was derived from viewing the auditory system as an information processing system and thus governed by the same description of distortion as all communication systems or channels. This led to using the Kullback-Leibler divergence (Kullback [1968], Johnson [1980], Bandyopadhyay & Young [2004]) instead of a signal to noise measure.

ITIM itself was built to span both general distortion dimensions, namely time and frequency. Previously, frequency domain distortions were quantized with the signal to noise ratio (SNR) across different bands; the Wiener filter is the classic maximal SNR filter. French & Steinberg [1947] first proposed a metric using the SNR in different independent frequency bands as a predictor of intelligibility. Kryter [1962a] gave a procedure for calculating the AI based on this, popularizing the tool for 40 years. Time distortions such as reverberation were not taken into account, and the AI has been largely superceded by techniques that encompass them. One of the initial metrics of time domain distortions on intelligibility was the MTF by Houtgast & Steeneken [1973]. Limited time and frequency domain distortions were combined into the STI. The STI extended the MTF's test signal to account for a wider range of distortions.

The STI test signal is a long-term average speech spectrum random signal, 100 % amplitude modulated by a 0.63 Hz to 12.5 Hz tone. Different frequency bands are switched on and off over the testing sequence to come up with an intelligibility score between zero and one. Inter-frequency band intermodulation sources can be

[1] This section is based on Bruce et al. [2002].

discerned, as long as the product does not fall into the testing band. Therefore, the STI allows for standard AI-frequency band weighted SNR effects, MTF-time domain effects, and some limited measurements of non-linearities. The STI shows a high correlation with empirical tests, and has been codified as ANSI standard S3.5-1997, ANSI [1997]. For general acoustics it is very good, but extending to an evaluation metric for hearing aid algorithms is not straight forward.

The STI does not accurately model intra-critical band masker non-linearities, phase distortions or the mechanisms of cochlear processing (outside of independant frequency bands). The STI is also not an acceptable measure of fidelity or predictor of processing efficacy or interactions. While the AI or STI can take into account threshold shifts in a hearing impaired individual, it cannot account for a hearing impaired persons suprathreshold degradations [van Schijndel et al., 2001]. Thus, intelligibility predictors as they are now implemented find use in more general acoustics than individual assessment.

By using the KLD as a measure of the divergence between a control signal (in this case any arbitrary sound, as opposed to the specially constructed sources for MTF based approaches) and a test signal made by passing the control through some distortion, a monolithic framework is derived.

3.1.1 Model Overview

Typically, researchers have attempted to model the acoustic environmental effects by a linear system representation.

$$x\left(t\right) = a\left(t\right) * s\left(t\right) + n\left(t\right) \tag{3.7}$$

That is, the received signal, $x(t)$, is a combination of additive noise $n(t)$ and the convolution of the linear impulse response, $a(t)$, (representing the environmental, multipath effects) and the input $s(t)$. Under processing, x, is mapped through a possibly non-linear function $f(.)$, to produce an output $y(t)$

$$y\left(t\right)=f\left(x\left(t\right),x\left(t-\tau_0\right),x\left(t-\tau_1\right),\ \ldots\right) \tag{3.8}$$

The nonlinearity and noise is often thought to produce a non-solvable condition, under normal processing constraints. But by using a standardized test signal one can marginalize most of the inherent pitfalls in the pragmatic system. Other possibilities for solving this type of problem are to introduce assumptions or relaxation of precision.

The ITIM instead treats the system in a general manner

$$y\left(t\right)=g\left(s\left(t\right)\right) \tag{3.9}$$

All additive, convolutive and nonlinear effects are treated together as a stochastic, nonlinear function g(.). Where g(.) is a measure of the distortion in the system from all affecters. This distortion function maps the control signal to a test space.

The mapping function was derived following Steeneken [1992]. The sounds files were taken from [van Schijndel et al., 2001], a Dutch Corpus, and are a sample of Dutch syllables, all of the consonant-vowel-consonant (CVC) form. The same 15 syllables are spoken by four females and four males. The initial sound files were sampled at 44.1 kHz, the rms power was normalized and each syllable/speaker sound file was upsampled to 500 kHz for processing by the auditory periphery model. The upsampled spectrum has resampling distortions under -120 dBc. Well under the dynamic range of the auditory system and the model.

Following Steeneken [1992], the syllables were filtered into 8 different passband conditions, these are given in table 3.5.

	Octave band centre frequency						
No	**125**	**250**	**500**	**1000**	**2000**	**4000**	**8000**
1	1	1	1	1	0	0	0
2	0	0	0	0	1	1	1
3	1	1	0	0	0	1	1
4	0	0	1	1	1	0	0
5	1	1	0	0	1	1	0
6	0	0	1	1	0	0	1
7	1	0	1	0	1	0	1
8	0	1	0	1	0	1	0

Table 3.5: The filtering conditions for ITIM's test distortion condition. A "1" in the column represents that the band was passed, a "0" that the band was filtered. Each filtered condition was based on a 1353 tap FIR.

After the spectral distortion, or filtering condition is applied, four different additive noise conditions were applied to each syllable. This additive noise was individualized per speaker by taking the spectrum of all their utterances and forming a long term average speech spectrum (LTASS). The shaped noise was then scaled and added to the normalized 500 kHz speech samples at 0 dB, 7.5 dB, 15 dB and infinite dB SNR samples.

This produced a training set with the accompanying intelligibility metrics provided in Steeneken [1992]. These 15 syllables by 8 speakers by 4 SNRs by 8 bandpass conditions, test sounds were input, along with the unaltered control sounds, into the Bruce et al. [2003] auditory model.

An AN response, or spike train, was derived for each stimulus for 20 representative frequencies. These frequencies started at 10 kHz, and moved down the cochlea in even increments according to Greenwood [1990] cochlear position-frequency function. This gave a close to constant relative bandwidth between frequencies above 1 kHz, and then a constant bandwidth representation below that. The presentation level was chosen to approximate the normal speaking level of 65 dBa.

The resulting 500 kHz sample rate spike trains were run 1000 times to create a probability of spiking profile. This probability train was down sampled to 1000 Hz, to coincide with the minimum refractory period of 1 ms. Here it should be evident that this representation is more akin to the rate coding model of brain activity than the spike timing paradigm. While this is clearly not the case for all sounds [Young & Sachs, 1979], the idea was to design a general framework that could be extended with synchrony neural codes.

Generally there are two spike trains at 20 different frequencies to be compared. The Control spike train, is the AN response for the noiseless stimulus, while the Test spike train is the AN response for the stimuli which have been transformed spectrally and had noise added. The KLD from the Control spike train to the Test spike train is:

$$D\left(p_t | p_c\right) = \int\limits_{\lambda} p_t \log \frac{p_t}{p_c} d\lambda \tag{3.10}$$

where the divergence, D, is the sum over the trial length, λ, of the Bernoulli trial probabilities of the test (denoted with subscript t) and the control (denoted with a subscript c) samples.

The weighted average of the divergence from the Control spike train to the Test spike train, and the divergence from the Test spike train to the Control spike train (since the KLD is non-symmetrical) is then

$$R\left(p_t, p_c\right) = \frac{D\left(p_t|p_c\right) D\left(p_c|p_t\right)}{D\left(p_t|p_c\right) + D\left(p_c|p_t\right)} \tag{3.11}$$

This average value was then normalized by the entropy of the Control spike train to arrive at a "unit entropy representation". This is a pragmatic assumption stemming from the notion that all syllables convey the same amount of information, and reduced the bias towards longer stimuli.

$$S\left(p_t, p_c\right) = {}^{R\left(p_t, p_c\right)}/_{H\left(p_c\right)} \tag{3.12}$$

This basic divergence is normalized by frequency band to give a zero mean, unit variance, named $\hat{S}$. This was done as low frequency bands are structured very differently from higher bands, giving a much higher divergence. The asymmetrical low pass structure of the basilar membrane was hypothesized as the main contributor for this effect.

$\hat{S}$ is transformed into an intelligibility number between 0 (for not intelligible) and 1 (for no intelligibility loss) by

$$I(BF) = 1/2 - 1/2 * \tanh(\hat{S}) \tag{3.13}$$

where the independent variable was changed from the test and control probability distributions to the Best Frequency (BF) moniker (BF ∈ [1,2,...,20]). The intelligibility of the resulting speech out of 100% was then a weighted summation of the individual frequency intelligibility factors:

$$I\% = \sum_{i=1}^{20} w_i I(i) \tag{3.14}$$

The frequency weights, w_i, were chosen to minimize

$$\min_{w} \sum_{i}^{SNR} \sum_{j}^{Env} \left(I_{ij}^{Steeneken} - \sum_{k=0}^{19} w_k I\%(k) \right) \tag{3.15}$$

3.1.2 Example calculation of the ITIM

To illustrate the above derivation, an example showing the various steps is included in this section. A typical trial run starts with the probability of spiking trains at the output of the model. For example a typical spike train pair of Control and Test is given in figure 3.1.

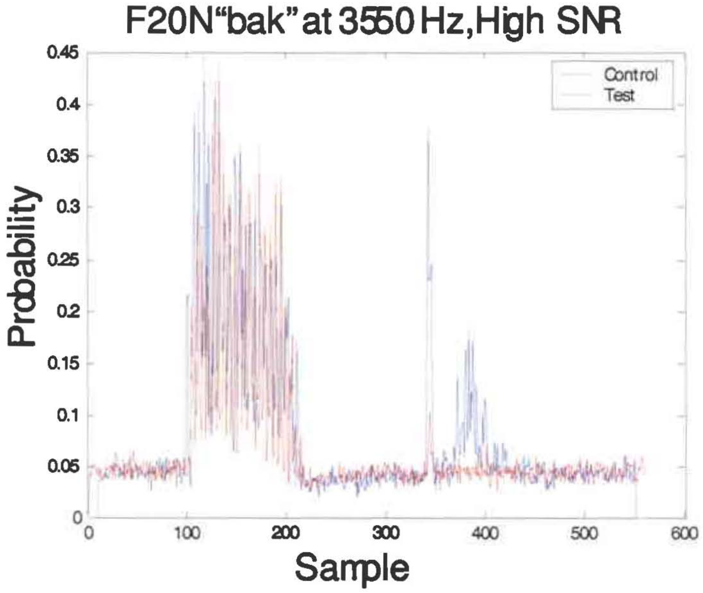


Figure 3.1: The AN response for the clean, Control stimulus, and the noisy, Test stimulus.

The model outputs are then used to calculate the divergence via equation 3.11. A typical result is shown in figure 3.2.

A plot of each stimuli's divergence, following the normalization by each frequency's mean and variance, shows a clear separation between the low SNR condition and high

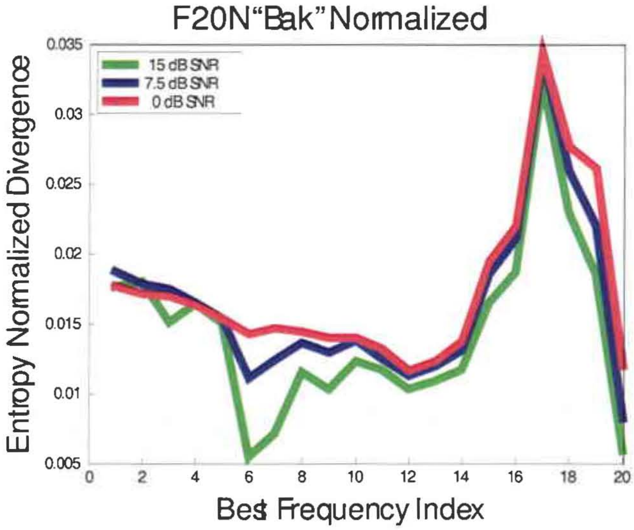


Figure 3.2: A characteristic graph of divergence versus frequency is computed by equation 3.11.

SNR, especially, in the lower frequency bands.

Following this, the intelligibility factor, in each frequency channel (ie. the value between zero and one) that represents how intelligible a particular channel is calculated via equation 3.13. This flips figure 3.3, the low SNR values (red) are now close to zero in most cases, while the high SNR approach one. This is seen in figure 3.4.

After the weighting structure is calculated through the adaptive least mean squares, the intelligibility predictor is shown in figure 3.5.

The difference between the predicted intelligibility and the empirical data has an RMS error of 11.1 %. This was a very promising first result. The STI best fit is around 8.4 %. Extensions of this work are possible by the introduction of high threshold/large dynamic range fibers, or general improvement to the way differences between spike trains are quantified. The auditory periphery model used shows a high validity for discrimination versus SNR as all channels showed an increased divergence with increasing SNR. The real problem with ITIM was in the way the KLD quantified

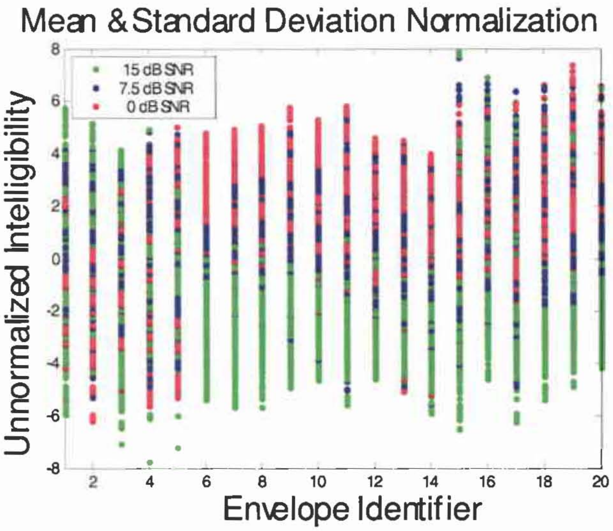


Figure 3.3: The scatter plot showing the distributions of intelligibility over frequency after the Gaussian normalization. Each point is the score for one example from the corpus. The point of interest from this slide is the grouping of the three SNR conditions.

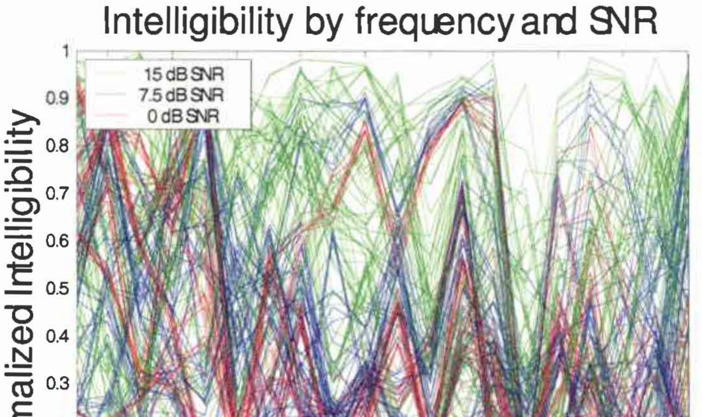


Figure 3.4: The intelligibility score per frequency channel in each frequency channel after normalization from equation 3.13. Each vertex is the score for one example from the corpus. The point of interest from this slide is the grouping of the three SNR conditions.

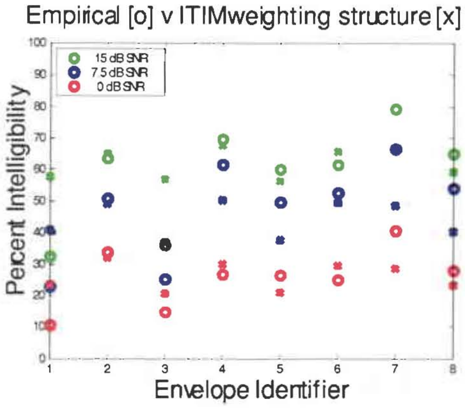


Figure 3.5: The empirical percent intelligibility scores are plotted as 'o's and the ITIM predicted scores are plotted as 'x's for the eight different envelope conditions and the three signal to noise ratios.

filtering conditions. Where additive noise was well captured, a better metric may be able to capture the loss of intelligibility brought on by attenuation in different frequency bands.

	Calculated CVC word score at SNR		
No	**15 dB**	**7.5 dB**	**0 dB**
1	57.7	40.9	23.5
2	64.9	48.9	32.0
3	56.8	36.9	20.5
4	67.6	50.1	29.9
5	56.3	37.5	21.0
6	65.8	49.4	29.4
7	66.2	48.3	28.5
8	59.3	40.3	23.1

Table 3.6: The ITIM predicted intelligibility scores tabulated from figure 3.5.

	CVC word score at SNR		
No	**15 dB**	**7.5 dB**	**0 dB**
1	32.5	22.9	10.7
2	63.6	50.7	33.7
3	36.2	25.2	14.8
4	69.8	61.6	26.7
5	60.0	49.6	26.6
6	61.6	52.5	25.1
7	79.5	66.4	40.6
8	65.1	53.9	27.9

Table 3.7: The empirical intelligibility scores tabulated from figure 3.5.

3.2 Further Development of a Neural Intelligibility Predictor

The major issue in this section in further development of an offline metric for evaluating speech enhancement and hearing compensation algorithms and that can be used as an error metric for the machine learning methods used later in this dissertation[2]. The Speech Transmission Index (STI) failed to account for masking effects that arise from the highly nonlinear cochlear transfer function. The proposed Neural Articulation Index (NAI) overcomes this by estimating speech intelligibility from the instantaneous neural spike rate over time, produced when a signal is processed by an auditory neural model. In highly rippled frequency transfer conditions the NAI's prediction error is 8% versus the STI's prediction error of 10.8%.

A wide range of intelligibility measures in current use rest on the assumption that intelligibility of a speech signal is based upon the sum of contributions of intelligibility within individual frequency bands, as first proposed by French and Steinberg [French & Steinberg, 1947]. This basic method applies a function of the Signal-to-Noise Ratio (SNR) in a set of bands, then averages across these bands to come up with a prediction of intelligibility. French and Steinberg's original Articulation Index (AI) is based on 20 equally contributing bands, and produces an intelligibility score between zero and one:

$$AI = \frac{1}{20}\sum_{i=1}^{20} TI_i \qquad (3.16)$$

where TI_i (Transmission Index i) is the normalized intelligibility in the i^{th} band. The

[2] This section is based on Bondy et al. [2004].

TI per band is a function of the signal to noise ratio or:

$$TI_i = \frac{SNR_i + 12}{30} \tag{3.17}$$

for SNRs between −12 dB and 18 dB. A SNR of greater than 18 dB means that the band has perfect intelligibility and TI equals 1, while an SNR under −12 dB means that a band is not contributing at all, and the TI of that band equals 0. The overall intelligibility is then a function of the AI, but this function changes depending on the semantic context of the signal.

Kryter validated many of the underlying AI principles [Kryter, 1962a]. Kryter also presented the mechanics for calculating the AI for different number of bands – 5,6,15 or the original 20 – as well as important correction factors [Kryter, 1962b]. Some of the most important correction factors account for the effects of modulated noise, peak clipping, and reverberation. Even with the application of various correction factors, the AI does not predict intelligibility in the presence of some time-domain distortions. Consequently, the Modulation Transfer Function (MTF) has been utilized to measure the loss of intelligibility due to echoes and reverberation [Houtgast & Steeneken, 1973]. Steeneken and Houtgast later extended this approach to include nonlinear distortions, giving a new name to the predictor: the Speech Transmission Index (STI) [Steeneken & Houtgast, 1980]. These metrics proved more valid for a larger range of environments and interferences.

Using a spiking model of the auditory periphery [Bruce et al., 2003] the Neural Articulation Index (NAI) is formed by describing distortions in the spike trains of different frequency bands. The spiking over time of an auditory nerve fiber for an undistorted speech signal (control case) is compared to the neural spiking over time for the same signal after undergoing some distortion (test case). The difference in the estimated instantaneous discharge rate for the two cases is used to calculate a neural equivalent to the TI, the Neural Distortion (ND), for each frequency band. Then the NAI is calculated with a weighted average of NDs at different Best Frequencies (BFs). In general detection theory terms, the control neuronal response sets some locus in a high dimensional space, then the distorted neuronal response will project near that locus if it is perceptually equivalent, or very far away if it is not. Thus, the distance between the control neuronal response and the distorted neuronal response is a function of intelligibility. Due to the limitations of the STI mentioned above it is predicted that a measure of the neural coding error will be a better predictor than SNR for human intelligibility word-scores.

3.2.1 Method

The auditory periphery model used throughout is from [Bruce et al., 2003]. The model is fully discussed in 2.1.4. The parameters of the synapse section of the model are set to produce adaptation and discharge-rate versus level behavior appropriate for a high-spontaneous-rate/low-threshold auditory nerve fiber. In order to avoid having to generate many spike trains to obtain a reliable estimate of the instantaneous discharge rate over time, the synaptic release rate, an approximation of the discharge rate ignoring the effects of neural refractoriness, is used instead.

The following formulation emulates most of the simulations described in Chapter 2 of Steeneken [1992]; it describes the full development of an STI metric from inception to end. For those interested, the following simulations try to map most of the second chapter, but instead of basing the distortion metric on a SNR calculation, the neural distortion is used.

There are two sets of experiments. The first, deals with applying a frequency weighting structure to combine the band distortion values, then there is an introduction of redundancy factors. The bands, chosen to match Steeneken [1992], are octave bands centered at [125, 250, 500, 1000, 2000, 4000, 8000] Hz. Only seven bands are used here. The Neural AI (NAI) for this is:

$$\mathrm{NAI} = \alpha_1 \cdot \mathrm{NTI}_1 + \alpha_2 \cdot \mathrm{NTI}_2 + \ldots + \alpha_7 \cdot \mathrm{NTI}_7 \ , \tag{3.18}$$

where α_i is the ith bands contribution and NTI$_i$ is the Neural Transmission Index in the ith band. All the αs sum to one, so each α factor can be thought of as the percentage contribution of a band to intelligibility. Since the NTI is between [0,1], it can also be thought of as the percentage of acoustic features that are intelligible in a particular band. The ND per band is the projection of the distorted (Test) instantaneous spike rate against the clean (Control) instantaneous spike rate:

$$\mathrm{ND} = 1 - \frac{\mathrm{Test} \cdot \mathrm{Control}^T}{\mathrm{Control} \cdot \mathrm{Control}^T}, \tag{3.19}$$

where Control and Test are vectors of the instantaneous spike rate over time, sampled at 22050 Hz. This type of error metric can only deal with steady state channel distortions, such as the ones used in Steeneken [1992]. ND was then linearly fit to resemble the TI equation 1-2, after normalizing each of the seven bands to have zero means and unit standard deviations across each of the seven bands. The NTI in the ith band was calculated as

$$\mathrm{NTI}_i = m \frac{\mathrm{ND}_i - \mu_i}{\sigma_i} + b \ . \tag{3.20}$$

NTI$_i$ is then thresholded to be no less then 0 and no greater then 1, following the

TI thresholding. In equation 3.20 the factors, m = 2.5, b = -1, were the best linear fit to produce NTI$_i$'s in bands with SNR greater than 15 dB of 1, bands with 7.5 dB SNR produce NTI$_i$'s of 0.75, and bands with 0 dB SNR produced NTI$_i$'s of 0.5. This closely followed the procedure outlined in section 2.3.3 of Steeneken [1992]. As the TI is a best linear fit of SNR to intelligibility, the NTI is a best linear fit of neural distortion to intelligibility.

The input stimuli were taken from a Dutch corpus [van Son et al., 2001], and consisted of 10 Consonant-Vowel-Consonant (CVC) words, each spoken by four males and four females and sampled at 44100 Hz. The Steeneken study had many more, but the exact corpus could not be obtained. 80 total words is enough to produce meaningful frequency weighting factors. There were 26 frequency channel distortion conditions used for male speakers, 17 for female and three SNRs (+15 dB, +7.5 dB and 0 dB). The channel conditions were split into four groups given in tables 3.8 through 3.11. Female speakers have little to no energy in the 125 Hz band. For female speakers, a subset of the filtering conditions are used, these are marked with an asterisk in table 3.8 through table 3.11.

In the above tables a one represents a passband and a zero a stop band. A 1353 tap FIR filter was designed for each envelope condition. The female envelopes are a subset of these because they have no appreciable speech energy in the 125 Hz octave band. Using the 40 male utterances and 40 female utterances under distortion and calculating the NAI following equation 3.20 produces only a value between [0,1]. To produce a word-score intelligibility prediction between zero and 100 percent, the NAI value was fit to a third order polynomial that produced the lowest standard deviation of error from empirical data. While Fletcher & Galt [1950] state that the relation between AI and intelligibility is exponential, Steeneken [1992] fits with a third order

	OCTAVE-BAND CENTRE FREQUENCY						
ID #_	**125**	**250**	**500**	**1K**	**2K**	**4K**	**8K**
1*	1	1	1	1	0	0	0
2*	0	0	0	0	1	1	1
3*	1	1	0	0	0	1	1
4*	0	0	1	1	1	0	0
5*	1	1	0	0	1	1	0
6*	0	0	1	1	0	0	1
7*	1	0	1	0	1	0	1
8*	0	1	0	1	0	1	0

Table 3.8: Rippled Envelope

	OCTAVE-BAND CENTRE FREQUENCY						
ID #_	**125**	**250**	**500**	**1K**	**2K**	**4K**	**8K**
9	1	1	1	0	0	0	0
10	0	1	1	1	0	0	0
11*	0	0	0	1	1	1	0

Table 3.9: Adjacent Triplets

	OCTAVE-BAND CENTRE FREQUENCY						
ID #_	**125**	**250**	**500**	**1K**	**2K**	**4K**	**8K**
12	1	0	1	0	1	0	0
13	1	0	1	0	0	1	0
14	1	0	0	1	0	1	0
15*	0	1	0	1	0	0	1
16*	0	1	0	0	1	0	1
17	0	0	1	0	1	0	1

Table 3.10: Isolated Triplets

	OCTAVE-BAND CENTRE FREQUENCY						
ID #_	125	250	500	1K	2K	4K	8K
18*	0	1	1	1	1	0	0
19*	0	0	1	1	1	1	0
20*	0	0	0	1	1	1	1
21	1	1	1	1	1	0	0
22*	0	1	1	1	1	1	0
23*	0	0	1	1	1	1	1
24	1	1	1	1	1	1	0
25	0	1	1	1	1	1	1
26*	1	1	1	1	1	1	1

Table 3.11: Contiguous Bands

polynomial, and we have chosen to compare to Steeneken [1992]. The empirical word-score intelligibility was from Steeneken [1992].

3.2.2 Results

3.2.2.1 Determining frequency weighting structure

For the first tests, the optimal frequency weights (the values of α_i from equation 3.20) were designed through minimizing the difference between the predicted intelligibility and the empirical intelligibility. At each iteration one of the values was varied up or down, and then the sum of the α_i was normalized to one. This is very similar to Steeneken & Houtgast [1980] whose final standard deviation of prediction error for males was 12.8%, and 8.8% for females. The NAI's final standard deviation of prediction error for males was 8.9%, and 7.1% for females. Full results are plotted in figure 3.6.

The frequency weighting factors are similar for the NAI and the STI. The STI

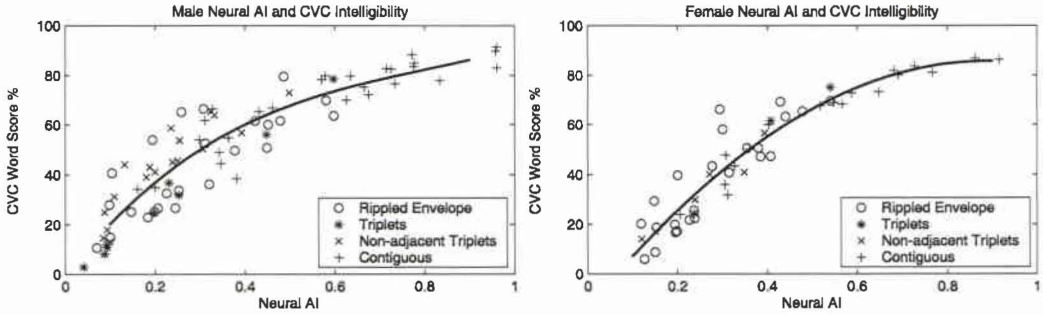


Figure 3.6: Relation between NAI and empirical word-score intelligibility for male (left) and female (right) speech with bandpass limiting and noise. The vertical spread from the best fitting polynomial for males has a s.d. = 8.9% versus the STI [5] s.d. = 12.8%, for females the fit has a s.d. = 7.1% versus the STI Steeneken & Houtgast [1980] s.d. = 8.8%

weighting factors from Steeneken [1992], which produced the optimal prediction of empirical data (male s.d. = 6.8%, female s.d. = 6.0%) and the NAI are plotted in figure 3.7.

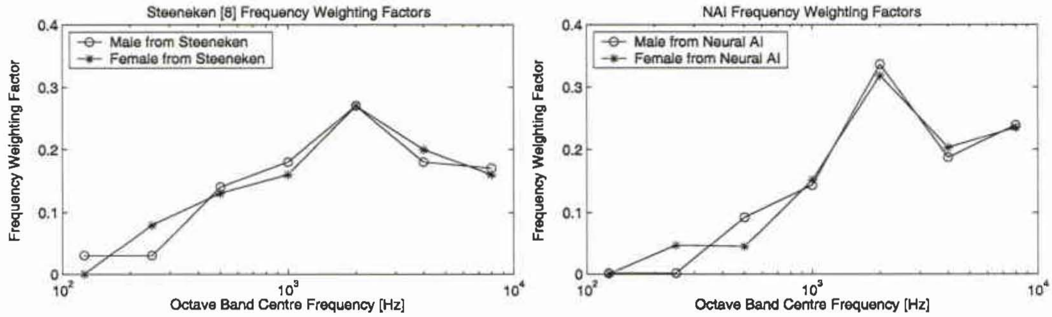


Figure 3.7: Frequency weighting factors for the optimal predictor of male and female intelligibility calculated with the NAI and published by Steeneken [1992]

As one can see, the low frequency information is tremendously suppressed in the NAI, while the high frequencies are emphasized. This may be an effect of the stimuli corpus. The corpus has a high percentage of stops and fricatives in the initial and final consonant positions. Since these have a comparatively large amount of high frequency signal they may explain this discrepancy at the cost of the low frequency weights. Steeneken [1992] does state that these frequency weights are dependant upon the conditions used for evaluation.

3.2.2.2 Determining frequency weighting with redundancy factors

In experiment two, rather then using equation 3.20 that assumes each frequency band contributes independently, we introduce redundancy factors. There is correlation between the different frequency bands of speech [Houtgast & Verhave, 1991], which tends to make the STI over-predict intelligibility. The redundancy factors attempt to remove correlated signals between bands. Equation 3.20 then becomes:

$$\mathrm{NAI}_r = \alpha_1 \cdot \mathrm{NTI}_1 - \beta_1\sqrt{\mathrm{NTI}_1 \cdot \mathrm{NTI}_2} + \alpha_2 \cdot \mathrm{NTI}_2 - \beta_1\sqrt{\mathrm{NTI}_2 \cdot \mathrm{NTI}_3} + ... + \alpha_7 \cdot \mathrm{NTI}_7\ , \tag{3.21}$$

where the r subscript denotes a redundant NAI and β is the correlation factor. Only adjacent bands are used here to reduce complexity. Replicating section 3.1 except using equation 3.21, the same testing, and adaptation strategy from before was used to find the optimal αs and βs. Results of this method for male and female speakers are in figure 3.8.

The frequency weighting and redundancy factors given as optimal in Steeneken, versus calculated through optimizing the NAI_r are given in figure 3.9.

The frequency weights for the NAI_r and STI_r are more similar than those calculated without redundancy factors. The redundancy factors are different though. The NAI redundancy factors show no real frequency dependence unlike the convex STI redundancy factors. This may be due to differences in optimization that were not clear in Steeneken [1992].

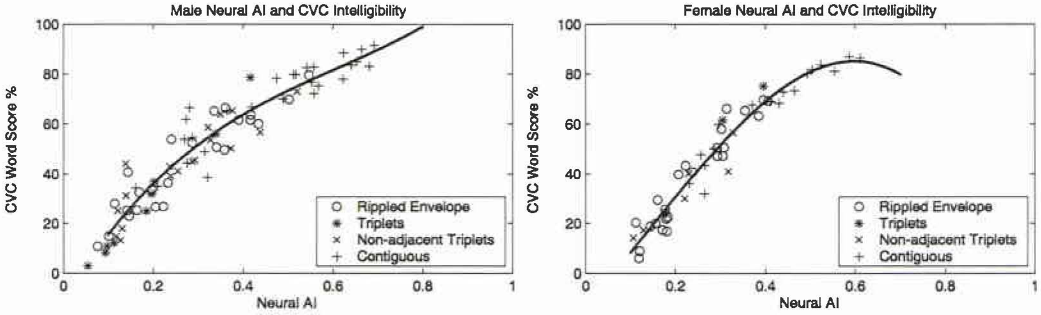


Figure 3.8: Relation between NAI$_r$ and empirical word-score intelligibility for male speech (right) and female speech (left) with bandpass limiting and noise with Redundancy Factors. The vertical spread from the best fitting polynomial for males has a s.d. = 6.9% versus the STI$_r$ [8] s.d. = 4.7%, for females the best fitting polynomial has a s.d. = 5.4% versus the STI$_r$ [8] s.d. = 4.0%.

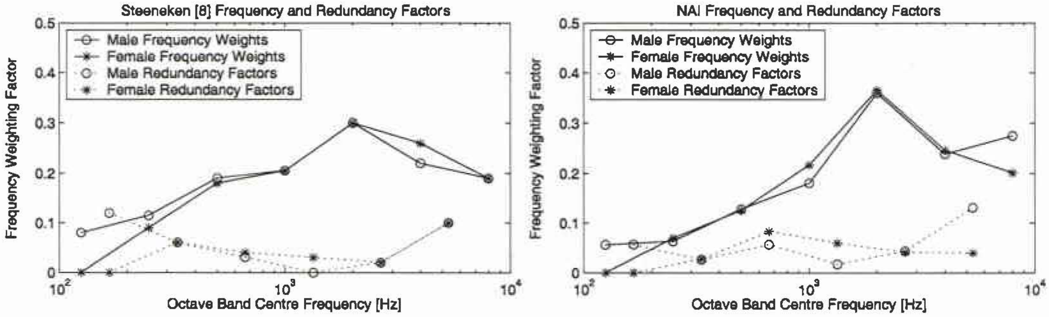


Figure 3.9: Frequency and redundancy factors for the optimal predictor of male and female intelligibility calculated with the NAI$_r$, from Steeneken [1992].

_	MALE EQ. 3	FEMALE EQ. 3	MALE EQ. 6	FEMALE EQ. 6
NAI	8.9 %	7.1 %	6.9 %	5.4 %
STI [5]	12.8 %	8.8 %		
STI [8]	6.8 %	6.0 %	4.7 %	4.0 %

Table 3.12: Standard Deviation of Prediction Error

The mean difference in error between the STI_r, as given in Steeneken [1992], and the NAI_r is 1.7%. This difference may be from the limited CVC word choice. It is well within the range of normal speaker variation, about 2%, the NAI and NAI_r are comparable to the STI and STI_r in predicting speech intelligibility.

3.2.3 Discussion, LTASS Extensions

The NAI provides a modest improvement over STI in predicting intelligibility. It is not a replacement for the STI for general acoustics since the NAI is much more computationally complex then the STI. The NAI's end applications are in predicting hearing impairment intelligibility and using statistical decision theory to describe the auditory systems feature extractors - tasks which the STI cannot do, but are available to the NAI.

While the AI and STI can take into account threshold shifts in a hearing impaired individual, neither can account for sensorineural, suprathreshold degradations [van Schijndel et al., 2001]. The accuracy of this model, based on cat anatomy and physiology, in predicting human speech intelligibility provides strong validation of attempts to design hearing aid amplification schemes based on physiological data and models [Sachs et al., 2002]. By quantifying the hearing impairment in an intelligibility metric by way of a damaged auditory model, one can provide a more accurate assessment of the distortion, probe how the distortion is changing the neuronal response and provide feedback for preprocessing via a hearing aid before the impairment. The NAI may also give insight into how the ear codes stimuli for the very robust, human auditory system.

In the following Chapters, specifically chapter 4 this derivation is used with other corpora. To produce a more structured approach, it is first obvious that the error in any band is the absolute value of one minus the correlation between the normal and impaired firing patterns and normalized by the average spike rate from the output of the normal auditory model. Results from equation 3.19 were always positive for the van Son et al. [2001] syllable corpus, so did not need the absolute value. Replacing NTI with ε; the error is calculated for each of the 7 frequency bands. The error in the i^{th} frequency band, for the j^{th} impaired condition is:

$$\varepsilon_{ij} = \left| 1 - \frac{\vec{x}_i \vec{y}_{ij}^{\,T}}{\vec{x}_i \vec{x}_i^{\,T}} \right| . \tag{3.22}$$

Where $\vec{x}$ is the normal auditory model instantaneous spiking rate vector, and $\vec{y}$ is the impaired auditory models instantaneous spiking rate vector over time. This metric cannot capture transient, or timing information of the auditory model because it cannot be coded through synchrony capture.

The difficulty in computing results over an entire corpus of sounds was minimized with using a frozen Long Term Average Speech Spectrum (LTASS) Gaussian noise. LTASS is spectrally steady-state signal, the metric from 3.22 can capture distortion in the response; coinciding with a statistically mean processing strategy. This is loosely equivalent to using a Signal-to-Noise Ratio (SNR) metric, since most of the power in an utterance is due to voiced speech, the SNR captures mostly effects of voiced speech while synchrony capture is very evident in the auditory nerve during voiced speech and is the main distortion mechanism for the NAI. Further research is necessary to deal with the auditory systems time-adaptive characteristics of the healthy cochlea, it is the subject of chapter 5.

The individual bands are then summed into a single error value with a weighting function following the STI [Steeneken & Houtgast, 1980] frequency importance weighting, but derived for the neural representation specifically [Bondy et al., 2004] and derived specifically for LTASS. The total error, in accordance with equation 3.18 is calculated using

$$\text{Error}_j = \sum_{i=1}^{N} \alpha_i \cdot \varepsilon_i \,, \tag{3.23}$$

where α_i's are the bands importance weighting functions (shown in Figure 3.10), and the ε_i's are calculated through Equation 3.22.

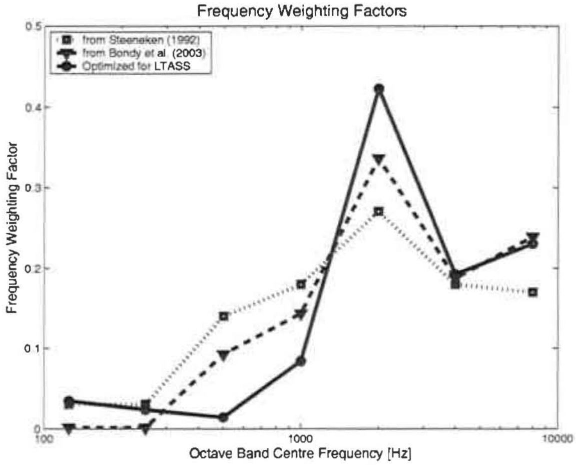


Figure 3.10: Frequency weighting factors used in calculating the neural distortion, as well as factors derived from a different stimuli Bondy et al. [2004] and an acoustic signal counterpart Steeneken & Houtgast [1980]. There are large differences between the importance of different frequencies for different corpuses.

Bondy et al. [2004] shows that this weighted sum of Hebbian error is a monotonic function of intelligibility. This is important for offline assessment because it gives a way to judge different hearing aid processing strategies against one another. The actual error will be a relative indicator between strategies under test, not an intelligibility value.

Chapter 4

Processing Blocks for Machine Learning

After successful completion of the intelligibility metric/error signal from Chapter 3 the next open question was providing evidence of the ability to drive optimization of a hearing aid speech processing block[1]. The most widely used hearing aids are simple linear processing blocks. Section 4.1 shows the development of the machine learning, linearly constrained hearing aid. The newest, industry accepted hearing aid are nonlinear, compression hearing aids. Section 4.2 details some interesting insights tied to developing the machine learning compression processing block. Finally, some researchers have suggested that the auditory nerve's coding mechanism may be a form of divisive normalization [Schwartz & Simoncelli, 2001]. An unsuccessful attempt was made to envelope this theoretical footing in hearing aid processing in Section 4.3.

Before delving into the processing though, the machine learning technique was designed to take into account different levels of hearing impairment. For this a set of

[1] This section is based on Bondy et al. [2004].

"offline" subjects was needed. Initial testing of the machine learning framework to establish the validity of the model-based approach was accomplished using idealized models of hearing impairment similar to those described in Byrne et al. [2001]. The different audiograms are given in Figure 4.1 A, and the individual contributions of inner and outer hair cell losses to each loss profile are given in Figure 4.1 B through F.

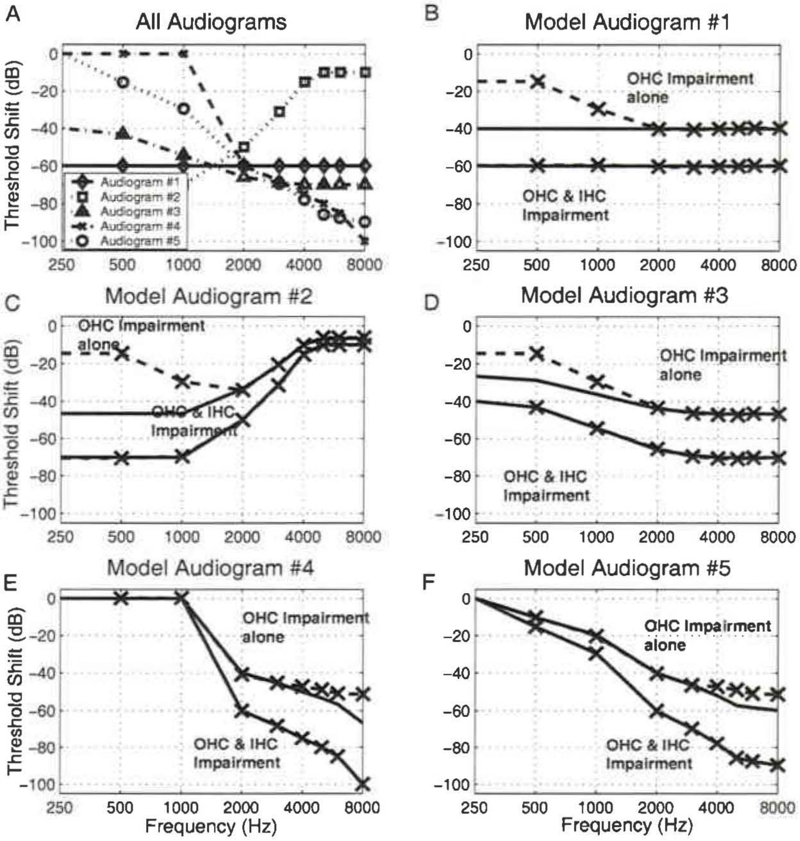


Figure 4.1: Hearing threshold shift for 5 candidate audiograms taken from Byrne et al. [2001]. Differences are due to approximating threshold shift with IHC and OHC damage. The dashed and X marked lines are ideal estimated audiograms, while the solid lines are the actual audiogram provided by the Bruce et al. [2003] model.

The impairments of inner and outer hair cells per frequency were calculated so that OHC impairment accounted for approximately 50-60% of the total threshold shift in dB [Moore et al., 2000]. The percent IHC loss was then adjusted to explain the remaining threshold shift. Loss profiles 3, 4, and 5 are indicative of presbyacucis or progressive sound-induced hearing loss and are more typical of the normal hearing loss pathology.

Unlike, the van Son et al. [2001] Dutch syllable corpus used in much of chapter 3, the stimulus used in this chapter was Gaussian noise shaped to have the same spectrum as the Long Term Average Speech Spectrum (LTASS, ?, combined data). A 200 millisecond LTASS input was sampled at 500 kHz and presented at 75 dB SPL into the normal model. The impaired model input would be processed by some test compensation strategy before being input into one of the impaired models, so the power level would fluctuate depending on the compensation strategy used. The high sample frequency is necessary for the Bruce et al. [2003] auditory model.

The output of the model was a time series, 230 ms long (the extra time versus the input could be used to judge offset effects), with a 22050 Hz sample rate, of instantaneous neural spike rates across 7 octave bands, starting at 125 Hz and ending at 8000 Hz. The neural best frequencies were chosen to mimic the Byrne et al. [2001]; (Figure 2) audiogram data points. A typical output of a normal and impaired auditory nerve fiber model with a BF of 250 Hz is given in Figure 4.2.

In general, it is hoped that encompassing the cochlear processing and impairment will circumvent the problems Fabry & van Tassell [1990] had with using the articulation index to fit hearing-aids.

The process of predicting efficacy of hearing aid algorithms is to take the LTASS stimulus, pass it through the normal auditory model and then take the same stimulus, preprocess it with the hearing-aid algorithm under test and pass it through an

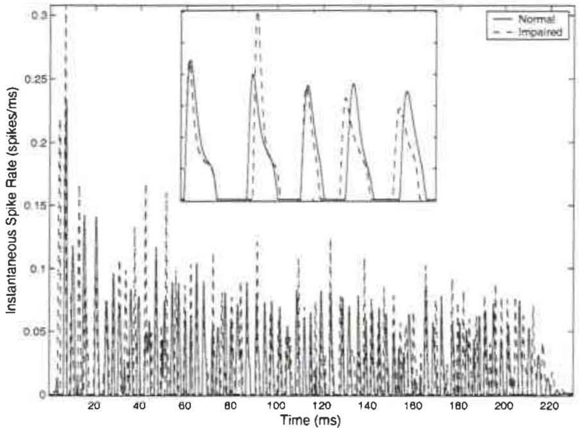


Figure 4.2: Time plot of instantaneous spiking rate versus time for the normal auditory model and the impaired auditory model. The impaired model is based on Audiogram #4 from Figure 4.1 with the half-gain rule applied (see Section 4.1) to the input. The inset shows a closeup, where synchrony is very evident, as well as some differences between the Normal and Impaired outputs.

impaired auditory model, whose audiogram loss follows one of the profiles in figure 4.1, then calculate the neural distortion following equation 3.22. This is repeated for several frequencies and the errors are summed following equation 3.23. The next section illustrates how this predictive measure closely fits empirical data, and subsequent sections extends this to train new processing strategies for hearing-aids.

Since the new compensation strategy relies heavily on neural network type training, and is in essence trying to re-establish normal neural activity, the general processing strategy was coined Neurocompensation. A Neurocompensator is any block whose weights are fitted to an individual's hearing loss through a training sequence that attempts to return the normal neural code. The training sequence is represented in figure 4.3.

The Neurocompensator, N_c, is trained on a set of input signals, supervised by the difference between the output across a set of frequencies of the normal auditory model, H, and the output of the impaired auditory model, $\hat{H}$. For each training iteration the

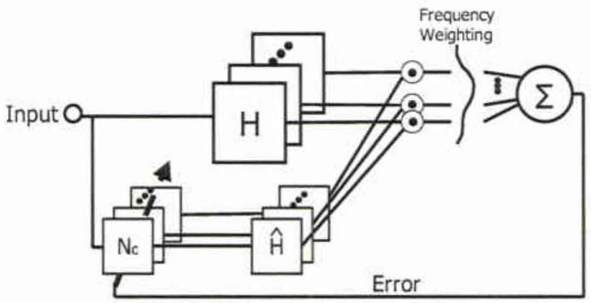


Figure 4.3: Block representation of the NeuroCompensator training sequence. The dot operator before the frequency weightings corresponds to Equation 3.23, the weighting operator corresponds to Equation 3.22. The normal and impaired auditory output is a set of the Auditory Model at different best frequencies, while the Neurocompensator is represented as a different preprocessor at the different frequencies, but this is not necessarily the case. There may be only one Neurocompensator preprocessing block.

Neurocompensator is adjusted by changing weights in its gain function to minimize the error signal. Training with LTASS noise will lead to a Neurocompensator that is optimal in the mean sense.

4.1 Optimal Linear Hearing Aid Processing

The validity of restoring normal auditory nerve activity through the use of supervised learning is tested by asking the question: would the standard fitting strategies, when applied to the input of the impaired model, result in optimal improvements in neural coding?

4.1.1 The Half Gain Rule

Early papers in audiology attempted to describe the amount of gain necessary for comfort and intelligibility. Markle & Zaner [1966] gave data showing how restoring normal hearing thresholds by setting the gain in each frequency band exactly equal to the threshold loss results in a signal too loud to be comfortable or intelligible. Rather than employing a one-to-one gain to threshold loss, Byrne & Fifield [1974] found that a 0.46-to-one gain to threshold shift was optimal. That is, for every 10 dB threshold shift, the gain for optimum intelligibility should be 4.6 dB. The Byrne & Fifield [1974] data is the basis for the widely used [Martin et al., 1998] fitting strategies from the National Acoustics Lab of Australia (NAL-1, NAL-R, NAL-RP, NAL-NL1 . . .).

The first experiment modeled the neural representation distortion introduced by setting different gains per dB threshold shift. Multiple ratios, R, are modeled with the gain in dB per dB of loss changing from 1:1, or 1 dB of gain per dB of threshold loss (mimicking Markle & Zaner [1966]) to 0:1, or no processing whatsoever.

The LTASS input stimulus and the error calculation is described at the end of chapter 3. The experiment is run for all five loss profiles with a sweep of the gain ratio, and the results are shown in figure 4.4, the y-axis is in model units, with larger values representing more distortion in the auditory nerve.

The vertical line drawn at the minimum error point for the average of the five model audiograms is 0.44 dB gain per dB threshold loss. Clearly this result is very close to empirical evidence (0.46). Model audiograms three, four and five are more indicative of typical hearing loss pathology and these have less individual error than the flat audiogram of loss profile one. Another important insight is that the more severe losses need higher gain. This is consistent with empirical data for fitting the profoundly impaired.

Brooks [1973] gives a possible theoretical footing for increasing gain at half the hearing threshold degradation by pointing out that the most comfortable level is approximately half way between threshold and the maximum tolerance level. Kates [1993] suggests that the half-gain rule is based on the natural compression ratio of

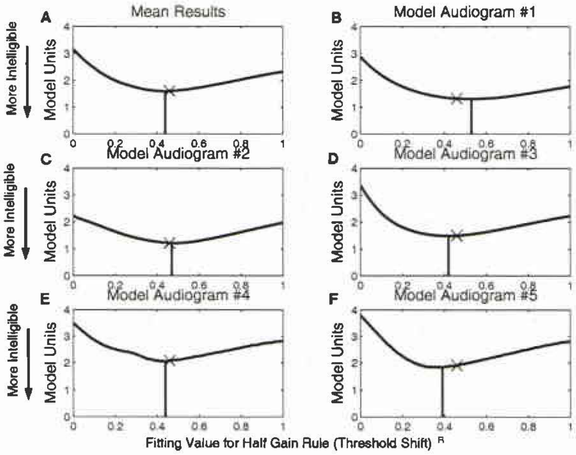


Figure 4.4: Neural error function versus gain ratio (R), showing a minimization of differences between the normal and impaired auditory models for a hearing-aid gain ratio. The vertical line is at the ratio (R) which minimizes the error curve, the X is at the value predicted by Byrne et al. [2001] data (0.46). The mean scheme in A minimizes neural differences between the normal and impaired at a ratio of gain to hearing loss of 0.44 which is very close to the empirical data of 0.46 dB of gain per dB of hearing loss. The Threshold Shift is raised to the power of the ratio since the fitting strategies correspond to a dB:dB gain. Graphs B-F show the error curves for each loss profile.

the active ear. These results show another possible scenario: people are trying to fit their gain response to a normal neural representation.

4.1.2 NAL-RP Gain Rate

The next experiment dealt with the prediction of linear fitting strategies. The formula used for the following examples is the NAL-RP [Byrne et al., 1990] formula, introduced as the NAL-R [Byrne & Fifield, 1974]. Without the profound loss additional gain factors NAL-RP is:

$$H_{3FA} = (H_{500} + H_{1000} + H_{2000})/3 \tag{4.24}$$

$$X = 0.15 * H_{3FA} \tag{4.25}$$

$$IG_i = X + R * H_i + k_i \tag{4.26}$$

Here H_i is the threshold shift measured at frequency i Hz. H_{3FA} is the average threshold shift at 500, 1000 and 2000 Hz, X is a gain factor across all frequencies, and R is the gain in dB for each dB of loss. In the NAL-RP formula, $R = 0.31$. The insertion gain at frequency i, IG$_i$, is made up of the constant gain factor X, 0.31:1 dB of gain per dB of threshold shift at that frequency, and a gain factor that is dependent upon the frequency, k$_i$, but not the audiogram. k$_i$ is described in Table 4.13.

	Frequency [Hz]						
	250	**500**	**1000**	**2000**	**3000**	**4000**	**6000**
k$_i$ [dB]	-17	-8	1	-1	-2	-2	-2

Table 4.13: Frequency shaping gain values for NAL-R

Following the Half-Gain rule experiment, the neural error was used to try to predict the constants in the NAL-RP formula. LTASS is input at 75 dB SPL, the same input and output frequencies, and the same loss profiles were used. This time the gain per dB of threshold shift in Equation 4.26 (0.31:1) is modelled. The curve in figure 4.5 is the neural distortion when the multiplier (R) is swept from zero to one, with the fitting strategy having a prescribed value of 0.31.

The minimum error of the multiplier for the strategy that attempts to restore neural firing patterns (0.34), based on these simulations, closely matches the NAL-RP fitting strategy's multiplier (0.31), derived through empirical evidence. Most of

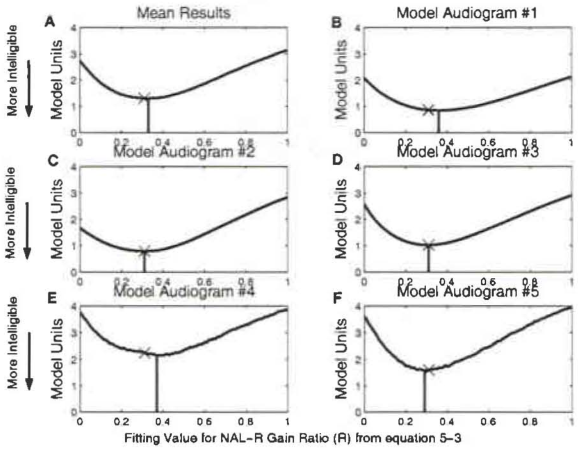


Figure 4.5: NAL-RP fitting strategy with the threshold shift versus gain ratio swept to see if the neural distortion error predicts empirical, published results. The X is at R = 0.31 and the vertical bar is at the minimum value of the error surface. The mean optimal value predicted by the neural distortion (in A) error is 0.34 dB gain per dB of hearing threshold shift versus the historical data of 0.31. The individual error functions for each model audiogram are given in parts B-F.

these curves have a lower minimization point with the NAL-RP formula than the Half-Gain rule as well, with the exception of loss profile #4.

Rankovic [1991] reports that people with profound hearing loss found that fitting with a high amount of gain in high loss frequency regions had their intelligibility reduced. Figure 4.5's higher tail towards the full 1:1 dB of gain per dB of loss agrees with Rankovic, if intelligibility is a monotonic function of neural distortion.

4.1.3 NAL-RP Frequency Shaping

The final modelling was for the NAL-RP frequency weighting factors, (k_i). A simultaneous optimization of the seven frequency gain factors, starting at unity gain, was carried out. All initial condition were as before. The gain per dB of threshold shift was the NAL-RP recommended 0.31:1, not 0.34:1 as recommended by the previous experiment; the remaining NAL-RP factors were used. It is hoped that the frequency weightings calculated would closely match NAL-RP's k_i factors. The optimized frequency weightings versus the original weightings are in figure 4.6. This is for loss profile #4, as a profile that is typical of moderate age or noise induced hearing loss.

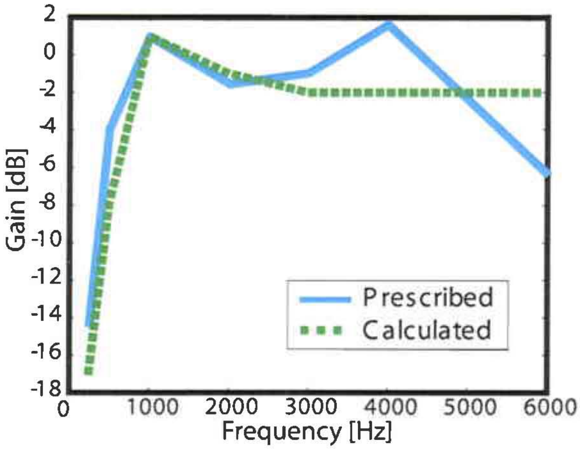


Figure 4.6: NAL frequency weightings calculated through neural error metric optimization and what was prescribed in the original NAL-RP strategy.

The calculated gain frequency weightings in figure 4.6 has had a small, flat gain shift (less than one dB, which could coincide with the different optimal gain ratios) applied to it before being plotted to center it on the NAL-RP gain curve, emphasizing the differences in shape.

The calculated gain frequency weightings and the prescribed weightings are clearly similar. The general low frequency attenuation, and the second formant range being emphasized is represented in both. There are differences including a lowering of the gain at the knee point of the audiogram and much lower high frequency gain. The knee point effect could be introduced by some nonlinearity between the normal hearing region and impaired region, or a model effect. The lower high frequency gain could be similar to how NAL-NL1 [Byrne et al., 2001] limits gains in highly damaged regions.

The above set of experiments started out simply and attempted to increase the complexity to test the basic assumption that offline modeling to return neural patterns to a hearing impaired auditory system corresponds to empirical data. This culminated in a metric that shows pronounced similarities to experimental data while being able to optimize multiple parameters. This section attempted to illustrate a connection between traditional, empirically derived hearing aid fitting strategies, and a new quantitative metric based on re-establishing normal neural representations in a hearing impaired individual. The neural error metric produced results very similar to empirical data, giving credence to the possibility of evaluating many fitting strategies quantitatively, and in corollary: the ability to calculate optimal characteristics in designing hearing-aid algorithms offline.

4.2 Optimal Compressive Hearing Aid Processing

The rationale behind including some sort of compressive preprocessing in a hearing-aid is the fact that an auditory system loses dynamic range due to sensorineural impairment[2]. Most researchers now agree that this loss is due to the destruction of hair bundles on outer hair cells (OHCs) in the cochlea. OHCs mechanically modulate the traveling wave of acoustic energy along the basilar membrane. This modulation acts as a nonlinear amplification at a particular frequency, and is also responsible for the suppression or contrast enhancement characteristics of a normal ear. Presumably, to restore normal hearing to a sensorineural impaired individual, there must then be some sort of compression in a hearing-aid.

[2] This section is based on Bondy et al. [2003].

Compression circuits in hearing-aids are characterized by time, intensity and frequency parameters. Individual parameters are selected based on reasons such as loudness normalization, discomfort avoidance, or dynamic range compression. Dillon gives an overview and tutorial on the competing rationales and characteristics Dillon [1996]. The degrees of freedom available to a hearing aid circuit designer make it infeasible to perform empirical intelligibility testing across all the possible parameters. Also, these studies look at what can be done to alleviate the symptoms of sensorineural hearing impairment, and do not address the core problem.

The true problem that needs to be modelled is how the compressive non-linearity of the cochlear amplifier, disturbed by sensorineural hearing loss, can be restored by signal processing in a hearing-aid. There is a complicated set of signal processing that is taking place in the cochlea that ultimately affects intelligibility. Quantitative evaluation of compression circuits in hearing-aids reduces the burden on empirical testing.

Quantitative analysis must also predict why there is such a large discrepancy between the hearing impaired and normal hearing person's ability to unmask competing speech. Understanding this disparity is key to building optimal compression circuits. Carhart & Tillman [1970] shows a SNR advantage between 12-15 dB for normal hearing people over hearing impaired people in identifying syllables in competing speech. Over time, testing methodologies have been refined, but results still show an enormous discrepancy between normal hearing and hearing impaired people's ability to understand speech against contending speech. Table 4.14 gives an overview of normal versus hearing impaired peoples ability to recognize target speech with a masking speaker.

To underscore table 4.14, in noise with a long term average speech spectrum

Study	Description	SRT Normal/ Impaired
Duquesnoy [1983]	20 elderly subjects with ski-slope high frequency loss; freefield; Competing @ 55 dBA	-17.6/-5.3
Festen & Plomp [1990]	20 mixed age and losses; monaural earphones; Competing @ 80 dBA	-11.4/-1.1
Hygge et al. [1992]	24 mixed age; freefield, binaural; Competing Speech.	-9.2/7.0 * SNR
Peters et al. [1998]	10 elderly subjects with ski- slope high frequency loss; monaural earphones; Competing @ 65 dBA	-11.9/0.8

Table 4.14: Intelligibility in speech and speech-like noise

(LTASS) the difference in SRTs between normal and impaired hearing individuals is only 2-5 dB [Glasberg & Moore, 1989].

It seems that to allow a sensorineural impaired person the ability to operate in the classical cocktail party in a way that approaches a normal hearing person, auditory compression must be understood in the competing speech regime. This will intertwine the counterbalanced processes of compression and suppression.

This section is also an advancement on the error metric provided in chapter 3. Because compression circuits are adaptive, a more speech-like corpus is necessary, but equation 3.22 averages over a stimulus, eliminating important temporal data. This is circumvented by an additional processing step for the NAI that relies on discharge rate contrasts to highlight important information. Fully, the acoustic signal is, as before, preprocessed with a compression algorithm, and then the representation of the auditory nerve activity is modeled. This modeling takes into account the cochlear compression that is being studied. The discharge rates over time and frequency are further processed by calculating regions of onset of activation, and clustering the onset data across time and frequency. This spectro-temporal fusion reveals a very different pattern between normal and impaired auditory representations and leads to a mapping between a normal and impaired hearing representation in this domain, to obtain a novel model of intelligibility.

Revisiting previous research, shows several attempts to produce a quantitative model to assess hearing aid performance or for hearing-aid circuit design. Fabry & van Tassell [1990] used the articulation index, Kates [1993] used a fairly simple compressing/suppressing auditory model, and Anderson [1994] used an invertible auditory model. *None of these attempts represent temporal information.* In this section timing information is introduced into the distortion metric. From the introductory discussion in chapter 2, the temporal modulations in competing speech are important in unmasking target speech in normal hearing people but are not accessible to hearing impaired people. The aim here is to provide an approach to encompass temporal information. Temporal information is lost with sensorineural impairment; present hearing-aid processing strategies do not address this, and neither do intelligibility predictors.

The first modelling experiment attempted to optimize different parameters of compression such as attack and release time, channel numbers, compression rates and thresholds for the loss profiles in figure 4.1. No consistent set of parameters resulted in enhanced predicted intelligibility across the impairments. Recent evidence points to a wider variation in efficacies for nonlinear fitting strategies between loss profiles than for linear fitting strategies [Gatehouse et al., 2005]. So the modelled results may actually be informative, but at the time there was no empirical data along these lines, most research had been focussed on deriving the single fitting framework. The inability to find quantitative data led to extending the intelligibility model to account for temporal mechanics in a qualitative way.

4.2.1 Extended Temporal Intelligibility Model

This model follows the process and data in Moore et al. [1999]. They carried out SRT tests on elderly hearing-impaired people with ski-slope, high-frequency loss with simulated linear, wide dynamic range compression (WDRC) and multiband (2, 4, 8) compression hearing-aids. The subjects were fitted using the “Cambridge” formula [Moore & Glasberg, 1998] for the linear condition. The compression ratio (CR) and threshold (CT) were determined by applying the following two constraints:

1. The gain in each channel for a 65 dB SPL, speech shaped, input noise is the same as in the linear condition.

2. The gain in each channel makes a 45 dB SPL speech signal in 65 dB SPL noise audible. That is when operating at 65 dB SPL, there is 20 dB of range between the output and 0 dB SL in each band.

The second constraint could not be held in all conditions. The offline subjects loss profile was the average loss profile of the 18 subjects from the study. The hearing loss in dB SPL, compression ratio and compression threshold per channel are listed in Table 4.15. The attack and release times are typical of fast compression: both were 8.2 ms. The output SRTs in competing speech for the unaided, linear, and eight-channel compression were reported as 0.5, -2.0, and -2.9 dB respectively [Moore et al., 1999].

Frequency	Hearing Loss	CR	CT [dB]
250	28	1.7	22.3
500	31	1.1	24.6
1000	38	1.3	16.1
2000	50	1.7	9.5
3000	59	2.4	7
4000	64	2.9	7
5000	66	2.9	7
6000	68	2.9	7

Table 4.15: Loss profile and parameters for a compression circuit.

While Moore, Peters and Stone used several different noise types, competing speech will be focussed on here, because of the large differences in intelligibility between stimulus types at the same SNR. 20 of the same HINT sentences [Nilsson et al., 1994], but recorded for multiple talkers [Trainor et al., 2004] are used. The auditory periphery model used throughout was taken from Bruce et al. [2003]. Figure 4.7 is an example of normal and damaged auditory responses from this model.

Effects of sensorineural impairment such as spreading in time and frequency are evident in the lack of separation between the lines representing formant frequencies. What is not obvious in this representation is how intelligibility is affected.

A representation of the acoustic waveform allowing grouping of onset cues was chosen as a way of identifying acoustic events that are perceptually relevant and may be the source of the intelligibility difference between normal hearing and hearing impaired people in competing speech. Onset characteristics of the auditory representation were calculated with a difference of exponentials filter, $h_1[n]$, in each frequency band

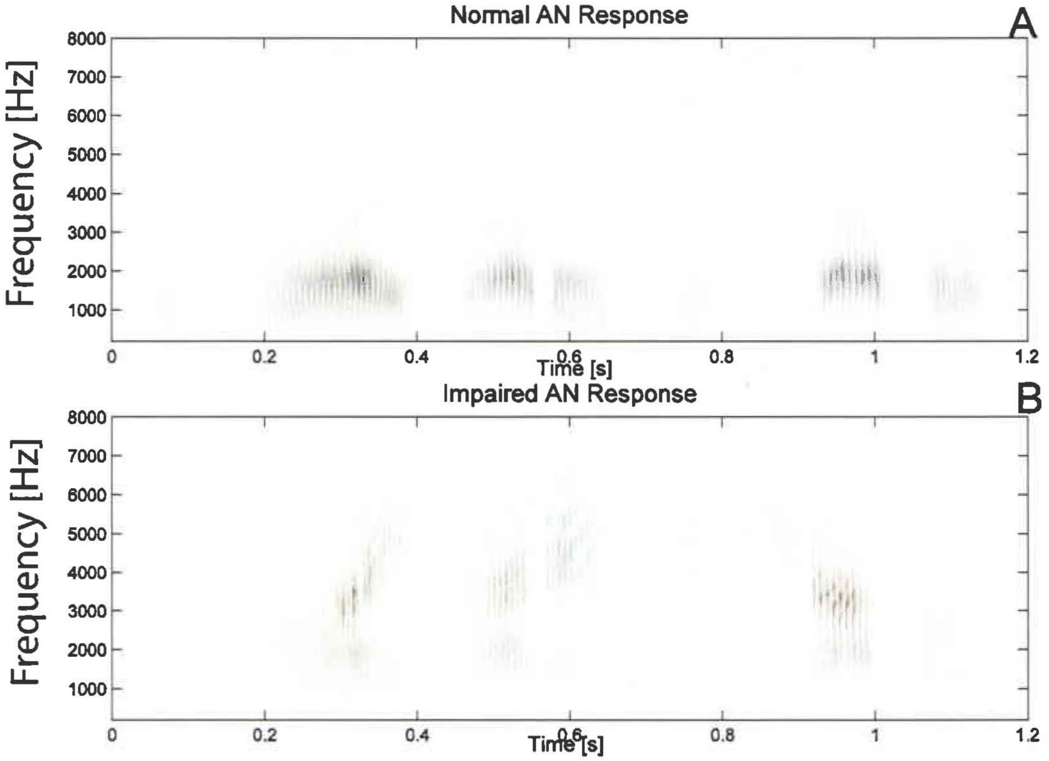


Figure 4.7: Auditory representation for the sentence "The boy got into trouble". Part A is from a normal hearing auditory model, B is from a sensorineural impaired auditory model. There is noticeable smearing in time and frequency as a result of the impairment.

$$h_1[n] = {}^{n}\!/_{\alpha_1^2} \exp^{-n/\alpha_1} - {}^{n}\!/_{\alpha_2^2} \exp^{-n/\alpha_2} . \tag{4.27}$$

α_1 and α_2 were selected to pass frequencies from 4 to 32 Hz. These frequencies contribute most to intelligibility, with a signal's fine temporal structure only adding a small amount to intelligibility [Drullman et al., 1994].

This onset data was then integrated over a typical acoustic event time window, $h_2[n]$, which had a 6 dB cutoff at 125 Hz. This integrator had a similar form to $h_1[n]$,

$$h_2(t) = {}^{t}\!/_{\alpha_3^2} \exp^{-t/\alpha_3} . \tag{4.28}$$

For a sampling rate of 11025 Hz, α_1 was 0.06, α_2 was 0.10, and α_3 was 0.001. An adaptive threshold and refraction operation was then applied. The threshold value was determined to produce some percentage of active events in the discretized time-frequency grid when the refractory period is 1 ms. An 1:1000 events-to-nonevent-ratio was selected arbitrarily (Many ratios clustered in experiments, and gave qualitatively similar results. Anything less then 1:100 clustered similarly). This produced a discrete event map such as the one given in Figure 4.8.

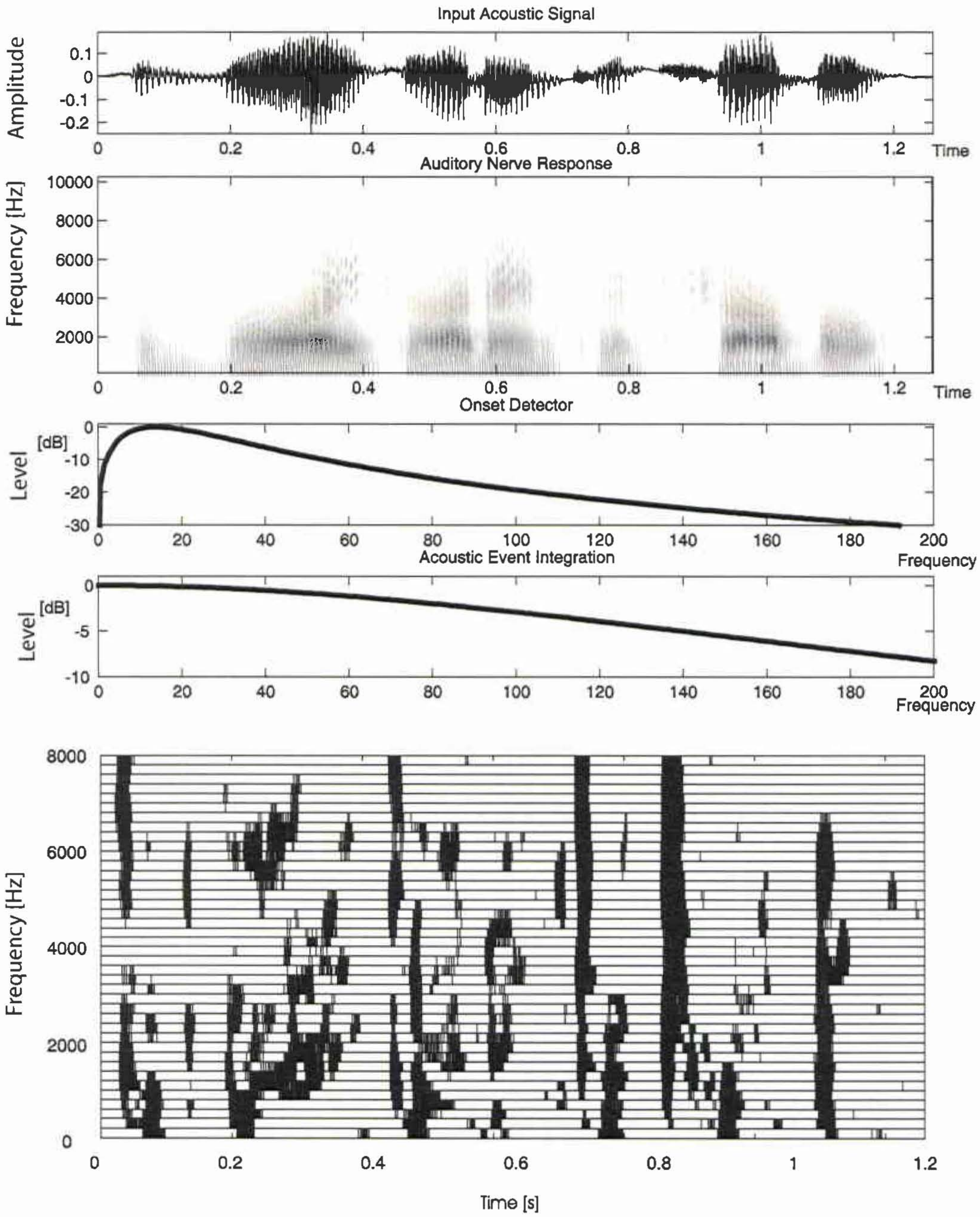


Figure 4.8: The input signal is represented on the auditory nerve. Then low frequency auditory information is extracted with an onset detector. The onset events are then integrated over a speech cue period. With thresholding and a refraction window, discrete events are finally mapped.

Figure 4.8 shows that important timing information is carried across multiple frequencies. To calculate perceptual relevance, a clustering algorithm using a hard-decision rule for class membership based on a Gaussian probability distribution assumption is applied. Taking the thresholded information from Figure 4.8, and making each event, k, a two dimensional sample in time (subscript t) and frequency (subscript f),$\vec{z}_k$={z_{tk}, z_{fk}}, the whole set of acoustic events is represented as **Z**. Starting with a limited number of possible classes, J, an iterative clustering algorithm, with death for small clusters is run [Duda & Hart, 1973]. A sample was assigned to class j when

$$\pi_j P(j|\vec{z}_k) > \pi_i P(i|\vec{z}_k), \quad \textit{all } i \neq j \tag{4.29}$$

where

$$P(j|\vec{z}_k) = \frac{1}{\sqrt{2\pi}\,|\det(\Sigma_j)|} \exp^{-\frac{1}{2}(\vec{z}_k - \vec{\mu}_j)\Sigma_j^{-1}(\vec{z}_k - \vec{\mu}_j)^T} \tag{4.30}$$

π_j, μ_j *and* $\Sigma_j$are the prior probability, mean and covariance statistics for class j, respectively. All samples were classified before the prior and statistics are recalculated in a batch mode. Classes with a low prior probability were pruned; in these examples, classes with less then half a percent of all the events were discarded. Classification and statistical updates were iterated until the priors stopped changing between iterations by more then two percent root-mean-square.

The classes were then split in half along the temporal axis and classification was again seeded and performed in the halved datasets to account for time warping or long pauses. This bifurcation helps competition and reduces reliance on initial conditions. The result of this clustering, using the onset data from figure 4.8 and bifurcated once with 50 initial classes is given in figure 4.9.

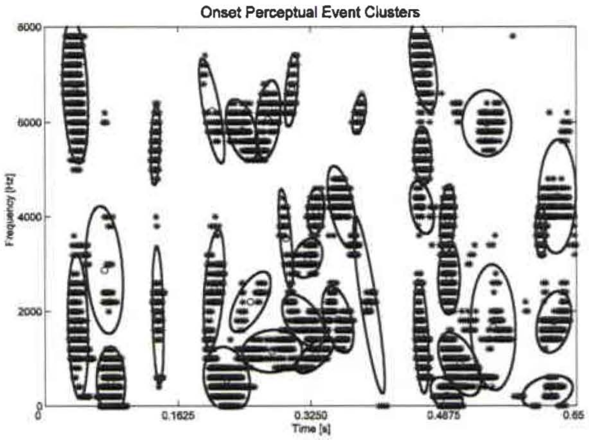


Figure 4.9: The normal hearing, perceptual clusters mean and two-standard deviation contours are plotted for the first half of the example sentence in steady state LTASS noise.

4.2.2 Qualitative Model for Temporal Information

The goal of this research was to be able to quantify effects on intelligibility of non-linear, dynamic algorithms for sensorineural impairment. The question looked at first was whether this "perceptually relevant" clustering produced distinctly different representations for the normal and impaired auditory models. Using the same Cambridge linear fitting strategy, with the simulated steeply sloped hearing loss as detailed in Table 4.15 the stimulus was presented to the normal and damaged auditory models. The normal model produced the clusters shown in figure 4.9 and the damaged model with preprocessing gain calculated by the Cambridge formula produced the clusters shown in figure 4.10.

For this example, the impaired model forms fewer classes and the variances of those classes are greatly enlarged, while entire onset cues for some phones are lost. The dotted lines representing the normal hearing clusters are sometimes far removed from the impaired clusters. These results are indicative of both spectral and temporal spreading. Table 4.16 highlights the general results.

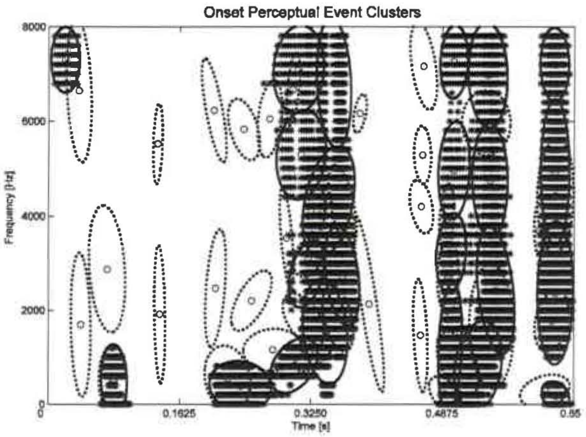


Figure 4.10: The sensorineural impaired perceptual clusters mean and two-standard deviation contours for a sentence in LTASS are plotted in solid lines. The normal clusters from figure 4.9 are plotted with dotted lines for reference.

Variable	Normal	Impaired Linear	Impaired 8-Channel
σ_t	10 ms	11 ms	12 ms
σ_f	398 Hz	503 Hz	517 Hz
Classes/second	53.8	35.5	32.6

Table 4.16: Differences in Normal and Impaired perceptual clustering in steady state noise

This general pattern should be indicative of differences between normal and impaired listeners on the order of 3 to 5 dB in SRT. This is the baseline deficit that hearing impaired people face in conditions without any temporally modulated noise. Another test versus empirical data is to judge the difference between linear and the 8-channel compression preprocessing. With the 8-channel compression circuit the impaired results are almost identical to the linear case. In only three of the twenty test sentences did compression produce a "phantom grouping", where a cluster was formed outside of a phone boundary. This is the expected result with speech presented at 65 dB SPL because it will very rarely go under the compression threshold with the windowed energy calculation used here. Compression circuits do not overly change the AN representation of onset cues.

So far all distortion conditions have been steady state noise. An important open question is how are these results affected by competing speech streams? Moore, Peters and Stone (1999) used a female talker whose long term average speech spectrum was modified to match the male targets as an interfering signal. The time envelope was basically undisturbed. Figure 4.11 shows the clustering that takes place in a normal hearing model for these data.

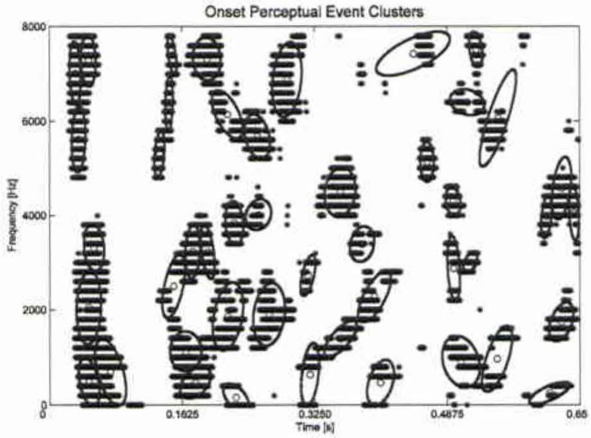


Figure 4.11: Normal clusters for the same input as figure 4.9 and 4.10 but with a competing speech masker. There are more classes with smaller variances between them than the normal hearing model clustering without competing speech.

The compression and suppression characteristics of a normal undamaged ear have clearly changed the representation between target speech and target speech with competing speech. The clusters are smaller and greater in number. This is not the case in the impaired auditory system's ability to cluster two speech signals as shown in figure 4.12.

Figure 4.12 is the grouping that takes place with 8-channel compression. Clearly the auditory system can not make use of the onset characteristics of a speech signal

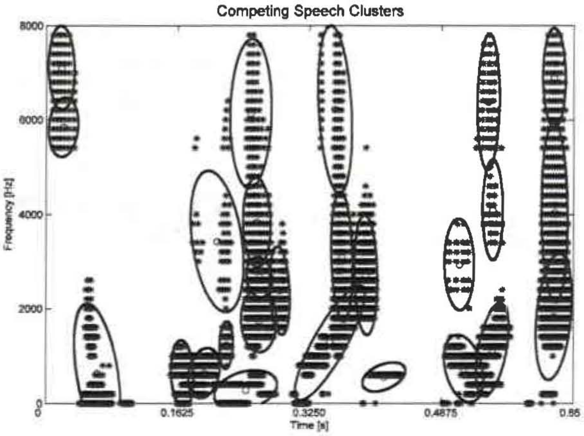


Figure 4.12: Impaired competing speech clusters for the same input as figures 4.9 and 4.10. There are roughly the same number of classes and those class variances remain high.

with this type of compression. While the normal ear responds with more specific groupings of acoustic events because of spectral-temporal suppression, the normal compression circuit does nothing to reestablish normal cochlear signal processing.

Table 4.17 is a comparison of the clustering statistics in competing speech.

Variable	NORMAL	Impaired Linear	IMPAIRED 8-CHANNEL
σ_t	10 ms	12 ms	12 ms
σ_f	348 Hz	555 Hz	573 Hz
J	70.8	38.4	34.1

Table 4.17: Differences in Normal and Impaired perceptual clustering in competing speech

Comparing table 4.17 to table 4.16 the data that jump out are the much smaller variances in the normal ear, the larger number of classes, while the statistics for the impaired ear remain remarkably similar. This is conceivably the reason why a normal hearing person has reduced SRT in competing speech versus steady noise;

−12 dB versus −4 dB; while a hearing impaired person does not see the same level of advantage; −2 dB versus 1 dB [Peters et al., 1998].

If intelligibility is the ability to group perceptually relevant acoustic cues while removing other events from different streams, normal hearing people have a clear advantage at the *segmentation* level. This result is clearly different from the articulation index (AI) or speech intelligibility index (SII). They calculate the intelligibility of a speech token based on the summation of signal-to-noise ratio (SNR) in a set of bands. The temporal information clustering criterion maintains that a more appropriate measure of the intelligibility of a speech token is the event-to-noise ratio. Above, the events will have some spectral-temporal mask that can be used to determine whether the acoustic cue is discriminable. This can test specific phones, while the AI and SII measures have an implicit assumption about the ensemble statistical structure of speech across frequency.

A novel way of representing acoustic material that qualitatively predicts intelligibility for a compression circuit in a hearing-aid in competing speech takes into account the discharge rate contrast. This representation is affected by time, intensity and frequency parameters. To make this into a more useful intelligibility metric it still needs validation against normal conditions, and a mapping between the clustered space and a scalar intelligibility value.

4.3 Divisive Normalization for Hearing Aids

Another avenue that was explored under the umbrella of nonlinear processing was divisive normalization[3]. Schwartz & Simoncelli [2001] suggested divisive normalization as the coding strategy that would optimally transform a set of coefficients from a linear auditory filterbank representation. Their derivation reduces dynamic range (compression, and desirable for limiting coding level issues) and increases independence (reducing coding redundancy). Extending section 4.1, where the formation of the metric and validation of the strategy was a linear processing block, this is a novel hearing aid algorithm.

The conceptual hearing aid processing block, or Neurocompensator, block is an attempt at spectral contrast enhancement following Schwartz & Simoncelli [2001]. The analytic equation is given in Equation 4.31.

$$G_i = \frac{v_i f_i^2}{\sum_j w_j f_j^2 + \sigma} \tag{4.31}$$

The gain at a frequency indexed by i, G_i, is a divisive function of the weighted (weighted by v_i) input power, f_i^2, at frequency index i, and the weighted sum (weighted by w_j) of all the frequencies power, f_j^2. σ is a term to ensure that G_i does not go to infinity. The weights, v_i and w_j, are trained in this Neurocompensator. The format of this example will produce a compensator that can apply level dependent gain, but not compression versus level, and ideally will produce some spectral contrast enhancement. The level dependent gain should produce a weighting that will show compression limiting.

The first step in training the Neurocompensator is a pre-processing stage where

[3] This section is based on Bondy et al. [2004].

the time signal is compartmentalized into time-overlapped windowed samples. These windowed samples are filtered into twenty frequency subbands, corresponding to the model bands that will be combined in the error signal, and the power is taken in each band (f_k where k = [1,2,...,20]). These are the statistics used as the input to the compensator model. A time series per frequency channel is derived, or G_i changes over time.

Each weight, G_i, is applied per time slice to the short-time Fourier transform and the inverse Fourier transform is taken. All the time-slices are assembled by overlapping and adding the processed windowed samples. The resulting time-domain waveform is the input to the damaged model. The input to the normal model can be thought of having G_i equals one over every frequency and every time-slice.

During the training phase, the v_i and w_j gain coefficients are adapted to minimize the error metric summed over all the time slices. The parameters of the compensator are optimized so that the output of $\hat{H}$ matches the output of H as closely as possible. Once the compensator is trained, the gain coefficients are set and it becomes the final stage of processing in a digital hearing aid, replacing the fitting strategy.

The Alopex algorithm [Unnikrishnan & Venugopal, 1994; Bia, 2001] was used to train the model weights via the error signal. Alopex is a stochastic optimization algorithm closely related to reinforcement learning and dynamic programming methods. It relies on the correlation between successive positive/negative weight changes and objective function changes from trial to trial to stochastically decide in which direction to move each weight.

Initial experiments were conducted with loss profile # 5. Instead of using LTASS noise, the dutch syllable "kas" from the van Son et al. [2001] corpus was chosen. This was because of the difficulty in compensating spectra of the stop /k/ and fricative

/S/ would give a more interesting example for contrast enhancement.

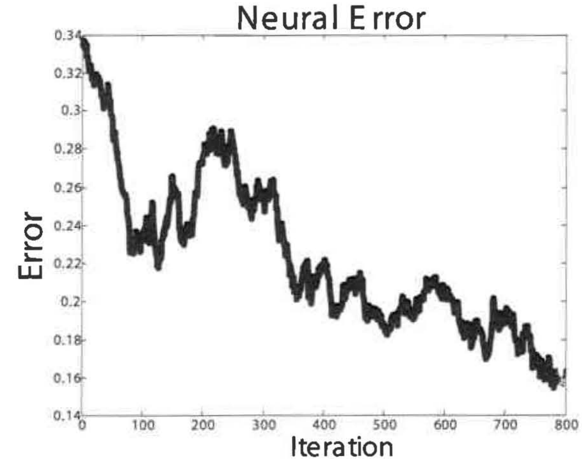


Figure 4.13: Training curve for the Neurocompensator for 800 iterations.

Figure 4.13 shows the error signal plotted versus iteration. Longer test results show that this error does not go to zero, meaning that better compensation models are needed, or that full restoration of the neural representation is not possible.

The idea behind the original processing block was that it would provide spectral unmasking, or provide some contrast enhancement for the ear. This is plainly evident when one compares the settled algorithm's spectrogram in Figure 4.14B to the input spectrogram in Figure 4.14A.

The dynamic range of both spectrograms is 60 dB. The NeuroCompensator has 35 dB more energy than the unprocessed input signal. Of special note is the second formant in the signal spectrum for the NeuroCompensator. It shows evidence of compelling lateral suppression that reduces the spectra above and below the formant. This would spread the response to the formant, and formant capture is very evident in the normal auditory system.

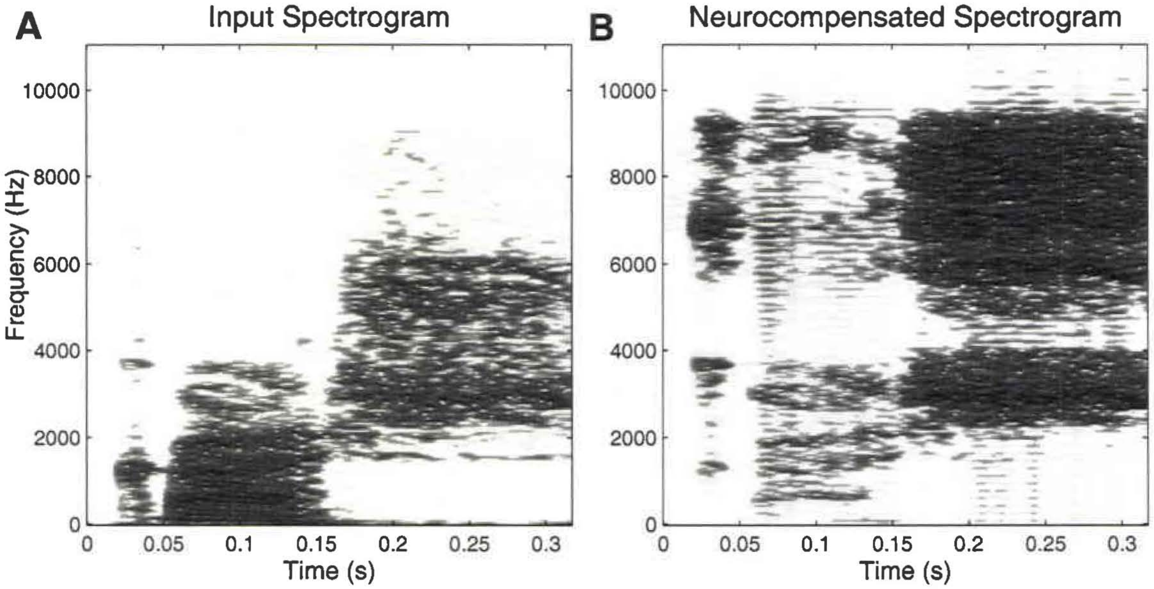


Figure 4.14: The unprocessed input spectrogram (A) and the spectrogram of the signal that would be presented to the hearing impaired ear after Neurocompensation (B).

Similarly to Kates [1993] the weights are dependent on the input stimulus, and should change over time mimicking the cochlea's cycle-to-cycle adaptive behaviour. At present, it is beyond the scope of the objective function to capture time adaptive and nonlinear, stimulus dependent effects. How the v_i and w_j change is a matter of future research.

This type of processing also introduces a gain dependent upon received level. An example of the weighting factors changing over time is given in Figure 4.15.

Figure 4.15 clearly shows an attempt to aid the transient response, or the '/k/' stop and the '/S/' fricative, and limit the voiced vowel '/a/'. This can be viewed as loudness equalization across time periods, but the present Neurocompensator trial does not have look-ahead or look-behind in time capabilities, so it should not be able to return proper time adaptive auditory processing that is lost with sensorineural

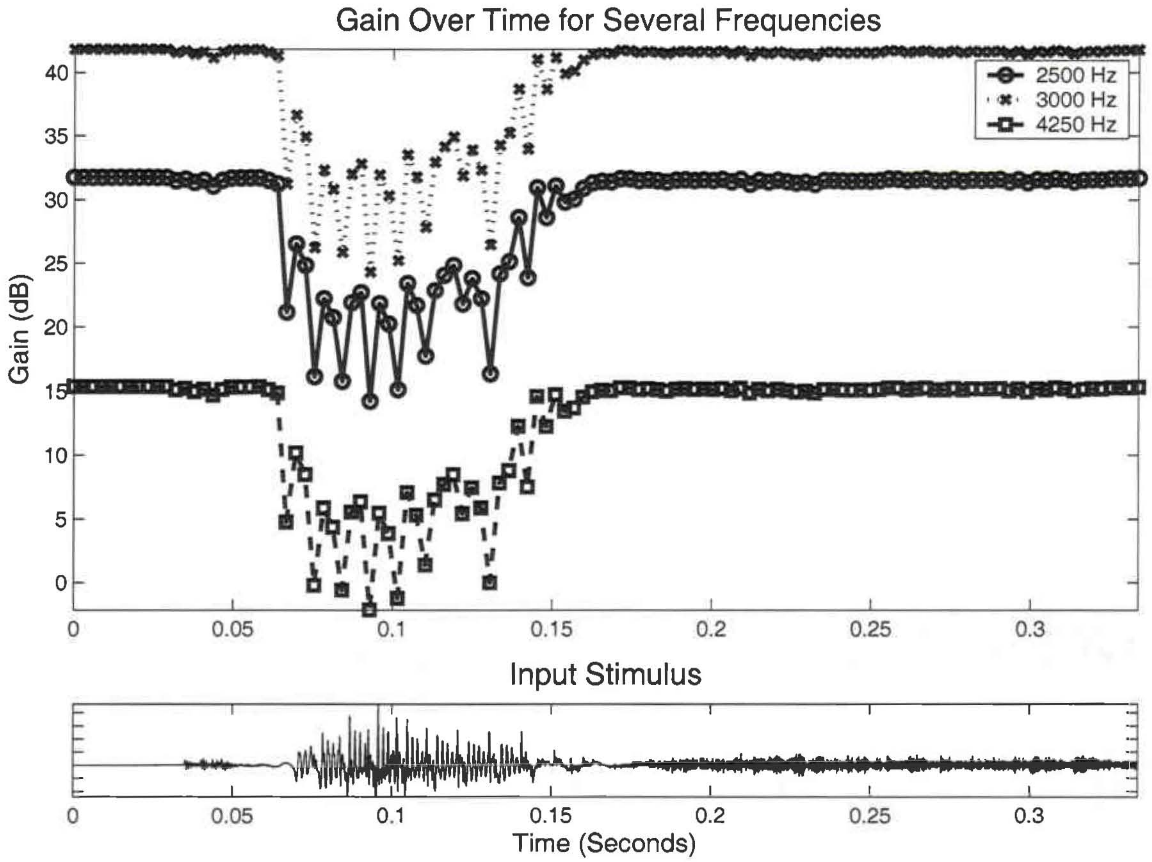


Figure 4.15: Gain over time for several frequencies of the trial Neurocompensator hearing impairment.

Chapter 5

Differences in Normal and Impaired AN Responses

The results from Chapter 4 form a basis for understanding hearing aid processing as an attempt to re-establish normal neural firing patterns. Chapter 4 indicates the possibility of optimizing hearing aid fittings per hearing loss pathology offline. It was not successful in adding new algorithm classes into the sensorineural impairment lexicon. By constructing the tools, or machine learning framework, in line with previous paradigms no novel insight could be developed. This is linked to the problems with the intelligibility predictors, where the initial assumptions about stationarity belie the actual problem.

In section 2.2, the reader was introduced to a large number of psychophysical measures that change with the advent of sensorineural hearing loss. This chapter hopes to integrate the root causes of the psychophysical symptoms, combining knowledge of important cochlear mechanisms and their effects on hearing in competing speech.

Due to the complexity of the problem this chapter remains largely theoretical to provide a foundation. Instead of continuing the assumption heavy intelligibility metric framework, this chapter derives novel insights into how the normal auditory periphery can parse difficult environments an *order of magnitude* better than the moderately impaired system in competing speech, and *several orders of magnitude* better then machine systems.

The chapter focusses on the degradation of audio coding brought on by specific types of cochlear damage. This shows new ways of quantifying sensorineural impairment, as it alters the neural coding of the acoustic environment.

An acoustic signal is often represented with a spectrogram; figure 5.1 is an example. This three dimensional representation (amplitude, time and frequency) is presented as an analogue to the AN representation. The cochlear processing that differentiates the AN representation from a spectrogram, namely compression, suppression and adaptation, are discussed in relation to coding in amplitude, frequency and time domains, respectively.

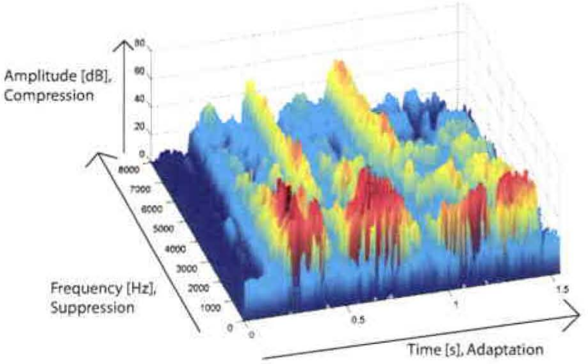


Figure 5.1: The 3-Dimensional representation of the auditory space that is fantastically well coded by normal cochlear processing. The spectrogram used is for the TIMIT sentence "How much allowance do you get?".

Understanding the changes in encoding this multidimensional space brought on by the loss of cochlear processing is fundamental in understanding hearing impairment as well as deriving new hearing aid algorithms.

In section 4.2, the problem of applying machine learning to nonlinear hearing aid or automatic gain control algorithms has already been delineated with the simple description of neural coding and distortion. No time-constants, channel counts or compression rates improved the predicted intelligibility. Section 5.1 deals with the loss of the cochlear compressive nonlinearity in a novel context. The highlight of section 5.1 shows how the desired gain in a hearing aid algorithm to produce the closest AN match is not well described by the acoustic signal power. The loss of compression cannot be fully described without understanding consequences further up the auditory system, and so an adaptive compression scheme is derived in Chapter 6.

Section 5.2 continues with the impact of suppression. Suppression, in the simplest terms, reduces neighbouring frequencies discharge rates when a central frequency is stimulated. An interesting motivation for this decorrelating process is the consequences on correlated firings (Hebbian mechanisms) in the auditory brain. In section 5.2 the loss of suppression engenders strong correlations between frequency channels.

The loss of contrast between frequencies from the loss of suppression also has a counterpart in temporal coding intrinsic to the loss of adaptation, especially fast adaptation. Section 5.3 highlights how onset adaptation is a powerful cue that highlights the first wavefront across frequencies. This temporal correlation across frequencies may be a key determiner to grouping frequency information. With the linearization that is a hallmark of sensorineural impairment this very important cue is lost, leading to deficits, not only in the hearing impaired person's ability to track temporal dips and the reduction of being able to use the first wavefront (a possible precursor to the precedence effect), but also to combine information across frequency channels (important for localization and dealing with reverberant environments). Hebbian learning dictates correlated firings are informative. This loss of correlated temporal-place firings may be important for describing many aspects of sensorineural impairment.

By viewing the auditory system as a bottom-up processor [Allen, 1994] the goal of this chapter is to motivate new insights into what processes cause something to be intelligible, how the loss of these processes affects intelligibility, and lastly to motivate the development of new hearing aid algorithms (Chapter 6).

An example of the discharge rate over time for a normal auditory nerve fiber, and a representative impaired auditory nerve fiber is in figure 5.2. The discharge rate of the normal auditory nerve is in blue, while the red curve is the impaired response (for loss profile #4 from figure 4.1) with NAL-R applied to the input acoustic waveform.

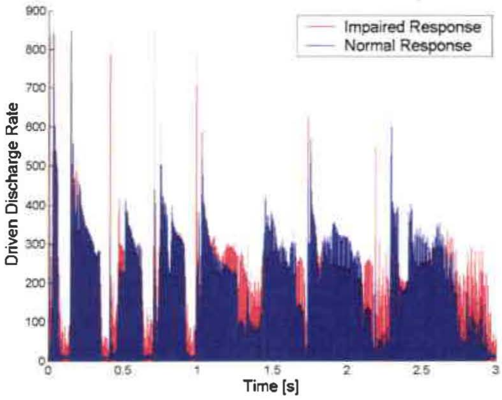


Figure 5.2: AN Response or discharge rate over time for the normal auditory nerve, and a representative auditory nerve for the TIMIT sentence "The dark, murky lagoon wound around for miles" at a BF of 750 Hz.

Generally the maximum discharge rate between the curves looks the same, and the general shape looks similar. A notable exception being the speech silences are captured better by the normal AN response. On closer inspection some interesting differences pop up. A zoomed in section of figure 5.2 of roughly 100 ms is given in figure 5.3.

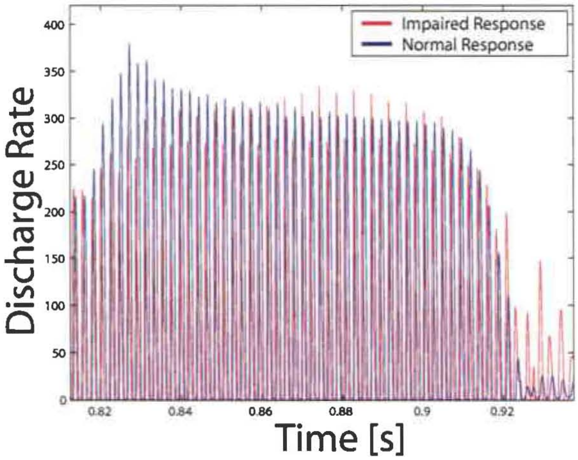


Figure 5.3: A snapshot of the AN response from figure 5.2, loosely corresponding to the /a/ in “lagoon”.

It is evident in figure 5.3 that the normal and impaired AN responses are about the same level on average, but they have distinct temporal qualities. The normal response quickly builds up at onset, while the more linear impaired response has a peak which captures a variation in the input acoustic waveform. Also the offset is much more distinct in the normal response then the impaired one. How important these effects are and trying to quantify what is important on the auditory nerve forms the basis of the remainder of this Chapter.

5.1 Differences in Compression Responses

Loudness recruitment (section 2.2.2) is often explained by an increasingly steep input level to rate function (rate-level function) brought on from sensorineural impairment[1]. Most compression circuits in hearing aids are really based on this assumption. In reality, Heinz & Young [2004] show that this steeper rate-level assumption does not hold. In general the damaged AN fibers resulted in rate-level slopes shallower than normal for tones. This is a very new and exciting result, directly affecting the understanding of loudness recruitment, and in turn having huge implications on hearing aid compression circuits.

Heinz & Young [2004] show discharge rate growth with sound level in the AN fibers of cats with noise induced hearing loss and a normal hearing, control group. They used a range of stimuli, including tones at an AN fiber's BF, fixed tones, broadband noise and a short CVC "besh". These stimuli produce a range of responses in normal hearing cats, but one clear consistency between normal hearing and hearing impaired cats is there is a reduction in the *range* of responses for the hearing impaired cats over the stimuli set.

The implied assumption in audition science is that loudness is proportional to the AN discharge rate, and the implied assumption in audiology is that hearing impaired people have steeper than normal discharge rates. This is almost always given as the reason behind loudness recruitment. Nonlinear hearing aid algorithms are predicated by this string of assumptions, but does not seem to hold in mammalian auditory systems. This important discrepancy is highlighted with the Heinz & Young [2004] data and the data replicated with the Bruce et al. [2003] model with a high spontaneous

[1] This section is based on Bondy & Bruce [2004b].

rate synapse model.

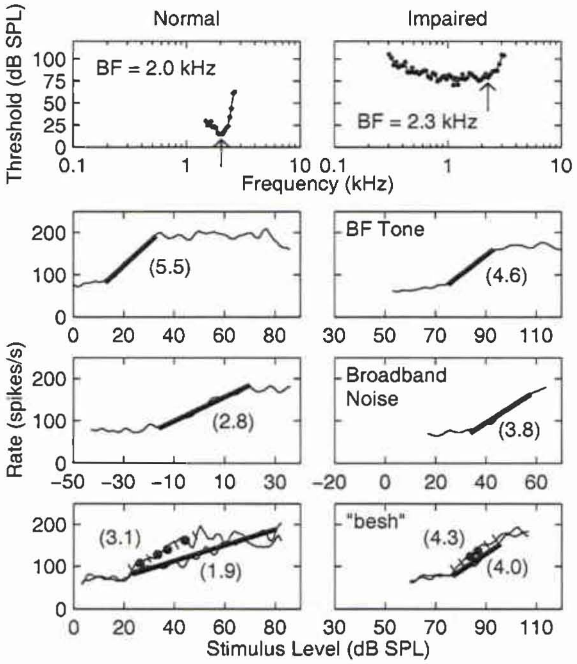


Figure 5.4: Example of an impaired AN fiber for which rate-level functions are similar for all frequencies. The normal hearing cats response is in the left column, the impaired fiber is in the right column. The top row is the tuning curve for the normal and impaired response. The three following rows are the rate-level response curves for the BF Tone, a broadband noise and the CVC "besh", respectively. Thick lines are the linear regression lines and their corresponding slopes are in brackets. In the bottom row, the fricative response (dotted line) and vowel response (solid line) are broken out versus overall vowel level. The normal fiber had a 2.3 kHz BF, 15 dB SPL threshold, 68.6 sp/s spont rate and 3.4 Q_{10}. The impaired fiber had a 2.3 kHz BF, 79 dB SPL threshold, 68.0 sp/s spont rate, and a Q_{10} of 1.0. Figure 6 taken from Heinz & Young [2004].

Figure 5.4 shows how the growth slopes of the impaired response to each stimulus have less deviation than the normal responses. The typical impaired nerve fiber that ellicited this response was broadly tuned with a moderate to severe threshold shift. To check the connection between the AN response modeled and the Heinz & Young [2004] data the three stimuli from figure 5.4 were input into the Bruce et al. [2003]

model. The broadband noise was luckily a frozen noise to reduce randomness. Figure 5.5 can be considered an extension of figure 5.4, with the two columns labelled normal simulated and impaired.

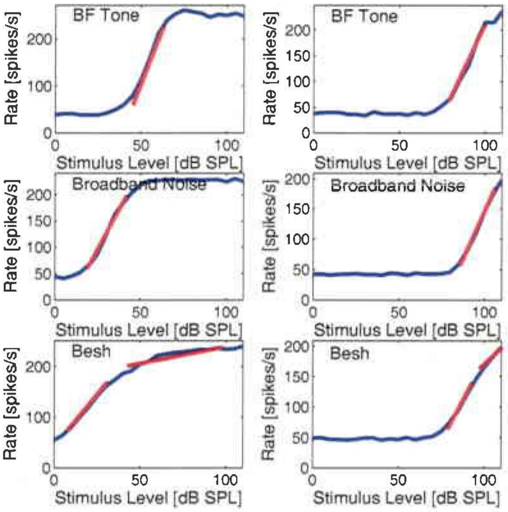


Figure 5.5: Example of the simulated impaired AN fiber rate-level functions. The three rows are the rate-level response curves for the BF Tone, a broadband noise and the CVC "besh", respectively. The normal auditory model produced slopes for the three stimuli of 8.1, 6.0 and 3.9/0.7 (low level slope over high level), while the impaired auditory model produced slopes of 7.4, 5.9 and 4.7/3.7.

Another class of rate-level functions were from nerve fibers that exhibited a growth of response that was shallower than normal. Heinz & Young [2004] says in general these impaired fibers generally still had a tight tuning curve, but exhibited a high threshold shift. This is generally thought to represent somewhat healthy OHC mechanisms, with highly impaired IHC mechanisms. Heinz & Young [2004] go on to say that this type of fiber is not only responsible for shallower responses, but can actually mimic normal rate-level operation. Figure 5.6 shows representative tuning curves and rate-level functions for the normal and impaired fibers that consistently produced shallower functions. The corresponding simulated AN response is in figure 5.7.

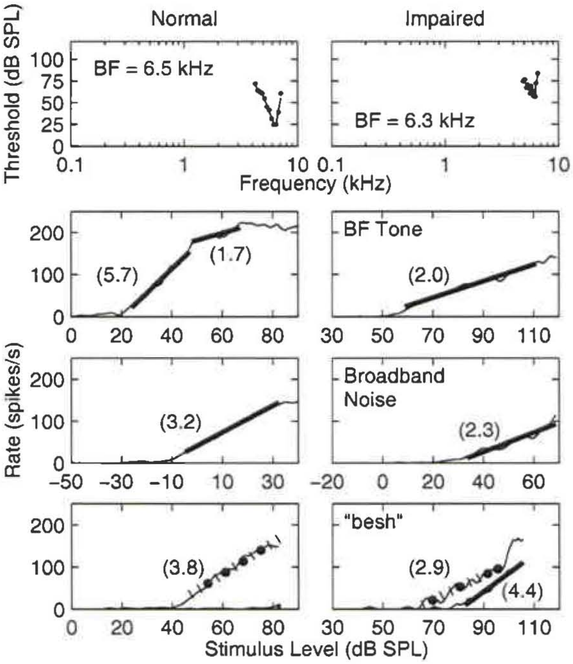


Figure 5.6: Example of an impaired AN fiber for which rate-level functions are shallower than normal. The normal hearing cats response is in the left column, the impaired fiber is in the right column. The top row is the tuning curve for the normal and impaired response. The three following rows are the rate-level response curves for the BF Tone, a broadband noise and the CVC "besh", respectively. Thick lines are the linear regression lines and their corresponding slopes are in brackets. In the bottom row, the fricative response (dotted line) and vowel response (solid line) are broken out versus overall vowel level. The normal fiber had a 6.5 kHz BF, 25 dB SPL threshold, 0.4 sp/s spont rate and 7.5 Q_{10}. The impaired fiber had a 6.3 kHz BF, 57 dB SPL threshold, 0.3 sp/s spont rate, and a Q_{10} of 7.6. Figure 7 taken from Heinz & Young [2004].

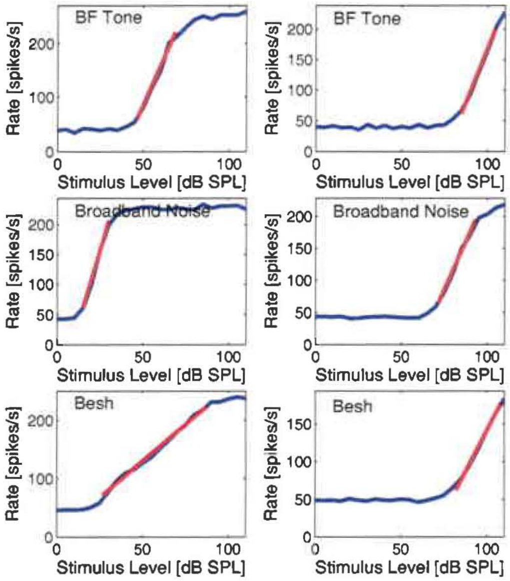


Figure 5.7: Example of the simulated impaired AN fiber rate-level functions in an attempt to produce shallower than normal rate-level curves. The three rows are the rate-level response curves for the BF Tone, a broadband noise and the CVC "besh", respectively. The normal auditory model produced slopes for the three stimuli of 6.4, 9.2 and 2.5, while the impaired auditory model produced slopes of 6.0, 5.6 and 4.0.

Some impaired fibers that were generally shallower than normal also had a distinctive high level extremely steep response. They retained a great deal of dynamic range, for the 20 dB-30 dB range over threshold, then hockey sticked.

The final possible change to rate level curves, were the steeper than normal rate-level curves. Which were largely the result of medium spontaneous rate fibers with severely elevated thresholds and broad tuning. Figure 5.8 shows the tuning curves and steeper rate-level functions. Figure 5.9 shows the corresponding simulated fiber rate-level curve that is steeper then normal.

There are many differences between normal and noise-induced hearing impaired rate-level curves, but to be useful for designing a single hearing aid compression circuit, these differences have to be consistent or their phenomenology must be understood. Table 5.18 shows the problems faced in dealing with returning a proper rate-level curve to the broadly tuned, hearing impaired AN.

Stimulus	Slope	Normal	Mild	Moderate/Severe
BF Tone	Low-level	7.1	5.3	5.4
	High-level	2.2	2.2	1.7
Broadband Noise	Low-level	5.0	4.7	4.5
	High-level	1.6	2.5	–
	Relative	0.7	0.9	0.8
Vowel	Low-level	3.6	3.6	4.9
	High-level	1.6	2.6	–
	Relative	0.5	0.7	0.9

Table 5.18: Summary of the effects of noise induced impairment on the response of AN rate-level

The mild and moderate/severe responses are statistically shallower than normal at low levels. The vowel response was statistically steeper for the moderate/severe AN, and in the high level and relative slopes for the broadband noise condition for

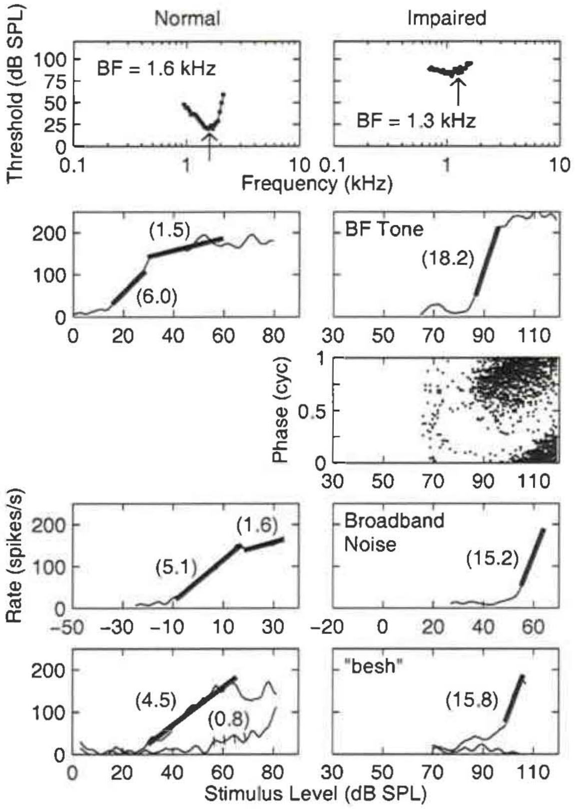


Figure 5.8: Example of an impaired AN fiber for which rate-level functions are steeper than normal. The normal hearing cats response is in the left column, the impaired fiber is in the right column. The top row is the tuning curve for the normal and impaired response. The three following rows are the rate-level response curves for the BF Tone, a broadband noise and the CVC "besh", respectively. Thick lines are the linear regression lines and their corresponding slopes are in brackets. In the bottom row, the fricative response (dotted line) and vowel response (solid line) are broken out versus overall vowel level. The normal fiber had a 1.6 kHz BF, 21 dB SPL threshold, 8.1 sp/s spont rate and 2.6 Q_{10}. The impaired fiber had a 1.3 kHz BF, 83 dB SPL threshold, 17.1 sp/s spont rate, and the Q_{10} was undefined. Figure 9 taken from Heinz & Young [2004].

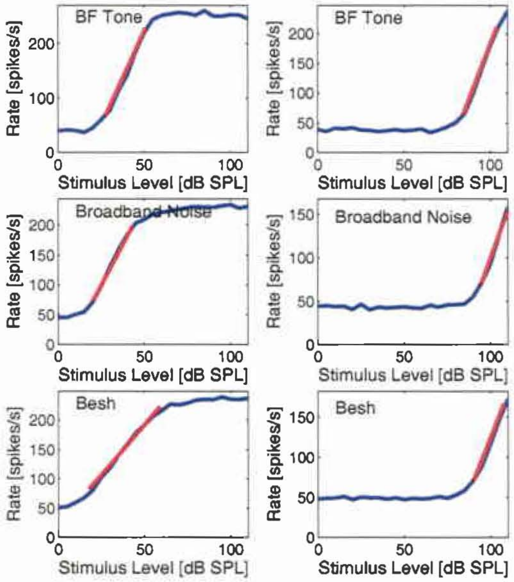


Figure 5.9: Example of the simulated impaired AN fiber rate-level functions in an attempt to produce steeper than normal rate-level curves. The three rows are the rate-level response curves for the BF Tone, a broadband noise and the CVC "besh", respectively. The normal auditory model produced slopes for the three stimuli of 7.1, 5.3 and 3.4, while the impaired auditory model produced slopes of 7.4, 5.7 and 5.5.

mildly impaired fibers. The mild loss also produced a steeper relative slope.

Explanations of several effects of sensorineural hearing loss (see Section 2.2) rely on steeper rate-level slopes after impairment. Heinz & Young [2004] show that rate-level curves are often shallower, and suggest a phenomenology of IHC impairment to account for this unexpected result. Because of the importance of the IHC transfer characteristics in determining rate coding effects, it seems vital to account for IHC impairment in interpreting sensorineural impairment.

Previously, it has been discussed that loudness is a function of the total rate of AN activity, if so, this interpretation has increased difficulty in dealing with the inconsistencies of rate-level curves with hearing impairment. Kiang et al. [1970] suggested that recruitment could result from the more rapid spread of excitation, but Heinz & Young [2004] says that broadened tuning does not have an appreciable effect on total AN activity. There are several other hypotheses for loudness recruitment, including more central effects; auditory brain neurons show recruitment effects, possibly stemming from synaptic plasticity effects [Popelar et al., 1987], and there is also the possibility that loudness is not a strong function of total average discharge rate.

One thing that is certain is that compression algorithms in hearing aids cannot simply rectify the BM compression of the impaired cochlea. This is an advantage for the machine learning framework previously presented because it already has embedded dependance on OHC and IHC impairment.

Just hair cell impairment speaks to the complexity of re-establishing normal amplitude coding in the impaired ear. Even without considering stria vascularis atrophy, or higher level cognitive effects, including loss of efferent connections to the cochlea from auditory brain centers, there are a wide range of responses just in looking at stimuli and hair cell impairments. One common hypothesis on the reduction to the range of slopes brought on by hearing impairment, is that feedback mechanisms mediate a response that matches rate-level growth to the acoustic environment. From the rate-level curves in figures 5.5, 5.7 and 5.9 this might not necessarily be the case, the Bruce et al. [2003] model does not include these feedback effects. Much of the variance between stimuli can be described by the difference between acoustic properties and hair cell impairment. To further illustrate this, the slopes for a range of inner and outer hair cell impairments were calculated for the three stimuli. Figures 5.10, 5.11 and 5.12 are graphs for the tone, noise and "besh" stimuli, respectively.

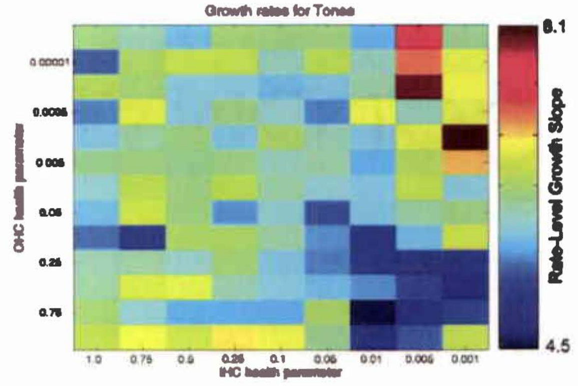


Figure 5.10: The rate-level growth slopes for a range of outer and inner hair cell losses and the tone stimulus. The regression plane slopes heavily uphill in the direction of outer hair cell loss, while increasing inner hair cell loss produces a greater range of responses, but not a consistent slope.

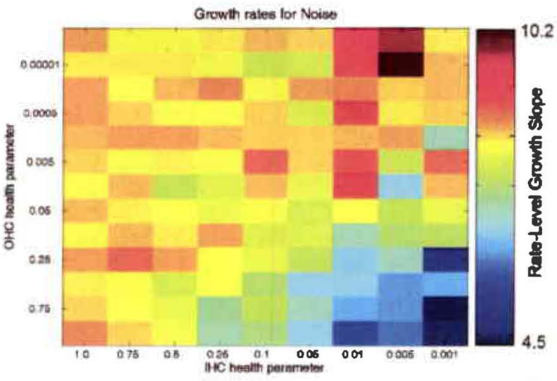


Figure 5.11: The rate-level growth slopes for a range of outer and inner hair cell losses and the broadband noise stimulus. The regression plane slopes similarly to the tone stimulus, but shows a greater range of slopes.

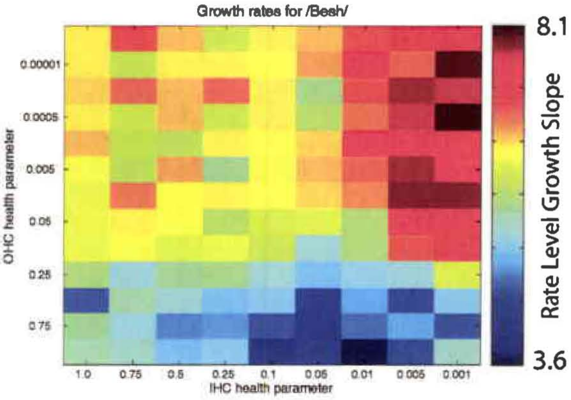


Figure 5.12: The rate-level growth slopes for a range of outer and inner hair cell losses and the "Besh" stimulus. The regression plane slopes much more for IHC damage than the other stimuli, as well, there is much less difference in the amount of variation among the impairments.

The common hypothesis that sensorineural impairment causes steeper rate-level growth curves does hold on average. It must be considered, that the average response is not necessarily the best way to fit a hearing aid. It is speculated that the average pathology of sensorineural impairment hinges on outer hair cell damage. Figures 5.13, 5.14 and 5.15 plot the resulting slopes for the tone, noise and "besh" stimuli, respectively, but only for OHC damage.

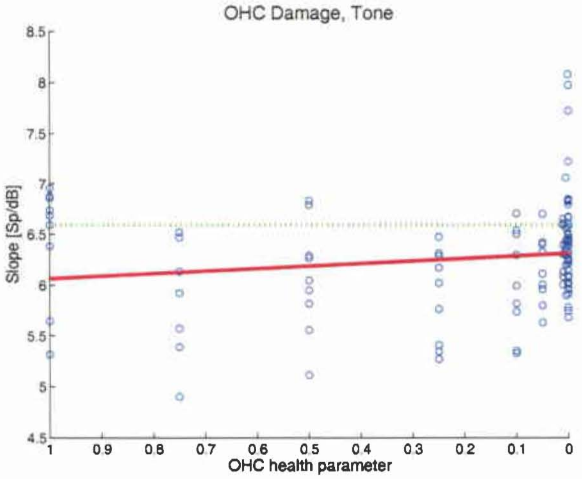


Figure 5.13: The slopes of rate-level growth for the tone stimulus, plotted versus OHC impairment. Slopes generally steepen versus increasing impairment, much in accord with accepted wisdom. More impairment results in a wider range of slopes. The green dotted line is the healthy cochlea slope, the red line is the linear regression line for all the different slopes and impairments.

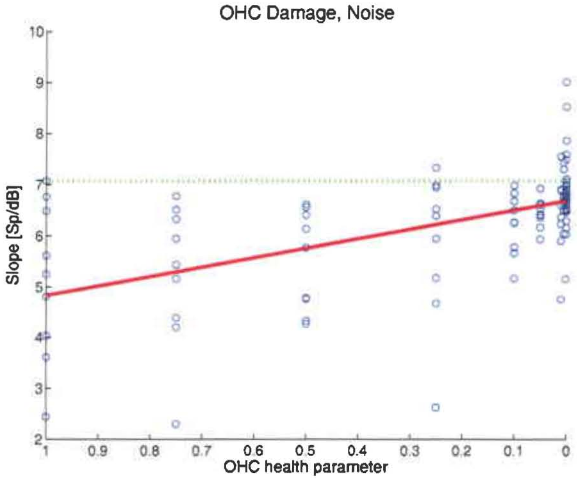


Figure 5.14: The slopes of rate-level growth for the noise stimulus, plotted versus OHC impairment. Slopes generally steepen versus increasing impairment, much in accord with accepted wisdom. More impairment results in a wider range of slopes. The green dotted line is the healthy cochlea slope, the red line is the linear regression line for all the different slopes and impairments.

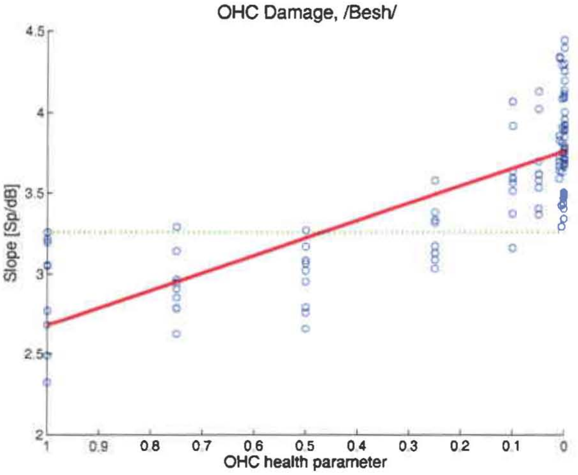


Figure 5.15: The slopes of rate-level growth for the "besh" stimulus, plotted versus OHC impairment. Slopes generally steepen versus increasing impairment, much in accord with accepted wisdom. More impairment results in a wider range of slopes. The green dotted line is the healthy cochlea slope, the red line is the linear regression line for all the different slopes and impairments.

Looking at the regression lines, all stimuli see steeper rate-level curves, but this is not necessarily the rule when IHC impairment is figured in. Several points are under the normal hearing slope. The relative increase in steepness for the broadband noise and the /besh/ stimuli are the same, and are both steeper than the tone increase. Again, OHC loss is often generalized as the cause of sensorineural hearing loss, and in turn, loudness recruitment from steeper rate level curves. As can be seen in the figures 5.16, 5.17 and 5.18 the resulting slopes for the tone, noise and "besh" stimuli, respectively, for IHC damage show no increase in steepness.

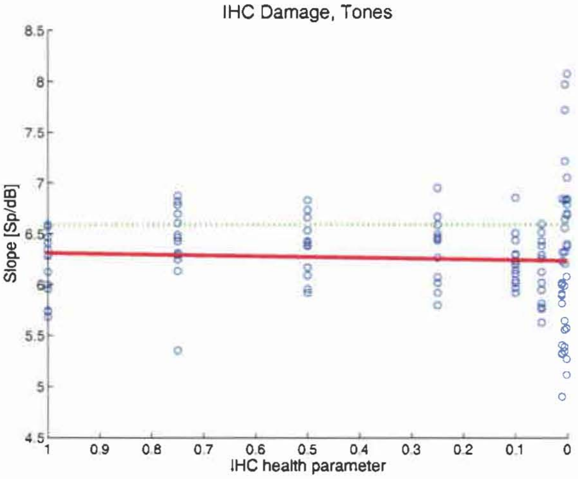


Figure 5.16: The slopes of rate-level growth for the tone stimulus, plotted versus IHC impairment. Slopes stay the same with increasing impairment, much in accord with accepted wisdom. More impairment results in a wider range of slopes. The green dotted line is the healthy cochlea slope, the red line is the linear regression line for all the different slopes and impairments.

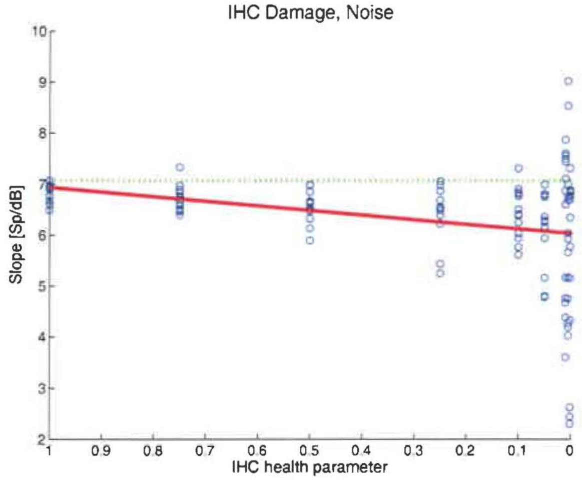


Figure 5.17: The slopes of rate-level growth for the noise stimulus, plotted versus IHC impairment. Slopes stay the same with increasing impairment, much in accord with accepted wisdom. More impairment results in a wider range of slopes. The green dotted line is the healthy cochlea slope, the red line is the linear regression line for all the different slopes and impairments.

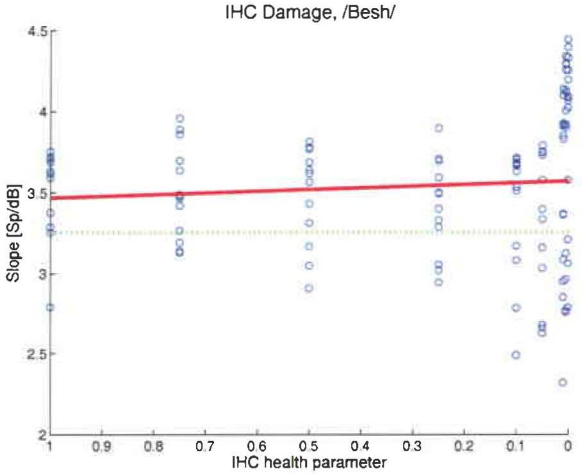


Figure 5.18: The slopes of rate-level growth for the "besh" stimulus, plotted versus IHC impairment. Slopes slightly steepen with increasing impairment, not predicted by presented theories. More impairment results in a wider range of slopes. The green dotted line is the healthy cochlea slope, the red line is the linear regression line for all the different slopes and impairments.

Interestingly, the "besh" stimulus is the only stimulus that shows appreciable gains in steepness. The appreciation in slope is five times less than seen with OHC impairment, and may arise from the statistical nature of the synapse model used or the automated computation of slope.

The steepening of the rate-level slopes used so often to explain loudness recruitment does not hold for all hearing loss pathologies. It would be a mistake to build hearing aid technologies on using the mean results from such a varied population. Noise induced hearing impairment generally damages OHCs more than IHCs, so the mean results do hold with the above simulation, but many novel insights are also brought to light. In general hair cell pathology can account for a wide range of rate level curves seen across different stimuli, without resorting to cognitive function, or brain feedback. A hearing aid compression circuit may need to understand the depths of this pathology to properly restore level perception. This would be further confounded with changes in a population of neurons, including the changes in threshold and spontaneous rate that are incumbent in sensorineural impairment Liberman & Dodds [1984b]. The changes in amplitude coding brought on by stimulus types may also drive hearing aids that intelligently make use of the acoustic ecology.

5.1.1 Simulation Experiments

To extend the above description of the complexity in dealing with loss of cochlear compression, there is an obvious need for a deeper understanding of how the rate-level function changes under sensorineural impairment. The first experiment to derive optimal gains over input powers for various stimuli used the TIMIT database [TIMIT, 1990]. The simulated loss profile used for this test is #4 from figure 4.1. The goal was to see what a complex loss of both IHC and OHC required to return the normal

AN rate response, and could a general gain rule be made for a specific loss.

To do this the acoustic input was separated into 25 bands. There were 10 bands of equal bandwidth under 1000 Hz, and then 15 bands spaced $1/6^{th}$ an octave for frequencies above that. The AN response was derived for a fiber with a BF at the center of each of the acoustic bands. The speech was normalized to 65 dB SPL before the Wiener & Ross [1946] outer ear transfer function was applied.

The initial gain before the impaired model was NAL-RP, as described in section 4.1. The optimal gains in the 25 channels were derived by adjusting the gains by 0.5 % over 500 iterations. Optimality was defined as producing the closest correspondence between the normal and impaired mean firing rates.

The mean firing rates were calculated by removing phase locking effects through taking the envelope of the AN response. The envelope was calculated through the linear interpolation between the peaks brought on by synchrony. The mean rate is then the average of this vector. While not precisely the mean rate, because this removes the off periods brought on by the IHC gating channel mechanism, it is monotonically related to the mean rate. That is, when comparing two mean rate vectors, the real mean rate of the discharge rate vector is larger when this measure is larger.

The gain was not calculated for one specific sentence, but rather for each of the phonemic segments in each sentence. TIMIT [1990] has been hand segmented into 63 phonemic categories that correspond to steady state spectra, and for the purposes outlined here correspond to approximately steady spectra.

The calculation of the mean rate then has two major sources of error. The first is that the response during a phonemic interval is somewhat altered by the acoustic signal preceding it, or the forward masking effect. The second is adaptation, especially fast onset adaptation biases the mean positively for the normal response, but adaptation is different in the damaged cochlea. This effect has a time constant of roughly 2 ms. These two sources of error are not important in the longer phonemic intervals.

The above experimental setup produces instantaneous gain changes at phonemic boundaries. To a listener's auditory system this would result in large pops, clicks, or overall frequency distortions. This experiment was designed to simulate the amplitude coding characteristics of the auditory system; temporal coding will be discussed in section 5.3.

For the TIMIT, "SX" training data, each phone segment has it's SPL calculated for each frequency band and then the gain in dB (versus NAL-RP, ie. if the optimized gain is equal to the NAL-RP prescription then gain = 0 dB) is plotted versus the input SPL. Each point in Figure 5.19 represents the gain for one frequency channel for one phonemic segment.

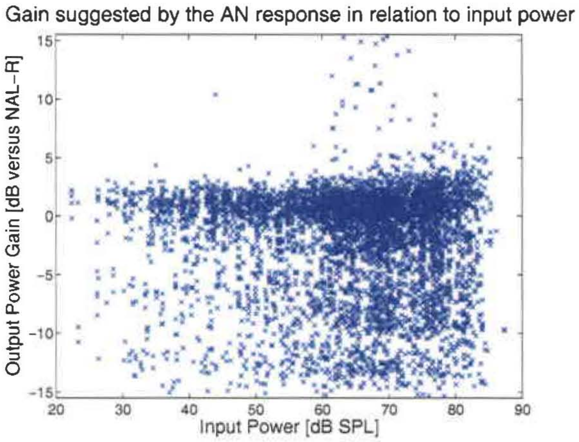


Figure 5.19: Scatter of optimal gains in different channels for the TIMIT "SX" database.

Obviously from the above graph, the mode response is a little over 0 dB, but the mean response is zero. An interesting result is the spread in the data. It seems that there is not a strong correlation between input power and optimal gain required to reestablish normal firing rates in a complex but completely realistic hearing loss.

If there is little correlation between short term input powers or syllabic SPL and required gain, what should hearing aid circuits use to drive their nonlinear processing? Is this the reason that there is little or no intelligibility gain from these circuits? To try to answer this very important question, the data from figure 5.19 can be broken out by frequency to see if there is a consistency between the spread of desired gains and the type of loss. The histograms for gains over NAL-RP versus input SPL were calculated for 25 frequency channels. To keep some level of compactness, only the channels centered on 250, 1050, 2200 and 2750 are shown in figures 5.20, 5.21, 5.22 and 5.23, respectively.

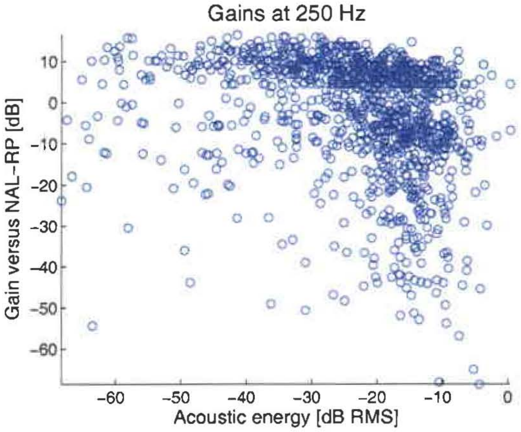


Figure 5.20: Gain over the standard NAL-RP prescription versus input power reported in DB RMS when the input signal is 65 dB SPL. This is the gains in a single frequency channel centered at approximately 250 Hz.

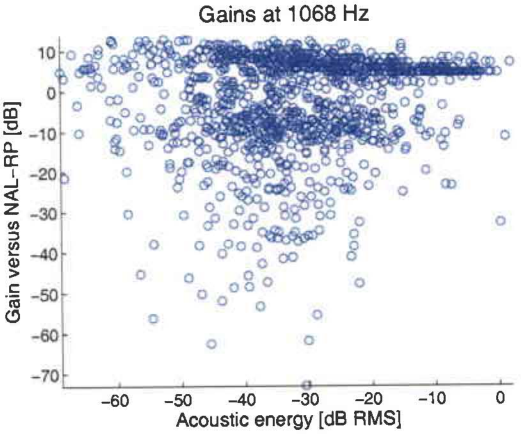


Figure 5.21: Gain over the standard NAL-RP prescription versus input power reported in DB RMS when the input signal is 65 dB SPL. This is the gains in a single frequency channel centered at approximately 1050 Hz.

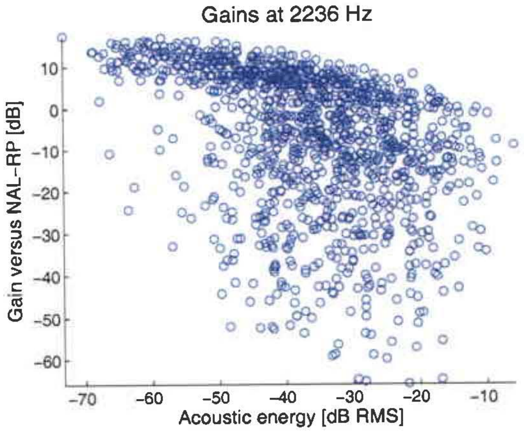


Figure 5.22: Gain over the standard NAL-RP prescription versus input power reported in DB RMS when the input signal is 65 dB SPL. This is the gains in a single frequency channel centered at approximately 2200 Hz.

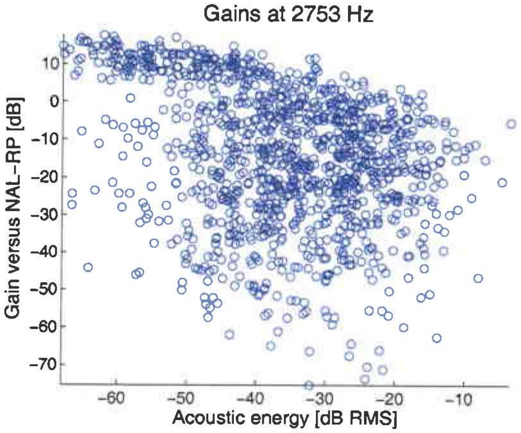


Figure 5.23: Gain over the standard NAL-RP prescription versus input power reported in DB RMS when the input signal is 65 dB SPL. This is the gains in a single frequency channel centered at approximately 2750 Hz.

There is still quite a spread in any single frequency channel. The marginal distributions, dependent upon input power in a channel are given in figures 5.24 - 5.27. Obviously, the previous scatter plots describe a complex stochastic process that might have a more discernable mode.

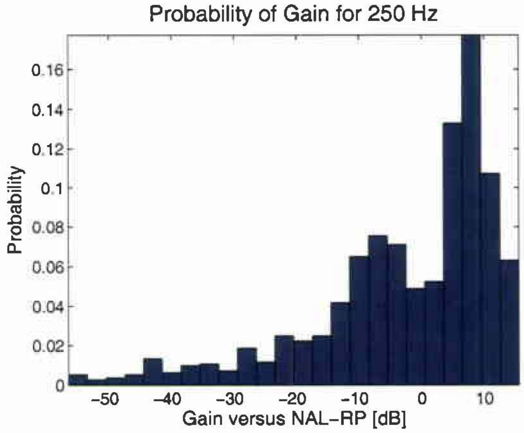


Figure 5.24: Marginal distribution of the gains over the standard NAL-RP prescription versus input power reported in DB RMS when the input signal is 65 dB SPL. This is for the frequency channel centered at approximately 250 Hz.

There is an interesting bimodality in these curves: one distribution looks tightly formed and centered with a gain above the NAL-RP prescription and the other centered well below the NAL-RP prescription and increasingly broad as frequency increases. The process making the broadly tuned mode was conjectured to be tied to the spread of masking; all the points from this distribution are attenuating, so it was conjectured that other frequency channels were responsible for excitation spread. Since the gain optimization for a channel was tied to the difference between overall excitation and the difference between the normal and impaired rate in that channel, spread of excitation would result in the gain being decreased in a channel without effecting the overall excitation. Because of the randomness of speech and the optimization process, negative gains in a channel with spread of excitation would become

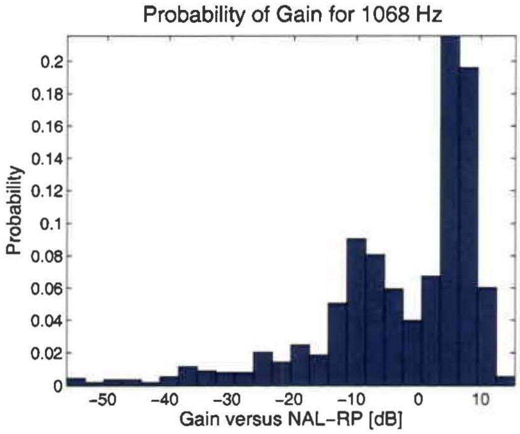


Figure 5.25: Marginal distribution of the gains over the standard NAL-RP prescription versus input power reported in DB RMS when the input signal is 65 dB SPL. This is for the frequency channel centered at approximately 1050 Hz.

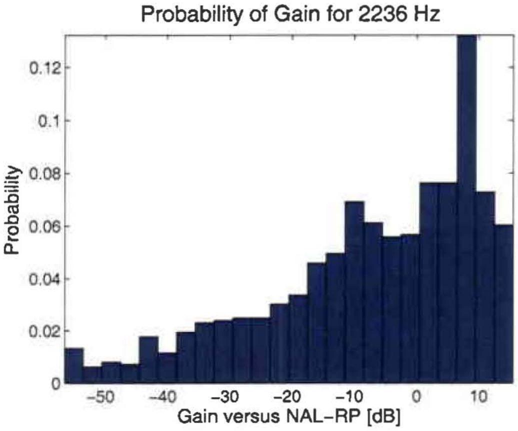


Figure 5.26: Marginal distribution of the gains over the standard NAL-RP prescription versus input power reported in DB RMS when the input signal is 65 dB SPL. This is for the frequency channel centered at approximately 2200 Hz.

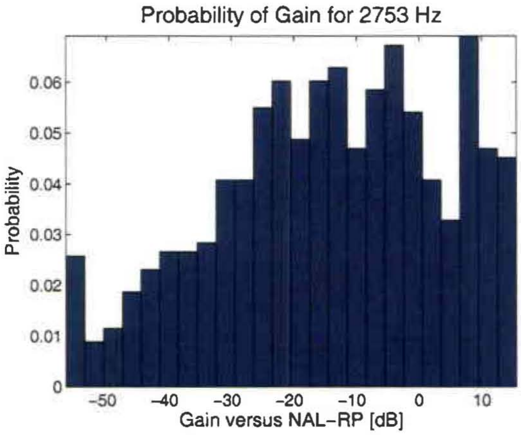


Figure 5.27: Marginal distribution of the gains over the standard NAL-RP prescription versus input power reported in DB RMS when the input signal is 65 dB SPL. This is for the frequency channel centered at approximately 2750 Hz.

increasingly random. Combined with the asymmetry in the spread of masking this means that high frequency channels are more likely to be captured by off channel signals.

The second distribution is a little more informative, it has much less spread at all frequencies, and correlates input and output power. Figures 5.28 to 5.31 plot the points that required gains above NAL-RP to equalize loudness.

This was the first machine learning experiment that successfully calculated compression characteristics, as the increasing compression rates match empirical data on intelligibility quite well. The compression ratios mimic empirical data, the larger the impairment, the higher the compression ratio is. Also, the compression ratio does not begin to elevate until well after the knee point of the loss profile. A second point is that compression is not the principle mode for the excitation optimization, attenuation is. Typically, the impaired auditory system sees too much activity: intelligent attenuation of frequency channels improves the neural representation and could possibly improve intelligibility.

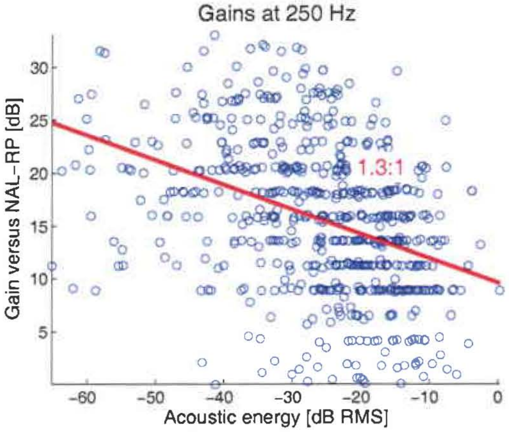


Figure 5.28: For the frequency channel centered at approximately 250 Hz there was a light compression where gain shrank at 1.3 dB per dB of input power, over the entire range of speech tested.

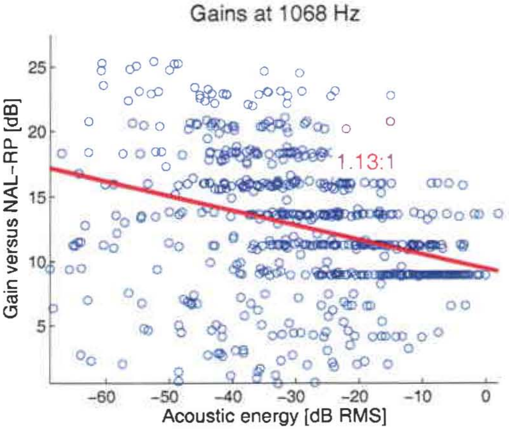


Figure 5.29: For the frequency channel centered at approximately 1050 Hz there was a light compression where gain shrank at 1.13 dB per dB of input power, over the entire range of speech tested.

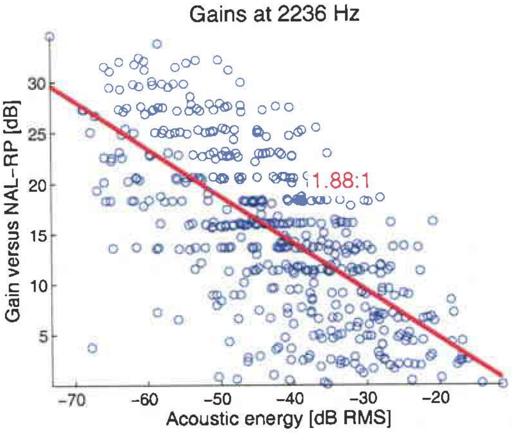


Figure 5.30: For the frequency channel centered at approximately 2200 Hz there was a light compression where gain shrank at 1.88 dB per dB of input power, over the entire range of speech tested.

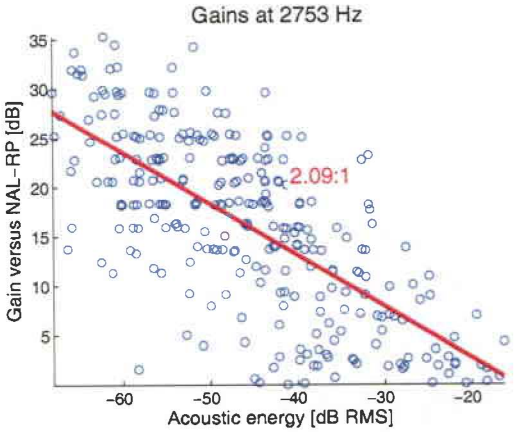


Figure 5.31: For the frequency channel centered at approximately 2750 Hz there was a light compression where gain shrank at 2.09 dB per dB of input power, over the entire range of speech tested.

5.2 Differences in Suppression Responses

The second audio coding dimension discussed in this chapter will be how the auditory periphery codes frequency[2]. While the many processes in the cochlea all overlap and have some importance in the frequency domain, the particular nonlinearity focussed on in this section is suppression. Most, if not all, hearing aid algorithms consider the loss of the suppression mechanism through sensorineural impairment to be negligible, while it is the assertion of this dissertation that not only is suppression keenly important, its effect on the frequency representation in the auditory brain can be clearly delineated through audio coding foundations.

Suppression is often glossed over because it changes the gain so little in comparison to compression. While compression can induce 50 to 60 dB of swing in cochlear gain, suppression accounts for small bands to either side of a suppressor and is often only a few dB. Figure 5.32 shows the limited amount of frequencies and amplitudes that are affected by the suppression mechanism.

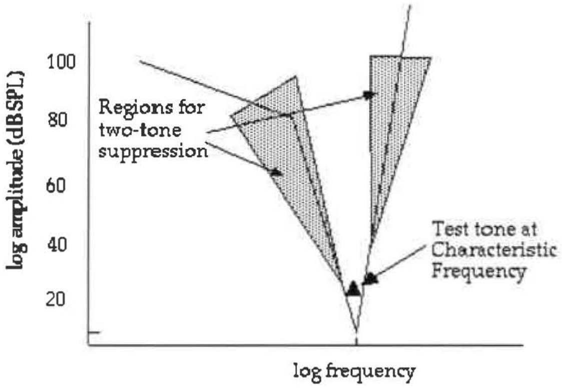

log amplitude (dB SPL)
100
80
60
40
20
Regions for two-tone suppression
Test tone at Characteristic Frequency
log frequency

Figure 5.32: The highlighted areas are the range of frequencies and levels that are suppressed by a tone at the peak.

[2] This section is based on Bondy & Bruce [2004a].

In the simplest terms possible, as the suppressor tone's level increases, it pushes down the AN response of neighbouring fibers. This describes a decorrelative process, which begs to understand suppression under the audio coding paradigm. This is an expansion of the introductory material on information theoretic coding given in section 5.1. It has previously been touched upon that the input-output relationship of the biological coding mechanism is based on function (hypothesized for high and low spont rate fibers) or theoretically defined by the statistics of its environment; optimally corresponding to the cumulative distribution function [Barlow, 1961]. The extension of this is that when coding the environment, the statistical redundancies should be minimized across channels. In the case of the cochlea, the frequency domain is separated into different AN fibers, each being a different frequency channel, often called the place coding theory of the cochlea.

Getting back to the impact on coding done by the cochlea, each IHC attaches to the auditory nerve and which in turn innervates the auditory brainstem. There are conservatively, 4000 IHCs with 10 to 20 nerve fibers innervating each one, each spiking up to 250 spikes/second (say on average 90), with each spike having somewhere between 3 to 8 bits of information [Rieke et al., 1997], so playing conservatively that's about 13Mbps (4000x10x90x3) for one ear into the auditory brainstem. This is only about two to three times less then the information throughput of the earliest stage of the vision system. There is a point to be made that if this information is highly redundant, then a lot of the brain's time is spent dealing with repetitive information.

The theory behind the above derivation is that for optimality, the information coded by each filter/frequency channel along the cochlea should be statistically independent. In most mammalian ears, broadly tuned, high frequency responses are located at the basal end of the cochlea, and decrease in frequency and bandwidth by progressing towards the apical end of the cochlea. The actual tradeoff between frequency and bandwidth is sublinear. Wavelet decompositions have linear frequency-bandwidth tradeoffs. In the biological case the lower frequency filters have a lower Q value (center frequency over bandwidth) than the higher frequency filters.

The question is then: does the empirical evidence of BM tuning correspond to the statistically optimal filters derived above? By applying the learning rule from independence maximization to a general filterbank and adapting the filterbank to make the outputs non-redundant and sparse, will the derived filterbank follow the empirical? Lewicki [2002] showed that the optimal filterbank does indeed follow biological data; results are shown in figure 5.33.

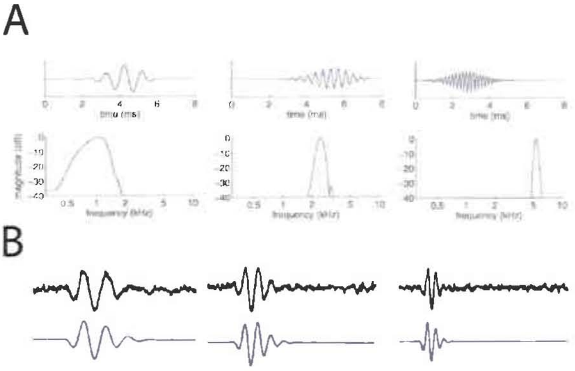


Figure 5.33: Filterbank derived by Lewicki [2002]. In A the top row shows three vectors from the A matrix, underneath them are the corresponding frequency responses. The top row of panel B has recorded AN response, revcor kernels, plotted versus the derived filterbank.

The optimally computed bases have progressively better Q factors as CF increases, matching the empirical measured curves quite well. Lewicki's work closely coincides with empirical data, and thus some statistical basis for evolved mechanism can be conjectured. The filterbank implementation of the auditory periphery may have evolved for redundancy removal, and higher brain centers might then "learn" by independance maximization approaches. It is clear that there is temporal and frequency localization for a filterbank that optimally encodes speech, which matches the biological system.

5.2.1 Simulation Experiments

How does sensorineural impairment affect the optimality of the filterbank; and how does suppression directly effect the coding of the auditory environment? The above derivation motivates looking at the loss of independence in place coding as central to understanding aspects of sensorineural impairment: a simpler form is needed to be usable however. In the simplest form, information is maximized if the joint distribution can be factorized by the constituent distributions,

$$P\left(C_{XX}\right) = \prod_{i=1}^{n} P\left(x_i\right) \tag{5.32}$$

where x_i is the discharge rate over time for the i^{th} AN fiber. Using the entire acoustic ensemble and each AN fibers distribution is intractable, independence would be impossible to calculate. Instead, one can look to see how the normal and the sensorineural impaired auditory periphery decorrelates the activity on the AN. This is much simpler, because decorrelation can be achieved with a simple linear transformation.

Correlations are captured by the auto-covariance matrix, it is a simple 2^{nd} order statistic that along with the mean produces sufficient statistics for gaussian distributions. While the AN rates are not Gaussian in speech, the covariance matrix is still the best estimator of the volume of the distribution. Equation 5.33 is the covariance calculation, with X representing the discharge rates over time vectors for an ensemble of fibers.

$$C_{XX} = E\left[(X-\mu)(X-\mu)^{T}\right] \tag{5.33}$$

Equation 5.33 is a simple transformation of the autocorrelation matrix, which contains the same information, or more precisely,

$$R_{XX}(i,j) = \frac{C_{XX}(i,j)}{\sqrt{C_{XX}(i,i)\,C_{XX}(j,j)}} \tag{5.34}$$

Figure 5.34 is the covariance matrix of a syllable when it is split into 25 separate frequency channels, with filters mimicking the distribution and bandwidth of the auditory system. There were 10 bands of equal bandwidth under 1000 Hz, and then 15 bands spaced $1/6^{th}$ an octave for frequencies above that. The AN response was derived for a fiber with a BF at the center of each of the acoustic bands. The speech was normalized to 65 dB SPL before the Wiener & Ross [1946] outer ear transfer function was applied. The van Son et al. [2001] corpus was used as a preliminary test set.

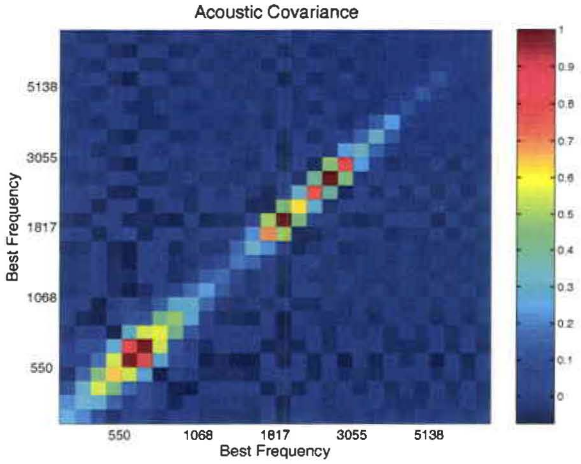


Figure 5.34: An acoustic covariance matrix for the syllable "bak".

It's obvious from Figure 5.34 that a linear filterbank devised to follow the cochlear representation has limited, yet appreciable correlations in adjacent channels. With nonlinear suppression, the covariance matrix produced with the discharge rates in the normal auditory periphery shows a reduction of the adjacent channel correlations at low frequencies, as seen in figure 5.35.

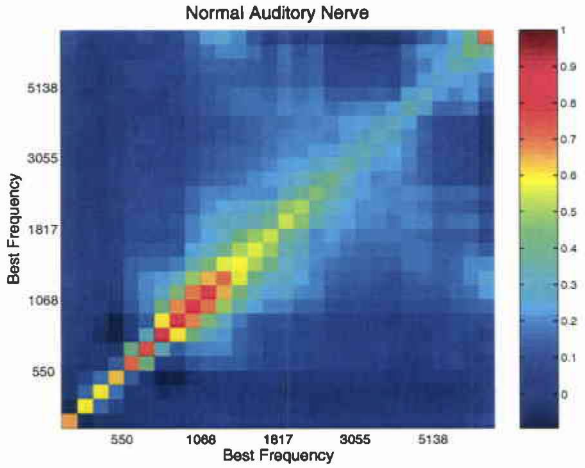


Figure 5.35: The discharge rate covariance matrix for the syllable "bak" from a normal auditory model.

From figure 5.35 there are across channel correlations from the production mechanisms and the timing of speech. A normal hearing person can reduce these, largely with the decorrelative, suppression mechanism, while the hearing impaired person must deal with correlations, such as those in figure 5.36.

While some of these correlations come from broadened tuning of neural fibers, the adjacent channels are now almost representing the same information, their correlation factors are approaching one. This is going to be important on two fronts. The first, is that if the impaired auditory system is encoding the same information on multiple channels with the same rate of information, then this redundancy reduces the salient information that higher brain centers can use for processing. The second, more interesting problem, stems from the consequences from Hebbian learning. A

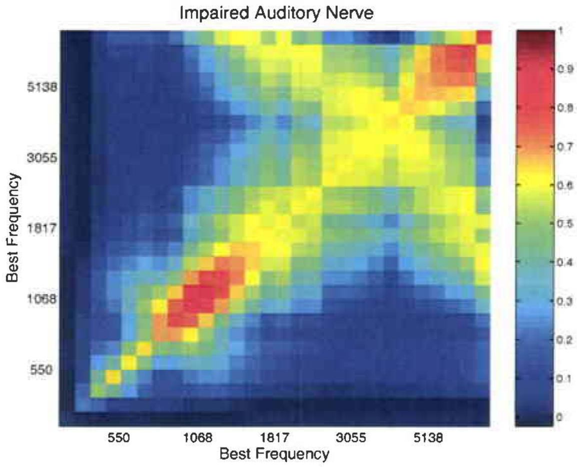


Figure 5.36: The discharge rate covariance matrix for the syllable "bak" from a sensorineurally impaired auditory model.

system built on the assumption of promulgating correlated activity, will therefore pass these meaningless correlations as important.

5.3 Differences in Adaptation Responses

An increase in a stimulus' level results in the discharge rate of an AN fiber quickly incrementing, well over the steady state response. After the initial maximum is reached there is a more gradual decay to the steady state level[3]. This effect is called adaptation and is actually several different processes combined. These processes have different time constants, the largest adaptation rates decay in 2 ms and about 40 ms, the fast and medium adaptation processes (other effects can be several seconds). On the flip side, a decrease in stimulus level results in the AN response becoming greatly depressed, much less then the steady state response for about the same time periods. Since the AN discharge rate is a positive only count, this leads to a asymmetry in the

[3] This section is based on Bondy & Bruce [2005].

onset and offset response, effectively the off or decrement response inhibits firing for an extended amount of time [Nelson & Carney, 2004].

There is some research suggesting adaptation arises at the IHC synapse. Some data does not show nonlinear growth and decay of the membrane potential leading to the Westerman & Smith [1988] or "three pool" adaptation model used in Bruce et al. [2003]. Nelson & Carney [2004] later improved on the model to take into account the asymmetry in increment and decrement response. Interestingly though, many mammalian species have now been shown to have adaptation processes intrinsic to the IHC, prior to the synapse [Choe et al., 1998; Manley et al., 2001; Fettiplace & Ricci, 2003]. And while the auditory modelling done by Bruce et al. [2003] relies on the synapse only adaptation, in reality there is a change to membrane potential incumbent upon sensorineural impairment. The hair cell gain function in the normal auditory model is a saturating nonlinearity that with damage becomes more linear, thus empirical data points to the IHC having some gain function. This in turn affects adaptation by reducing the size of a differential that is the input into the three pool adaptation model. Less differential means less adaptation; adaptation is a function of both levels to either side of the discontinuity, as well as frequency.

Adaptation has been suggested as an enhancement to the neural representation of rapid intensity transients; those that are perceptually important in speech and music. Additionally, this chapter makes a case for the fast adaptation responses to be key to segmenting acoustic cues.

Figure 5.37 is the envelope of the AN response made by linearly interpolating between the peaks in the synchrony response of the 750 Hz simulated fiber. The stimulus was a TIMIT sentence, while only the response for the vowel /a/ is emphasized.

Clearly there is a large difference at the start of and end of this segment. The

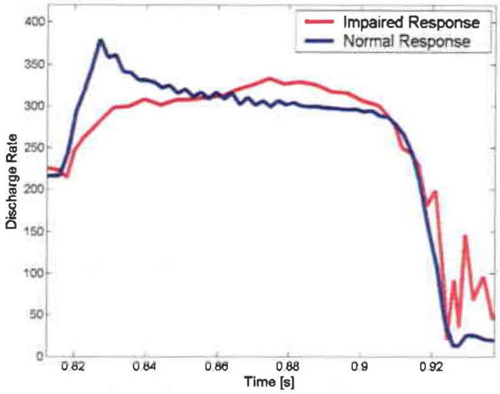


Figure 5.37: The envelope of the AN responses for the normal and impaired auditory periphery from figure 5.3.

normal AN response peaks very quickly after the phonemic segment change and then has very evident rate suppression at the end of the phonemic segment. Contrast this to the more linear response that the impaired response shows, where the peak in the response is actually capturing a variation in spoken level, and is not consistent with the phone boundary.

This chapter focuses on fast onset adaptation as a marker for changes in the acoustic stimulus brought on by the phonetic structure of speech. For unvoiced sounds, or fast consonants, high CF fibers produce the majority of adaptation peaks, and voiced sounds elicit the adaptation spike in lower CF fibers. The novel interpretation is that onset adaptation not only enhances the representation of rapid transients in speech, but it is a key mechanism in the grouping of auditory stimuli in time at different frequencies. The normal ear correlates firings across frequencies for onsets in a very small time window, giving Hebbian cues that can be used to reduce entropy in higher auditory brain centers.

Gockel et al. [2003] produced an interesting insight into how important it is for an auditory stimulus to have its frequency components "lined up" in time, or that its first wavefront is "in phase". They matched the loudness and masking response for three stimuli with varying degrees of phase alignment. The base stimulus was a harmonic tone complex with the various frequency components added in phase (CPH). This produced a high trough to crest ratio. The next stimulus was again a harmonic tone complex, yet this time with the various frequency components added with random phase (RPH). This produced a much smaller trough to crest. The final stimulus was noise, with the same bandwidth as both tone complexes (Noise).

5.3.1 Fast Adaptation on Normal ANs

Figure 5.38 highlights the results of a loudness matching experiment between the three stimuli. The tone complexes had a fundamental frequency of 62.5 Hz, and were filtered to between the 10^{th} harmonic (625 Hz) and 5 kHz. In this experiment from Gockel et al. [2003], the control stimulus was played, then one of the remaining two classes was played. The second class was adjusted in RMS level until the subject concluded that each stimulus had equal loudness. Figure 5.38 plots the difference in RMS between the stimulus at equal loudness.

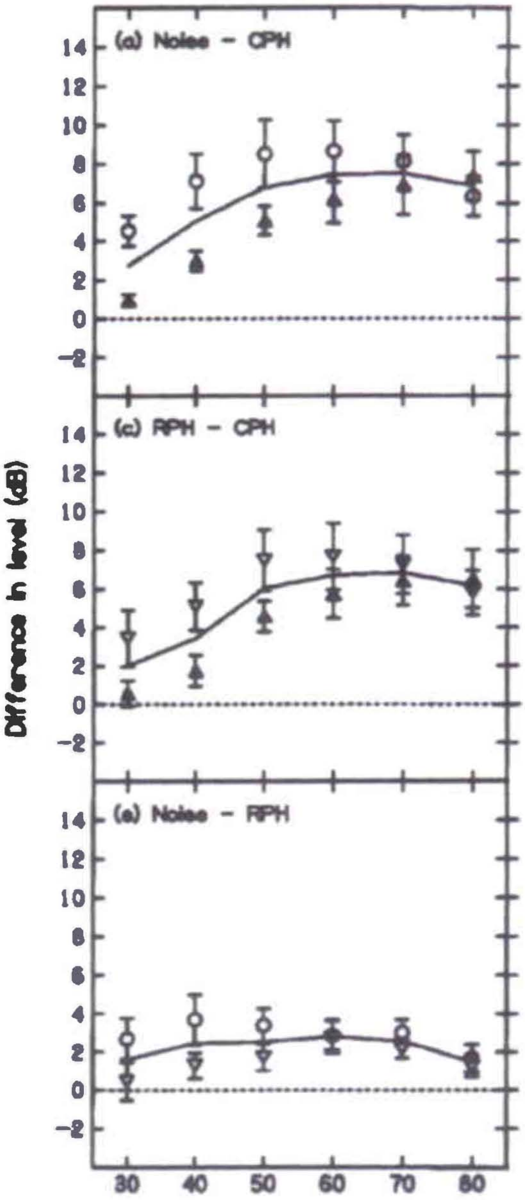


Figure 5.38: Figure 2 from Gockel et al. [2003]. Each plot is the dB difference between the RMS levels of a stimulus pair versus the control level. The first stimulus in the pairing is the one that needed gain. Open circles, upper pointing triangles and lower pointing triangles represent when the noise, CPH, and RPH was varied in level, respectively. The solid line is the average of the two control cases for each stimuli pair.

At all control levels, subjects were biased towards setting the variable stimulus at a level higher than required for equal loudness. This is typical of loudness matching experiments, and this bias decreased with increasing level. Interestingly, for the same RMS level, the CPH stimulus was louder then both the Noise and RPH stimuli. In fact, the RPH and Noise stimuli show close proximity, only needing about 2 dB gain in the Noise stimulus. In general the CPH stimulus is much louder at equivalent RMS levels than the RPH and Noise stimuli, while the RPH and Noise stimuli are about the same.

While it is not news that the average loudness model is inaccurate, the amount of fine timing information that affects loudness is. Another newsworthy illustration that came out of Gockel et al. [2003] was the differences in forward masking the three stimuli showed. Most masking models are predicated on louder stimuli producing more masking.

Figure 5.39 sums up a forward masking experiment carried out in Gockel et al. [2003]. In the experiment, a 208 ms masker of one of the three stimuli was used to forward mask a tone pulse. The tone was chosen from the set of 702, 1114, 1768, 2806 and 4454. Average forward masking for the RPH and Noise stimuli was approximately equal, showing similarities to the loudness experiment. Not consistent with the loudness experiment was that the CPH stimulus was a less effective masker then either the Noise or RPH stimulus.

Obviously, there is an enormous deficit in the ability of the CPH stimuli to mask a signal. Strangely, the loudest stimulus is the weakest masker. This may mimic sensorineural hearing loss, which shows unusual loudness growth that cannot be accounted for with excitation modeling, and intrinsic masking whose psychophysics quantities are well in excess of excitation modeling. It may be that these models do

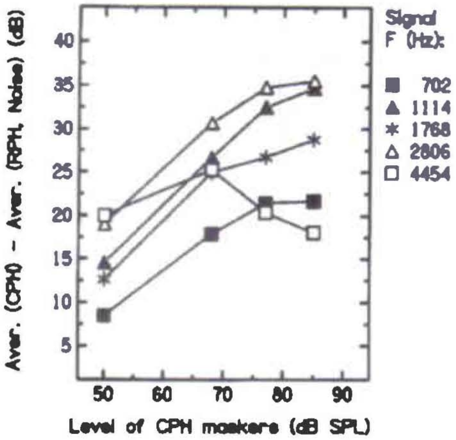


Figure 5.39: Figure 6 from Gockel et al. [2003]. The average gain that had to be applied to the CPH masker to make it as effective as the Noise or RPH maskers plotted versus the necessary level of the CPH masker. The five lines represent the five masked tones pulses from the legend.

not take into the temporal qualities of the auditory system. Specifically most excitation experiments have been done with the linearized Patterson et al. [1988] auditory model, a model without the active gain and segmentation mechanism of adaptation.

5.3.2 Simulation Experiments

What does happen when adaptation is taken into account with the three stimuli from Gockel et al. [2003]? And, is there a connection between these classes of stimuli and the problems associated with sensorineural hearing loss? To answer the first question, one begins with looking at the particulars of the stimuli. Figures 5.40, 5.41 and 5.42 are the amplitude-time graphs of the CPH, RPH and Noise stimuli, respectively.

Clearly, the CPH stimulus has much more temporal information than the other two, but the RPH stimulus also can be considered to have more than the Noise

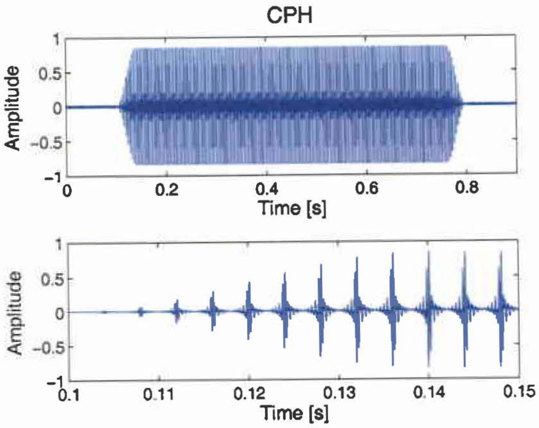


Figure 5.40: The CPH stimulus has a very large trough to crest because of the harmonic periodicity.

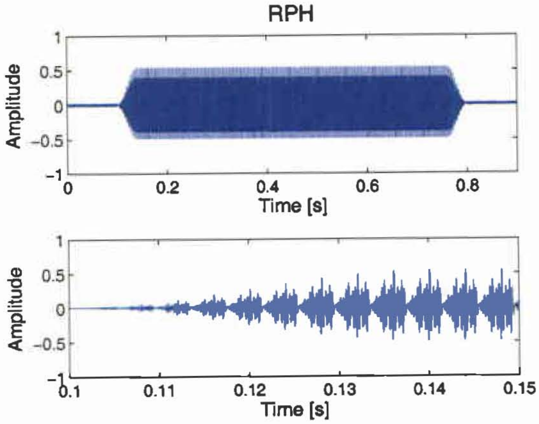


Figure 5.41: The RPH stimulus trough to crest is greatly reduced. Depending on the randomization of the tones in the complex this number can change.

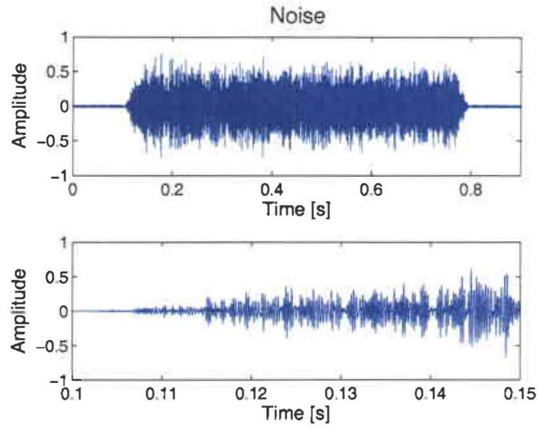


Figure 5.42: The Noise stimulus does not have a particular periodicity, but its maximum amplitude is on average quite close to the RPH stimulus.

stimulus. To see how a linear filterbank analyzes the components figure 5.43, 5.44 and 5.45 show the spectrograms of the raised cosine onset ramp along with several milliseconds of the steady state stimulus.

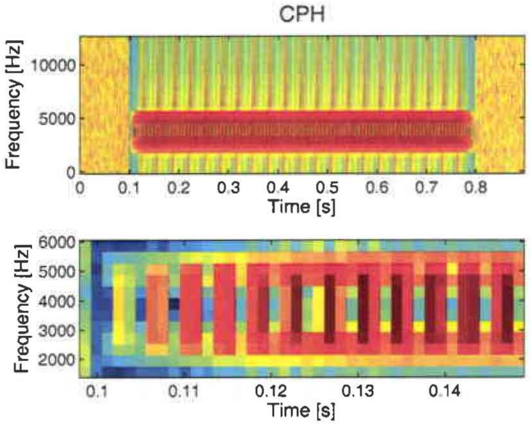


Figure 5.43: The CPH spectrogram. When all the tones in the complex are in phase the maximum value is very evident.

Here the temporal qualities of each frequency channel are more or less apparent. Because a linear frequency decomposition, by the FFT, makes an essential tradeoff

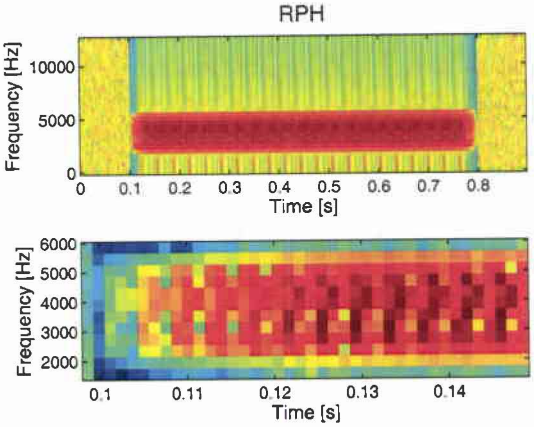


Figure 5.44: The RPH spectrogram. The complex lacks the beat rhythm of the CPH tone, but still retains some pulsing behaviour

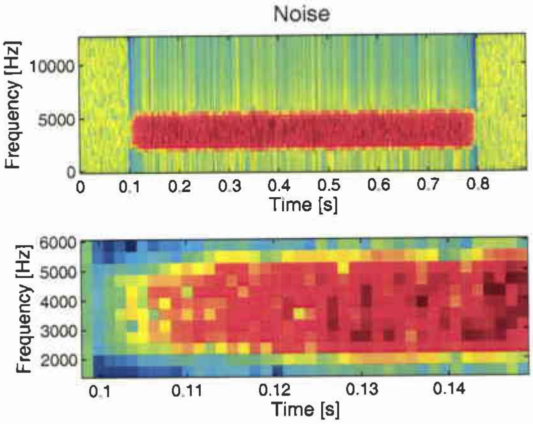


Figure 5.45: The Noise spectrogram has no pulses because it lacks any semblance of a comb spectrum.

between temporal and frequency resolution there is some uncertainty in this representation. This can minimally be overcome with a wavelet-like decomposition, but the auditory system seems to be able to analyse frequencies and process them within one half of the excitatory phase of an acoustic stimulus. That is to say, that the necessary quantities that provide distinction between the Gockel et al. [2003] stimuli are not well analyzed with normal auditory analysis tools. If instead one looks at the neurogram response of the stimuli using the fully nonlinear Bruce et al. [2003] model the resolution shows very interesting detail. Figures 5.46, 5.47 and 5.48 are the neurograms replicating the spectrogram figures, above. The heterogenous rates are plotted in 8 ms windows. Because of this windowing, there is almost no difference between these three figures. This one of the important factors in dealing with sensorineural impairment, simply by windowing information is lost.

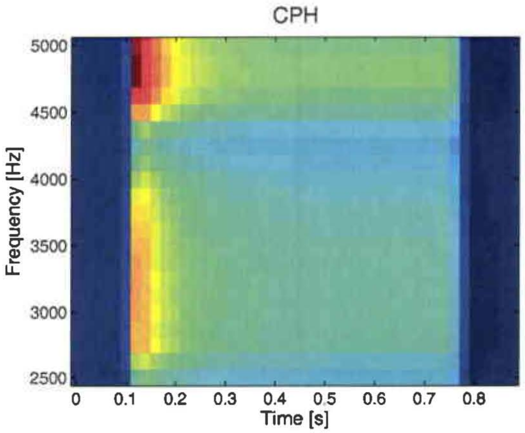


Figure 5.46: The CPH Neurogram. Adaptation has greatly emphasized the onset, with the Neurogram reacting to the raised cosine onset ramp.

This raised the question of what type of analysis could be done on the nonlinear neurograms that could emphasize the onset differences. By using the poisson point

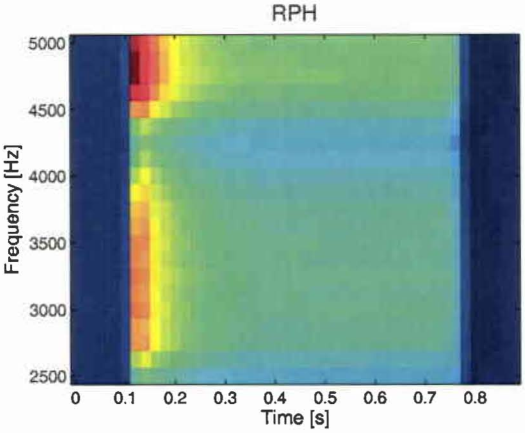


Figure 5.47: The RPH Neurogram. Adaptation has again emphasized the onset, but unlike the CPH Neurogram, or the spectrogram, there is a loss of coherence at the input. Because of the spectro-temporal resolution tradeoff of the FFT, there is some "smearing" in time, making differences in onsets that the normal auditory system picks up upon lost with inaccurate analysis.

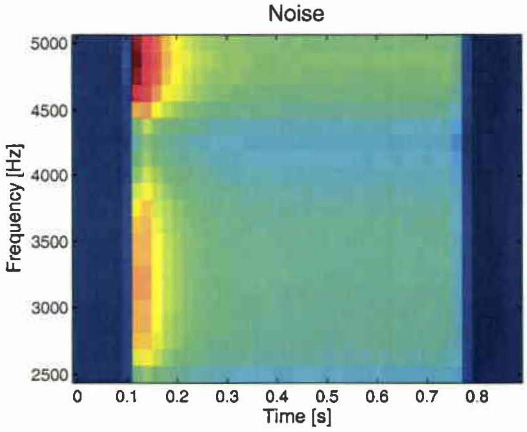


Figure 5.48: The Noise Neurogram looks startling similar to the RPH neurogram. The similarities showing far clearer than linear spectro-temporal analysis, and giving a clear connection why both stimuli react the same in psychophysical experiments.

process approximation of the spiking process, with randomization of the phase alignments for the RPH stimulus and random Noise stimulus the maximal AN response was determined. The highest probability of AN response, taking into account spiking properties shows much more difference than the neurograms. Figures 5.49, 5.50 and 5.51 are the probability distributions that the maximal spiking probability occurs at a specific time.

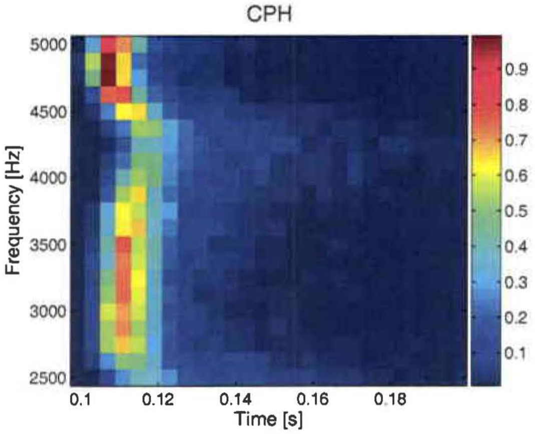


Figure 5.49: The CPH probability of maximal response. Adaptation has greatly emphasized the onset, differentiating it from the RPH and Noise stimuli.

The Noise and RPH maximum onset responses are almost indistinguishable, while the CPH response is sharper with less probability of having the maximum captured by acoustic or neural randomness. Fast adaptation produces a more specific indicator of where an onset occurs for the CPH stimulus. Figure 5.52 shows the probability mass function (pmf) of the three stimuli. Entropy is often used to quantify how well a pmf describes an event. Entropy is

$$H(N) = -\sum_{w=1}^{N} P(w) \log_2 P(w) \tag{5.35}$$

The entropy of an ensemble N, is the negative sum of the probabilities of each event in

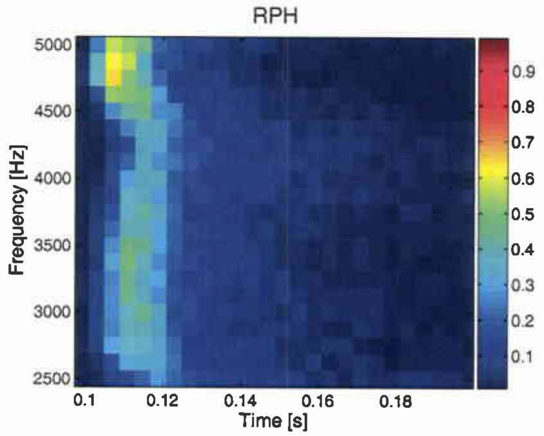


Figure 5.50: The RPH probability of maximal response. Adaptation has less of an effect on the onset, producing a maximal onset response very similar to the Noise stimulus.

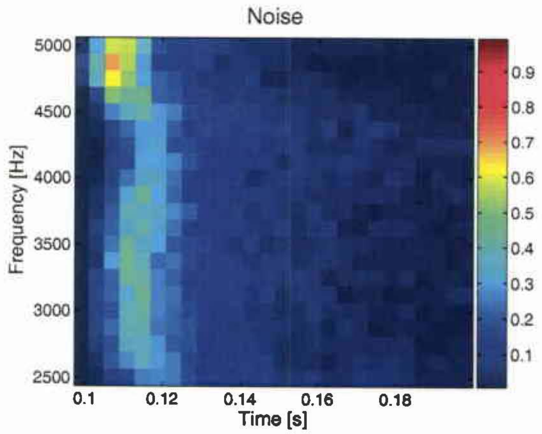


Figure 5.51: The Noise probability of maximal response. Adaptation has less of an effect on the onset, producing a maximal onset response very similar to the RPH stimulus.

the ensemble. A uniform distribution has maximum entropy, while the deterministic distribution has zero entropy.

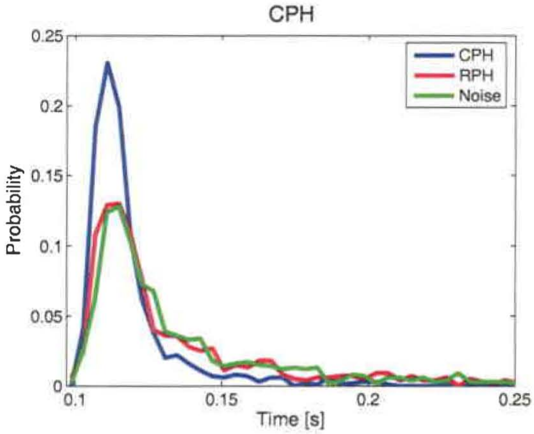


Figure 5.52: The maximal onset response for the CPH, RPH and Noise stimuli used in Gockel et al. [2003]. The CPH has far less entropy then the RPH and Noise stimuli, which are approximately equal.

If the auditory system is able to key on the fast adaptation marker imprinted on the AN response, it would have no problem coding the CPH marker. It is a lower entropy source, as shown in figure 5.53

A low entropy feature is less susceptible to noise, and since the brain reduces entropy as sensory input is processed to higher levels, it is much easier to *code* and *group* without losing information. The probability distributions for the onset characteristics of the randomly phased tone complex and the noise are strikingly similar, to the extent that they can provide a qualitative answer to the similarities in the psychophysics.

Onset characteristics came up several times in the discussion of the symptoms of psychophysics in section 2.2. They included the problems the sensorineurally impaired have with using the precedence effect, the deficit they see in reverberation and overall dealing with the competing speech that may arise from the inability to segment the

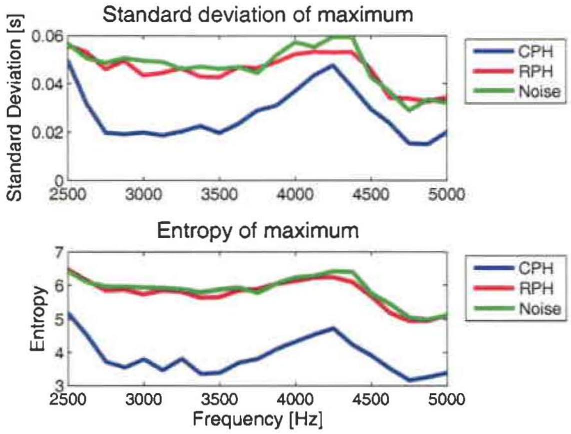


Figure 5.53: (Top) The standard deviations in the maximal response of the CPH, RPH and Noise stimuli used in Gockel et al. [2003]. (Bottom) The CPH stimulus has far less entropy then the RPH and Noise stimuli, which are approximately equal.

acoustic environment properly. The connection to Gockel et al. [2003] is that, while the normal ear nicely highlights onset cues and parcels the time-frequency plane, the loss of adaptation diminishes the ability to correctly landmark changes in spectral content. Since voiced speech is made from harmonic complexes, the spectral onsets to voiced speech in the normal auditory system with strong adaptation response will be aligned like the CPH stimuli making it a stronger cue for combining information across frequencies. With the diminished adaptation response brought on by sensorineural impairment, the onset burst may not occur at the same location across frequencies, making the impaired auditory systems onset representation naturally similar to the RPH or Noise stimuli. Compounding the loss of adaptation is that hearing aids reduce the trough-to-crest ratios because of high frequency amplification.

To illustrate this hypothesis the onset representation was calculated from the AN response of the TIMIT 'SX' training sentences over the 8 different regions for each phonemic class. The AN representation was calculated through the Bruce et al. [2003] model for a normal hearing auditory system and for losses like those used in Byrne et al. [2001]. The impaired auditory system had NAL-RP applied before being input into the model. A set of spectro-temporal response functions (STRFs) were made following the process in section 4.2.1. In short this set of STRFs were differential operators with different temporal passbands, following from

$$h_1[n] = {}^{n}/_{\alpha_1^2} \exp^{-n/\alpha_1} - {}^{n}/_{\alpha_2^2} \exp^{-n/\alpha_2} . \tag{5.36}$$

Unlike section 4.2.1, α_1 and α_2 were selected to pass a range of temporal modulation from sub-millisecond to 20 ms. A range of responses is shown in figure 5.54.

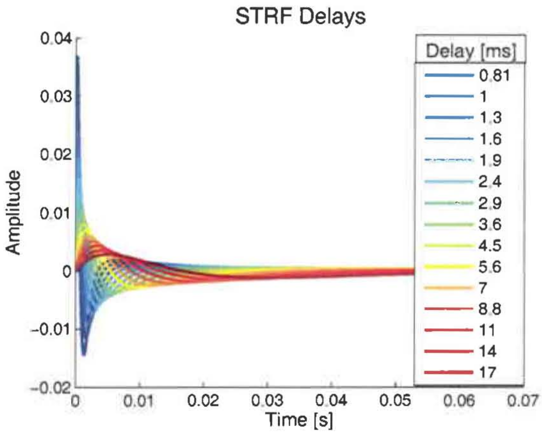


Figure 5.54: A snapshot of the temporal integrators used to probe differences of onsets encoded on the normal and impaired AN. The dark blue are faster than fast adaptation, and the dark red are slower but still faster than syllabic integration.

These integrators show the largest variance of AM information as measured by entropy. To produce the pmf used to calculate entropy, the dynamic range after an integrator was applied to either the normal or impaired AN was adjusted to 30 equally spaced rates. Figure 5.55 details the resulting average entropy for the TIMIT 'SX' training corpus.

The fast adaptation marker seems to lose a lot of its information with hearing

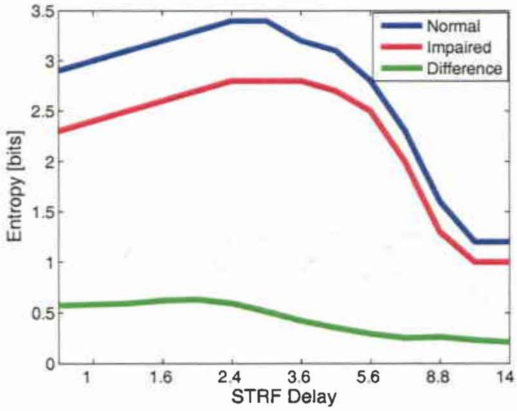


Figure 5.55: Both the Normal AN entropy rate (blue) and the Impaired AN entropy rate (red) peak around 2 ms, in line with fast adaptation being a very informative marker. The 2 ms marker is also the point where the Normal AN has the largest advantage over the Impaired AN, the difference between the two rates is in green.

impairment. Consistent with psychophysical experiments which show slow AM difference limens being relatively similar in normal hearing and hearing impaired, if the above graph was extended to slower rates, both entropy rates would become equal.

Since the fast adaptation marker is capable of providing the most information, it was studied further. The 2 ms STRF was then applied to the TIMIT "SX" corpus as detailed above. In a phonemic class an adaptation peak was identified as any rate more than three standard deviations above mean rates with zero rates discarded for that particular phone slice. AN responses are not Gaussian (they are much more compact), so the percent activation was much less than 0.1 %, instead of about 1 % if AN responses were Gaussian. This produced figures 5.56 to 5.59.

What is extremely interesting in figures 5.55 to 5.59 is that the normal cochlea encoded a feature much more robustly, with a lower entropy of representation, in a high entropy dimension. The only way for this flip-flop to occur is if the nonlinearity in the healthy cochlea is specially suited to robustly encode onset statistics.

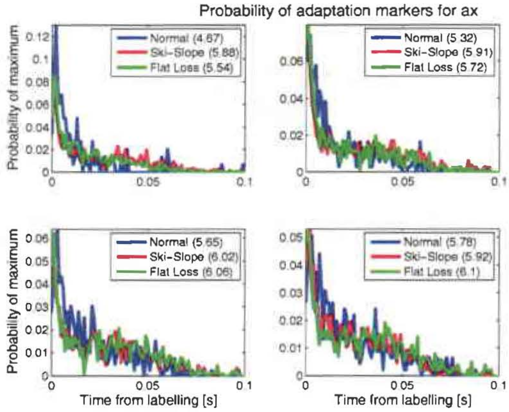


Figure 5.56: Four probabilities versus time from labelled marker of maximal adaptation response for a soft a. The top left graph is low frequency, top right is a higher frequency, the bottom left is even higher and the bottom right is the highest frequency. The blue curve is the normal cochleas response, red is for a ski-slope loss and the green curve is for a 60 dB flat loss. The number in the legend is the entropy of the corresponding curve.

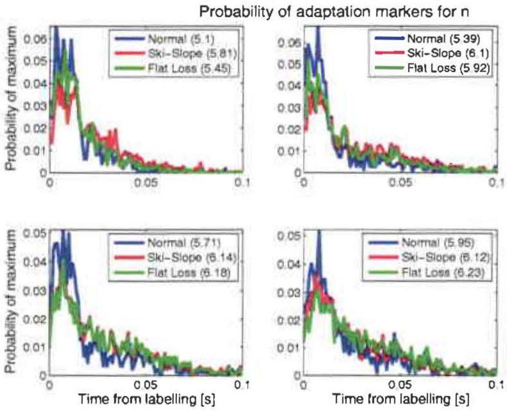


Figure 5.57: Four probabilities versus time from labelled marker of maximal adaptation response for a nasal, /n/. The top left graph is low frequency, top right is a higher frequency, the bottom left is even higher and the bottom right is the highest frequency. The blue curve is the normal cochleas response, red is for a ski-slope loss and the green curve is for a 60 dB flat loss. The number in the legend is the entropy of the corresponding curve.

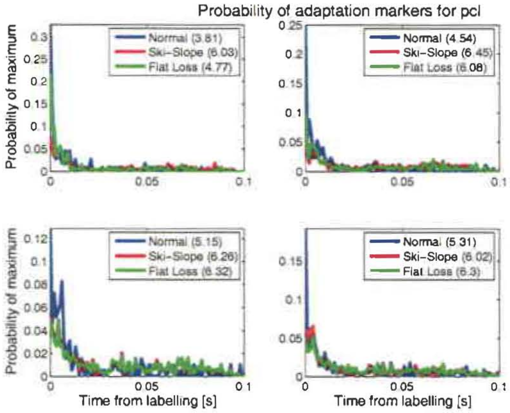


Figure 5.58: Four probabilities versus time from labelled marker of maximal adaptation response for the closure proceeding p. The top left graph is low frequency, top right is a higher frequency, the bottom left is even higher and the bottom right is the highest frequency. The blue curve is the normal cochleas response, red is for a ski-slope loss and the green curve is for a 60 dB flat loss. The number in the legend is the entropy of the corresponding curve.

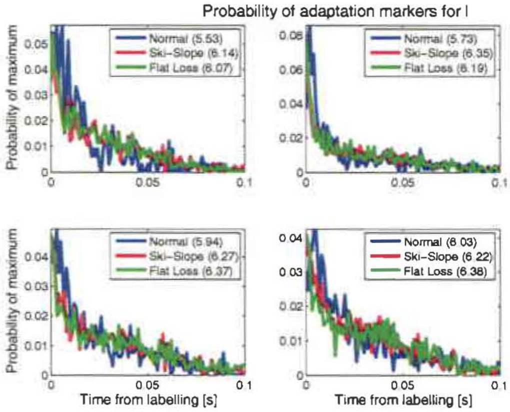


Figure 5.59: Four probabilities versus time from labelled marker of maximal adaptation response for glide, /l/. The top left graph is low frequency, top right is a higher frequency, the bottom left is even higher and the bottom right is the highest frequency. The blue curve is the normal cochleas response, red is for a ski-slope loss and the green curve is for a 60 dB flat loss. The number in the legend is the entropy of the corresponding curve.

The importance of the onset statistics can be described by comparing the onsets of the normal auditory representation in figure 5.60 and the impaired auditory representation, preprocessed with NAL-RP in figure 5.61. This is the amalgam of all the training data for the phoneme "v". The probability of the maximal AN response occurring at time after the acoustic segmentation marker is plotted versus frequency. There is a huge loss in the quality of onset information in the impaired auditory channels; the normal auditory systems probabilities are much more in line than the impaired ears onset information.

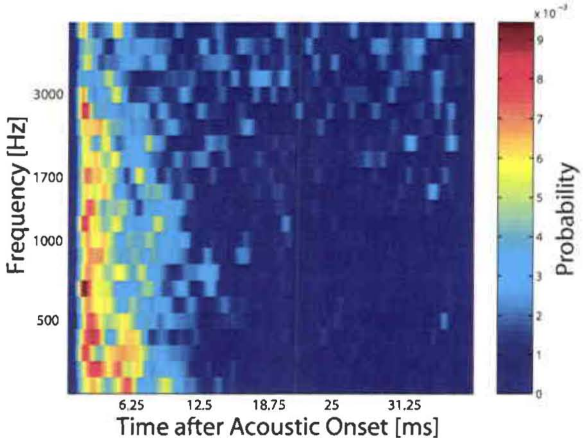


Figure 5.60: Maximal onset map for all the TIMIT acoustic segments labelled "v" when simulated with normal auditory periphery.

These adaptation differences stem from both inner and outer hair cell loss. The loss of inner hair cell gain linearizes the saturating nonlinearity in the inner hair cell, which in turn produces a less peaked, longer activity cycle. The phase response of the basilar membrane is also distorted producing a different traveling wave in the normal and impaired cochlea. Colloquially, one seems to see a loss of contrast between the different segments of an acoustic waveform, coded on the AN.

This is where Gockel et al. [2003] is very interesting. Figure 5.60 is visually

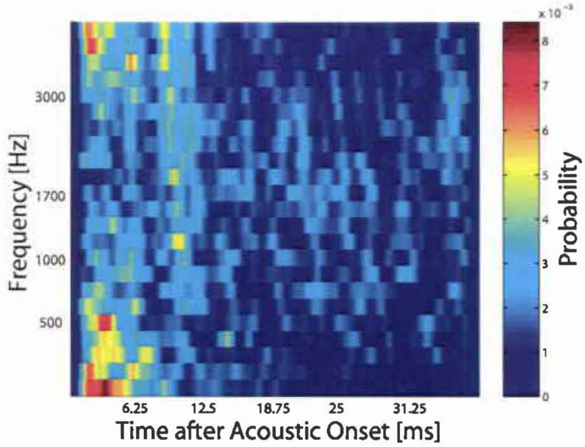


Figure 5.61: Maximal onset map for all the TIMIT acoustic segments labelled "v" when simulated with an impaired auditory periphery with NAL-R as preprocessing.

similar to 5.46, while figure 5.61 corresponds closely to 5.47. So the normal auditory system can make use of the natural phase information in speech, while the impaired auditory system loses that information embedded in the first wavefront because of the loss of adaptation. Naturally, the impaired auditory system will have all the idiosyncrasies associated with the random phase tone complex: unusual loudness perception and a heightened susceptibility to masking. These qualitatively match the effects of sensorineural impairment: loudness growth is accelerated for a period a little above threshold, and sensorineurally impaired people have a tremendous problem unmasking, especially temporally modulated maskers.

Chapter 6

Discussion and Future Work

6.1 Review

Chapter 3 dealt with a novel intelligibility metric that arose from the building up the standard AI. There are many different predictive measures that stem from this, all really being based on SNR. The SNR gain after processing does not accurately reflect the real benefit for hearing impaired people. Many researchers are wary of using SNR benefits without human testing. Chapter 3 started as an attempt to move away from the perceptually irrelevant SNR basis to intelligibility prediction for the hearing impaired to a measure encompassing their cochlear loss. This led to the development of the NAI in section 3.2.

There has been one other metric that attempted to provide a neural equivalent. Elhilali et al. [2003] introduced the spectro-temporal modulation index (STMI) as an intelligibility predictor based on the neural representation. The STMI can be described as a straightforward application of the MTF stimulus like that used in the STI, but put through a simple auditory model. The auditory model used in the STMI lacks the ability to model the differences between the normal and impaired cochlea. Many other applications of intelligibility predictors lack the dynamic nonlinear effects of the cochlea. But these are essential, as they are the pivotal differences between the normal cochleas and hearing impaired ones. Many researchers discard these differences because of the inherent difficulty in dealing with nonlinear phenomena.

For simple linear analysis the NAI works as well as the AI, SII, or STI and opened an avenue for machine learning to train hearing aid algorithms. This was explored in chapter 4. In early works based on this idea by Anderson [1994]; Rankovic [1991]; Kates [1993] their learning algorithm minimized mean square error, or SNR. Chapter 4 goes past these prior attempts by encompassing the impairment. Arising from these experiments was the insight that the differences between the normal and impaired auditory nerve responses underwent large changes during the course of a single sentence. At some points in time both the normal and impaired auditory responses were very similar, while at others they were completely different. This is where a new paradigm for intelligibility metrics had to be derived. Different parts of speech have to be judge differently.

While most of Chapter 5 follows from this, some recent data is worth mentioning. Kates & Arehart [2005] present an intelligibility predictor based on the coherence between a control signal and the signal under test. While equivalent to the SNR to within a linear transformation and almost mathematically equivalent to the NAI's distortion metric, they introduce an additional step. Kates & Arehart [2005] separate the stimulus into time segments and calculate the distortion in each segment. The segments are then averaged together with their importance based on their envelope amplitude. The envelope amplitude is suggested to be a major determining factor in finding where speech energy is. In chapter 5 there is a tacit assumption that different time segments of speech have different importance to intelligibility, and that they have different coding mechanisms. How to deal with this is dealt with further in the next section.

6.2 Future Work

This section details two novel possibilities for hearing aid processing that is derived from the framework in Chapter 5. The first processing block attempts to adaptively re-establish aggregate spiking rates on the hearing impaired AN. Section 6.3 uses the results derived in section 5.1 to train a nonlinear network. A similar nonlinear network is trained in section 6.4 this time with a main goal of reducing inter-channel correlation stemming from section 5.2. This is seen to be very similar to applying psychophysical masking from standard audio applications. Both sections 6.3 and 6.4 are preliminary results and require future study.

6.3 Compression: Adaptive Networks

Section 5.1 delineated that the aggregate loudness hypothesis does not produce the simple (normal dynamic range):(impaired dynamic range) compression ratio. In reality there is a complex interplay between the pathology and the acoustic input level, plus other factors that are as yet to be determined. The first step in realizing an adaptive nonlinear hearing aid algorithm along the lines detailed in section 5.1 is model verification.

The best type of hypothesis testing would involve human intelligibility testing. Future work could include using the Iowa consonant test [Tyler et al., 1987]. The

Iowa consonant test is made up of aCa tokens, or random consonants embedded with an /a/ preceding and following to act as a carrier. Consonants have a much more varied temporal structure, so this is an ideal place to start testing for benefits.

If the AN aggregate loudness hypothesis does hold under human testing, then more simulations open up the possibility of training adaptive networks. To try to identify an intelligent strategy, the optimal excitation gains by phone were compiled for the TIMIT "SX" corpus. Gains over NAL-RP versus input SPL were partly able to describe the empirical gain ratios derived from human testing. Since the input spectrum is a function of phonemic category, it was theorized that the input acoustic waveform shape may help reduce the variations seen across level. If this is the case, then figures 6.1 - 6.6 should show consistent differences in the distribution for a phonemic family.

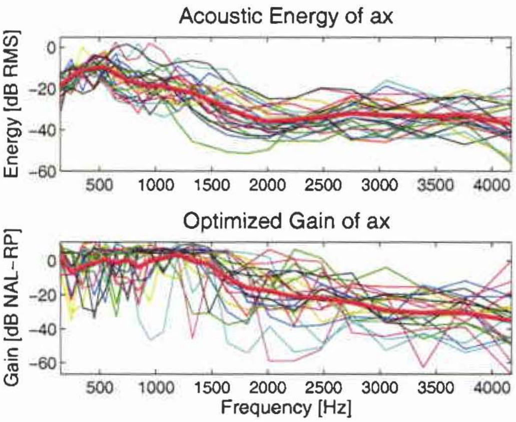


Figure 6.1: Optimal gains for a soft a. The top plot is the set of the acoustic db RMS across frequencies, the heavy red line is the mean acoustic input spectrum. The bottom graph is the set of optimal gains versus NAL-RP for all of the input slices. The heavy red line in the bottom graph is the mean of the derived gains.

From the graphs by phone, it is clear that the mean representation loses the pertinent information to produce optimality. In much the same way that the early

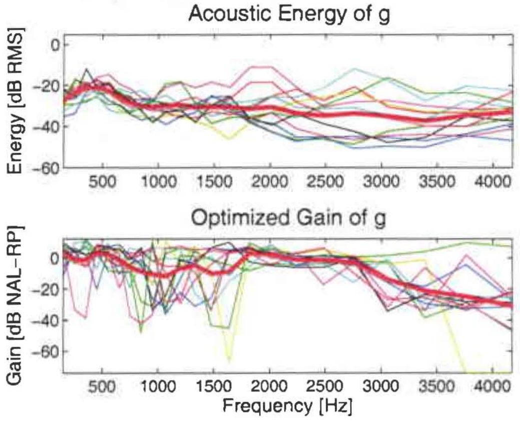


Figure 6.2: Optimal gains for /g/. The top plot is the set of the acoustic db RMS across frequencies, the heavy red line is the mean acoustic input spectrum. The bottom graph is the set of optimal gains versus NAL-RP for all of the input slices. The heavy red line in the bottom graph is the mean of the derived gains.

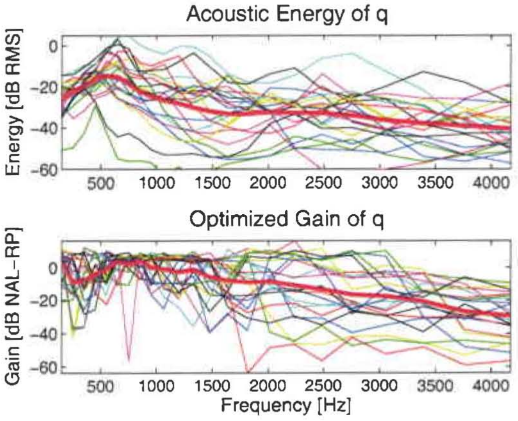


Figure 6.3: Optimal gains for q. The top plot is the set of the acoustic db RMS across frequencies, the heavy red line is the mean acoustic input spectrum. The bottom graph is the set of optimal gains versus NAL-RP for all of the input slices. The heavy red line in the bottom graph is the mean of the derived gains.

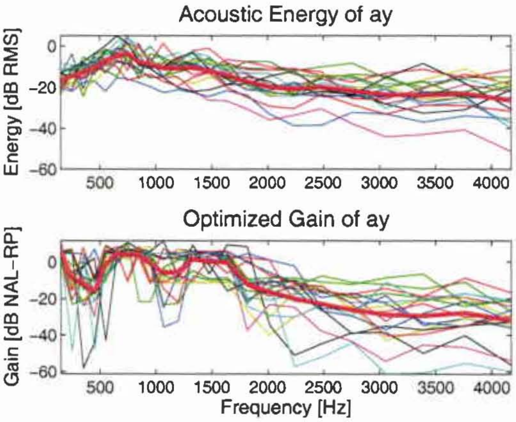


Figure 6.4: Optimal gains for ay. The top plot is the set of the acoustic db RMS across frequencies, the heavy red line is the mean acoustic input spectrum. The bottom graph is the set of optimal gains versus NAL-RP for all of the input slices. The heavy red line in the bottom graph is the mean of the derived gains.

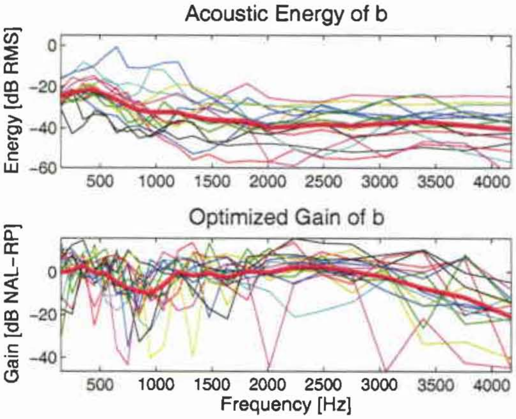


Figure 6.5: Optimal gains for b. The top plot is the set of the acoustic db RMS across frequencies, the heavy red line is the mean acoustic input spectrum. The bottom graph is the set of optimal gains versus NAL-RP for all of the input slices. The heavy red line in the bottom graph is the mean of the derived gains.

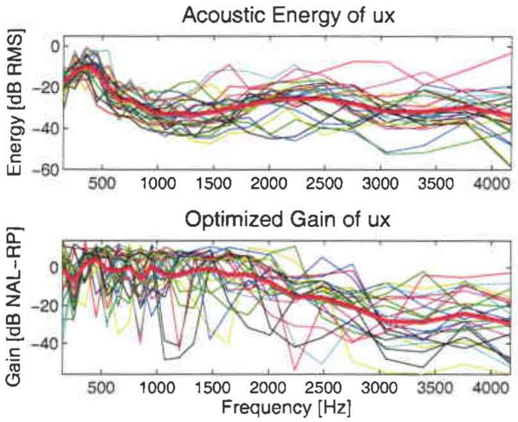


Figure 6.6: Optimal gains for ux. The top plot is the set of the acoustic db RMS across frequencies, the heavy red line is the mean acoustic input spectrum. The bottom graph is the set of optimal gains versus NAL-RP for all of the input slices. The heavy red line in the bottom graph is the mean of the derived gains.

NAI type developments simply averaged across a set of information, producing a smooth surface, this averaging operation removes the interesting aspects of hearing and replaces it with a manageable, yet ultimately trivial response. Obviously, there is not enough information to prescribe adequately what should be done to derive the nonlinear gain to reestablish normal firing rates in the sensorineural impaired ear if one takes averaged phone classes. So a network may be able to derive the necessary spectrum analysis to gain shape synthesis operation.

6.4 Suppression: Adaptive Networks

From section 5.2 it should be obvious that something needs to be done to reduce stimulus dependent correlations in different frequency channels. Not only does the loss of suppression and spread of excitation lead to redundant information and a diminishing throughput of information on the AN, but the effect of introducing spurious correlations creates confounding onsets.

Section 6.3 delineated processing based on the aggregate loudness hypothesis; this section will follow another hypothesis, the correlation reduction hypothesis proposed in section 5.2. Like, section 6.3, the first step in realizing an adaptive de-correlating hearing aid algorithm is model verification.

Again, human testing could be accomplished with the hVd corpus from Hillenbrand et al. [1995]. Vowels are made of harmonics, and every repetition beyond a pitch period decorrelates the constituent frequencies.

If human testing shows improvement over linear processing, then an adaptive scheme similar to 6.3 can be built. This section details an attempt to decorrelate the AN response as an initial step in hearing aid processing. The following algorithm is not based on maximizing intelligibility but in minimizing the amount of redundant information in the auditory system. Intelligibility turns out to be a poor predictor of hearing aid use [Bentler et al., 1993], and counter-intuitively of benefit; most hearing-aid users left to their own devices choose much less high frequency gain than is normally prescribed. One of the best predictors of the self-assessed benefit by sensorineural impaired people is the *acceptable noise level* [Nabelek et al., 1991, ANL;], which can be thought of as a measure of the hearing effort a hearing impaired person is willing to make.

The decorrelative hearing aid algorithm is theorized to reduce hearing effort by reducing confounding correlated structures introduced by sensorineural hearing loss while keeping the those intrinsic to the acoustic stimulus. The initial supervised learning signal, based on intelligibility, is replaced with driving the impaired frequency-covariance matrix towards the normal frequency covariance matrix

$$E\left(\text{Normal}, \text{Impaired}\right) = sqrt\left(\left|\left|C_{\text{Normal}} - C_{\text{Impaired}}\right|\right|_{\text{Frob}}\right) \tag{6.37}$$

The covariance matrix is formed by taking the discharge rate vectors of 25 frequency channels and multiplying that by itself transposed, then normalizing by the number of time samples. There were 10 bands of equal bandwidth under 1000 Hz, and then 15 bands spaced $1/6^{th}$ an octave above that. The AN response was derived for a fiber with a BF at the center of each of the acoustic bands. The speech was normalized to 65 dB SPL before the Wiener & Ross [1946] outer ear transfer function was applied. Figure 6.7 is an example of how the covariance matrix evolves from in the normal auditory system, to a moderate presbycusis AN representation with NAL-RP, and how proper adaptive shaping can produce a covariance with less redundancy even with hearing impairment.

The main diagonal for each covariance matrix is the average driven discharge rate in that channel. The other diagonals show how correlated one channel is to another. The error metric is function of the main diagonal, or the spectral shape of the input, as well as the redundancy across channels, in the off diagonal matrix elements.

Initial tests minimizing the off-diagonal elements lead to a decorrelating frequency response, as the null matrix is an obvious solution. By keeping the main diagonal information, the error metric from equation 6.37 also attempts to "recenter" the peaks in the impaired excitation response, where formants may have moved from the spreading excitation brought on by sensorineural impairment.

For the impaired listener, a typical hearing aid gain was set following the NAL-RP prescription, as described in section 4.1. Obviously, the covariance matrix will show a great deal more inter-channel correlation. Figure 6.7 is the resulting matrix derived from the dutch syllable "bak", with a NAL-RP fitting through the impaired audiogram #4 from figure 4.1.

Finally, for the decorrelative hearing aid algorithm the gains in 25 channels were derived by adjusting the initial NAL-R prescribed gains by 0.5 % over 200 iterations, minimizing equation 6.37 by a stochastic optimization. Optimality was defined as producing the closest correspondence between the normal and impaired 2^{nd} order statistics.

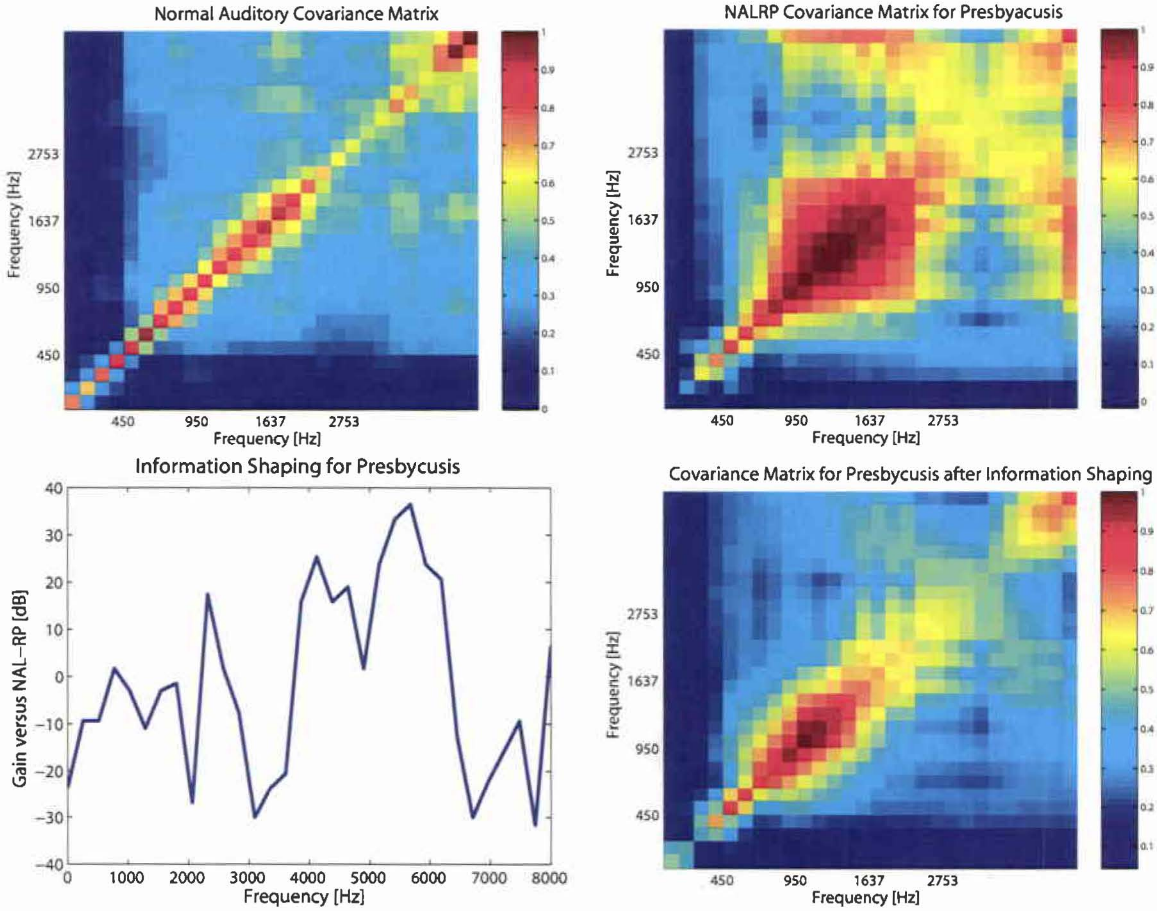
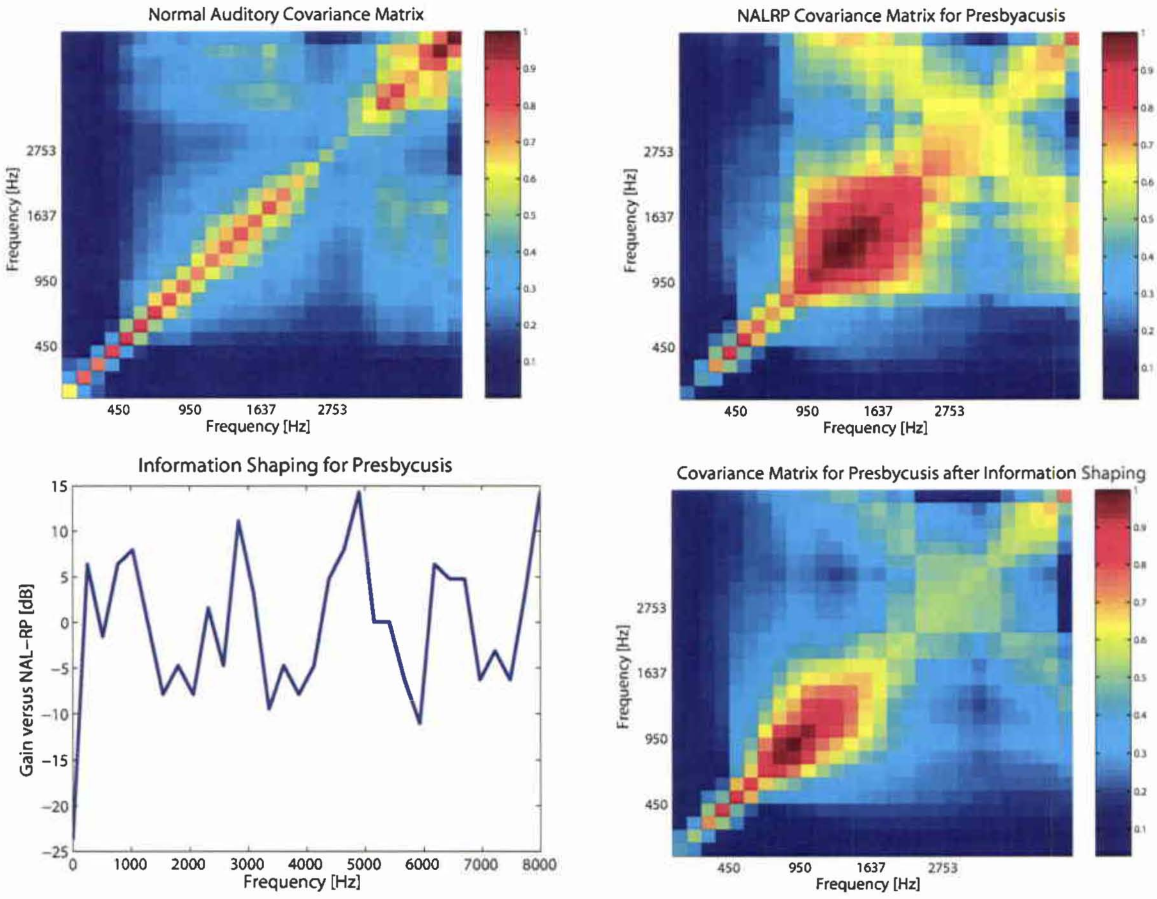


Figure 6.7: Top left is the normal auditory covariance matrix of the syllable 'bak', top right is the same syllable when simulated with an impaired auditory model with NAL-RP preprocessing, bottom left is the frequency shaping derived to minimize redundancy for the hearing impaired and bottom right is the resulting decorrelated covariance matrix for the same loss as seen in the top right.

In the normal response from figure 6.7, suppression actively attenuates correlation in adjacent channels. The minor diagonals have very low coefficients. The decorrelative coding mechanisms in the ear in the impaired ear are greatly reduced. This coupled with the spread of excitation brought on by sensorineural impairment gives quite large inter-channel correlations. There is also a spread of excitation along the main diagonal and the spectral peak-to-trough is reduced. It is speculated that this loss of suppression can be thought of as the spectral equivalent to the loss of temporal adaptation. Both are active processes that in essence are drawing attention to changes in the auditory input. Both enhance the representation of changes in the acoustic signal in the auditory brain, the vision counterpart would be contrast enhancing, lateral inhibitory cells.

Several further examples of the normal, impaired with NAL-R preprocessing, and impaired with optimal correlation reduction are given in Figures 6.8 to 6.9



Normal Auditory Covariance Matrix

NALRP Covariance Matrix for Presbyacusis

Information Shaping for Presbycusis

Covariance Matrix for Presbycusis after Information Shaping

Figure 6.8: Top left is the normal auditory covariance matrix of the syllable 'drup', top right is the same syllable when simulated with an impaired auditory model with NAL-RP preprocessing, bottom left is the frequency shaping derived to minimize redundancy for the hearing impaired and bottom right is the resulting decorrelated covariance matrix for the same loss as seen in the top right.

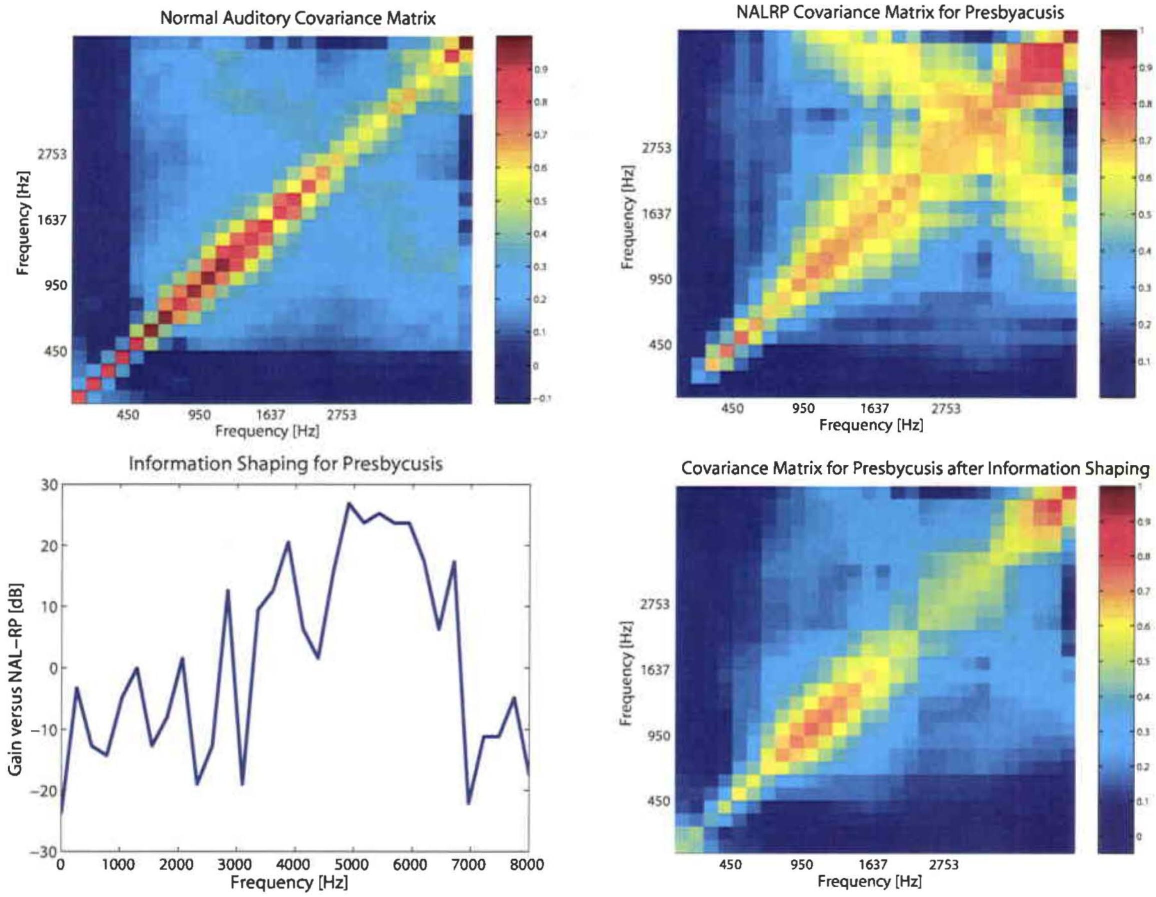


Figure 6.9: Top left is the normal auditory covariance matrix of the syllable 'fluit', top right is the same syllable when simulated with an impaired auditory model with NAL-RP preprocessing, bottom left is the frequency shaping derived to minimize redundancy for the hearing impaired and bottom right is the resulting decorrelated covariance matrix for the same loss as seen in the top right.

For each stimuli the changes in correlative structure after are quite striking. To make a successful algorithm though, the input must be analyzed and the proper gain strategy calculated. The resulting gain shapes for other syllables, the input spectrogram, optimal decorrelative matrix and output spectrogram must all be embedded in a successful adaptive network. Obviously, the changes in the decorrelative gain shape is a complex function of the input, but it is also a complex function of the loss. Loss profile #1 from figure 4.1 optimal gains were much different then those reported above. To make a useful algorithm, it is necessary to apply machine learning to a network that can find all the statistically significant information from the various input stimuli.

This was done by stacking all the individual short word spectra together as an input to the network. Each spectrum was calculated as a 32 bin periodogram, using Welch's method. The decorrelation targets were linearly interpolated from 25 channels spaced similarly to human cochlea place mapping to 32 linear spaced channels to match the input size. The target network was a 32x64x32 feedforward network, the training procedure is shown in figure 6.10.

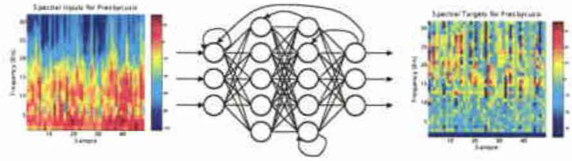

Spectral Inputs for Presbycusis
Sample
Spectral Targets for Presbycusis
Sample

Figure 6.10: An example of the type of adaptive network that may be able to embed the adaptive decorrelating process lost with sensorineural impairment.

To apply the trained network to running speech, the periodogram is calculated in short time windows, and used as an input to the network. The resulting vector of channel gains is then low passed filtered by channel, to reduce fast gain fluctuations, or popping. The aim of this section was to highlight an attempt to return the audio coding that is done in the normal auditory periphery. One of the key nonlinearities responsible for the great job of coding in the ear is suppression, which was shown to be a decorrelative process that enhances the contrast between adjacent frequency bands.

Chapter 7

Conclusion

This dissertation started off with trying to introduce the reader to a problem that is neither well defined, nor answerable with the tools that presently exist. Chapter 2 provided a primer on the symptoms of hearing impairment, and up to this point hearing aid signal processing has largely been focussed on dealing with one symptom or the other. Chapter 3 details the development of a novel intelligibility metric that would be able to compare hearing aid algorithms offline, with attempts to prove out the machine learning framework in chapter 4.

This is where the dissertation moved from empirical modelling to theoretical modelling. There are no data or quantitative statistics on the adaptive nonlinear mechanics of the healthy and impaired cochlea, especially when the interplay of pathology with acoustics is taken into consideration. Simply put, there needed to be a distillation of the numerous symptoms and how to combat them from the introduction. This gave rise to the key for the remaining chapters. The differences between the normal and sensorineural impaired cochlea can be summed up by saying that the impaired cochlea operates more linearly; the normal functioning cochleas adaptive nonlinearity is the optimal transformation for dealing with the statistics of speech.

The theoretical framework in Chapter 5 studied the key differences between the normally operating and the sensorineural impaired cochlea. Chapter 5 discusses how important these nonlinearities are by polling the normal and impaired cochleas auditory coding characteristics, instead of numerical modelling.

The discussion and future work in chapter 6 discussed how to adapt a hearing aid's signal processing to address the core adaptive nonlinear problem of sensorineural impairment. These new processing strategies are key in mimicking the normal auditory periphery's dynamic nonlinearities.

The conclusion to this dissertation is not an ending, as the body of work raises more questions then it started out with. It is hoped that the new modelling techniques suggested in this dissertation can be trialled in human tests. Or other cochlear modelling ideas might be incorporated to improve the health of the hearing impaired populace.

Bibliography

Allen, J. B. (1990). Loudness growth in 1/2-octave bands (lgob); a procedure for the assessment of loudness. *J. Acoust. Soc. Am.*, *88*, 745–753.

Allen, J. B. (1994). How do humans process and recognize speech? *Speech and Audio Processing, IEEE Transactions on,*, *2*(4), 567–577.

Anderson, D. V. (1994). Model based development of a hearing aid. Master's thesis, Brigham Young University, Provo, Utah.

Anderson, D. V., Harris, R. W., & Chabries, D. (1995). Evaluation of a hearing compensation algorithm. *IEEE ASSP-95*, 3531–3533.

ANSI (1997). *ANSI S3.5-1997 Methods for calculation of the speech intelligibility index.* New York: American National Standards Institute.

Attneave, F. (1954). Some information aspects of visual perception. *Psychological Review*, *61*, 183–193.

Bandyopadhyay, S. & Young, E. D. (2004). Discrimination of voiced stop consonants based on auditory nerve discharges. *J. Neurosci.*, *24*, 531–541.

Barlow, H. (1961). *Possible principles underlying the transformation of sensory messages.* Cambridge MA: MIT Press.

Becker, S. (1996). Mutual information maximization: Models of cortical self organization. *Network: Computation in Neural Systems*, *7*.

Becker, S. & Bruce, I. C. (2002). Neural coding in the auditory periphery:insights from physiology and modeling lead to a novel hearing compensation algorithm. In *Workshop on Neural Information Coding*, Les Houches France.

Bentler, R. A., Niebuhr, J., Getta, C., & Anderson, C. (1993). Longitudinal study of hearing aid effectiveness. ii. subjective measures. *Journal of speech and hearing research*, *36*, 820–831.

Bia, A. (2001). Alopex-b: A new, simpler but yet faster version of the alopex training algorithm. *International Journal of Neural Systems, Special Issue on Non-gradient optimization methods*, 497–507.

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

Bondy, J. & Bruce, I. (2004a). Degradation of acoustic coding in the auditory nerve by sensorineural impairment. In *Canadian Acoustical Association, Acoustics week in Canada*, Ottawa.

Bondy, J. & Bruce, I. (2004b). Machine learning and the auditory nerve. In *International Hearing Aid Research Conference (IHCON)*, Lake Tahoe.

Bondy, J. & Bruce, I. (2005). Cochlear nonlinearities and temporal neural clues. In *Toronto Auditory Temporal Processing Symposium.*

Bondy, J., Bruce, I. C., Becker, S., & Haykin, S. (2004). Predicting speech intelligibility from a population of neurons. In Thrun, S., Saul, L., & Schoelkopf, B. (Eds.), *Advances in Neural Information Processing Systems 16.* MIT Press.

Bondy, J., Bruce, I. C., Dong, R., Becker, S., & Haykin, S. (2003). Modeling intelligibility of hearing-aid compression circuits. *Signals, Systems & Computers, 2003 The Thrity-Seventh Asilomar Conference on,*, *1*, 720–724.

Brooks, D. (1973). Gain requirements of hearing aid users. *Scand. Audiol.*, *2*, 199.

Bruce, I. C. (2004). Physiological assessment of contrast-enhancing frequency shaping and multiband compression in hearing aids. *Physiological Measurement.*

Bruce, I. C., Bondy, J., Haykin, S., & Becker, S. (2002). A physiologically based predictor of speech intelligibility. In *International Conference on Hearing Aid Research*, Lake Tahoe, CA.

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

Bruce, I. C., Young, E., & Sachs, M. (1999). Modification of an auditory periphery model to describe the effects of acoustic trauma on auditory nerve response. In *22nd annual ARO mid-winter meeting*, St. Petersburg Beach, FL.

Bunnell, H. T. (1990). On enhancement of spectral contrast in speech for hearing-impaired listeners. *J. Acoust. Soc. Am.*, *88*(6), 2546–2556.

Buus, S., Florentine, M., & Redden, R. (1982). The sisi test: A review. part ii. *Audiology*, *21*, 365–385.

Byrne, D., Dillon, H., Ching, T., Katsch, R., & Keisder, G. (2001). Nal-nl1 procedure for fitting nonlinear hearing aids: characteristics and comparisons with other procedures. *J. Am. Acad. Audiol.*, *12*, 37–51.

Byrne, D. & Fifield, D. (1974). Evaluation of hearing aid fittings for infants. *Brit. J. Audiol.*, *8*, 47–54.

Byrne, D., Parkinson, A., & Newall, P. (1990). Hearing aid gain and frequency response requirements for the severely/profoundly hearing impaired. *Ear and Hearing*, *11*, 40–49.

Carhart, R. & Tillman, T. (1970). Interaction of competing speech signals with hearing losses. *Arch. Otolaryngol.*, *91*, 273–279.

Chabries, D., Anderson, D., Stockham, T., & Christiansen, R. (1995). Application of a human auditory model to lousness compensation and hearing compensation. *IEEE ASSP-95*, 3527–3530.

Choe, Y., Magnasco, M. O., & Hudspeth, A. (1998). A model for amplification of hair-bundle motion by cyclical binding of ca2+ to mechanoelectrical-transduction channels. *Proceedings of the National Acadamy of Science of the USA*, *95*, 15321–15326.

Dallos, P., Popper, A. N., & Fay, R. R. (Eds.). (1996). *The Cochlea.* Springer-Verlag.

Davis, H. & Silverman, S. (1970). *Hearing and Deafness.* Holt, Rinehart and Winston.

Delgutte, B. (1980). Representation of speech-like sounds in the discharge patterns of auditory-nerve fibers. *J Acoust Soc Am*, *68*, 843–857.

Delgutte, B. & Kiang, N. Y. S. (1984). Speech coding in the auditory nerve: Iv. sounds with consonant-like dynamic characteristics. *J. Acoust. Soc. Am.*, *75*(3), 897–918.

Depireux, D., Simon, J., Klein, D., & Shamma, S. (2001). Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex. *J. Neurophysiol.*, *85*(3), 1220–1234.

Dillon, H. (1996). Compression? yes, but for low or high frequencies, for low or high intensities, and with what response times? *Ear. Hear.*, *17*, 287–307.

Drullman, R., Festen, J., & Plomp, R. (1994). Effect of reducing slow temporal modulations on speech reception. *J. Acoust. Soc. Am.*, *95*(5), 2670–2680.

Duda, R. & Hart, P. (1973). *Pattern classification and scene analysis.* Wiley.

Duquesnoy, A. (1983). Effect of a single interfering noise or speech source upon the binaural sentence intelligibility of aged persons. *J. Acoust. Soc. Am.*, *74*(3), 739–743.

Elhilali, M., Chi, T., & Shamma, S. A. (2003). A spectro-temporal modulation index (stmi) for assessment of speech intelligibility. *Speech Communications*, *43*(2), 331–348.

Fabry, D. & van Tassell, D. (1990). Evaluation of an articulation-index based model for predicting the effects of adaptive frequency response hearing aids. *Journal of speech and hearing research*, *33*, 676–689.

Festen, J. & Plomp, R. (1983). Relations between auditory functions in impaired hearing. *J. Acoust. Soc. Am.*, *73*, 652–662.

Festen, J. & Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *J. Acoust. Soc. Am.*, *88*, 1725–1736.

Fettiplace, R. & Ricci, A. (2003). Adaptation in auditory hair cells. *Current opinion in Neurobiology*, *13*, 446–451.

Fettiplace, R., Ricci, A. J., & Hackney, C.

Fitzgibbons, P. J. & Wightman, F. L. (1982). Gap detection in normal and hearing-impaired listeners. *J. Acoust. Soc. Am.*, *72*(3), 761–765.

Fletcher, H. (1940). Auditory patterns. *Rev. Mod. Phys.*, *12*, 47–65.

Fletcher, H. & Galt, R. (1950). The perception of speech and its relation to telephony. *J. Acoust. Soc. Am.*, *22*, 89–151.

Florentine, M., Fastl, H., & Buus, S. (1988). Temporal integration in normal hearing, cochlear impairment, and impairment simulated by masking. *J. Acoust. Soc. Am.*, *73*, 961–965.

French, N. & Steinberg, J. (1947). Factors governing the intelligibility of speech sounds. *J. Acoust. Soc. AM.*, *19*, 90–119.

Gatehouse, S., Naylor, G., & Elberling, C. (2005). Linear and non-linear hearing aid fitting – patterns of benefit. *submitted to International journal of audiology.*

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

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

Glasberg, B. & Moore, B. (1992). Effects of envelope fluctuations on gap detection. *Hear. Res.*, *64*, 81–92.

Glowatzki, E. (2004). The auditory periphery. *Course Notes*, 1–30.

Gockel, H., Moore, B. C. J., Patterson, R. D., & Meddis, R. (2003). Louder sounds can produce less forward masking: Effects of component phase in complex tones. *J. Acoust. Soc. Am.*, *114*(2), 978–990.

Goodman, A. (1965). Reference zero levels for pure tone audiometer. *ASHA*, *7*, 262–263.

Grant, K. (1987). Frequency modulation detection by normally hearing and profoundly hearing impaired listeners. *J. Speech Hear. Res.*, *30*, 558–563.

Gratton, M., Schmiedt, R., & Schulte, B. (1996). Age-related decreases in endocochlear potential are associated with vascular abnormalities in the stria vascularis. *Hear. Res.*, *102*((1-2)), 181–190.

Greenwood, D. (1990). A cochlear frequency-position function for several species - 29 years later. *J. Acoust. Soc. Am.*, *87*, 2592–2605.

Hall, J. & Fernandez, M. (1983). Temporal integration, frequency resolution, and off-frequency listening in normal hearing and cochlear impaired listeners. *J. Acoust. Soc. Am.*, *74*, 1172–1177.

Hall, J., Tyler, R., & Fernandez, M. (1984). Factors influencing the masking level difference in cochlear hearing impaired and normal hearing listeners. *J. Acoust. Soc. Am.*, *74*, 1172–1177.

Harris, R. & Swenson, D. (1990). Effects of reverberation and noise on speech recognition by adults with various amounts of sensorineural hearing impairment. *Audiology*, *29*(6), 314–321.

Hausler, R., Colburn, H. S., & Marr, E. (1983). Sound localization in subjects with impaired hearing. *Acta Otoloaryngol. Suppl.*, *400*, 1–62.

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

Heinz, M. G. & Young, E. D. (2004). Response growth with sound level in auditory nerve fibers after noise-induced hearing loss. *J. Neurophysiology*, *91*, 784–795.

Hellman, R. & Meiselman, C. (1993). *J. Acoust. Soc. Am.*, *93*, 966–975.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of american english vowels. *J. Acoust. Soc. Am.*, *97*(5), 3099–3111.

Hoekstra, A. & Ritsma, R. (1977). Perceptive hearing loss and frequency selectivity, 263–271.

Houtgast, T. & Steeneken, H. (1973). The modulation transfer function in room acoustics as a predictor of speech intelligibility. *Acustica*, *28*, 66–73.

Houtgast, T. & Verhave, J. (1991). A physical approach to speech quality assessment: correlation patterns in the speech spectrogram. *Proc. Eurospeech 1991*, 285–288.

Hygge, S., Rnneberg, J., Larsby, B., & Arlinger, S. (1992). Normal hearing and hearing-impaired subjects' ability to just follow conversation in competing speech, reversed speech and noise backgrounds. *J. Speech Hear. Res.*, *35*, 208–215.

Irwin, R. & McAuley, S. F. (1987). Relations among temporal acuity, hearing loss, and the perception of speech distorted by noise and reverberation. *J Acoust Soc Am.*, *81*(5), 1557–1565.

Johnson, D. H. (1980). The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones. *J. Acoust. Soc. Am.*, *68*(4), 1115–1122.

Julicher, F., Andor, D., & Duke, T. (2001). The physical basis of two-tone interference in hearing. *Proc. Natl. Acad. Sci. USA*, *98*, 9080–9085.

Kates, J. (1993). Toward a theory of optimal hearing aid processing. *J. of Rehab. Res.*, *30*(1), 39–48.

Kates, J. M. & Arehart, K. (2005). Coherence intelligibility junk. *J. Acoust. Soc. Am.*, *My Butt*, 8934–298489.

Kiang, N., Moxon, E., & Levine, R. (1970). Auditory nerve activity in cats with normal and abnormal cochleas, 241–273.

Kinkel, M., Kollmeier, B., & Holube, I. (1991). Binaurales horen bei normalhorenden und schwehorigen. i. mebethods and mebergebnisse. *audiologishe akustik*, *6/91*, 192–201.

Kryter, K. D. (1962a). Method for the calculation and use of the articulation index. *J. Acoust. Soc. Am.*, *34*(11), 1689–1697.

Kryter, K. D. (1962b). Validation of the articulation index. *J. Acoust. Soc. Am.*, *34*(11), 1698–1702.

Kullback, S. (1968). *Information Theory and Statistics.* New York: Dover Publications.

Leek, M. R. & Summers, V. (1993). Auditory filter shapes of normal-hearing and hearing-impaired listeners in continuous broadband noise. *J. Acoust. Soc. Am.*, *94*, 3127–3137.

Leeuw, A. & Dreschler, W. A. (1994). Frequency-resolution measurements with notched noise for clinical purposes. *Ear Hear.*, *15*, 240–255.

Lewicki, M. (2002). Efficient coding of natural sounds. *Nature Neuroscience online.*

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

Liberman, M. C. & Dodds, L. W. (1984b). Single-neuron labelling and chronic cochlear pathology. ii. stereocilia damage and alterations of spontaneous discharge rates. *Hearing Research*, *16*, 43–53.

Liberman, M. C. & Kiang, N. Y. (1978). Acoustic trauma in cats. cochlear pathology and auditory-nerve activity. *Acta Otolaryngol. Suppl.*, *358*, 1–63.

Linsker, R. (1992). Local synaptic learning rules suffice to maximize mutual information in a linear network. *Neural Computation*, *4*, 691–702.

Manley, G., Kirk, D., Koppl, C., & Yates, G. (2001). In vivo evidence for a cochlear amplifier in the hair-cell bundle of lizards. *Proc Natl Acad Sci USA*, *98*, 2826–2831.

Markle, D. & Zaner, A. (1966). The determination of 'gain requirements' of hearing aids: A new method. *J. Audiol. Res.*, *6*, 371–377.

Martin, F., Champlin, C., & Chambers, J. (1998). Seventh survey of audiometric practices in the united states. *J. Am. Acad. Audiol.*, *9*, 95–104.

Miller, G. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 81–97.

Miller, R. L., Schilling, J. R., Franck, K. R., & Young, E. D. (1997). Effects of acoustic trauma on the representation of the vowel /e/ in cat auditory nerve fibers. *J. Acoust. Soc. Am.*, *101*, 36023616.

Moore, B. (1995). *Perceptual Consequences of Cochlear Damage*. Oxford: Oxford University Press.

Moore, B. & Glasberg, B. (1986). Comparison of frequency selectivity in simultaneous and forward masking for subjects with unilateral cochlear impairments. *J. Acoust. Soc. Am.*, *80*, 93–107.

Moore, B. & Glasberg, B. (1998). Use of a loudness model for hearing-aid fitting. i. linear hearing aids. *Br. J. Audiol.*, *32*, 317–335.

Moore, B., Glasberg, B., Donaldson, E., McPherson, T., & Plack, C. (1989). Detection of temporal gaps in sinusoids by normally hearing and hearing-impaired subjects. *J. Acoust. Soc. Am.*, *85*, 1266–1275.

Moore, B., Glasberg, B., Hess, R., & Birchall, J. (1985). Effects of flanking noise bands on the rate of growth of loudness of tones in normal and recruiting ears. *J. Acoust. Soc. Am.*, *77*, 1505–1515.

Moore, B., Huss, M., Vickers, D., Glasberg, B., & Alcantara, J. (2000). A test for the diagnosis of dead regions in the cochlea. *Br. J. Audiol.*, *34*(4), 205–224.

Moore, B., Peters, R., & Stone, M. (1999). Benefits of linear amplification and multichannel compression for speech comprehension in backgrounds with spectral and temporal dips. *J. Acoust. Soc. Am.*, *105*(1), 400–411.

Moore, B., Schailer, M., Hall, J., & Schooneveldt, G. (1993). Comodulation masking release in subjects with unilateral and bilateral cochlear hearing impairment. *J. Acoust. Soc. Am.*, *93*, 435–451.

Nabelek, A. K., Tucker, F., & Letowski, T. (1991). Toleration of background noises: Relationship with patterns of hearing aid use by elderly persons. *Journal of speech and hearing research*, *34*, 679–685.

Nelson, P. C. & Carney, L. H. (2004). A phenomenological model of peripheral and central neural responses to amplitude modulated tones. *J. Acoust. Soc. Am.*, *116*(4), 2173–2186.

Nilsson, M., Soli, S., & Sullivan, J. (1994). Development of the hearing in noise test for measurement of speech reception thresholds in quiet and in noise. *J. Acoust. Soc. Am.*, *95*(2), 1085–1099.

Nobili, R., Mammano, F., & Ashmore, J. (1998). How well do we understand the cochlea? *Trends in Neuroscience*, *21*, 159–167.

Nolte, J. (1993). *The Human Brain, An Introduction to its Functional Anatomy 3rd Ed.*, (pp. 213). Mosby-Year Book, Inc.

Nordlund, B. (1964). Directional audiometry. *Acta Otolaryngol.*, *57*, 1–18.

O'Loughlin, B. J. & Moore, B. (1981). Off-frequency listening: effects on psychoacoustical tuning curves obtained in simultaneous and forward masking. *J. Acoust. Soc. Am.*, *69*, 1119–1125.

Pascoe, D. (1978). An approach to hearing aid selection. *Hearing Instruments*, *29*, 12–16.

Patterson, R., Nimmo-Smith, I., Holdsworth, J., & Rice, P. (1988). Implementing a gammatone filter bank. *SVOS Final Report: The Auditory Filter bank.*

Payton, K., Uchanski, R., & Braida, L. (1994). Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing. *J Acoust Soc Am.*, *95*(3), 1581–1592.

Penner, M. (1972). Neural or energy summation in poisson counting model. *J. Math. Psych.*, *9*, 286–293.

Peters, R., Moore, B., & Baer, T. (1998). Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people. *J. Acoust. Soc. Am.*, *103*(1), 577–587.

Popelar, J., Syka, J., & Berndt, H. (1987). Effect of noise on auditory evoked responses in awake guinea pigs. *Hearing Research*, *26*, 239–247.

Rankovic, C. (1991). An application of the articulation index to hearing aid fitting. *Journal of Speech and Hearing Research*, *34*, 391–402.

Reed, R. D. & Marks, R. J. (1998). *Neural Smithing: Supervised learning in feedforward artificial neural networks.* Cambridge, Mass.

Rieke, F., Warland, D., de Ruyter van Steveninck, R., & Bialek, W. (1997). *Spikes:Exploring the Neural Code.* MIT Press.

Roberts, R., Koehnke, J., & Besing, J. (2003). Effects of noise and reverberation on the precedence effect in listeners with normal hearing and impaired hearing. *Am J Audiol.*, *12*(2), 96–105.

Robinson, D. W. & Dadson, R. (1956). A re-determination of the equal-loudness relations for pure tones. *Brit. J. Appl. Phys.*, *7*, 166–181.

Ruggero, M. & Rich, N. (1987). Timing of spike initiation in cochlear afferents: Dependence on site of innervation. *J. Neurophysiol.*, *58*, 379–403.

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

Sachs, M. B. & Kiang, N. Y. S. (1968). Two-tone inhibition in auditory-nerve fibers. *J. Acoust. Soc. Am.*, *43*, 1120–1128.

Sachs, M. B., Winslow, R., & Sokolowski, B. H. A. (1989). A computational model for rate-level functions form cat auditory nerve fibers. *Hearing Research*, *41*, 61–70.

Schwartz, O. & Simoncelli, E. (2001). Natural sound statistics and divisive normalization in the auditory system. In Leen, T., Dietterich, T., & Tresp, V. (Eds.), *Advances in Neural Information Processing Systems 13*, (pp. 166–172). MIT Press.

Sek, A. & Moore, B. (1995). Frequency discrimination as a function of frequency measured in several ways. *J. Acoust. Soc. Am.*, *97*, 2479–2486.

Slepecky, N., Hamernik, R., Henderson, D., & Coling, D. (1982). Correlation of audiometric data with changes in cochlear hair cell stereocilia resulting from impulse noise trauma. *Acta. Otolaryngol.*, *93*, 329–349.

Small, A. (1959). Pure tone masking. *J. Acoust. Soc. Am.*, *31*, 1619–1625.

Smith, I. D. K. (1980).

Smoorenburg, G. (2004). Big ass study of compression. *IHCON*.

Smoski, W. & Trahiotis, C. (1986). Discrimination of interaural temporal disparities by nomral hearing listeners and listeners with high frequency sensorineural hearing loss. *J. Acoust. Soc. Am.*, *79*, 1541–1547.

Steeneken, H. (1992). *On measuring and predicting speech intelligibility.* PhD thesis, University of Amsterdam.

Steeneken, H. & Houtgast, T. (1980). A physical method for measuring speech-transmission quality. *J. Acoust. Soc. Am.*, *67*, 318–326.

TIMIT (1990). Timit acoustic-phonetic continuous speech corpus. *Speech Disc 1-1.1*(NTIS Order No. PB91-5050651996).

Trainor, L., Sonnadara, R., Wiklund, K., Bondy, J., Gupta, S., Becker, S., Bruce, I., & Haykin, S. (2004). Development of a flexible, realistic hearing in noise test environment (r-hint-e). *Signal Processing*, *84*(2), 299–309.

Tyler, R., Preece, J., & Lowder, M. (1987). The iowa audiovisual speech perception laser videodisc. *Laser Videodisc and Laboratory Report, Dept. of Otolaryngology, Head and Neck Surgery, University of Iowa Hospital and Clinics, Iowa City.*

Ukrainec, A. & Haykin, S. (1996). A modular neural network for enhancement of cross-polar radar targets. *Neural Networks*, *9*, 143–168.

Unnikrishnan, K. & Venugopal, K. (1994). Alopex: A correlation-based learning algorithm for feedforward and recurrent neural networks. *Neural Computation*, *6*(3), 469–490.

Uttley, A. (1970). *Information transmission in the nervous system.* London: Academic press.

van Rullen, R. & Thorpe, S. (2002). Surfing a spike wave down the ventral stream. *Vision Research*, *42*, 2593–2615.

van Schijndel, N., Houtgast, T., & Festen, J. (2001). Effects of degradation of intensity, time, or frequency content on speech intelligibility for normal hearing and hearing-impaired listeners. *J. Acoust. Soc. Am.*, *110*(1), 529–542.

van Son, R., Binnenpoorte, D., van den Heuvel, H., & Pols, L. (2001). The ifa corpus: a phonemically segmented dutch "open source" speech database. In *Proc. Eurospeech 2001*, Aalborg Denmark.

Verschuure, J. (1981). Pulsation patterns and nonlinearity of auditory tuning. ii. analysis of psychophysical results. *Acustica*, *49*, 296–306.

von Bekesy, G. (1960). *Experiments in hearing.* McGraw-Hill, Inc.

Westerman, L. & Smith, R. L. (1988). A diffusion model of the transient response of the cochlear inner hair cell synapse. *J. Acoust. Soc. Am.*, *83*, 2266–2276.

Wiener, F. & Ross, D. (1946). The pressure distribution in the auditory canal of a progressive sound field. *J. Acoust. Soc. Am.*, *18*(2), 401–408.

Wightman, F., McGee, T., & Kramer, M. (1977). Factors influencing frequency selectivity in normal and hearing impaired listeners, 295–306.

Yost, W. (1974). Discrimination of interaural phase differences. *J. Acoust. Soc. Am.*, *55*, 1299–1303.

Yost, W. & Nielsen, D. (1977). *Fundamentals of hearing.* Holt, Rinehart and Winston.

Young, E. & Sachs, M. (1979). Representation of steady-state vowels in the temporal aspects of the discharge patterns of the populations of auditory nerve fibers. *J. Acoust. Soc. Am.*, *66*(5), 1382–1403.

Zhang, X., Heinz, M., Bruce, I. C., & Carney, L. (2001). A phenomenological model for the responses of auditory-nerve fibers: I. nonlinear tuning with compression and suppression. *J. Acoust. Soc. Am.*, *109*(2), 648–670.

Zurek, P. & Formby, C. (1981). Frequency discrimination ability of hearing impaired listeners. *J. Speech Hear. Res.*, *24*, 108–112.

Zwisklocki, J. (1960). Theory of temporal auditory summation. *J. Acoust. Soc. Am.*, *32*, 1046–1060.