2011

DEVELOPMENT AND EVALUATION OF ENVELOPE, SPECTRAL AND TIME ENHANCEMENT ALGORITHMS FOR AUDITORY NEUROPATHY

Benjamin Robert Morgan

DEVELOPMENT AND EVALUATION OF ENVELOPE, SPECTRAL AND TIME ENHANCEMENT ALGORITHMS FOR AUDITORY NEUROPATHY

(Spine title: Speech enhancement for auditory neuropathy)

(Thesis format: Monograph)

by

Benjamin Morgan

Graduate Program in
Engineering Science
Electrical and Computer Engineering

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Engineering Science

School of Graduate Studies
The University of Western Ontario
London, Ontario, Canada

THE UNIVERSITY OF WESTERN ONTARIO

School of Graduate and Postdoctoral Studies

CERTIFICATE OF EXAMINATION

Supervisor:	**Examining Board:**
Dr. Vijay Parsa	Dr. Prudence Allen
	Dr. James Lacefield
	Dr. Ilia Polushin

The thesis by

Benjamin Robert Morgan

entitled:

Development and Evaluation of Envelope, Spectral and Time Enhancement Algorithms for Auditory Neuropathy

is accepted in partial fulfillment of the
requirements for the degree of
Master of Engineering Science

Date	Abbas Samani, Chair of the Thesis Examination Board

Abstract

Auditory neuropathy (AN) is a hearing disorder that reduces the ability to detect temporal cues in speech, thus leading to deprived speech perception. Traditional amplification and frequency shifting techniques used in modern hearing aids are not suitable to assist individuals with AN due to the unique symptoms that result from the disorder. This study proposes a method for combining both speech envelope enhancement and time scaling to combine the proven benefits of each algorithm. In addition, spectral enhancement is cascaded with envelope and time enhancement to address the poor frequency discrimination in AN. The proposed speech enhancement strategy was evaluated using an AN simulator with normal hearing listeners under varying degrees of AN severity. The results showed a significant increase in word recognition scores for time scaling and envelope enhancement over envelope enhancement alone. Furthermore, the addition of spectral enhancement resulted in further increase in word recognition at profound AN severity.

Keywords

Auditory neuropathy, hearing disorders, digital signal processing, speech processing, envelope enhancement, time scale modification, spectral enhancement, sentence recognition

Acknowledgements

There are many individuals who have contributed both directly and indirectly to my entire university life. From the beginning of my undergraduate degree to the final days of my Master's degree, I have been extremely fortunate to have the support of friends, colleagues and supervisors every step of the way.

First, I would like to thank my supervisor, Dr. Vijay Parsa, for his consistent and caring guidance and instruction. Ever since taking his 4th year digital signal processing course in undergrad, I have been drawn to the excitement, challenge and boundless potential that exists in this field of research. His ability to bridge the gap between engineering and other fields has been inspiring. Furthermore, I would like to thank Chris Allan for her kind help with the ethics approval process and Dr. Prudence Allen and Dr. David Purcell for being on my advisory committee.

Next many thanks go out to my engineering student colleagues, including lab buddies Julie Ladiges, Jonathan Pietrobon and Arvind Venkatasubramanian, for the countless study sessions, DSP talks and thesis support groups.

There are too many special friends and family members who have contributed in some way to this thesis. Special mention is required for my parents, Nancy and Dale (an electrical engineer who may have had some influence on this degree choice), sisters, Sara and Lisa, Uncle Jim (who helped with many a calculus problem and is also an electrical engineer), grandparents, and all other family members.

Finally, roommate Daniel Hall and friend Stephanie Varriano deserve honourable mention for listening to a plethora of engineering stories and providing infinite emotional support to a brain typically lost in algorithms and bots.

TABLE OF CONTENTS

List of Tables

List of Figures

Chapter 1

1 Introduction

1.1 Human Hearing

Hearing involves the perception of sound through an auditory system. The human auditory system is composed of the outer, middle and inner ears, as well as the central auditory nervous system. Figure 1-1 [1] illustrates the basic anatomy of the human ear.

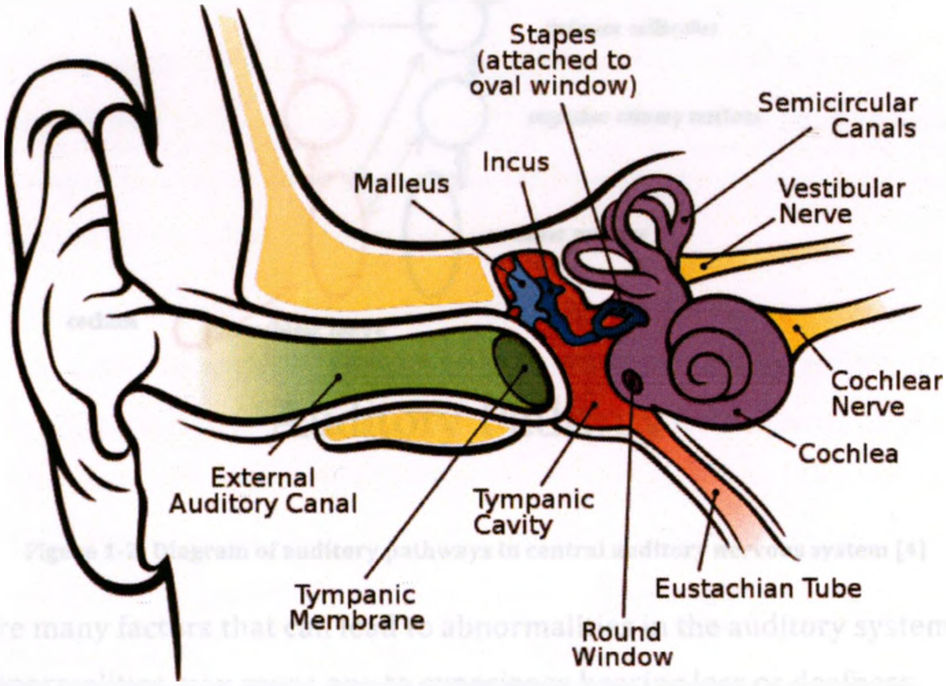

Figure 1-1: Anatomy of the human ear [1]

The outer ear consists of the pinna (the visible part of the ear) and external auditory canal, which leads to the tympanic membrane (eardrum). Next, the middle ear consists of the tympanic membrane and three middle ear bones: the malleus, incus and stapes. This anatomy provides an effective means to deliver sound to the inner ear. Neural processing begins in the inner ear. The inner ear can be divided into three parts: the semicircular canals, the vestibule and the cochlea. The cochlea contains the primary auditory organ of the ear [2] and it is here that vibrations are converted to electric signals that travel to the central auditory nervous system. The central auditory nervous system transmits and processes electrical stimuli from the inner ear. The flow of information starts in the auditory nerve and then travels to the brainstem and then to the auditory cortex [3]. A simplified diagram of the auditory pathways is provided in Figure 1-2 [4].

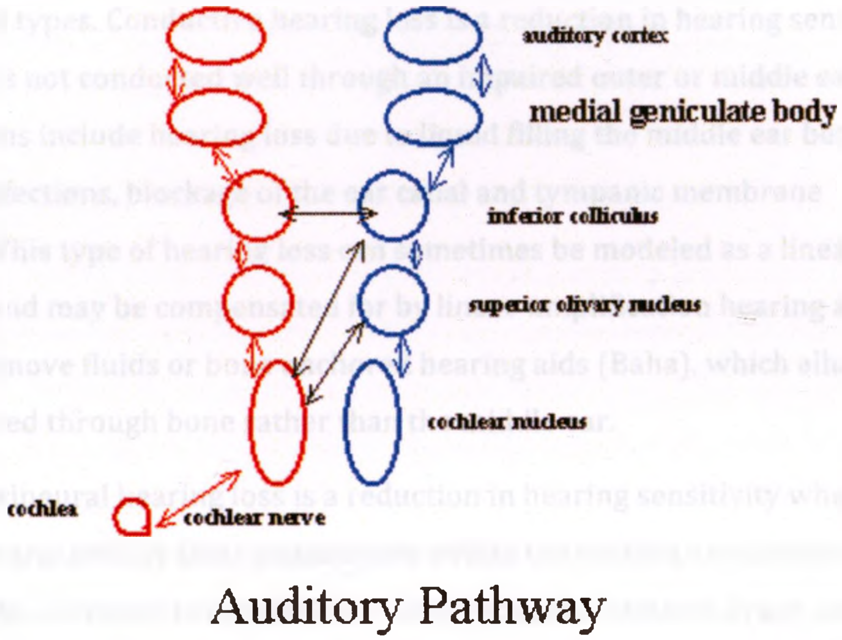

Figure 1-2: Diagram of auditory pathways in central auditory nervous system [4]

There are many factors that can lead to abnormalities in the auditory system and these abnormalities may cause one to experience hearing loss or deafness.

1.2 Hearing Disorders

Over the last 20 years, the hearing loss population has been increasing steadily. In Canada, over 1 million adults have reported having a hearing-related disability [5]. Between the years 2000 and 2004, the hearing loss population in the United States grew 9.9% compared to a 6.8% increase in US households [6]. In 2004, that equated to over 31 million people in the US and that number has been projected to increase to 53 million people by the year 2050 [6]. These numbers cost the US more than $56 billion annually [7]. To improve communication for these affected people, much research and effort has been put toward assistive listening devices. These devices range from hearing aids to cochlear implants and have achieved impressive results for many different types and severities of hearing disorders.

1.2.1 Common Forms of Hearing Loss

Hearing loss has traditionally been classified into conductive and sensorineural types. Conductive hearing loss is a reduction in hearing sensitivity when sound is not conducted well through an impaired outer or middle ear [8]. Common forms include hearing loss due to liquid filling the middle ear because of middle ear infections, blockage of the ear canal and tympanic membrane perforation. This type of hearing loss can sometimes be modeled as a linear attenuation and may be compensated for by linear-amplification hearing aids, surgery to remove fluids or bone anchored hearing aids (Baha), which allow sound to be conducted through bone rather than the middle ear.

Sensorineural hearing loss is a reduction in hearing sensitivity when the sensory or neural cells or their connections within the cochlea are absent or not functioning [8]. Common causes include noise exposure, ototoxic drugs and aging [7]. The result is an abnormal amplification function provided by the outer hair cells in the cochlea. Therefore hearing aids with nonlinear compression circuits can be produce significant benefits for individuals with sensorineural hearing loss, especially those demonstrating loudness recruitment.

However, in the last 20 years, a new type of hearing disorder has been discovered that does not fall into either of the above categories. Auditory neuropathy (AN) consists of normal or near-normal cochlear function, but abnormal auditory nerve responses. In many cases, patients report: "I can hear but do not understand" [7]. Auditory neuropathy will be discussed in detail next.

1.3 Auditory Neuropathy

Auditory neuropathy is a hearing disorder that affects the timing of neural activity in the auditory pathway and disrupts temporal aspects of auditory perception [9]. Unlike the most common type of hearing loss which is caused by damage to the outer hair cells, AN can result from damage to the inner hair cells, the synapse between the inner hair cells and the auditory nerve, and/or the auditory nerve or brainstem pathways. However, the exact causes or treating methods are not well understood. The impact on an individual with AN is decreased speech recognition abilities.

1.3.1 Background Information

An interesting condition consisting of absent auditory brain-stem responses (ABR) but near-normal audiograms was first reported in the early 1980s [10],[11],[12]. Around this time, it was estimated that this was the case in 1% of the clinical population and 12-14% of those who would otherwise have been thought to have a severe-to-profound cochlear hearing loss [13]. Currently, it is estimated that 10% of children seen with severe-to-profound deafness may have a neural rather than hair cell disorder [14],[15].

Over the last 30 years, a number of names have been assigned to this condition, including *paradoxical, brain-stem auditory processing syndrome, central auditory dysfunction* and *neural synchrony disorder*. However, in the last 10 years the term *auditory neuropathy* has been adopted by most of the field.

Before these discoveries were made, absent ABRs meant clinicians concluded definitively that the individual was "deaf," resulting in people with AN, who actually had normal cochlear function and were responsive to sound, being misdiagnosed and treated incorrectly (as they were assumed to have conventional hearing loss). Fortunately, the addition of otoacoustic emission (OAE) testing (a test of cochlear function) in the 1990s has helped clinicians with the diagnosis of AN [13].

AN can occur in absence of any apparent medical problem, but has also been linked to a variety of medical conditions, including infectious processes, immune disorders and various genetic or syndromal conditions. Most cases are congenital or occur at an early age, but it has been reported that an onset of AN can occur at any age [16]. In addition, some cases are transient or intermittent, some change little over time and some worsen [13].

The exact physiological causes remain unknown despite a significant amount of research over the last 30 years. Much of this can be attributed to the fact that AN is a heterogeneous disorder that is classified by common characteristics in auditory function, making isolation of precise etiologies very difficult. The common psychoacoustic effects of the disorder have been closely examined and the results are explained in Section 1.3.3.

1.3.2 Definition and Diagnosis

In order to be clinically diagnosed as having auditory neuropathy an individual must meet all of the following criteria [16]:

1. Evidence of poor auditory function (hearing) in at least some situations or for some stimuli regardless of pure tone thresholds must be demonstrated.
2. Evidence of poor auditory neural function must be demonstrated. At a minimum, the patient must have elevated or absent auditory brain-stem reflexes and abnormality of the ABR. A severe case would show no clear ABR waveform to any click or stimuli, while a mild case might have a poor ABR morphology or abnormal peak latency for fast clicks.
3. Evidence of normal hair cell function must be demonstrated. Most patients have OAEs, but in the small percentage that do not, the cochlear microphonic (CM) can be used to evaluate cochlear function. One of these two readings must be present to be considered for AN.

In general, it has been summarized that most individuals with AN exhibit the following characteristics:

- Elevated thresholds on pure-tone audiogram by air and bone conduction
- Very poor speech discrimination for degree of loss (in comparison to audiogram), particularly in noise
- No acoustic reflex in any configuration for any stimuli
- No ABR even with stimuli well above detection threshold
- Evidence of large CM in auditory brain-stem response recordings
- Present OAEs to low-level stimuli.

The combined effects of these symptoms on auditory processing and speech recognition are evaluated in the next section.

1.3.3 Psychoacoustics and Speech Perception

When developing assistive hearing devices, it is critical to understand the psychoacoustic effect caused by the given disorder. In [17] and [18], in depth studies were conducted to describe the psychoacoustic profile of people with AN. The goals of these studies were to characterize the functional capabilities in people with AN (*i.e.* understand why they can hear sounds but not understand speech), develop behavioral tests that can help determine underlying physiological mechanisms and differentiate hearing loss of different origins, and provide guidance for designing auditory prosthesis and rehabilitation strategies [17], [18]. These studies addressed these goals through an evaluation of intensity, frequency and temporal processing parameters and their results are discussed next.

1.3.3.1 Temporal Processing

In order to describe the temporal processing of people with AN, their temporal integration, gap detection and temporal modulation transfer data were collected [17]. In addition, audiograms were measured. The results give rise to a number of important observations.

Audiogram:

First, an audiogram plots pure tone thresholds for different frequencies and is a standard way of representing someone's hearing abilities. The results indicated that the AN patients who participated in this study generally had moderate to severe hearing loss at low frequencies and mild to moderate hearing loss at high frequencies.

Temporal Integration:

Temporal integration functions are used to determine the subject's intensity threshold to detect noise bursts of different durations. The significance of this test in relation to auditory neuropathy is that it provides insight into whether their speech recognition deficits are related to their inability to hear short-duration sounds. However, the results in [17] and [18] show that most subjects have normal or near-normal temporal integration functions allowing one to speculate that this particular psychoacoustic dimension need not be focused on while developing assistive algorithms for AN.

Gap Detection:

Gap detection thresholds are a measure of the subject's ability to detect short silence intervals (gaps) in acoustic signals. The aforementioned studies reveal that, in contrast to temporal integration, gap detection thresholds are uniformly impaired in the AN patients. It is intuitive to note the diminishing effects impaired gap detection would have on speech perception, particularly for discrimination at high syllabic rates. This idea will be explored in Chapter 3, particularly for the development of time enhancement algorithms.

Temporal Modulation:

Temporal modulation thresholds show the sensitivity to slow and fast temporal fluctuations. In other words, it is a measure of the threshold for detecting changes in the amplitude of a sound as a function of modulation frequency. The aforementioned study demonstrated a consistent impairment across all AN subjects for both slow and fast temporal fluctuations. This information is also very important when designing assistive algorithms for AN patients as the goal is to accommodate for these psychoacoustic impairments.

1.3.3.2 Frequency Processing

Frequency processing in AN subjects was examined in [18]. A frequency discrimination task was carried out and the results showed that, in general, people with AN had very poor frequency discrimination at low frequencies (<2000 Hz), but above 4000 Hz their results were indistinguishable from the normal hearing (NH) listeners. Discrimination in the middle frequency region (1000-3000 Hz) was also impaired.

The significance of this poor frequency discrimination in speech perception is that it may pose a problem for discriminating second formant frequencies of two spectrally closely spaced vowels [18]. Furthermore, poor discrimination at lower frequencies (<2000 Hz) may be related to the role of temporal cues in pitch encoding of low frequencies [18].

1.3.3.3 Intensity Processing

Loudness growth measures and intensity discrimination data were collected for subjects with AN in [18]. It was concluded by the authors that intensity processing is likely not a major factor in contributing to poor speech perception in people with AN.

1.3.3.4 Conclusions from Psychoacoustic Analysis

Speech understanding depends on the processing of subtle speech cues in the signal. The studies conducted in [17] and [18] suggest that temporal processing (particularly modulation thresholds and gap detection) and frequency discrimination are areas of significant impairment for people with AN, and that these are potentially the major factors in their poor speech processing. Therefore when designing assistive algorithms to increase speech intelligibility for these people, a method to bring the temporal and frequency characteristics of speech into within their thresholds should be explored.

1.3.4 Use of Traditional Amplification Techniques

There is considerable controversy over the use of conventional hearing aids for people with AN. Although some studies have shown that 50% of affected children benefit from conventional amplification hearing aids, others have shown detrimental effects, including loss of OAEs (some without any change in pure-tone sensitivity) and permanent threshold shifts [19]. Unfortunately, it has not been possible for researchers to predict which children would benefit from amplification.

As a compromise, some have suggested that if more powerful hearing aids are required for children with AN, they should be worn only for limited periods and only in one ear. This would prevent permanent threshold shifts in at least one ear of the child [19]. However, another study showed that if the hearing aids are fitted conservatively and with careful attention to parameter verification, threshold shifts can be avoided in all but those with severe-to-profound hearing loss [19].

In summary, when compared to other techniques such as cochlear implantation, hearing aids are less intrusive and studies have shown some success in avoiding permanent threshold shifts. It is generally recommended to trial hearing aids as a first step before considering cochlear implantation, and it has been shown to be just as effective for some children [20].

1.3.5 Use of Cochlear Implantation

Cochlear implants are devices that bypass the inner ear and provide direct electrical stimulation to the auditory nerve. It involves a surgical implantation of an array of electrodes into the cochlea, thus is invasive and expensive. Cochlear implantation is routinely performed on patients with sensorineural losses where the cochlea is the primary site of dysfunction [21]. It has the potential to be an excellent assistive device for AN if the source of dysfunction is at the inner hair cells or synapse because it bypasses these stages and directly stimulates the auditory nerve. Also, in the case of neural demyelination or axonal loss, the cochlear implant may be more effective than the hearing aid because electrical stimulation has been known to produce more synchronized neural activity than any acoustic stimulation [18]. If a normalized firing pattern is induced in the auditory nerve through electrical stimulation, the result should be better speech perception for people with AN. Furthermore, electrical stimulation has shown potential to promote neural survival which may in turn restore temporal encoding. Nonetheless, the lack of understanding of the physiological causes for AN mean that cochlear implantation does not guarantee improved speech perception. Pathology of the auditory nerve is suggested as a possible root cause for abnormal ABRs and in this case, the usefulness of cochlear implants is questioned [21].

Despite the potential benefits of cochlear implants, drawbacks also exist. These drawbacks include high cost, standard surgical risks, cochlear damage resulting from insertion of the electrode array (thus destroying residual hearing) [22] and no guarantee of oral speech communication skills development [21].

In summary, some studies have shown considerable speech perception improvement with cochlear implantation, while others have shown no improvement. In some cases, conventional hearing aids have been just as effective [20]. It remains unclear what distinguishes poor implant users from successful users. Therefore, as mentioned in the previous section, the best current practice is to have a trial period with conventional amplification methods before implantation is considered [20].

1.4 Problem Statement

Many people with auditory neuropathy suffer from very poor speech discrimination, especially in the presence of noise. The greatest shortcoming of conventional hearing aids being used for AN patients is that they do not attempt to compensate for the impaired temporal processing. It has been shown that temporal processing deficits are likely a major contributor to poor speech discrimination for these people. Cochlear implantation shows some promise, yet success remains unpredictable and surgical procedures are best avoided if possible. Thus, a need exists for specialized algorithms to be researched and developed that could be implemented on a hearing aid platform to achieve greater and more consistent performance.

Currently, temporal envelope enhancement algorithms, which exaggerate the temporal peaks and valleys of a speech signal, have shown promising results in AN patients [9], [23], [24]. Furthermore, the insertion of silence intervals between consonant-vowel pairs and formant transitions in consonants showed improved speech perception in subjects with AN [25]. In addition, spectral enhancement algorithms have been shown to be beneficial to some cochlear implant and hearing aid users [26], [27]. These algorithms increase the spectral contrast, thus allowing for increased contrasts between speech formants.

However, no known algorithm has been developed or tested which accounts for the multiple impairments proven to be common in many AN patients, namely: poor modulation thresholds, gap detection thresholds and frequency discrimination.

1.5 Proposed Solution and Objectives

The proposed solution is to develop and test algorithms that provide enhancement for the three areas of impairment listed above (modulation thresholds, gap detection thresholds and frequency discrimination) in an attempt to achieve greater improvement than envelope enhancement and silence-insertion achieve as stand-alone algorithms. The proposed solution combines envelope enhancement, time enhancement, and spectral enhancement strategies into a realizable algorithm that could potentially be implemented in a portable assistive device.

More specifically, the objectives are:

- Development of algorithms which combine benefits from envelope enhancement, time enhancement and spectral enhancement in different combinations
- Subjectively test these algorithms for sentence-level speech perception with normal hearing listeners and an AN simulator
- Compare the results obtained using the proposed algorithms to those garnered using the envelope enhancement alone in [9]
- Suggest realtime implementation strategies for the proposed algorithm as well as provide ideas for future work and development.

1.6 Thesis Layout

Chapter 2 *Literature review*: current envelope enhancement, time enhancement and spectral enhancement algorithms are described, as well as their results and achievements

Chapter 3 *Implementation*: development of the proposed algorithm is described, as well as the methodology used to recreate the algorithms described the literature review

Chapter 4 *Subjective data collection and analysis*: the testing procedures for normal hearing listeners (using an AN simulator) are described and the results are analyzed for statistical significance and explained

Chapter 5 *Conclusions and future work*: conclusions from the study are stated, future development techniques are suggested and ideas for practical realtime implementation in a portable assistive hearing device are explored

Chapter 2

2 Literature Review

This chapter discusses three speech enhancement algorithms *viz.* envelope, time and spectral enhancement, and their application to improving speech recognition in people with auditory neuropathy. Envelope enhancement and duration modification of consonant-vowel pairs and formant transitions in consonants have both shown favorable results for increasing word recognition in people with AN. Spectral enhancement has not yet knowingly been applied to an AN study.

2.1 Envelope Enhancement

Most of the work completed in the field of AN speech enhancement has been related to temporal envelope enhancement (EE) [9], [23], [24]. These studies have shown an increase in word identification scores when the envelope of the speech was enhanced.

2.1.1 Principles

As mentioned in Section 1.3.3, studies indicate that there are three major psychoacoustic impairments that likely contribute the most to degraded speech perception in people with AN: temporal modulation thresholds, gap detection thresholds and frequency discrimination. Envelope enhancement primarily focuses on accounting for poor modulation thresholds by reinforcing temporal speech cues.

Low frequency modulations are prominent in the temporal envelope of continuous speech. In all frequency bands, the most dominant modulation frequencies are 3-4 Hz (the average syllabic rate in speech) [28]. Studies have revealed that speech processed to have minimal spectral information is still intelligible given that these low-frequency modulations (2-50 Hz) are preserved

[29]. This emphasizes the importance of the temporal envelope to speech recognition.

Although envelope enhancement has been shown to degrade speech quality in quiet for normal hearing listeners, it has shown positive results when presented with background noise [30] and very promising results when tested on AN patients.

Essentially, envelope enhancement exaggerates the temporal envelope of the signal. Considering the frequency threshold for amplitude modulation is typically inversely proportional to the modulation depth, it is intuitive that increasing the modulation depth of the speech envelope would improve the temporal processing for people with AN.

Different expansion schemes have been tested, but a 'power law expansion scheme' has demonstrated the most consistent results [9], [29], [31]. Power law expansion implies raising the envelope of the signal to varying powers depending on its amplitude at a given point in time. Effectively, this produces a modified consonant to vowel ratio as portions of the signal with higher input amplitudes generally correspond to vocalic components of the sentence and portions with lower input amplitude typically correspond to consonant regions. As a result, envelope enhancement has also been shown to improve the identification of consonants [23]. An example of an envelope-enhanced signal is shown in Figure 2-1.

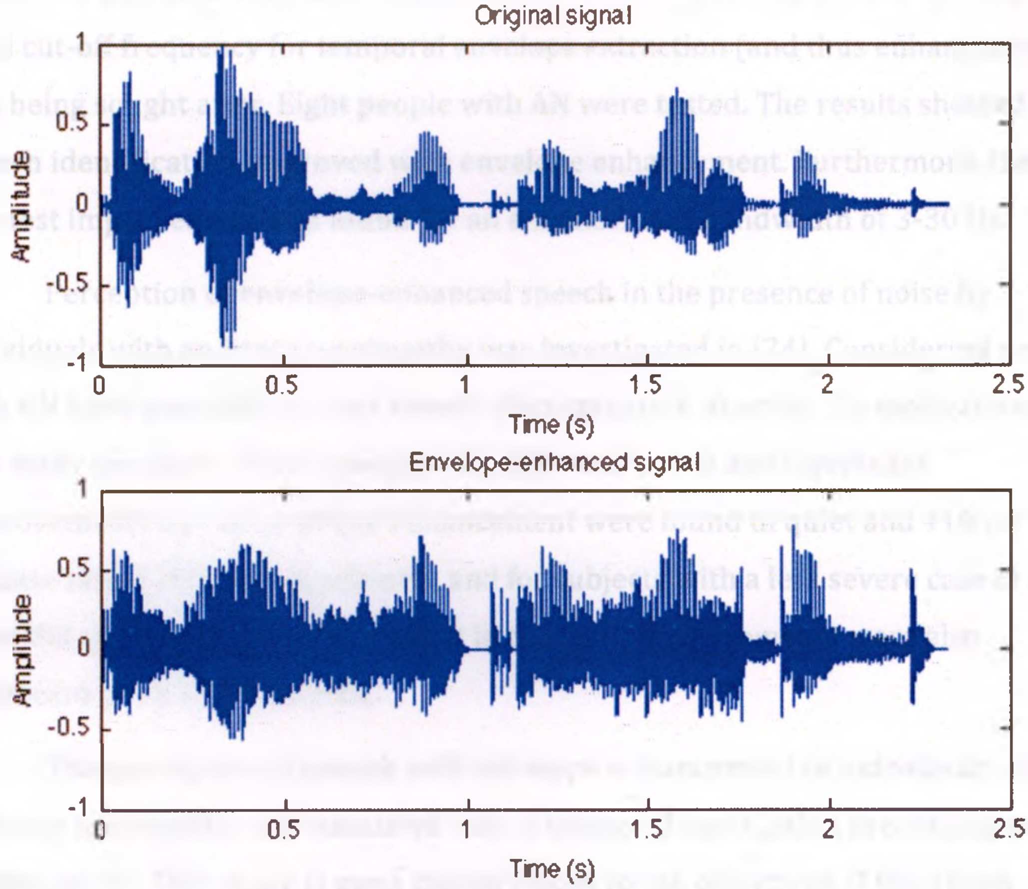

Figure 2-1: Comparison of speech signals with and without envelope enhancement for the utterance, "the wife helped her husband"

The procedures used to expand the sentences in Chapter 4 are derived from [31] and their implementation is described in Chapter 3.

2.1.2 Published Results

A survey of the literature revealed mixed performance results for a variety of expansion schemes. The only known approach used for AN testing was demonstrated by Narne and Vanaja in [9], [23], [24] (for various conditions). Therefore only the results from these studies will be discussed.

The effect of envelope enhancement on speech perception in individuals with auditory neuropathy was investigated in [23]. More specifically, the objectives were to investigate the ability of individuals with AN to identify consonant-vowel (CV) stimuli for different modulation bandwidths of enhancement. In other words, an ideal cut-off frequency for temporal envelope extraction (and thus enhancement) was being sought after. Eight people with AN were tested. The results showed that speech identification improved with envelope enhancement. Furthermore, the greatest improvement was found for an enhancement bandwidth of 3-30 Hz.

Perception of envelope-enhanced speech in the presence of noise by individuals with auditory neuropathy was investigated in [24]. Considering people with AN have particularly poor speech discrimination in noise, the motivations for this study are clear. Fifteen people with AN were tested and significant improvements due to envelope enhancement were found in quiet and +10 dB signal to noise ratio (SNR) for all subjects, and for subjects with a less severe case of AN (who did not exhibit the a floor effect in the test), improvements were also significant at +5 and 0 dB SNR.

The perception of speech with envelope enhancement in individuals with auditory neuropathy and simulated loss of temporal modulation processing was studied in [9]. This study is most closely linked to the objectives of this thesis. Hence, it is discussed below in greater detail than the previous two studies. Results from two experiments were reported in [9]. In Experiment I, an AN simulator was used to test the effectiveness of the envelope enhancement on 12 normal hearing listeners. The parameters of the AN simulator were adjusted to simulate mild, moderate, severe and profound degrees of neuropathy. The test stimuli consisted of bi-syllabic words in Kannada (a language spoken in a southern state of India). Speech scores were calculated by counting the number of words correctly repeated and converting to a percentage of total words presented. Results revealed a significant main effect of degree of AN and a significant interaction between the degree of AN and stimuli (processed vs. unprocessed) and a significant difference between mean identification scores across all degrees of simulation.

In the second experiment, 12 people with AN and 12 normal hearing listeners were recruited to compare the results of envelope-enhanced speech to unprocessed speech. Word recognition scores were obtained using the same test stimuli from Experiment I. Statistical analysis of the recognition data showed a significant improvement in speech scores for envelope-enhanced stimuli with AN subjects, but no significant differences were found with normal hearing subjects.

2.1.3 Significance to Thesis

Envelope enhancement has been shown to benefit word identification for people with AN. However, to date it has only been evaluated at word level and not sentence level for AN subjects. Therefore to make further contribution to the field in this research area, the thesis will apply envelope enhancement to sentence-level speech perception tasks and use it as a yardstick for comparing the performance of other enhancement techniques presented in this thesis, in isolation as well as in combinations. In order to compare results to [9], the subjective data analysis described in Chapter 4 will have a similar approach.

2.2 Time Enhancement

Time enhancement in this thesis refers to enhancing or changing the duration of a speech signal (or portions of the speech signal) such that temporal speech cues become more obvious to detect. One known study has analyzed the effects of duration modification on speech discrimination for people with AN [25]. This section describes how time enhancement applies to AN and contains a summary of [25].

2.2.1 Principles

Impaired gap detection thresholds have been established as one of the major psychoacoustic shortcomings in people with AN [17], [18]. It has been shown that individuals with AN typically require silent periods greater than 20 ms compared to less than 5 ms for normal hearing listeners for detecting gaps between noise bursts [20]. In speech, these short silence intervals correspond to silence periods in stop consonants. As demonstrated previously, the detection and processing of temporal cues are critical for speech discrimination. Even for normal hearing listeners, if speech is played at a speed that is beyond their gap detection thresholds, the listener will not detect critical temporal cues. Important cues include formant transitions, voice onset time (VOT) and duration of silent intervals before and after words, or before and after sounds within a word [32]. Studies have demonstrated poor detection for certain stop-consonant pairs in people with AN and it has been speculated that short VOTs were most likely the cause of error. Figure 2-2 illustrates the VOT for the word /paw/.

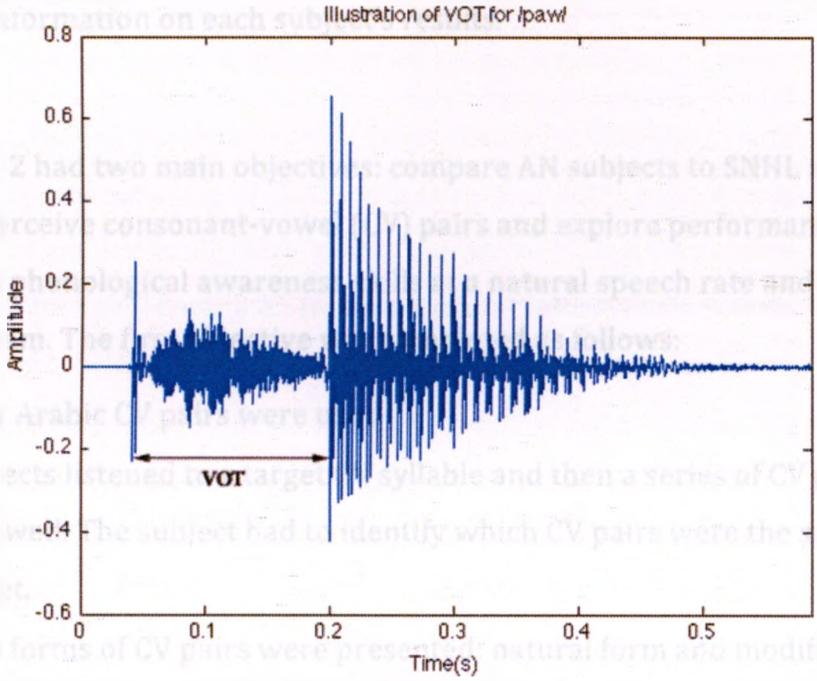

Figure 2-2: Illustration of VOT for /paw/

Furthermore, sentence recognition scores of clear speech (with an average duration of 3.3 seconds) and conversational speech (with an average duration of 1.5 seconds) were compared with AN subjects in [33]. The longer-duration clear speech showed significantly greater intelligibility, suggesting that drawn out speech cues and enhanced temporal modulations provide significant benefit to AN patients [25].

2.2.2 Published Results

As mentioned above, [25] is the only known duration modification study for AN. The goal of this study was to investigate the ability of AN subjects to perceive temporally modified CV pairs. Two tests were completed with a study group consisting of 14 subjects with AN and a control group of 14 subjects with moderate, bilateral, symmetrical sensorineural hearing loss (SNHL).

Test 1:

The goal of test 1 was to evaluate the hearing status of both study groups to determine their temporal processing ability (through pure tone gap detection). This was used to determine suitability for the experimental tasks and to provide insightful information on each subject's results.

Test 2:

Test 2 had two main objectives: compare AN subjects to SNHL subjects in ability to perceive consonant-vowel (CV) pairs and explore performance of AN subjects on phonological awareness skills at a natural speech rate and a prolonged speech stream. The first objective was completed as follows:

- Four Arabic CV pairs were used
- Subjects listened to a target CV syllable and then a series of CV pairs followed. The subject had to identify which CV pairs were the same as the target.
- Two forms of CV pairs were presented: natural form and modified form.
- In the natural form, only pauses between CV pairs were expanded (interstimulus interval (ISI)) and the consonant durations were unchanged. Thus the VOTs were around zero for voiced stops (/gi/, /do/) and around 30 ms and for voiceless stops (/ki/, /to/). The ISI changed from 1000 ms to 100 ms with 100 ms difference between each step (i.e. a stream of words with decreasing ISI between each word was presented until a certain 'percent correct' threshold was met).
- In the modified form, formant transitions occurring in consonants and the pauses between CV pairs were prolonged (i.e. VOT for stop consonants and fricative vowel gap for fricatives were expanded). An undocumented time scale modification was used to prolong the CV stimuli. The ISI changed from

1000 ms to 300 ms with 100 ms difference between each step. The prolongations had the following values: 250, 200, 150, 100, 80, 60, 40, and 20 ms. The modifications were presented in 8 steps, with the first step consisting of 1000 ms ISI and 250 ms prolongation, and the last step consisting of 300 ms ISI and 20 ms prolongation. The ISI and prolongation times decreased until a certain threshold of 'percent correct' was met.

There were two salient results from this test: (a) the AN subjects performed most poorly on discriminating stop consonants, and (b) the time-modified form showed a significant increase in performance for all CV pairs for the AN subjects.

Phonological awareness skills were evaluated through rhyme detection, segmentation and blending tests. Speech stream prolongation was applied to the segmentation and blending tests, where the prolongation was achieved by inserting a silence gap (ranging from 0.25 to 2 seconds) at intersyllabic points. Results showed that the SNHL subjects completed the task down to an ISI of 300 ms for all CV pairs. In addition, AN subjects scored quite poorly for all skills at natural speech rates while SNHL subjects were near-perfect. Interestingly, inserting a silence gap between syllables produced a significant improvement for the segmentation and blending skills.

2.2.3 Significance to Thesis

Time scale modification of CV pairs and insertion of silence intervals at intersyllabic points have shown significant benefits for subjects with AN [25]. Furthermore, studies have shown that speaking slowly in general improves speech recognition for AN subjects. However, similar to envelope enhancement, sentence-level testing of time-expansion schemes have not been tested for AN applications and the time scale modification schemes used in [25] are primitive and require manual insertion of silence intervals for words and prolongation of consonants for CV pairs. Furthermore, testing in combination with other temporal enhancement schemes such as envelope enhancement have not been evaluated. Therefore, the thesis will explore the use of sentence-level time scale modification algorithms that automatically expand the duration of sentences while preserving key temporal cues. Furthermore, these time enhancements will be combined with other forms of enhancement in an attempt to obtain maximal benefits.

2.3 Spectral Enhancement

Poor frequency discrimination at lower frequencies (<2000 Hz) is a common characteristic of people with auditory neuropathy. Spectral enhancement (SE) is a process that emphasizes the spectral peaks and valleys (as opposed to emphasizing the temporal peaks and valleys in envelope enhancement) to aid with frequency discrimination. However, no known studies have evaluated the benefit of spectral enhancement with auditory neuropathy subjects. The next section describes the principle of spectral enhancement and its application to the thesis.

2.3.1 Principles

Detecting the spectral shapes of speech signals, particularly of formants, is very important in speech discrimination [34]. The extraction of spectral information is particularly critical when speech is degraded by noise, as the noise may mask the important spectral cues. Normal hearing listeners have an impressive ability to extract this information, but the ability decreases rapidly with hearing loss [27]. Thus, algorithms that emphasize the dominant frequencies (usually formants) and reduce the remaining frequencies have been proposed as a form of speech enhancement for individuals with poor frequency discrimination. If auditory systems in individuals with poor frequency discrimination can be described as convolving the spectrum with a smoothing function, then SE can be described as a partial deconvolution process [35]. The benefit of SE has been evaluated on both hearing aid and cochlear implant users.

Mixed results have been found for the effectiveness of SE and the benefit appears to be quite contingent on parameter selection. However, most research studies show slightly increased speech recognition scores in noise [36]. In [34], SE was shown to improve vowel perception in most of their hearing aid subjects (primarily with cochlear hearing loss) and in [35], some sentence-level improvements in SNHL subjects were demonstrated.

Cochlear implant users require a much higher SNR to match the performance of NH listeners on speech discrimination in noise. SE was applied as an enhancement scheme for CI users in [26], [27]. These studies used a new approach for spectral enhancement that mimics two-tone suppression, a spectral-sharpening phenomenon found in the auditory system.

2.3.2 Two-Tone Suppression

It has been shown that a combination of SE and temporal compression yield improved speech perception in noise for hearing aid patients [35]. The approach in [26] aimed to improve upon the previous work by implementing an algorithm that inherently performs frequency-dependent syllabic compression, as well as spectral sharpening.

The outer hair cells (OHCs) in the cochlea perform non-linear processing to provide the necessary temporal and spectral resolution [27]. Complex interactions between the OHCs and the basilar membrane result in a nonlinear phenomenon called two-tone suppression. The effect of two-tone suppression is a decrease in the evoked response to a tone in the presence of a second tone. It is considered the primary mechanism for spectral enhancement in the auditory system and is thought to improve the SNR of stronger components [27]. It has been found that compression algorithms inhibit the perception of vowels due to the inherent degradation of spectral contrast: a weak tone at one frequency is strongly amplified to the same level as a weakly amplified strong tone at a different frequency. With two-tone suppression, the compression is prevented from degrading spectral contrast in regions close to a strong spectral peak, but audibility can still be increased at regions distant from the spectral peak [26]. Figure 2-3 is an excerpt from [27] that illustrates the spectral effect of two-tone suppression on a speech signal. Results from [27] indicated a significant improvement in the recognition of both phonemes and sentences in noise for cochlear implant users.

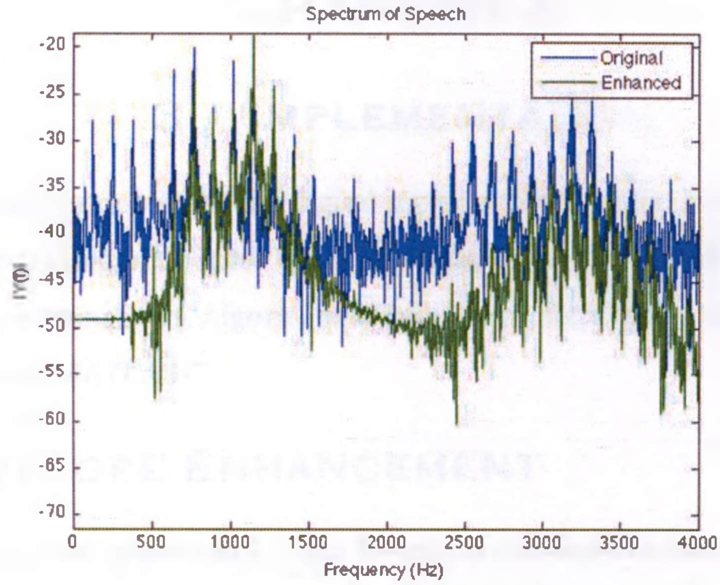

Figure 2-3: Effect of spectral enhancement on the spectrum of a vowel

2.3.3 Significance to Thesis

No known studies have evaluated the benefit of spectral enhancement in auditory neuropathy. However, poor frequency discrimination has been found in many people with AN. AN research has shown that nasal sounds are difficult to perceive due to the impaired ability to use the low frequency spectral cues, whereas fricatives (/s/, /sh/) were the easiest to perceive as a result of preserved accurate high frequency pitch discrimination in subjects with AN [25].

The two-tone suppression technique that combines frequency-dependent compression with spectral sharpening may be beneficial in accounting for the poor spectral discrimination in people with AN, in reducing the effect of spectral degradation caused by envelope enhancement, and in further enhancing temporal cues (through compression) as a preprocessor for the envelope enhancement algorithm. These questions will be explored through an implementation of the two-tone suppression approach in combination with envelope and time enhancements algorithms.

CHAPTER 3

3 IMPLEMENTATION

This chapter describes the digital signal processing (DSP) algorithms used for the subjective testing in Chapter 4 *viz.* envelope, time and spectral enhancement, as well as their combinations. Algorithm development, debugging and testing was completed using MATLAB-7.

3.1 ENVELOPE ENHANCEMENT

As described in Section 2.1, the benefit of envelope enhancement for speech perception in AN has been verified for CV pair and bi-syllabic word recognition. The goal of the present study was to evaluate the effectiveness of envelope enhancement on sentence-level speech perception for people with AN as well as use the results as a base for comparison with other proposed solutions. The algorithms described in [9], [23], [24], [30] were modified slightly for the purpose of this study.

3.1.1 IMPLEMENTATION OF ORIGINAL ALGORITHM

The envelope enhancement algorithm implemented in [9] is depicted in the block diagram shown in Figure 3-1.

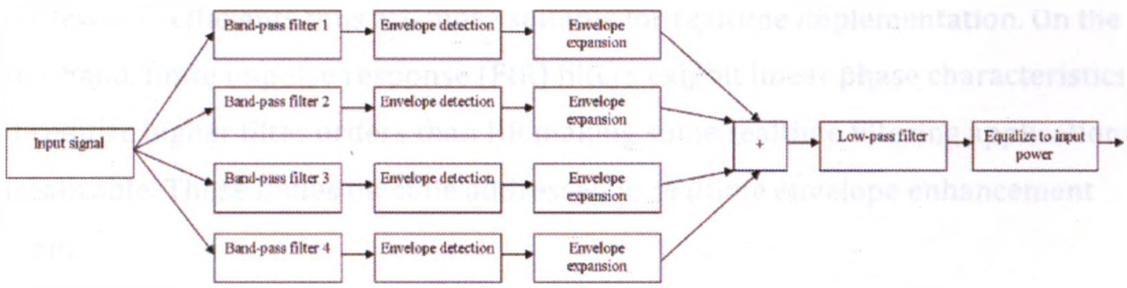

Figure 3-1: Block diagram of envelope enhancement algorithm

Filter bank:

The algorithm first divided the speech signal into a specified number of bands (in this case four) to provide frequency-dependent processing. In [9], it is claimed that the type of filters used was 3rd order Butterworth. However, an odd order bandpass digital Butterworth filter cannot be designed. Therefore 6th order Butterworth bandpass filters were designed (it was assumed this was the intention of the authors). This proved to provide robust and accurate results. The cut-off frequencies for the bandpass filters were specified as 150-550 Hz, 550-1550 Hz, 1550-3550 Hz and 3550-8000 Hz.

Envelope extraction:

Next, the envelope of the signal was extracted through full-wave rectification followed by lowpass filtering. A 1st order Butterworth filter was used with a cut-off frequency of 32 Hz, as [23] showed this provided optimal results. Although it was not specified, careful attention was given to the filter delay to ensure that the extracted envelope did not lag the actual envelope of the signal. A comparison between an envelope extracted using zero-phase filtering and one filtered without phase correction is shown in Figure 3-2. This figure suggests that the algorithm may miss critical short-duration peaks and cues, effectively nullifying important characteristics of temporal enhancement, if the phase delay is not accounted for. Infinite impulse response (IIR) filters exhibit nonlinear phase delay functions, but have fewer coefficients thus are more suitable for realtime implementation. On the other hand, finite impulse response (FIR) filters exhibit linear phase characteristics, but require higher filter orders than IIR making some realtime filtering applications unrealizable. These issues must be addressed in realtime envelope enhancement systems.

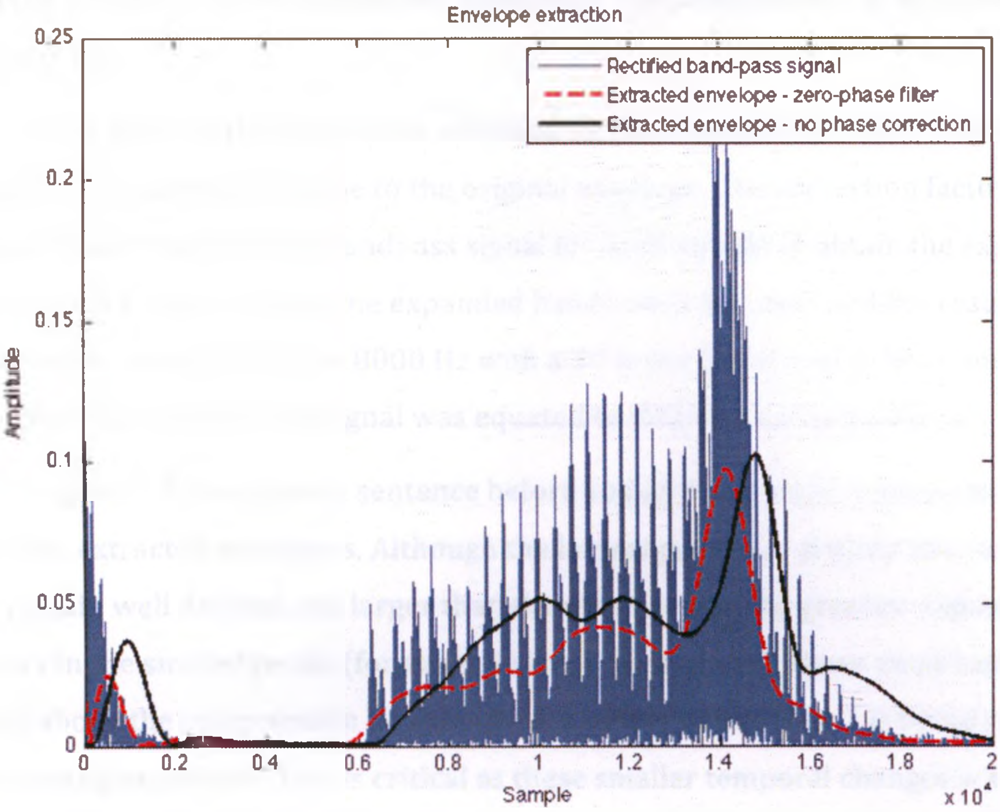

Figure 3-2: Comparison of envelope extraction with zero-phase filter and with no phase correction

Envelope expansion:

The extracted envelope was expanded by raising the envelope to the power *K*. *K* is calculated for each sample (in each band separately) through the following exponential function:

$$K = e^{\frac{-(E_i - E_{min})}{\tau}}(K_{max} - K_{min}) + K_{min} \tag{3-1}$$

where K_{min} = 0.3, K_{max} = 4, E_{min} is the minimum amplitude of the signal, E_i is the instantaneous amplitude value and τ is a time constant for the exponential. *K* and E_{min} are calculated for each band independently. The decreasing exponential is a function of E_{min}. As such, K_{max} is applied to the lowest amplitudes and K_{min} is applied to the largest. Considering all waveform amplitudes are less than 1, applying a power of 4 to the minimum amplitudes results in near-zero amplitude. Likewise, applying a power of 0.3 increases the amplitude. The selection of τ is discussed in Section 3.1.2.

Next, a correction factor was obtained for each sample by calculating the ratio of the expanded envelope to the original envelope. The correction factor was multiplied with the original bandpass signal for each sample to obtain the expanded signal for each band. Finally, the expanded bands were summed and the resulting signal was lowpass filtered to 8000 Hz with a 3rd order Butterworth filter and the RMS power of the resulting signal was equated to that of the original signal.

Figure 3-3 illustrates a sentence before and after envelope enhancement as well their extracted envelopes. Although the largest peaks are slightly attenuated, they remain well defined and larger than the other peaks. The greatest expansion appears in the smaller peaks (for example, at t=1.6 seconds). These amplitudes are clearly above the compression threshold, i.e. a power of less than 1 is being applied, thus causing expansion. This is critical as these smaller temporal changes are very likely to go undetected by individuals with AN due to their poor temporal processing. The masking due to AN of these low-power speech cues is very clear when viewing simulated AN waveforms in Chapter 4. It is evident from Figure 3-3 that the enhanced signal has exaggerated temporal envelope speech cues.

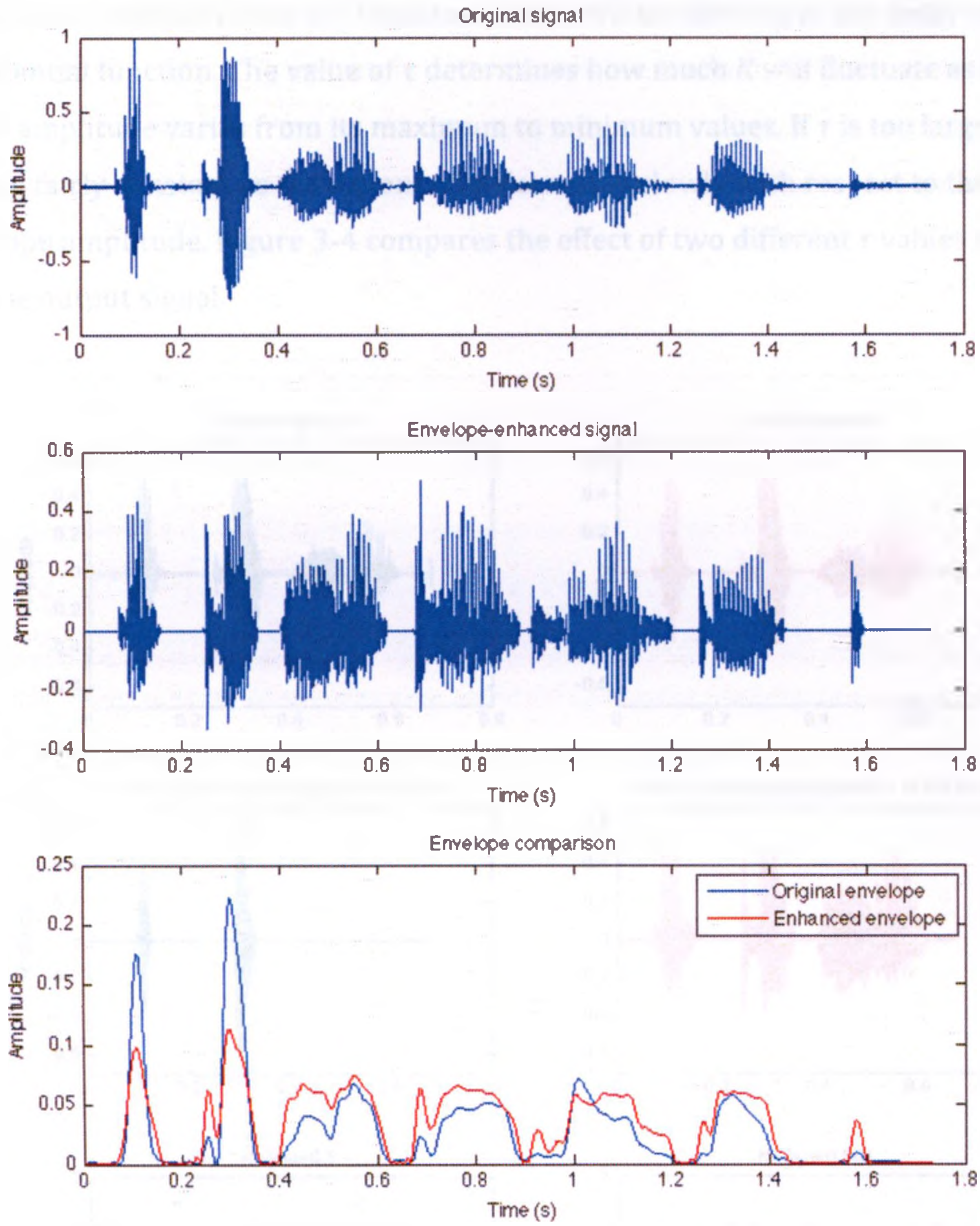

Figure 3-3: Effect of envelope enhancement on speech envelope for the utterance, "the picture came from a book"

3.1.2 Selection of Time-Constant

In [9], a value of 0.5 is used for τ. However, it was found that this was too large for all stimuli that were tested in the present study. The units of τ are the same as the units for the amplitude of the envelope. The term *time-constant* for this parameter is adopted from the literature because it is referring to the decay of an exponential function. The value of τ determines how much K will fluctuate as the signal amplitude varies from its maximum to minimum values. If τ is too large, K will remain fairly constant as the exponential decreases slowly with respect to the envelope amplitude. Figure 3-4 compares the effect of two different τ values on K and the output signal.

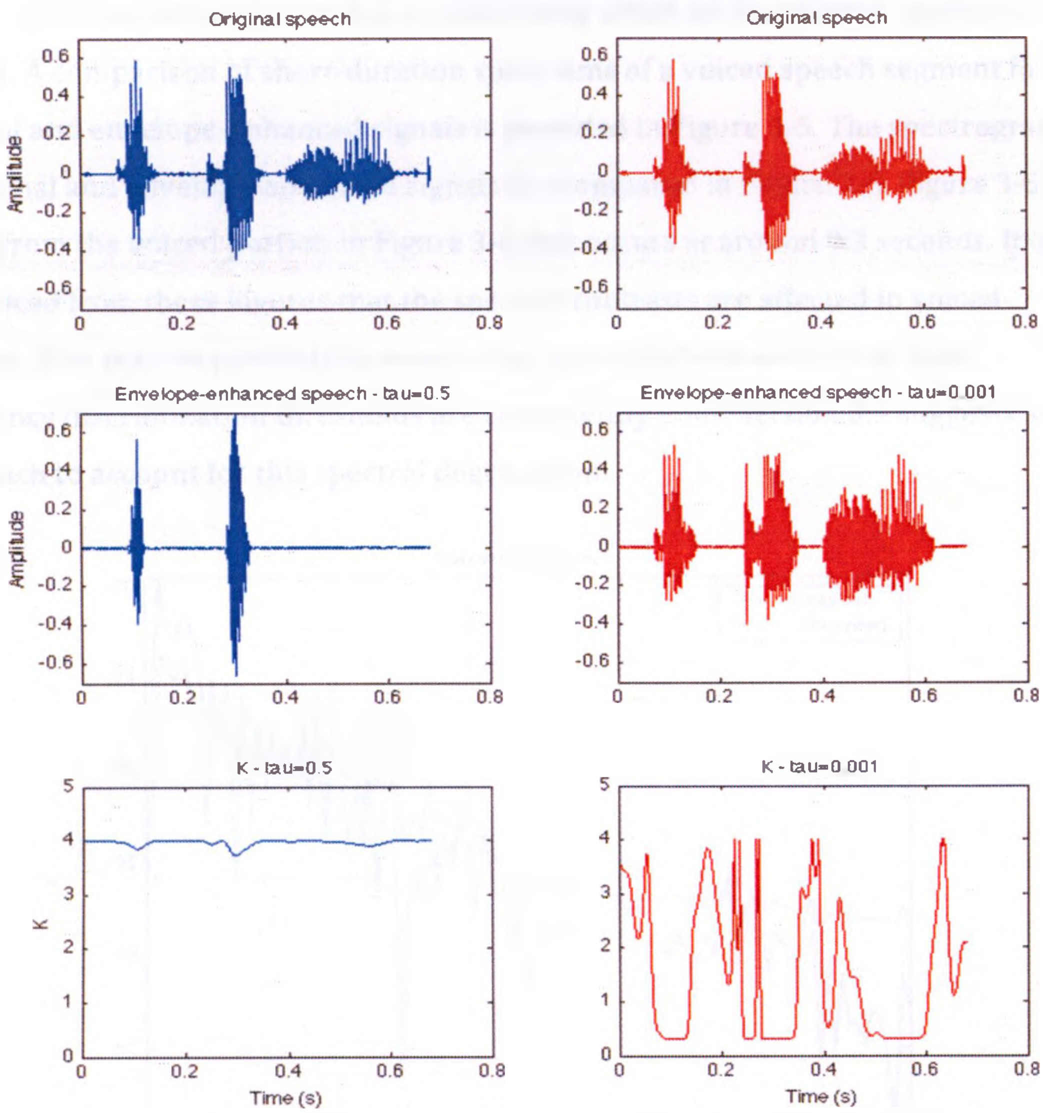

Figure 3-4: Effect of time-constant selection on envelope enhancement for the utterance, "the picture"

It is evident from Figure 3-4 that for τ=0.5, K remains close to its maximum value of 4 and a large expansion is applied to the entire signal resulting in a dramatic suppression of everything except for the two most powerful portions.

Through experimentation, a value of 0.001 for τ proved to consistently produce large variation in K for the sentences used for subjective testing in Chapter 4.

3.1.3 Effect on Spectral Content

Envelope enhancement has an interesting effect on the spectral content of speech. A comparison of short-duration spectrums of a voiced speech segment for original and envelope-enhanced signals is provided in Figure 3-5. The spectrograms of original and envelope-enhanced signals are compared in Figure 3-6. Figure 3-5 is made from the voiced portion in Figure 3-6 that occurs at around 0.3 seconds. It can be noticed from these Figures that the spectral contrasts are affected in voiced regions. This may be particularly concerning for individuals with AN as their frequency discrimination thresholds are consistently poor. Section 3.5 suggests an approach to account for this spectral degradation.

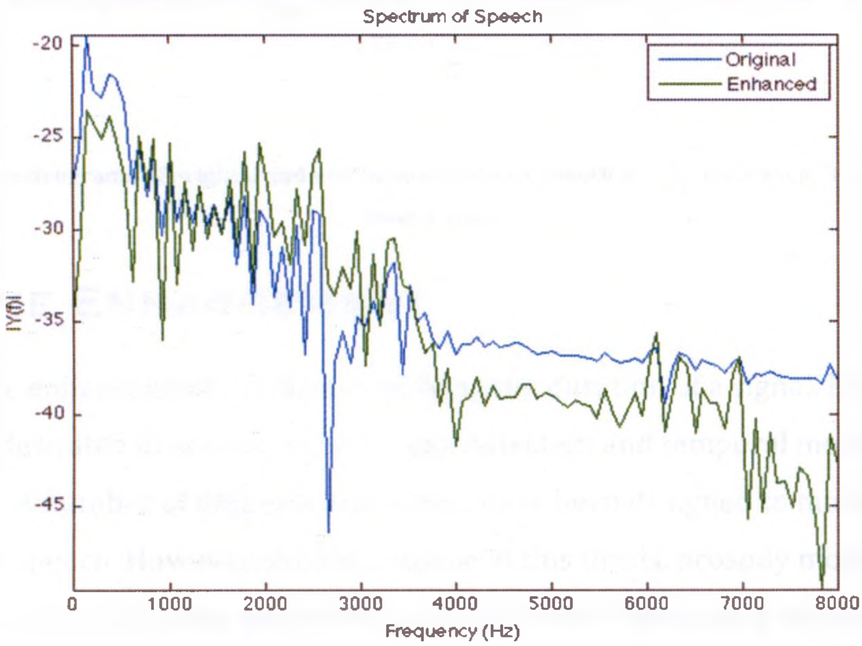

Figure 3-5: Comparison of short-duration spectrums for before and after envelope enhancement

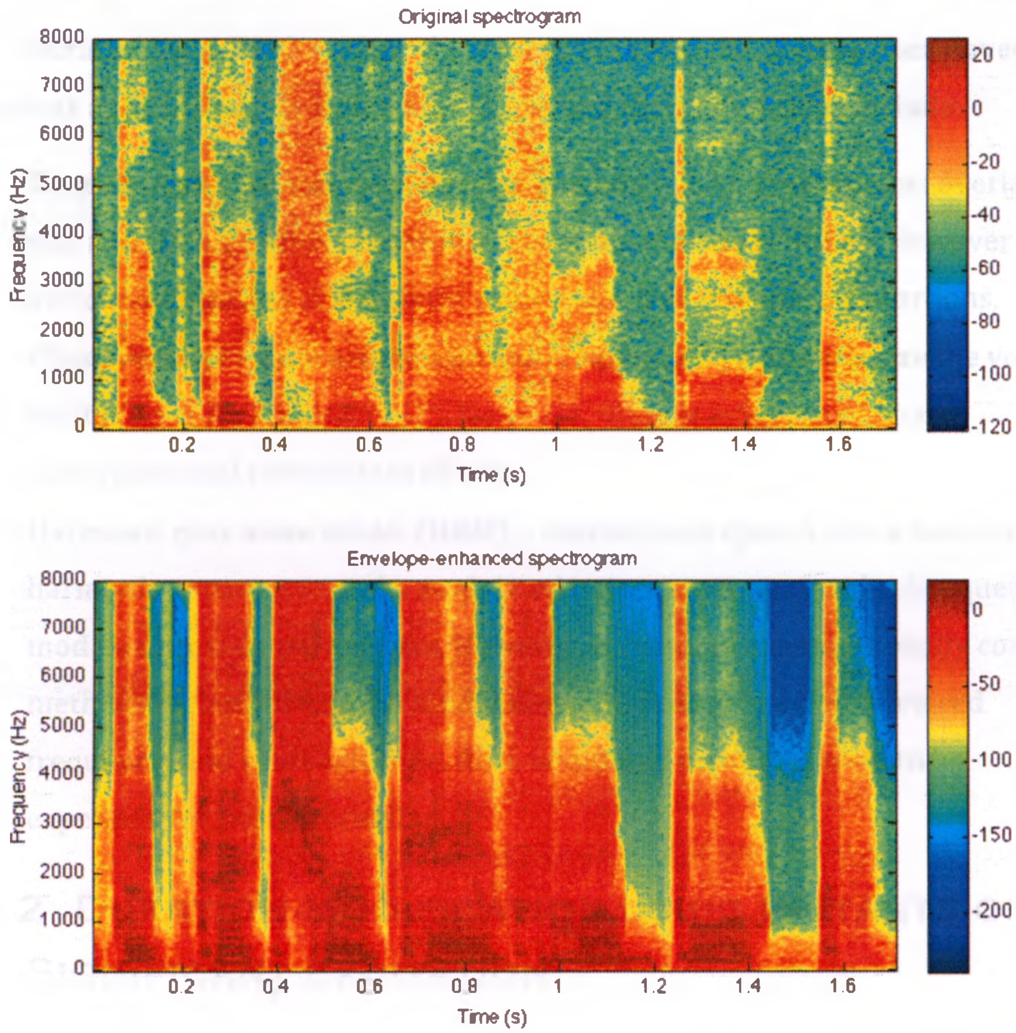

Figure 3-6: Spectrograms of original and envelope-enhanced speech for the utterance, "the picture came from a book"

3.2 Time Enhancement

Time enhancement involves modifying the duration of a signal. In the case of AN, this is desirable to account for poor gap detection and temporal modulation thresholds. A number of different algorithms have been designed to modify the duration of speech. However, for the purpose of this thesis, prosody modification using instants of significant excitation and vowel onset points was chosen as the most suitable for AN speech enhancement.

3.2.1 Existing Duration-Modification Algorithms

Duration modification algorithms for speech and music have been developed for various applications. Common algorithms and their limitations include:

- Time domain – for example: overlap and add (OLA), synchronous overlap and add (SOLA), pitch synchronous overlap and add (PSOLA). However, direct modification of the signal causes spectral and phase distortions.
- Phase vocoder – uses amplitudes and frequencies to characterize the voice excitation and vocal tract [37]. However, there is evidence of phase distortions and reverberant effects.
- Harmonic plus noise model (HNM) – decomposes speech into a time-varying harmonic component and a modulated noise component and subsequently modifies prosody parameters. However, it requires computationally complex methods for estimating the fundamental frequency, maximum voiced frequency, and synthesis time. Also, additional resources need to be expended for post-processing.

3.2.2 Duration Modification using Instants of Significant Excitation

It is assumed that phase distortions are particularly detrimental to speech perception in people with AN due to their poor frequency discrimination. It is also assumed that preservation of important temporal cues is extremely important when modifying the duration of speech because poor temporal processing has been suggested as the most impacting characteristic leading to poor speech perception. Therefore the selection of a time scale modification algorithm for this thesis depended heavily on these two factors. The algorithm presented in [38], which is an extension of [37], focuses on detecting and preserving important temporal aspects of speech, as well as minimizing spectral distortions.

In order to preserve temporal aspects, glottal closure (GC) instants (also known as instants of significant excitation or epochs) are calculated and used to manipulate (extend) the linear prediction (LP) residual. In turn, the LP residual is used to excite a time-varying filter with coefficients derived from the original speech signal. Vowel onset points (VOPs) are used to preserve the CV transition regions, consonant-consonant transition regions and consonant regions of the LP residual. The procedure is outlined as follows [38]:

1. Preemphasize the speech signal.
2. Compute the LP residual.
3. Determine the epochs from the LP residual using group delay analysis.
4. Determine VOPs from the Hilbert envelope of the LP residual, marking the consonant and transition regions.
5. Calculate new epoch locations with respect to the desired duration modification factor.
6. Determine the nearest original epoch to each new epoch.
7. Update the LP residual using the new epoch locations.
8. Calculate the new VOPs and replace the scaled transition and consonant regions with the original LP residual regions indicated by the VOPs from step 4.
9. Update the filter coefficients according to the new, scaled LP residual.
10. Excite the updated filter with the modified LP residual to produce the synthesized, duration-modified speech.

Distortion is reduced by operating on the LP residual signal [37] as opposed to directly manipulating the signal because there is less correlation between samples in the LP residual. Notable aspects of the algorithm are described in detail below. The algorithm was implemented in MATLAB with use of the VOICEBOX speech processing toolbox [39].

<u>*LP Analysis:*</u>

First, linear predictive coefficients (LPCs) are obtained. LPCs are filter coefficients that represent time varying vocal tract system characteristics by modeling it as an all-pole filter. Typically, to encode speech in LPCs, two mutually exclusive excitation functions are required to model voiced and unvoiced regions of speech. However, for the purpose of this system, no differentiation is made between these two regions [37]. The residual signal from LP analysis is found through inverse filtering the original signal with the LPCs. Tenth-order LP analysis is performed with a frame size of 20 ms, a frame shift of 5 ms and a Hamming window. A sample rate of 8000 Hz is used.

Computation of original and new epochs:

For unvoiced speech, epochs occur at random instants, but for voiced speech regions, the epochs represent instants of GC where the residual error is large. In this case, the time between epochs corresponds to the pitch period of the voiced speech [37]. Epochs are found by performing group delay analysis on the LP residual. Group delay analysis for epoch extraction is derived and discussed in [40]. Essentially, group delay can be calculated as the derivative of the phase function as follows:

$$\tau(\omega) = -\phi'(\omega) = \frac{X_R Y_R + X_I Y_I}{X_R^2 + X_I^2} \tag{3-2}$$

where $X(\omega)$ and $Y(\omega)$ are the Fourier transforms (using discrete Fourier transform (DFT)) of the windowed signal $x(n)$ and $nx(n)$ respectively, $X_R + jX_I = X(\omega)$ and $Y_R + jY_I = Y(\omega)$, $\phi(\omega)$ is the derivative of the phase function of $X(\omega)$ and $\tau(\omega)$ is the group delay function. After the group-delay function is calculated, smoothing is performed with three-point median filtering to get the phase slope function. Epochs are hypothesized at each positive zero crossing of the smoothed phase slope function [37]. For the thesis, the VOICEBOX toolbox was used to perform group-delay analysis and determine the epoch locations in the original speech signal.

For duration modification, a modification factor, β=1.5, is chosen. Figure 3-7 illustrates the calculation technique for finding new epochs at β=1.5. In this Figure, each epoch has an associated epoch interval (the number of samples between it and the previous epoch) and is plotted on the y-axis. The original epochs are marked as red "o"s and the red line represents a linear interpolation through each of the original epochs. Mapped original epochs are found by simply multiplying the original epoch locations by β. These mapped original epochs are marked as a blue "o"s and the blue line represents a linear interpolation through each mapped epoch. The blue line represents the desired epoch interval plot, as the new epochs should appear on this time-expanded version of the original epoch interval plot (epoch intervals represent pitch period and the pitch period should remain constant throughout duration modification). The new epoch locations are marked as a pink "x." These are found through the following procedure:

- The first new epoch is assigned the same position as the first epoch from the original signal. The desired epoch interval at that time (shown by the blue line) is used to determine the location of the next epoch (by simply adding the epoch interval to the current epoch location).
- The blue line at that next location on the x-axis is used in turn to determine the location of the third new epoch. This procedure is continued until the end of the desired epoch plot is reached.

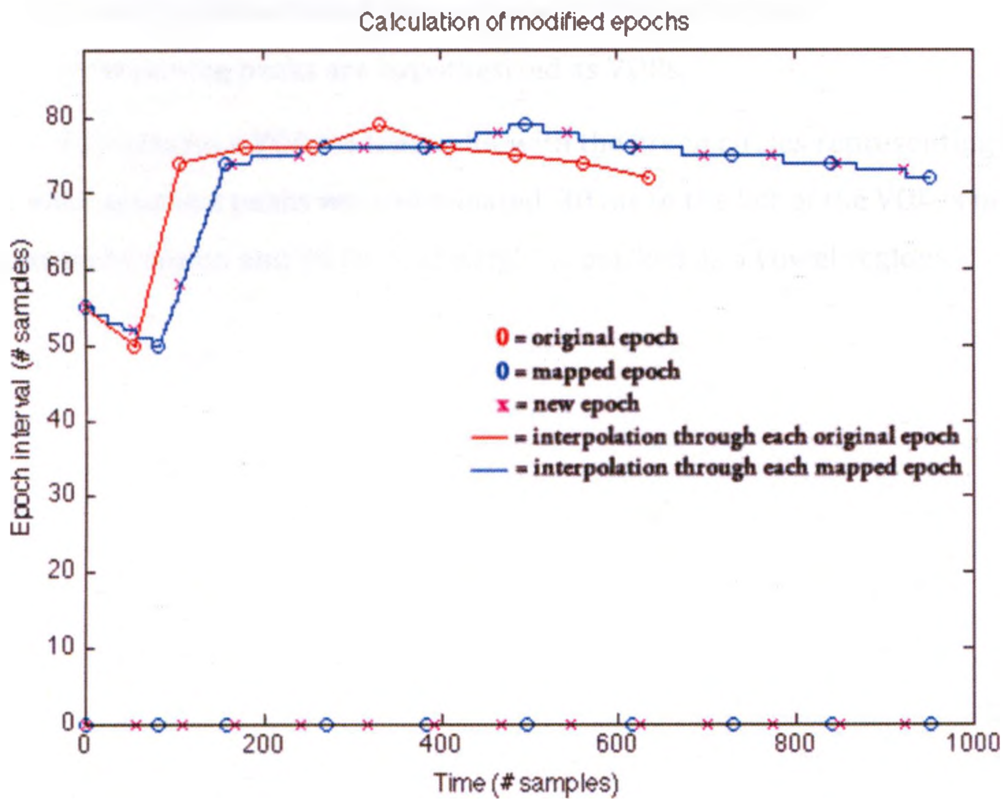

Figure 3-7: Illustration of epoch calculation

Calculating original and new VOPs:

In most cases, VOPs correspond to the transition from a consonant to the following vowel. Preserving this transition is important for preserving well-defined temporal cues. To detect the VOPs, the Hilbert envelope of the LP residual is first computed. It is then convolved with a Gabor filter (with spatial spread of the Gaussian σ=100, frequency of modulating sinusoid ω=0.0114 and filter length n=800) to produce a smoothed VOP evidence plot. Peaks in the VOP evidence plot represent potential VOPs. Spurious peaks are eliminated with the following procedure:

- If there is no negative region with reference to the next peak, the peak is determined to be spurious.
- If a peak is within 50 ms of the next peak, eliminate it as spurious.
- For the sentences used in this thesis, it was found through experimentation that eliminating peaks with amplitude of less than 15% of the max evidence plot amplitude increased the accuracy of VOP detection.
- The remaining peaks are hypothesized as VOPs.

Figure 3-8 illustrates a VOP evidence plot with the green circles representing VOP events after spurious peaks were eliminated. 30 ms to the left of the VOP is marked as a consonant region and 30 ms to the right is marked as a vowel region.

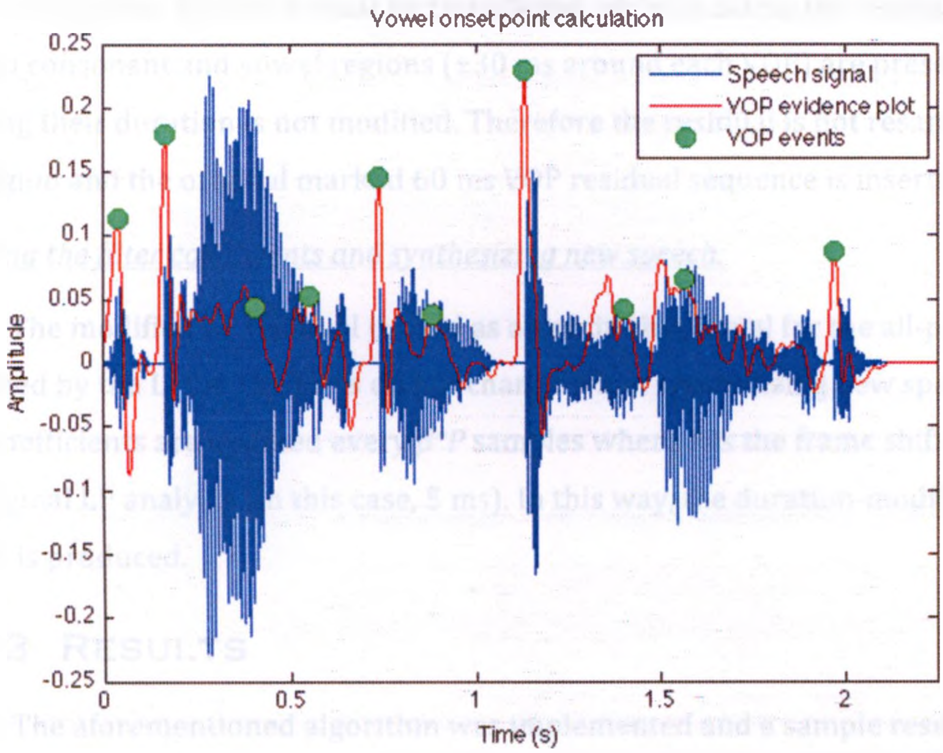

Figure 3-8: Illustration of vowel onset point detection for the utterance, "the car is going too fast"

Updating the LP residual:

The modified epoch sequence is used to alter the duration of the LP residual. Referring to Figure 3-7, the mapped original epochs ("o") closest to the new epochs ("x") are determined (shown on the x-axis). Each "o" has an associated original LP residual section equal in length to the value of the original epoch interval at that location. In other words, the locations of the "o"s are used to extract the corresponding sequence from the original residual. Each of these sequences corresponds to one pitch period.

When creating the new LP residual, the original residual segments are placed starting at the location of each new epoch. In the case of increasing the length of the residual (β=1.5), the value of the desired epoch interval is larger than the corresponding original epoch interval. Thus, resampling between each new epoch location is performed to create a modified residual sequence of the desired length.

In addition, the VOPs must be considered while updating the residual. The marked consonant and vowel regions (±30 ms around each VOP) are preserved, meaning their duration is not modified. Therefore the residual is not resampled in this region and the original marked 60 ms VOP residual sequence is inserted.

Updating the filter coefficients and synthesizing new speech:

The modified LP residual is used as an excitation signal for the all-pole filter described by the LPCs. The LPCs do not change when synthesizing new speech. The filter coefficients are updated every $\beta*P$ samples where P is the frame shift used in the original LP analysis (in this case, 5 ms). In this way, the duration-modified speech is produced.

3.2.3 Results

The aforementioned algorithm was implemented and a sample result is illustrated in Figure 3-9. It can be seen that there is little temporal or spectral distortion despite an increased duration by a factor of 1.5.

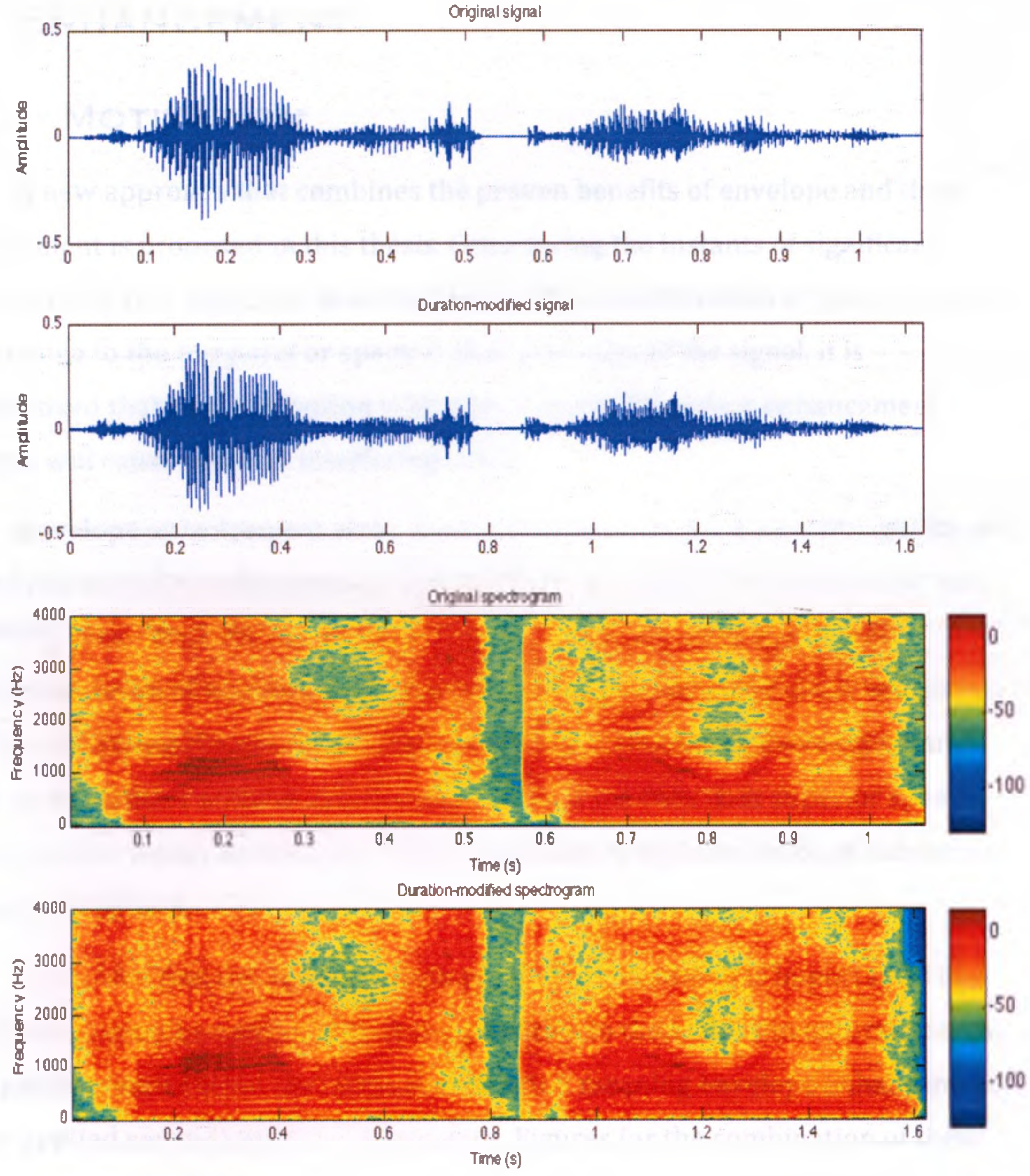

Figure 3-9: Temporal and spectral results of duration modification for expansion factor of β=1.5 for the utterance, "flowers grow in the"

3.3 Combined Envelope and Time Enhancement

3.3.1 Motivation

A new approach that combines the proven benefits of envelope and time enhancement is proposed in this thesis. Considering the instants of significant excitation and VOP approach described for duration modification of speech causes little change to the temporal or spectral characteristics of the signal, it is hypothesized that its combination with other time-independent enhancement schemes will cause minimal interfering effects.

Envelope enhancement alone does not address the documented inability of individuals with AN to discriminate speech with high syllabic rates. Likewise, time stretching alone does not sharpen the temporal contrast, meaning that the modulation depth of a speech signal may remain below the threshold of the person with AN, rendering the speech indecipherable. If the benefits of these algorithms prove to be additive without interference between the two, then superior speech discrimination would be achieved. This proposition is explored through subjective testing in Chapter 4.

The order that the algorithms are applied to the stimuli is important. Time modification must be applied first as it is sensitive to distortions in natural speech. The speech remains clean after time modification meaning envelope enhancement can be applied second without consequence. Figures for the combination of these two algorithms are not provided, as there is no visibly observable difference in time and envelope-enhanced speech and envelope-enhanced speech alone.

3.3.2 Using Vowel Onset Points as Cues for Envelope Modifications

The concept of using VOPs as cues for envelope enhancement was explored. Envelope enhancement inherently causes some unwanted spectral degradation that may, in some cases, lead to decreased intelligibility. Therefore, in an attempt to reduce this degradation, envelope enhancement was applied only to the VOP region (30 ms before and after the VOP location) as this is typically identified as the consonant region and consonant-vowel transition region.

In [25], CV transitions were shown to be significantly problematic for people with AN, and [23] concluded that an improved consonant-to-vowel ratio was one of the significant benefits derived from envelope enhancement. Thus, it appears plausible that applying envelope enhancement only to the VOP region would preserve the benefits of envelope enhancement as well as preserve the spectral content of the speech in remaining regions.

Upon implementation of this approach, distortions and transients were found at the interface of enhanced and clean speech. Interpolation across the interface, convolution with smoothing functions and median filtering were applied in an attempt to decrease the contrast between the processed and unprocessed speech. However, time limitations, unsatisfactory speech quality and the number of algorithms that already required subjective testing eliminated the implementation of this approach in Chapter 4 evaluations. The potential of this novel idea requires future work and consideration and will be discussed further in the future work section in Chapter 5.

3.4 Spectral Enhancement

The companding strategy described in [26], [27] was implemented in MATLAB and used to evaluate the effectiveness of spectral enhancement for enhancing speech recognition in people with AN.

3.4.1 Companding Architecture and Implementation

Figure 3-10 displays the block diagram of the companding architecture described in [26] where n_1 is the compression index set to 0.3, n_2 is the expansion index set to 1, filter F represents a wideband pre-filter, and filter G represents a narrowband post-filter. Both of these filters have the same resonant frequency in the same channel. The subscript for each filter represents the channel number.

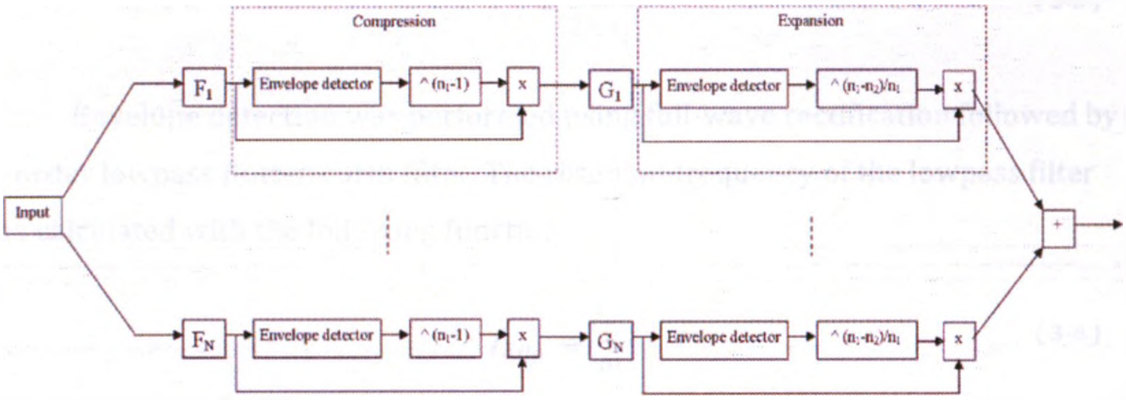

Figure 3-10: Block diagram of companding algorithm

For implementation, 50 channels were used, as in [26] (the authors chose 50 to reflect the typical processing of a cochlear implant), and bandpass filters *F* and *G* were described by the following functions:

$$F_i'(s) = \left(\frac{2\left(\frac{\tau_i}{q_1}\right)s}{\tau_i^2 s^2 + 2\left(\frac{\tau_i}{q_1}\right)s + 1} \right)^2 \tag{3-3}$$

$$G_i'(s) = \left(\frac{2\left(\frac{\tau_i}{q_2}\right)s}{\tau_i^2 s^2 + 2\left(\frac{\tau_i}{q_2}\right)s + 1} \right)^2 \tag{3-4}$$

where the subscript *i* refers to the channel index, $F_i(s) = F_i'^2(s)$ and $G_i(s) = G_i'^2(s)$, and q_1 and q_2 are filter parameters set to 2 and 12, respectively [27]. To create $F_i(s)$ and $G_i(s)$, $F_i'(s)$ and $G_i'(s)$ are each cascaded with themselves. The bilinear transform was used to derive digital versions of the aforementioned filters. Furthermore, in order to reduce interference across channels, zero-phase filtering was used. The resonant frequencies for each channel were logarithmically spaced between 250 and 4000 Hz. The resonant frequency, $f_{r,i}$, is related to τ_i by the following function:

$$f_{r,i} = \frac{1}{2\pi\tau_i} \tag{3-5}$$

Envelope detection was performed using full-wave rectification followed by a 1st order lowpass Butterworth filter. The resonant frequency of the lowpass filter was calculated with the following function:

$$f_{ED,i} = \frac{f_{r,i}}{\omega} \tag{3-6}$$

where ω was chosen to be 40 [26].

3.4.2 Two-Tone Suppression Fundamentals

What differentiates the companding architecture from traditional compression strategies and allows for two-tone suppression is the narrow post-filter. A high-level description, using Figure 3-11, of how this strategy results in two-tone suppression is provided below [26].

Assume *F* is broad and almost perfectly flat, while *G* is sharply tuned. A sinusoid, A_1, is at the resonant frequency of the channel and a sinusoid of larger amplitude, A_2, is at a different frequency. After filtering by *F*, A_1 and A_2 are plotted in Figure 3-11. The gain of the compression block is determined by the envelope detector, which is most heavily influenced by the stronger sinusoid, A_2. A_2 is transformed to B_2 and A_1 is transformed to C_1. G heavily suppresses A_2 since it is off the resonant frequency, meaning C_1 will be the only sinusoid passing through *G*. C_1 is then expanded to get D_1. Therefore A_1 has been suppressed to D_1 by an off-frequency strong tone, A_2. B_1 illustrates how the amplitude of A_1 would be unaffected by companding if A_2 had not been present. The stronger tone has the effect of suppressing the weaker tone, showing the spectral enhancement produced by companding. It must be noted here that an analytical proof of the spectral enhancement achieved by the companding architecture is given in [26].

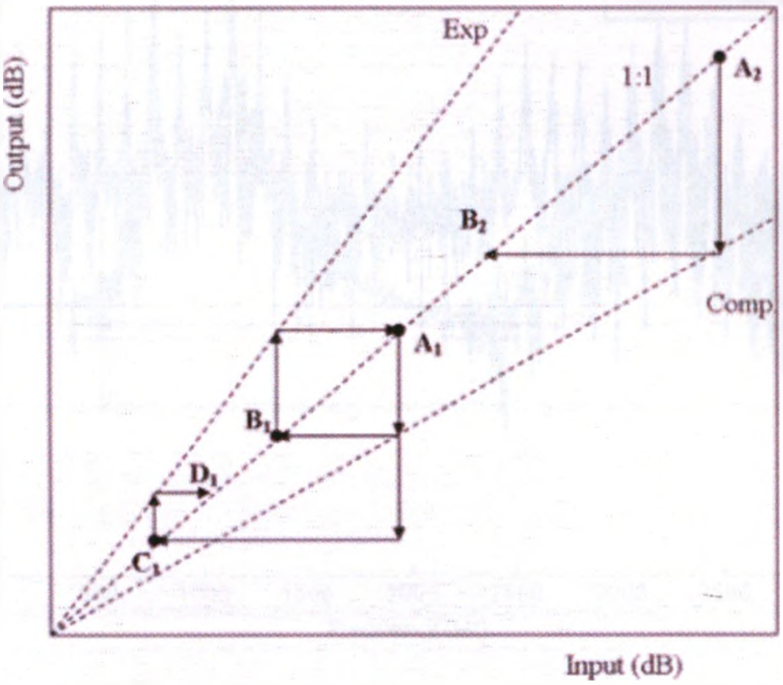

Figure 3-11: Graphical illustration of companding algorithm

3.4.3 Results

The companding algorithm was successfully implemented in MATLAB and its effectiveness was evaluated through visual inspection. Figure 3-12 demonstrates the performance of the algorithm on a vowel sound with additive white noise (the algorithm was designed partly for front-end noise suppression in CIs). Spectral sharpening of the two formant frequencies is clearly visible. However, the stimuli used in Chapter 4 are clean speech, thus Figure 3-13 demonstrates the performance of the algorithm on a bi-syllabic word in clean speech. The results are not as prominent as the previous plot but some spectral sharpening is observed at the formants, and higher, lower power frequencies are attenuated.

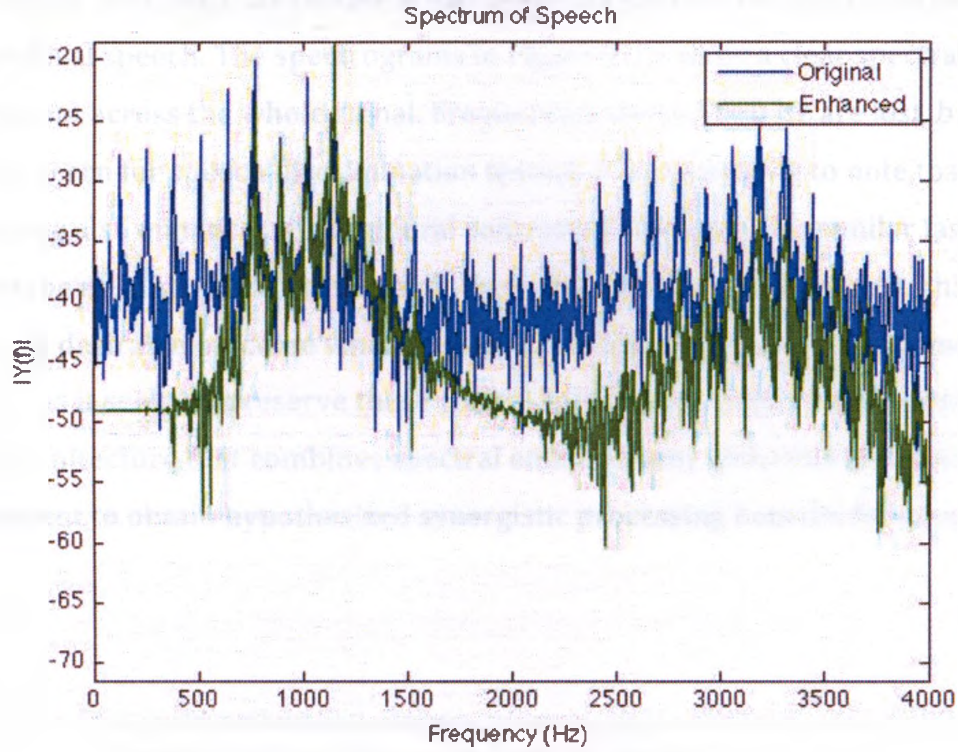

Figure 3-12: Spectrum of vowel with additive white noise for original and spectral-enhanced signals

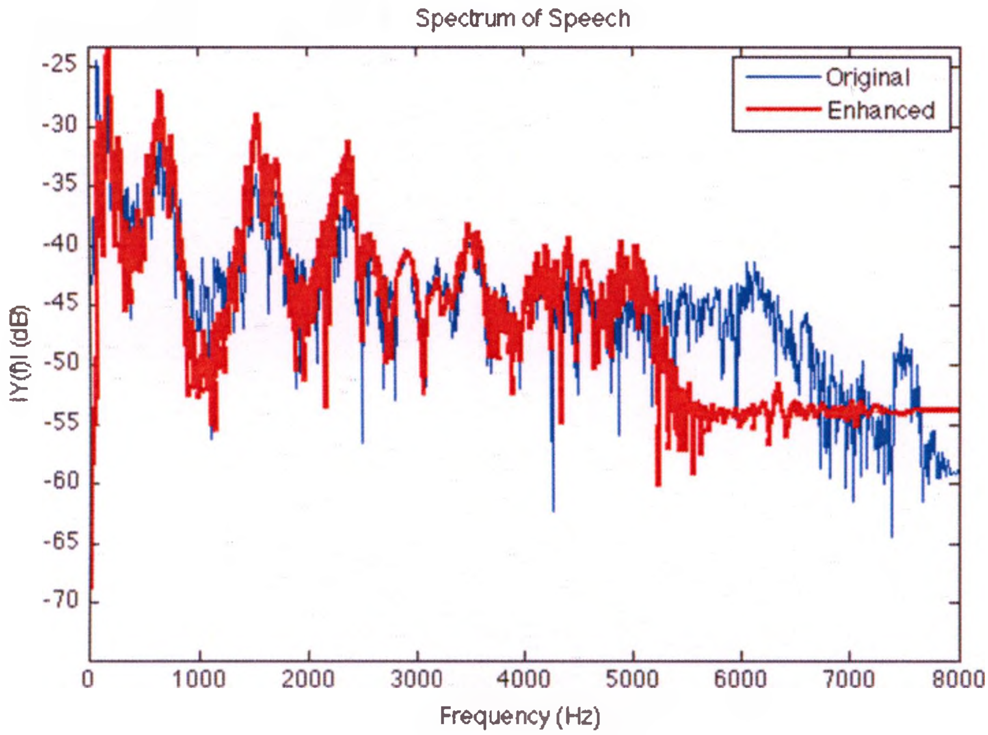

Figure 3-13: Spectrums of the word 'ashil' in quiet

Figure 3-14 plots the temporal and spectral effects of the algorithm for sentence-level speech. The spectrograms in Figure 3-14 show a clear spectral enhancement across the whole signal. Frequencies above 5000 Hz are lost, but this is not a concern for speech discrimination testing. It is interesting to note that companding also enhances the temporal contrast of the signal in a similar fashion to the envelope enhancement. Particularly, the vowel onsets are sharpened. This effect may have a desirable outcome when combined with the VOP time enhancement algorithm as it seeks to preserve these critical speech cues. Section 3.5 describes the specific architecture that combines spectral enhancement with time and envelope enhancement to obtain hypothesized synergistic processing benefits for people with AN.

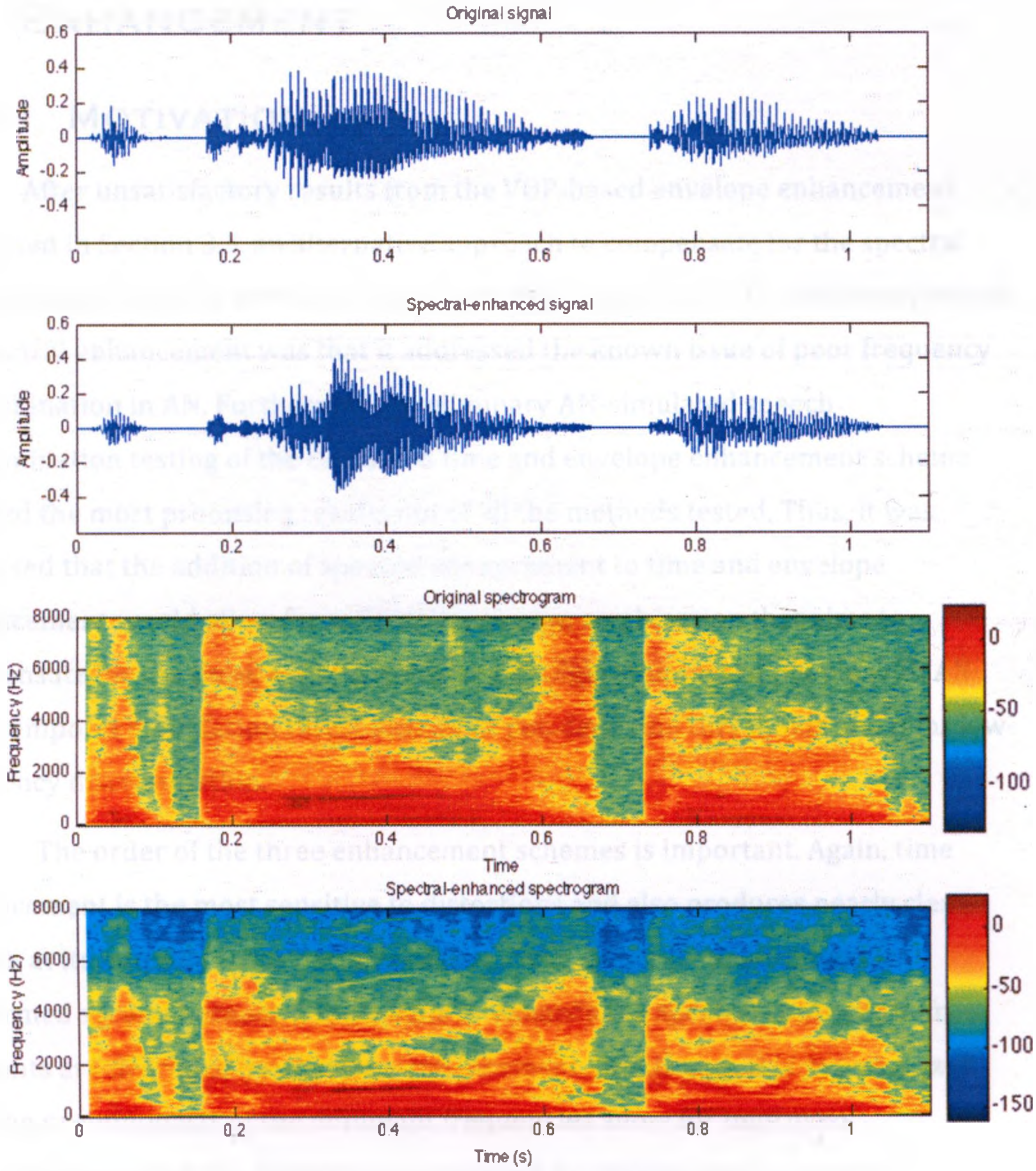

Figure 3-14: Temporal and spectral effects of spectral enhancement for the utterance, "the car is going"

3.5 Combined Envelope, Spectral and Time Enhancement

3.5.1 Motivation

After unsatisfactory results from the VOP-based envelope enhancement described in Section 3.3, an alternative approach to compensate for the spectral degradation caused by envelope enhancement was explored. The additional benefit of spectral enhancement was that it addressed the known issue of poor frequency discrimination in AN. Furthermore, preliminary AN-simulated speech discrimination testing of the combined time and envelope enhancement scheme showed the most promising results out of all the methods tested. Thus, it was proposed that the addition of spectral enhancement to time and envelope enhancement would allow for a signal processing combination that aims to compensate for the three major psychoacoustic shortcomings of people with AN: poor temporal modulation thresholds, poor gap detection thresholds and poor low-frequency discrimination.

The order of the three enhancement schemes is important. Again, time enhancement is the most sensitive to distortions and also produces nearly clean speech at its output, thus is performed first. Next, spectral enhancement is performed second because it is sensible to enhance the spectral contrasts before formants are degraded by envelope enhancement. In this way, it is hypothesized that the pre-emphasis of the dominant frequencies allow for improved discrimination after the distortions introduced by envelope enhancement. Consequently, envelope enhancement is applied to the signal last.

In Chapter 4, the performance of this scheme is compared to the aforementioned methods.

CHAPTER 4

4 SUBJECTIVE DATA COLLECTION AND ANALYSIS

Two experiments were completed in order to subjectively evaluate the performance of the algorithms described in Chapter 3. Experiment I compared unprocessed speech, envelope enhancement, time enhancement and time + envelope enhancement. Experiment II evaluated the addition of spectral enhancement to time + envelope enhancement. Both these experiments were conducted with normal hearing listeners after processing stimuli with an auditory neuropathy (AN) simulator. A brief description of the simulator is given below.

4.1 AUDITORY NEUROPATHY SIMULATOR

An AN simulator was used to simulate the effect of auditory neuropathy on speech such that normal hearing (NH) listeners could be recruited for experimentation. Speech stimuli were clean speech sentences taken from the hearing in noise test (HINT) database [41]. All processing was completed in MATLAB.

4.1.1 IMPLEMENTATION

The simulator was adopted from [9] with modifications to parameters and filtering added to obtain meaningful results with the stimuli used in Experiments I and II. Figure 4-1 displays the block diagram of the simulator used for experimentation.

First, the input signal, resampled to a sampling rate of 16 kHz, was divided into 16 ⅓-octave bands using a filter bank with a minimum frequency of 140 Hz and a maximum frequency of about 5700 Hz. Similar to the envelope enhancement scheme, 6th-order Butterworth bandpass filters were used. Each band was operated on independently and then summed to produce the final output.

Envelope detection was performed by taking the square of the Hilbert envelope followed by lowpass filtering. In [9], a 1st-order Butterworth filter was used whose cutoff frequency was derived from the temporal modulation functions of individuals with AN. Cutoff frequencies representing varying severities of AN were defined in [9] and are presented in Table 4-1.

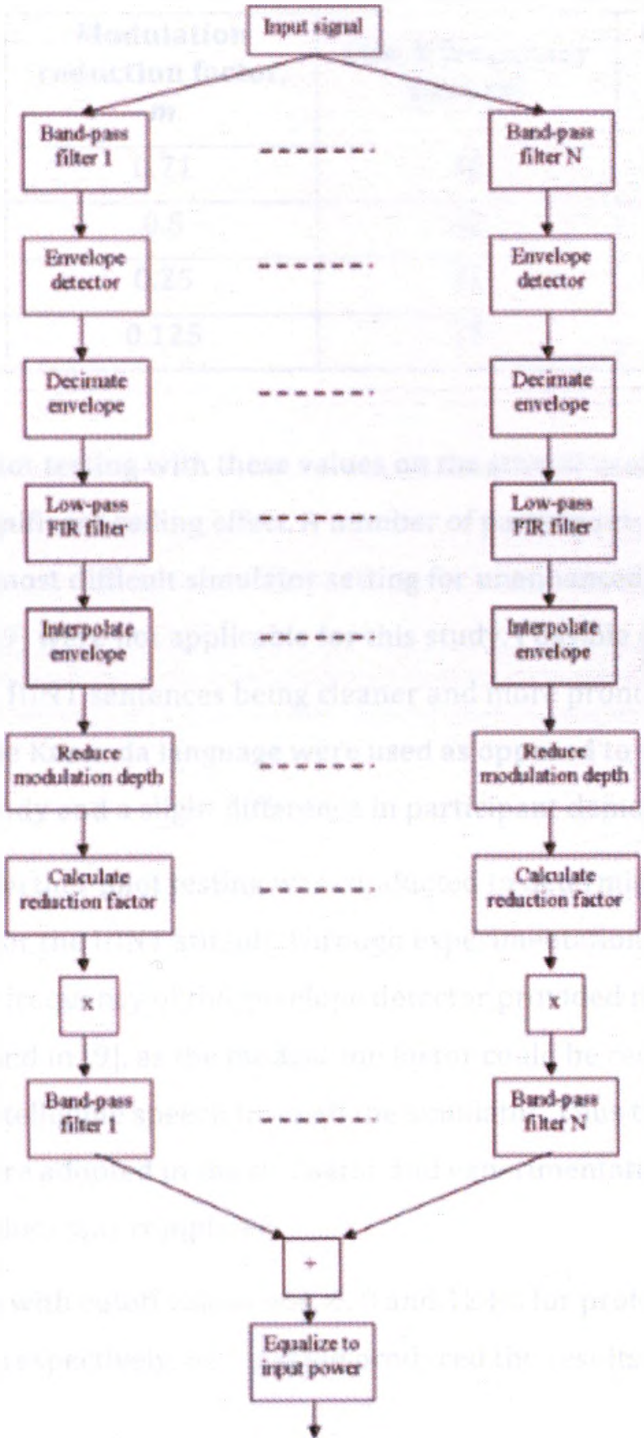

Figure 4-1: Block diagram of AN simulator

Table 4-1: Comparison of parameter values used in AN simulator

Degree of AN	Modulation reduction factor, *m*	Cutoff frequency (Hz) [9]	Cutoff frequency (Hz) used in study
Mild	0.71	42	8
Moderate	0.5	32	6
Severe	0.25	22	4
Profound	0.125	15	2

However, pilot testing with these values on the stimuli used in Experiments I and II showed a significant ceiling effect. A number of participants exceeded an 80% word score on the most difficult simulator setting for unenhanced speech. Clearly the parameters in [9] were not applicable for this study. Possible discrepancies in results include: the HINT sentences being cleaner and more pronounced, in [9], bi-syllabic words of the Kannada language were used as opposed to complete English sentences in this study and a slight difference in participant demographic.

As a result, further pilot testing was conducted to determine appropriate simulator settings for the HINT stimuli. Through experimentation, it was found that lowering the cutoff frequency of the envelope detector provided results more similar to those found in [9], as the modulation factor could be reduced to near-zero and still produce intelligible speech through the simulator. Thus the modulation factors from [9] were adopted in the simulator and experimentation with different cutoff frequency values was completed.

Pilot testing with cutoff values of 4, 6, 8 and 12 Hz for profound to mild simulator settings, respectively, on 5 people produced the results shown in Figure 4-2.

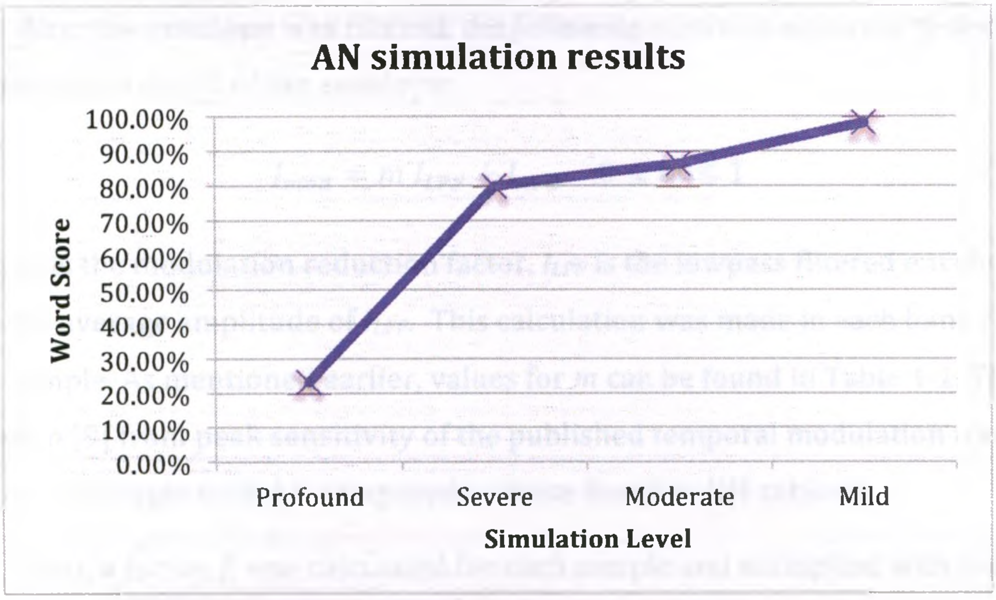

Figure 4-2: Simulation results for unenhanced speech with cutoff frequencies of 4,6,8 and 12 Hz

An average of about 80% word score was achieved at the severe setting, revealing that even lower cutoff frequencies were required to avoid ceiling effects, especially for enhanced versions of the stimuli which should score higher than the unenhanced stimuli shown here. Further pilot testing revealed that cutoff frequencies of 2, 4, 6 and 8 Hz provided a set of results that spanned from low word recognition scores to ceiling. The parameters used in Experiments I and II are summarized in Table 4-1.

Considering the low cutoff frequencies required for lowpass filtering in extracting the envelope, a different filtering scheme was required to produce a sharp magnitude response. The following procedure was performed:

1. Decimate the squared Hilbert envelope by a factor of 64 (down from a sampling rate of 16 kHz).
2. Create a lowpass FIR filter of order 135 and select the appropriate cutoff frequency (depending on the desired simulator setting). As discussed in Section 3.1.1, zero-phase filtering was performed to avoid imparting any phase delay on to the envelope.
3. Upsample the envelope by a factor of 64, to the original sampling rate of 16 kHz.

After the envelope was filtered, the following equation was used to decrease the modulation depth of the envelope:

$$I_{mod} = m\, I_{LPF} + I_{AVG} \quad 0 \leq m \leq 1 \tag{4-1}$$

where m is the modulation reduction factor, I_{LPF} is the lowpass filtered envelope and I_{AVG} is the average amplitude of I_{LFP}. This calculation was made in each band for every sample. As mentioned earlier, values for m can be found in Table 4-1. They are derived in [9] from peak sensitivity of the published temporal modulation transfer functions of people with AN compared to those found in NH subjects.

Next, a factor, f, was calculated for each sample and multiplied with the original bandpass filtered signal to apply the modulation reduction:

$$f = \sqrt{\frac{I_{mod}}{I_{org}}} \tag{4-2}$$

where I_{org} is the original intensity envelope of the band, before lowpass filtering.

Finally, the results of each band were filtered through their original bandpass filter, summed together and the RMS power of the sum was equated to the original signal.

4.1.2 Results

Application to envelope enhancement:

Examples of AN-simulated speech for different severities with and without envelope enhancement can be found in Figure 4-3. It is apparent that the exaggerated envelope allows for some temporal shaping to exist even after the most severe simulation level. These preserved temporal cues are essentially those that allow for increased speech understanding even at challenging AN simulator degrees.

Application to spectral enhancement:

A spectral comparison of AN-simulated speech with and without spectral enhancement is provided in Figure 4-4. Spectral degradation is inherently caused by the modulation reduction. This effect may be comparable to the poor frequency discrimination found in people with AN. Formants are more well-defined after processing through spectral enhancement, thus it was hypothesized that this would lead to increased word recognition scores in testing.

Application to time enhancement:

The effectiveness of the simulator in evaluating the benefit of time enhancement is not fully understood. To simulate the effect of poor gap detection thresholds, a more appropriate test may be decreasing the duration of the stimuli and then applying processing to slow it down to time-enhance the signal. However, the simulator may give insight into the effectiveness of time enhancement because it slows down the temporal modulations and that may prove to be beneficial at understanding the temporally smeared speech. An ideal test for time enhancement (and all enhancement schemes for that matter) involves recruitment of AN subjects; however, as discussed in the following section, this was not feasible.

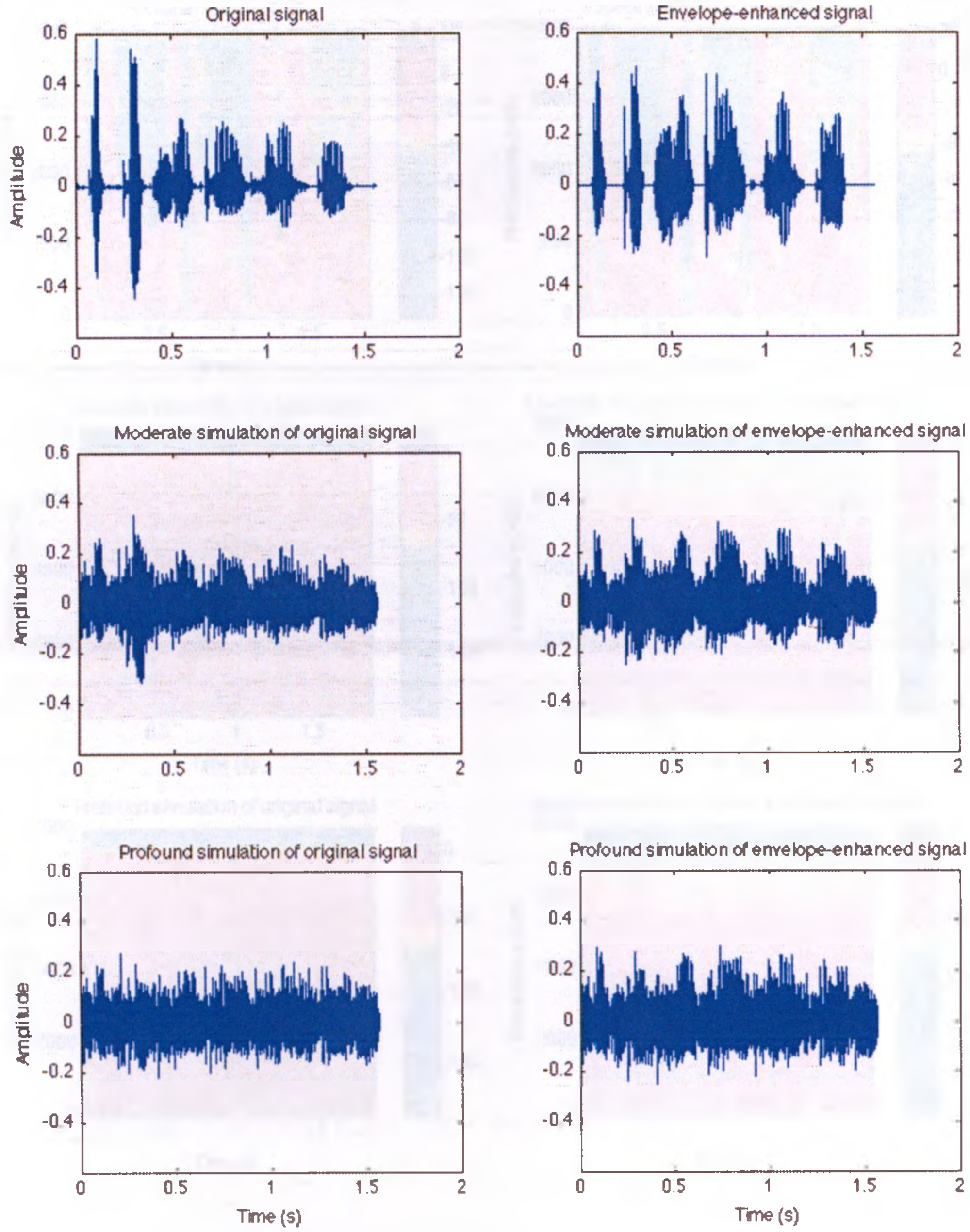

Figure 4-3: Simulated waveforms of unprocessed and envelope-enhanced speech for the utterance, "the car is going"

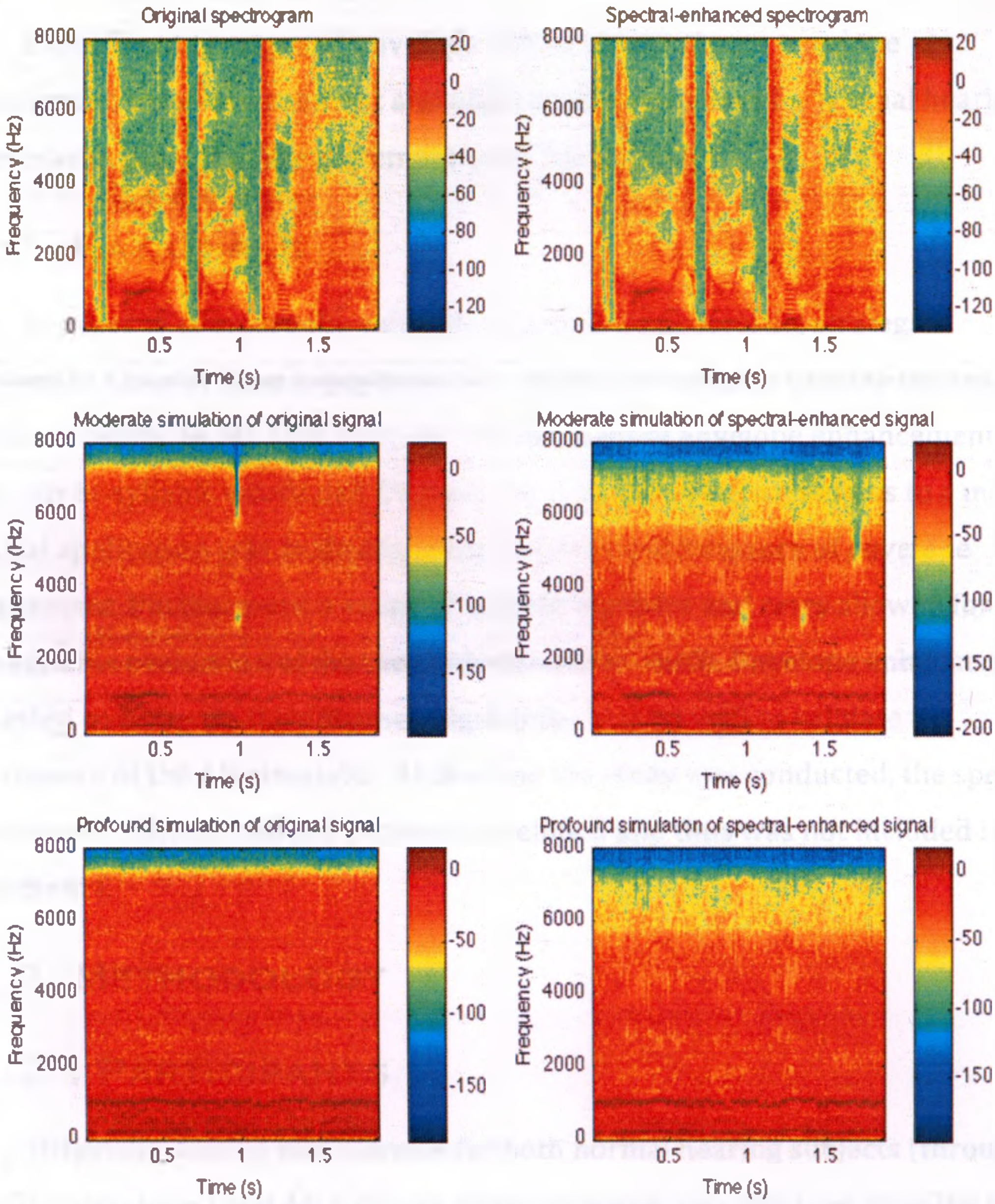

Figure 4-4: Spectrograms of simulated unprocessed and spectral-enhanced speech for the utterance, "the car is going"

4.2 Experiment I

Experiment I compared envelope enhancement, time + envelope enhancement, time enhancement and unprocessed sentences for normal hearing listeners with the use of the aforementioned AN simulator.

4.2.1 Rationale

In order to compare the value of the proposed processing strategies described in Chapter 3, an experiment was needed to compare them to the current published results. In [9], [23], [24], the effectiveness of envelope enhancement was shown for bi-syllabic words and CV pairs. An extension of these studies to a more practical application was evaluating sentence-level performance of envelope enhancement. Furthermore, a range of syllable numbers and rates allowed for a more suitable evaluation of the time enhancement benefit. Envelope enhancement was tested to compare the proposed algorithms to it, as well as validate the performance of the AN simulator. At the time the study was conducted, the spectral enhancement scheme had not yet been developed and thus was not included in Experiment I.

4.2.2 Methodology

4.2.2.1 Participants

Originally, testing was planned for both normal hearing subjects (through use of the simulator) and AN subjects. Ethics approval was obtained from the UWO Research Ethics Board (REB) for the testing of 16 normal hearing subjects and 16 AN subjects. The inclusion criteria required participants to be aged between 18-60 years old and be English first-language. For NH subjects, no hearing impairment or abnormality could be present in either ear and screening was completed for 20dB HL pure tones up to 4000 Hz (the sampling rate of the stimuli was 8000 Hz). AN subjects must be diagnosed as having AN by an audiologist or ENT. The number of participants was chosen using Horatio software for sample size estimation [42]. See Appendix A for a copy of the REB ethics approval.

Ethics approval was also obtained from the Lawson Clinical Research Impact Committee (CRIC) for the recruitment of AN patients from their computerized database. However, a lack of available AN subjects that met the inclusion criteria and a compressed testing timeline meant that testing these subjects was no longer feasible for this thesis. Considering AN has only begun to be diagnosed recently, there are very few adults diagnosed with AN. The majority of diagnoses occur in children, most of who are still minors. Therefore only the 16 NH listeners completed Experiment I with use of the AN simulator.

The ages of the 16 NH subjects ranged from 19-55; however, most were aged 20-24 and were students at the University of Western Ontario. Screening was completed (with a calibrated audiometer) to ensure each subject had pure tone sensitivity of at least 20 dB HL at octave frequencies between 500-4000 Hz for both ears.

4.2.2.2 Stimuli and Processing

Stimuli:

The speech stimuli used for Experiment I were taken from the hearing in noise test (HINT) database. The database consists of 28 wordlists of 10 sentences spoken by a male talker, each recorded at a sampling rate of 44.1 kHz.

Processing:

Processing was required to evaluate the four enhancement algorithms at four simulator settings each, creating 16 unique processing conditions, as outlined in Table 4-2. In addition, an unprocessed control condition with no simulation was included, making a total of 17 processing conditions.

Table 4-2: Processing conditions, not including control condition

	Mild	Moderate	Severe	Profound
Envelope (EE)	✓	✓	✓	✓
Time + envelope (TE+EE)	✓	✓	✓	✓
Time (TE)	✓	✓	✓	✓
Unprocessed	✓	✓	✓	✓

Therefore 17 unique wordlists were required for each subject (one for each condition). The same 17 wordlists were used for each subject, but the condition applied to each wordlist was randomized.

To create the database of processed speech stimuli, each HINT sentence was first resampled to 16 kHz. The enhancement algorithm was then applied (either EE, TE+EE, TE or unprocessed/no enhancement). As the TE condition operated at a sampling rate of 8 kHz, the 16 kHz was resampled to 8 kHz, processed by the TE algorithm, after which the signal was resampled back to 16 kHz. The AN simulator was then applied to the processed speech stimulus at the desired degree of AN severity. The simulator output was resampled to 8 kHz, scaled appropriately, and stored in .wav file. The final database contained the 170 sentences (17 wordlists) processed for each of the 17 conditions such that the testing software could select the condition randomly for each sentence. The RMS power of every file in the database was equated and scaled to maximize dynamic range.

4.2.2.3 Procedure

Presentation level, equipment and location:

Participants listened to the stimuli through Sennheiser HDA 200 headphones in a double-walled, acoustically treated sound booth at the National Centre for Audiology (NCA). Stimuli were presented at 8 kHz through a USBPre external sound card and were routed through a GSI 61 clinical audiometer for digital volume control. Presentation levels through the headphones were calibrated with the head and torso simulator (HATS) as shown in Figure 4-5 and SpectraPLUS software [43] for a white noise stimulus at the same power as the sentence database. Before testing commenced, participants were presented with a sentence at 60 dB HL and asked if it was at a comfortable level. Necessary adjustments were made for each subject; however, all subjects requested levels within the range of 55-65 dB HL and volume levels were consistent across all stimuli presented to each subject.

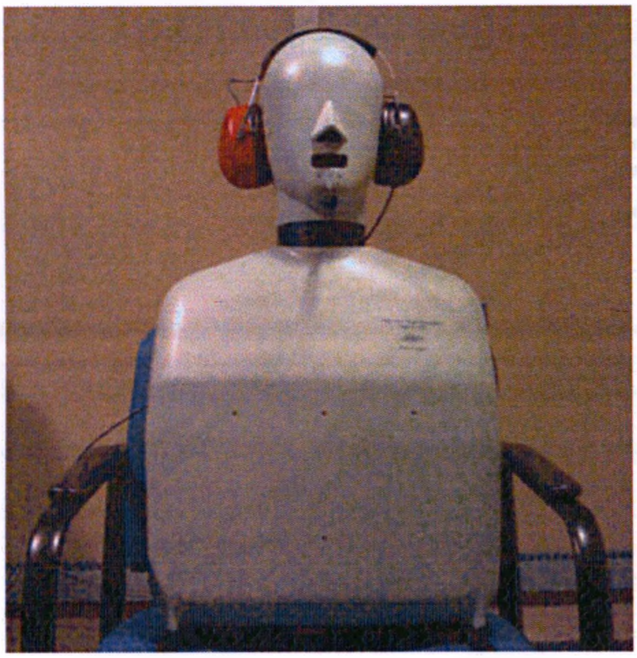

Figure 4-5: Hearing-level calibration using HATS

Testing process:

Testing software was developed in MATLAB to present randomized stimuli and record the word scores for each condition. Subjects were asked to focus on listening to each sentence as it was played back and attempt to repeat as many words as possible, even though some may seem very distorted, and that there was no penalty for incorrect guessing. The number of keywords correctly repeated for each sentence was entered into MATLAB, referring to the HINT manual of keywords for each sentence. A percent word score for each condition was then calculated by dividing the total number of correctly repeated keywords in the wordlist by the total keywords in that wordlist. Subjects listened to all 17 wordlists consecutively.

4.2.3 Results

Figure 4-6 summarizes the data collected from Experiment I. The word scores from all subjects at each condition were averaged and plotted against simulator level. Each enhancement scheme is colour-coded for easy performance comparison. The condition with no enhancement and no simulation is not shown because it scored 100% for all subjects (as expected for NH listeners).

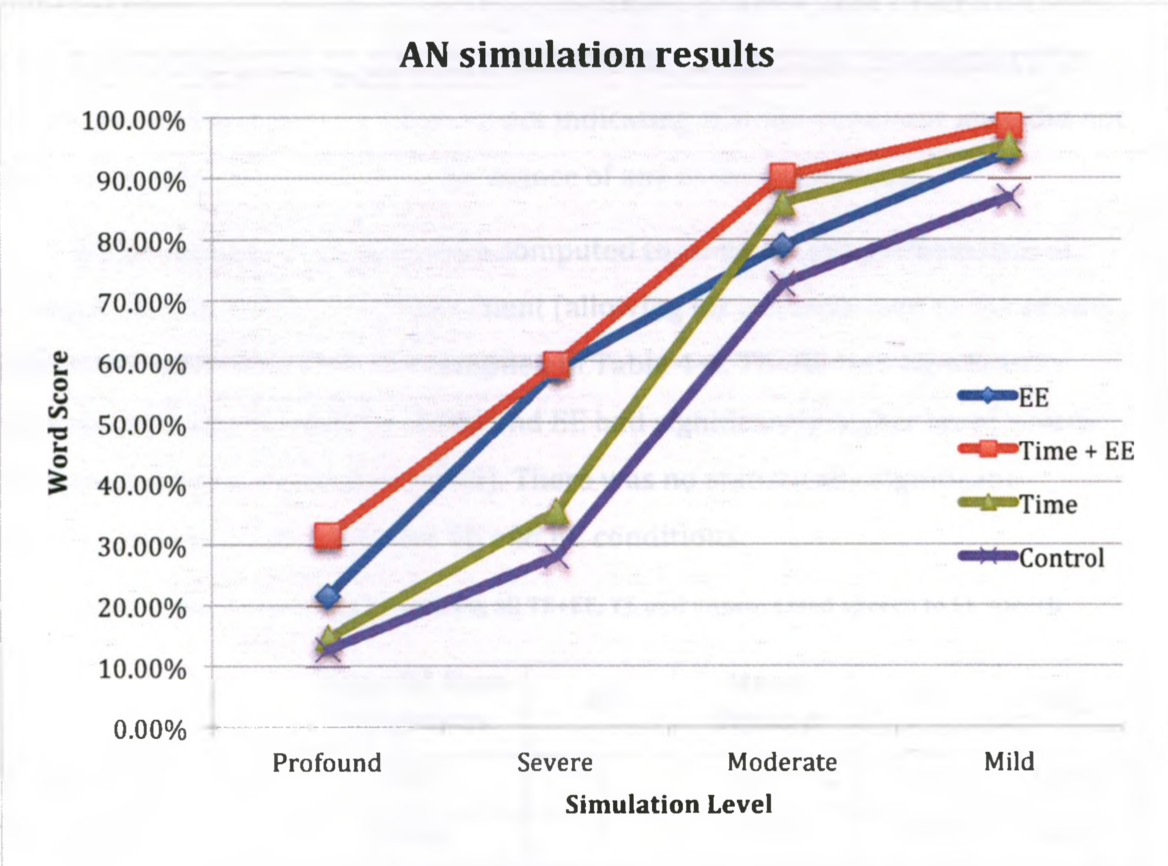

Figure 4-6: Experiment I results comparing EE, TE+EE, TE and unprocessed speech

Table 4-3 shows the standard deviations for each algorithm. Note that the values are not displayed as percentages.

Table 4-3: Standard deviations for each algorithm

EE	TE+EE	TE	Control
0.142	0.116	0.084	0.102

4.2.4 Statistical Analysis

A two-way repeated measure of analysis of variance (ANOVA) was performed using SPSS software to compare the performance of each processing scheme (4 levels) across each level of simulation (4 levels). The assumption of sphericity ($p>0.05$) was met for each of the three effects in the model (algorithm, simulator and the interaction between algorithm and simulator). There was a significant main effect of algorithm ($F(3, 13)=16.84$, $p<0.001$) and simulation level ($F(3, 13)=205.59$, $p<0.001$). In addition, there was no significant interaction ($F(9, 7)=3.29$, $p>0.05$) between the two factors indicating that the simulator level did not significantly affect the relative performance of any of the algorithms.

Within-subject contrasts were computed to compare the performance of each algorithm to envelope enhancement (allowing for a comparison to the results found in [9]) and the results are compiled in Table 4-4. TE+EE had significantly higher word scores than EE ($p<0.05$) and EE had significantly higher word scores than the unenhanced speech ($p<0.05$). There was no statistically significant difference between the scores for TE and EE conditions.

Table 4-4: Simple contrasts comparing all TE+EE, TE and unprocessed speech to EE speech

	Type III Sum of Squares	df	Mean Square	F	Sig.
TE+EE vs. EE	*0.07*	*1*	*0.07*	*5.512*	*0.033*
TE vs. EE	0.046	1	0.046	3.914	0.067
Unprocessed vs. EE	0.277	1	0.277	18.168	0.001

Because there was no interaction between algorithm and simulator level, each algorithm was collapsed and averaged across all simulator levels for each subject. This allowed for a false data rate (FDR) comparison between each algorithm. *FDR analysis is not as strict as Bonferroni because it uses a varying alpha* level. As such, it reduces type II error without worsening type I error and can be described as a balance between making too many false discoveries and missing real differences due to being too conservative [44]. The results from the FDR procedure are provided in Table 4-5. To perform FDR, pair-wise t-tests were conducted for every combination of two algorithms and the significance values from each test were recorded (shown in the column labeled *p-value* in Table 4-5). Next, each comparison was sorted by their significance level and corrected alpha levels were computed (shown in the column labeled *Sig.*). For each row, a statistically significant difference was deemed to exist if *p-value* < *Sig.* The results from this test indicated a significant difference between all combinations of algorithms except TE and EE. SPSS outputs from Experiment I can be found in Appendix B.

Table 4-5: FDR comparison of each algorithm

Algorithm 1	Algorithm 2	p-value	Sig.
TE+EE	TE	0	0.008
TE+EE	Unprocessed	0	0.017
EE	Unprocessed	0.001	0.025
TE	Unprocessed	0.016	0.033
EE	TE+EE	0.033	0.042
EE	TE	0.067	0.050

4.2.5 Discussion

The results of Experiment I demonstrated the significant benefit of combining time and envelope enhancement over envelope enhancement alone across all simulator levels. This suggests that the benefits of these two algorithms are additive and may effectively target separate psychoacoustic impairments.

Although time enhancement alone performed significantly better than the unenhanced speech, there was no significant difference between it and envelope enhancement. Individual cases of AN may require time enhancement to account for their psychoacoustic impairments while others may not. Furthermore, it is clear from Figure 4-6 that time enhancement alone performs poorly at profound and severe simulations levels and shows little benefit over the unenhanced speech. This suggests that envelope enhancement is critical at these levels to account for the very smeared temporal envelopes and the benefits of time enhancement are masked for the most part. However, as shown by the performance of TE+EE at the profound simulation level, it appears that the envelope enhancement unmasked the time-stretched temporal cues and thus maximum benefit was achieved.

The result that EE performed significantly better than unprocessed speech is consistent with the results from [9] and the simulator had a comparable effect on word scores across its four settings indicating that the modified simulator described here functioned as expected.

The best processing scheme for a given case of AN may have to be evaluated on a case-by-case basis. However, the simulations conducted in this study suggest that a combined algorithm of time and envelope enhancement provides the most benefit to speech discrimination in people with AN.

4.3 Experiment II

Experiment II evaluated the addition of spectral enhancement to the most successful algorithm from Experiment I, the time + envelope enhancement. This combined algorithm is denoted as TE+SE+EE.

4.3.1 Rationale

Section 2.3 outlined the motivations for testing spectral enhancement for AN and Section 3.5 outlined the motivations for combining it with time and envelope enhancement. Spectral enhancement was not seriously explored as an enhancement scheme until after testing for Experiment I had begun. In addition, Experiment I was already comparing four processing schemes. Therefore, upon completion of Experiment I, preliminary pilot testing of TE+SE+EE began. 12 of the 16 subjects from Experiment I were available to participate in Experiment II (new subjects could not be recruited due to ethics limitations).

Testing was limited to one new algorithm due to a limited number of wordlists available in the HINT database (wordlists from Experiment I could not be reused as Experiment II involved the same participants). It was hypothesized that the benefits of SE were independent from the benefits of EE and TE and in the interest of evaluating the most effective algorithm possible, TE+SE+EE was chosen as the one algorithm to be tested in Experiment II.

4.3.2 Methodology

4.3.2.1 Participants

The subjects from Experiment I were asked to return to do Experiment II, but only 12 were available to participate within the scheduled time frame. New subjects could not be recruited due to a cap of 16 NH subjects in the original ethics approval. The timeline did not permit a resubmission to the ethics board to request recruitment of new subjects.

4.3.2.2 Stimuli and Processing

New wordlists from the HINT database were used in Experiment II. Processing was completed in the same manner as Experiment I with the TE+SE+EE algorithm used for simulation preprocessing with the same simulator settings. For 6 of the returning participants, TE+EE was tested in addition to TE+SE+EE to allow for a direct comparison between the two algorithms.

4.3.2.3 Procedure

The same experimental procedure and explanations from Experiment I were completed for the TE+SE+EE-processed speech and TE+EE processed speech for 6 of the subjects. The RMS levels of the stimuli database were equalized, scaled for maximum dynamic range and headphone output power from the audiometer was once again calibrated using the HATS.

4.3.3 RESULTS

Figure 4-7 displays the average word scores across the 12 participants for TE+SE+EE plotted on top of the results from Experiment I for the same 12 participants.

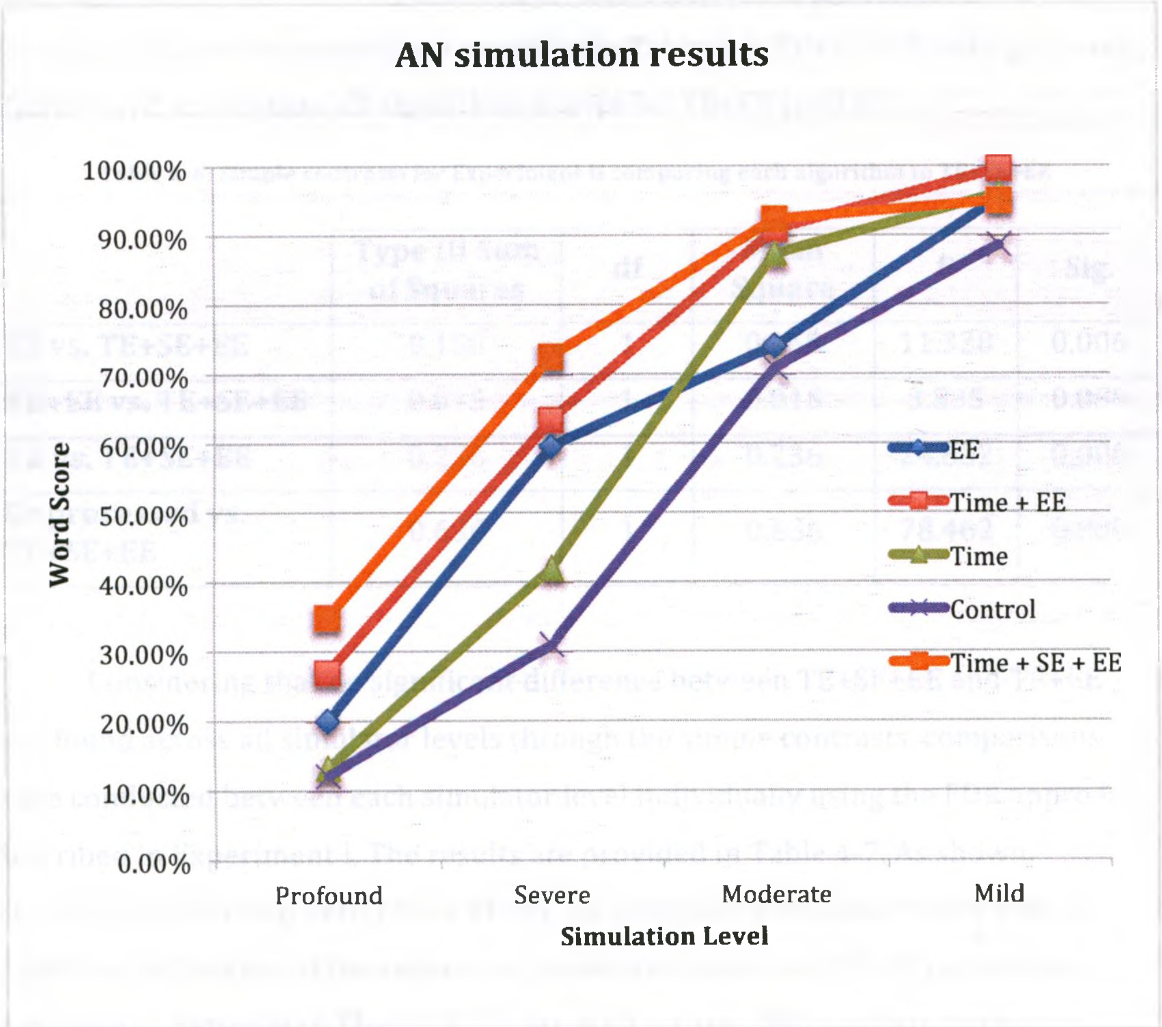

Figure 4-7: Experiment II results comparing the addition of TE+SE+EE to the results from Experiment I

4.3.4 STATISTICAL ANALYSIS

First, the data obtained from the 12 returning subjects was used to perform a two-way repeated measure ANOVA to compare the performance of each processing scheme (5 levels, including TE+EE+SE) across each level of simulation (4 levels).

The assumption of sphericity (p>0.05) was met for algorithm and simulator; however, there were insufficient residual degrees of freedom to perform tests of interaction between algorithm and simulator. There was a significant main effect of algorithm ($F(4,8)=18.67$, $p<0.001$) and simulation level ($F(3, 9)=178.14$, $p<0.001$).

Within-subject contrasts were computed to compare the performance of each algorithm to SE+EE+TE (allowing for a comparison to the results from Experiment I) and the results are compiled in Table 4-6. TE+SE+EE had significantly higher word scores than all algorithms except for TE+EE ($p<0.05$).

Table 4-6: Simple contrasts for Experiment II comparing each algorithm to TE+SE+EE

	Type III Sum of Squares	df	Mean Square	F	Sig.
EE vs. TE+SE+EE	0.158	1	0.158	11.338	0.006
TE+EE vs. TE+SE+EE	0.015	1	0.015	3.585	0.085
TE vs. TE+SE+EE	0.236	1	0.236	24.832	0.000
Unprocessed vs. TE+SE+EE	0.636	1	0.636	78.462	0.000

Considering that no significant difference between TE+SE+EE and TE+EE was found across all simulator levels through the simple contrasts, comparisons were conducted between each simulator level individually using the FDR approach described in Experiment I. The results are provided in Table 4-7. As shown, TE+SE+EE performed better than TE+EE for profound simulation, there was no significant difference at the severe and moderate levels and TE+EE performed significantly better than TE+SE+EE at the mild setting. SPSS outputs from Experiment II can be found in Appendix B.

Table 4-7: FDR comparison of TE+SE+EE and TE+EE for each simulator setting for a null hypothesis predicting no significant difference between algorithms and an alternative hypothesis predicting a significant difference between algorithms

Simulator Level	p	Sig
Mild	0.01	0.0125
Profound	0.014	0.025
Severe	0.275	0.0375
Moderate	0.743	0.05

4.3.5 Discussion

First, it must be noted that the participants completed this task on a different day than Experiment I meaning that there was no randomization of enhancement scheme. Practice effects may be present because all the participants had already completed Experiment I before Experiment II (although new wordlists were used). There were not enough wordlists available in the HINT database to retest all the algorithms with TE+SE+EE included.

With that in consideration, Figure 4-7 does illustrate some clear benefit of TE+SE+EE over TE+EE at the more severe simulation levels. Again, this suggests that the individual benefits of TE, EE and SE are additive with little interference between them. Surprisingly, due to a large standard deviation at the severe simulation level, there was no statistically significant difference between TE+EE and TE+SE+EE, although the average word scores were quite separated. The fact that TE+EE performs better at the mild simulation level suggests that more natural sounding speech performs better at low levels of simulator distortion.

More data is required to make a better judgment on the effectiveness of spectral enhancement for speech discrimination in AN. The combination of Experiment II results, the visual enhancements to simulated stimuli shown in Figure 4-4 and audible improvements when directly comparing TE+EE and TE+SE+EE processed sentences suggest benefits of spectral enhancement that motivate future analysis. Of course, as for the evaluation of all enhancement schemes, the most informative and authoritative data would come from subjective evaluation on people with AN.

Chapter 5

5 Conclusions and Future Work

5.1 Summary and Contributions

This thesis investigated the benefit of combining temporal envelope enhancement, time scale modification and spectral enhancement strategies in improving speech understanding capabilities for persons with auditory neuropath disorder. Algorithm development and subjective evaluation have shown that these separate enhancement strategies targeting the three main psychoacoustic impairments in people with auditory neuropathy (modulation thresholds, gap detection thresholds and frequency discrimination) may be combined in various forms to produce superior word discrimination benefits than as standalone enhancement schemes.

Experiment I demonstrated, through the use of a simulator and normal hearing study participants, improved performance over the published envelope enhancement algorithm by combining time and envelope enhancement. It appeared that the benefits of envelope and time enhancement were additive and do not negatively interfere with each other. Envelope enhancement and time enhancement alone were also evaluated in the subjective study, as well as a control condition of unenhanced speech. All algorithms showed significant improvement over the unenhanced speech emphasizing the importance of assistive devices that target these psychoacoustic impairments for people with AN.

Experiment II explored the addition of spectral enhancement to the best performing algorithm from Experiment I, the time + envelope enhancement. Although time and ethics approval restrictions limited a thorough analysis, promising improvements, especially at severe and profound simulation levels, were demonstrated by the time + spectral + envelope enhancement algorithm. Again, it was worth noting how these enhancement schemes, which target separate psychoacoustic impairments, can be combined with little negative interference.

In conclusion, these promising improvements on simulated word recognition warrant further testing with AN subjects. Furthermore, these contributions provide more incentive for the development of assistive devices employing these algorithms. As outlined in Chapter 1, in many cases, conventional hearing aids provide no assistance to AN patients and cochlear implants require surgical procedures, are cost and resource consuming and are not suitable for all patients.

5.2 Future Work

5.2.1 Subjective Testing on Subjects with Auditory Neuropathy

In order to most effectively determine the benefit of each processing scheme, subjective evaluations using the HINT stimuli must be completed on AN subjects. The simulator provides intuition into what algorithms provide the best enhancement, but more conclusive evidence only can be drawn from subjects with the disorder. In addition, the effectiveness of the simulator for evaluating time enhancement is not entirely known.

5.2.2 Integrate VOPs into Envelope and Spectral Enhancement

It is believed that there is potential in the processing approach described in Section 3.2.2 that limits envelope enhancement to the consonant and consonant-vowel transition regions, as determined by the vowel onset points calculated in the time enhancement algorithm. An improved method for smoothly transitioning between envelope-enhanced and unenhanced regions must be developed to obtain beneficial results from this approach.

In a similar manner, applying spectral enhancement to only the vowel regions of speech, where formant discrimination is most critical, may prove beneficial. These speech cue-dependent processing schemes may provide a significant improvement in speech quality, as in most cases, less processing equates to more natural sounding speech and less interference between algorithms (although little interference was observed).

5.2.3 Binaural Effects of Proposed Algorithms

The binaural effects of these algorithms are largely unknown, thus work should be completed to determine the effect of independent processing in each ear, whether processing should be symmetric in each ear, as well as the general effect of AN on binaural processing. In a normal hearing listener, binaural processing (efficiently combining and processing signals from left and right ears) leads to a number of benefits such as sound localization, and speech understanding in challenging environments. There is evidence that AN persons do exhibit impairments in binaural processing. It is therefore not clear if the same processing parameters are utilized for both left and right inputs. Additional research is warranted in this area.

5.2.4 Realtime Implementation of Proposed Algorithms

Realtime implementation of the proposed algorithms would require modifications to the processing schemes.

Envelope enhancement:

It appears that the envelope enhancement algorithm would not be too computationally expensive to run on a realtime platform. Special attention would have to be given phase corrections in the envelope detector, and the minimum envelope value, E_{min}, would have to be updated as a long-term average, or may even be assumed as zero. It may be useful for the user to have control over adjusting the maximum and minimum powers of expansion, K_{max} and K_{min}, as well as the exponential time constant, τ, to tune the algorithm to their individual needs and impairments. Less expansion results in more natural sounding speech.

Time enhancement:

Implementing time enhancement in real time poses an interesting problem. If the output duration is longer than the input, then the lag between input and output of the assistive hearing device would become larger and larger and this would cause a significant effect on how the user links visual information to what they're hearing. A possible approach may be to make use of a voice activity detector to allow the output to "catch up" to the input during silence intervals. Furthermore, tuning of the expansion factor, β, may be very important. The required β for a given case of AN may depend on two factors: the gap detection thresholds of the user and the rate of input speech. Therefore perhaps β could be controlled adaptively depending on the syllabic rate (or rate of VOPs since they're already being calculated) to allow for the minimum time expansion (according to gap detection thresholds) that provides the necessary time enhancement (less time expansion means less input-output lag).

Spectral enhancement:

The companding algorithm was designed for realtime implementation, thus there should be little to account for. However, it may be beneficial once again for the user to adjust and tune the filter parameters, q_1 and q_2, and the expansion/compression coefficients, n_1 and n_2 for their individual needs and speech environments.

Bibliography

[1] L. Chittka and A. Brockmann, "Perception Space—The Final Frontier," *PLoS Biology*, vol. 3, no. 4, p. e137, 2005.

[2] W. A. Yost, "Structure of the inner ear and its mechanical response," in *Fundamentals of hearing: an introduction*, 5th ed., Burlington, MA: Elsevier Inc., 2007, pp. 83-102.

[3] W. A. Yost, "The central auditory nervous system," in *Fundamentals of hearing: an introduction*, 5th ed., Burlington, MA: Elsevier Inc., 2007, pp. 223-248.

[4] "File:Aud pathway.png - Wikipedia, the free encyclopedia." [Online]. Available: http://en.wikipedia.org/wiki/File:Aud_pathway.png. [Accessed: 07-Jun-2011].

[5] "A profile of disability in Canada." Statistics Canada, 2002.

[6] S. Kochkin, "MarkeTrak VII: Hearing loss population tops 31 million people," *Hear Rev*, vol. 12, no. 7, pp. 16–29, 2005.

[7] F. G. Zeng, "Auditory neuropathy: why some hearing-impaired listeners can hear but do not understand and how can DSP technology help them," in *IEEE Signal Processing Society, Ninth DSP (DSP 2000) Workshop, Hunt, TX*, 2000.

[8] B. Stach, "The nature of hearing impairment," in *Clinical Audiology: An Introduction*, Singular Publishing Group, Inc., 1998, pp. 89-116.

[9] V. K. Narne and C. S. Vanaja, "Perception of speech with envelope enhancement in individuals with auditory neuropathy and simulated loss of temporal modulation processing," *International Journal of Audiology*, vol. 48, no. 10, pp. 700-707, Jan. 2009.

[10] D. Worthington and J. Peters, "Quantifiable hearing and no ABR - paradox or error?," *Ear and Hearing*, vol. 1, no. 5, pp. 281-285, 1980.

[11] H. Davis and S. Hirsh, "A slow brain stem response for low-frequency audiometry," *Audiology*, vol. 18, pp. 445-461, 1979.

[12] M. Lenhardt, "Childhood Central Auditory Processing Disorder With Brainstem Evoked Response Verification," *Archives of Otolaryngology*, vol. 107, no. 10, pp. 623-625, Oct. 1981.

[13] N. Kraus, "Auditory neuropathy: an historical and current perspective," in *Auditory neuropathy: a new perspective on hearing disorders*, Singular Thomson Learning, 2001, pp. 1-14.

[14] N. Kraus, O. Ozdamar, L. Stein, and N. Reed, "Absent auditory brain stem response: peripheral hearing loss or brain stem dysfunction?," *Laryngoscope*, vol. 94, pp. 400-406, Mar. 1984.

[15] G. Rance et al., "Clinical findings for a group of infants and young children with auditory neuropathy," *Ear and Hearing*, vol. 20, no. 3, p. 238, 1999.

[16] Y. Sininger and S. Oba, "Patients with auditory neuropathy: who are they and what can they hear?," in *Auditory neuropathy: a new perspective on hearing disorders*, Singular Thomson Learning, 2001, pp. 15-26.

[17] F. G. Zeng, S. Oba, S. Garde, Y. Sininger, and A. Starr, "Temporal and speech processing deficits in auditory neuropathy," *Neuroreport*, vol. 10, no. 16, p. 3429, 1999.

[18] F.-G. Zeng, S. Oba, S. Garde, Y. Sininger, and A. Starr, "Psychoacoustics and speech perception in auditory neuropathy," in *Auditory neuropathy: a new perspective on hearing disorders*, Singular Thomson Learning, 2001, pp. 141-164.

[19] B. Cone-Wesson, G. Rance, and Y. Sininger, "Amplification and rehabilitation strategies for patients with auditory neuropathy," in *Auditory neuropathy: a new perspective on hearing disorders*, Singular Thomson Learning, 2001, pp. 233-250.

[20] G. Rance and E. J. Barker, "Speech perception in children with auditory neuropathy/dyssynchrony managed with either hearing aids or cochlear implants," *Otology & Neurotology*, vol. 29, no. 2, p. 179, 2008.

[21] P. Trautwein, J. Shallop, L. Fabry, and R. Freidman, "Cochlear implementation of patients with auditory neuropathy," in *Auditory neuropathy: a new perspective on hearing disorders*, Singular Thomson Learning, 2001, pp. 203-232.

[22] W. Gstoettner et al., "Cochlear Implant Deep Electrode Insertion: Extent of Insertional Trauma," *Acta Otolaryngologica*, vol. 117, no. 2, pp. 274-277, 1997.

[23] V. K. Narne and C. S. Vanaja, "Effect of envelope enhancement on speech perception in individuals with auditory neuropathy," *Ear and Hearing*, vol. 29, no. 1, p. 45, 2008.

[24] V. K. Narne and C. S. Vanaja, "Perception of Envelope-Enhanced Speech in the Presence of Noise by Individuals with Auditory Neuropathy," *Ear and Hearing*, vol. 30, no. 1, p. 136, 2009.

[25] D. M. Hassan, "Perception of temporally modified speech in auditory neuropathy," *International Journal of Audiology*, vol. 50, no. 1, pp. 41-49, Jan. 2011.

[26] L. Turicchia and R. Sarpeshkar, "A bio-inspired companding strategy for spectral enhancement," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 2, pp. 243-253, Mar. 2005.

[27] A. Bhattacharya and F.-G. Zeng, "Companding to improve cochlear-implant speech recognition in speech-shaped noise," *The Journal of the Acoustical Society of America*, vol. 122, no. 2, p. 1079, 2007.

[28] R. Drullman, J. M. Festen, and R. Plomp, "Effect of temporal envelope smearing on speech reception," *Acoustical Society of America*, vol. 95, no. 2, pp. 1053-1064, Feb. 1994.

[29] Q. J. Fu and R. V. Shannon, "Effects of amplitude nonlinearity on phoneme recognition by cochlear implant users and normal-hearing listeners," *The Journal of the Acoustical Society of America*, vol. 104, p. 2570, 1998.

[30] C. Lorenzi, F. Berthommier, F. Apoux, and N. Bacri, "Effects of envelope expansion on speech recognition," *Hearing Research*, vol. 136, pp. 131-138, 1999.

[31] F. Apoux, N. Tribut, X. Debruille, and C. Lorenzi, "Identification of envelope-expanded sentences in normal-hearing and hearing-impaired listeners," *Hearing research*, vol. 189, no. 1-2, pp. 13–24, 2004.

[32] M. Fink, J. Churan, and M. Wittmann, "Assessment of auditory temporal-order thresholds-A comparison of different measurement procedures and the influences of age and gender," *Restorative neurology and neuroscience*, vol. 23, no. 5, pp. 281–296, 2005.

[33] F. G. Zeng and S. Liu, "Speech perception in individuals with auditory neuropathy," *Journal of Speech, Language, and Hearing Research*, vol. 49, no. 2, p. 367, 2006.

[34] B. A. M. Franck, C. S. G. M. van Kreveld-Bos, W. A. Dreschler, and H. Verschuure, "Evaluation of spectral enhancement in hearing aids, combined with phonemic compression," *The Journal of the Acoustical Society of America*, vol. 106, p. 1452, 1999.

[35] T. Baer, B. C. J. Moore, and S. Gatehouse, "Spectral contrast enhancement of speech in noise for listeners with sensorineural hearing impairment: effects on intelligibility, quality, and response times," *Journal of rehabilitation research and development*, vol. 30, pp. 49–49, 1993.

[36] H. Dillon, "Speech cue enhancement," in *Hearing Aids*, Australia: Boomerang Press, 2001, p. 205.

[37] K. S. Rao and B. Yegnanarayana, "Prosody modification using instants of significant excitation," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, no. 3, pp. 972-980, May. 2006.

[38] K. S. Rao and B. Yegnanarayana, "Duration modification using glottal closure instants and vowel onset points," *Speech Communication*, vol. 51, no. 12, pp. 1263–1269, 2009.

[39] M. Brookes, "VOICEBOX: Speech Processing Toolbox for MATLAB," *VOICEBOX: Speech processing toolbox for MATLAB*. [Online]. Available: http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html. [Accessed: 30-May-2011].

[40] S. Murthy and B. Yegnanarayana, "Robustness of group-delay-based method for extraction of significant instants of excitation from speech signals," *Speech and Audio Processing, IEEE Transactions on*, vol. 7, no. 6, pp. 609–619, 1999.

[41] M. Nilsson, S. D. Soli, and J. A. Sulivan, "Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise," *J Acoust Soc Am.*, vol. 95, no. 2, pp. 1085-1099, Feb. 1994.

[42] C. Lee, "Horatio." [Online]. Available: http://publish.uwo.ca/~cjlee/horatio/index.html. [Accessed: 02-Jun-2011].

[43] Pioneer Hill Software, "Real Time Analyzer : Vibration Analysis : FFT Spectrum Analyzer : Vibration Testing : SpectraPlus.com." [Online]. Available: http://www.telebyte.com/pioneer/. [Accessed: 02-Jun-2011].

[44] Y. Benjamini, D. Drai, G. Elmer, N. Kafkafi, and I. Golani, "Controlling the false discovery rate in behavior genetics research," *Behavioural brain research*, vol. 125, no. 1-2, pp. 279–284, 2001.

APPENDIX A – ETHICS APPROVAL NOTICE

Western

Office of Research Ethics

The University of Western Ontario
Room 5150 Support Services Building, London, ON, Canada N6A 3K7
Telephone: (519) 661-3036 Fax: (519) 850-2466 Email: ethics@uwo.ca
Website: www.uwo.ca/research/ethics

Use of Human Subjects - Ethics Approval Notice

Principal Investigator: Dr. V. Parsa
Review Number: 17668E **Review Level:** Expedited
Review Date: December 15, 2010 **Approved Local # of Participants:** 32
Protocol Title: Subjective evaluation of auditory neuropathy speech enhancement algorithms.
Department and Institution: Communication Sciences & Disorders, University of Western Ontario
Sponsor: Ontario Research Fund - Research Excellence
Ethics Approval Date: February 11, 2011 **Expiry Date:** September 30, 2011
Documents Reviewed and Approved: UWO Protocol. Letter of Information and Consent. Email.
Documents Received for Information:

This is to notify you that The University of Western Ontario Research Ethics Board for Health Sciences Research Involving Human Subjects (HSREB) which is organized and operates according to the Tri-Council Policy Statement: Ethical Conduct of Research Involving Humans and the Health Canada/ICH Good Clinical Practice Practices: Consolidated Guidelines; and the applicable laws and regulations of Ontario has reviewed and granted approval to the above referenced study on the approval date noted above. The membership of this REB also complies with the membership requirements for REB's as defined in Division 5 of the Food and Drug Regulations.

The ethics approval for this study shall remain valid until the expiry date noted above assuming timely and acceptable responses to the HSREB's periodic requests for surveillance and monitoring information. If you require an updated approval notice prior to that time you must request it using the UWO Updated Approval Request Form.

During the course of the research, no deviations from, or changes to, the protocol or consent form may be initiated without prior written approval from the HSREB except when necessary to eliminate immediate hazards to the subject or when the change(s) involve only logistical or administrative aspects of the study (e.g. change of monitor, telephone number). Expedited review of minor change(s) in ongoing studies will be considered. Subjects must receive a copy of the signed information/consent documentation.

Investigators must promptly also report to the HSREB:

a) changes increasing the risk to the participant(s) and/or affecting significantly the conduct of the study;
b) all adverse and unexpected experiences or events that are both serious and unexpected;
c) new information that may adversely affect the safety of the subjects or the conduct of the study.

If these changes/adverse events require a change to the information/consent documentation, and/or recruitment advertisement, the newly revised information/consent documentation, and/or advertisement, must be submitted to this office for approval.

Members of the HSREB who are named as investigators in research studies, or declare a conflict of interest, do not participate in discussion related to, nor vote on, such studies when they are presented to the HSREB.

[redacted]

Chair of HSREB: Dr. Joseph Gilbert
FDA Ref #: IRB 00000940

Ethics Officer to Contact for Further Information		
☐ Janice Sutherland (jsutherl@uwo.ca)	☐ Elizabeth Wambolt (ewambolt@uwo.ca)	☑ Grace Kelly (grace.kelly@uwo.ca)

This is an official document. Please retain the original in your files.

cc: ORE File

UWO HSREB Ethics Approval - Initial
V.2008-07-01 (rptApprovalNoticeHSREB_Initial) 17668E Page 1 of 1

Appendix B – SPSS Outputs

Experiment I Repeated Measures SPSS Output

General Linear Model

Within-Subjects Factors

Measure:MEASURE_1

Algorithm	Simulator	Dependent Variable
1	1	EE_Profound
	2	EE_Severe
	3	EE_Moderate
	4	EE_Mild
2	1	TE_EE_Profound
	2	TE_EE_Severe
	3	TE_EE_Moderate
	4	TE_EE_Mild
3	1	TE_Profound
	2	TE_Severe
	3	TE_Moderate
	4	TE_Mild
4	1	Unprocessed_Profound
	2	Unprocessed_Severe
	3	Unprocessed_Moderate
	4	Unprocessed_Mild

Multivariate Tests[b]

Effect		Value	F	Hypothesis df	Error df	Sig.
Algorithm	Pillai's Trace	.795	16.840[a]	3.000	13.000	.000
	Wilks' Lambda	.205	16.840[a]	3.000	13.000	.000
	Hotelling's Trace	3.886	16.840[a]	3.000	13.000	.000
	Roy's Largest Root	3.886	16.840[a]	3.000	13.000	.000
Simulator	Pillai's Trace	.979	205.591[a]	3.000	13.000	.000
	Wilks' Lambda	.021	205.591[a]	3.000	13.000	.000
	Hotelling's Trace	47.444	205.591[a]	3.000	13.000	.000
	Roy's Largest Root	47.444	205.591[a]	3.000	13.000	.000
Algorithm * Simulator	Pillai's Trace	.809	3.286[a]	9.000	7.000	.065
	Wilks' Lambda	.191	3.286[a]	9.000	7.000	.065
	Hotelling's Trace	4.224	3.286[a]	9.000	7.000	.065
	Roy's Largest Root	4.224	3.286[a]	9.000	7.000	.065

a. Exact statistic

b. Design: Intercept

Within Subjects Design: Algorithm + Simulator + Algorithm * Simulator

Mauchly's Test of Sphericity[b]

Measure:MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[a]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Algorithm	.959	.576	5	.989	.971	1.000	.333
Simulator	.505	9.369	5	.096	.693	.806	.333
Algorithm * Simulator	.016	48.702	44	.348	.595	.963	.111

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Mauchly's Test of Sphericity[b]

Measure:MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[a]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Algorithm	.959	.576	5	.989	.971	1.000	.333
Simulator	.505	9.369	5	.096	.693	.806	.333
Algorithm * Simulator	.016	48.702	44	.348	.595	.963	.111

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

b. Design: Intercept

Within Subjects Design: Algorithm + Simulator + Algorithm * Simulator

Tests of Within-Subjects Effects

Measure:MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Algorithm	Sphericity Assumed	1.347	3	.449	17.749	.000
	Greenhouse-Geisser	1.347	2.913	.462	17.749	.000
	Huynh-Feldt	1.347	3.000	.449	17.749	.000
	Lower-bound	1.347	1.000	1.347	17.749	.001
Error(Algorithm)	Sphericity Assumed	1.138	45	.025		
	Greenhouse-Geisser	1.138	43.702	.026		
	Huynh-Feldt	1.138	45.000	.025		
	Lower-bound	1.138	15.000	.076		
Simulator	Sphericity Assumed	21.998	3	7.333	170.151	.000

	Greenhouse-Geisser	21.998	2.078	10.584	170.151	.000
	Huynh-Feldt	21.998	2.419	9.094	170.151	.000
	Lower-bound	21.998	1.000	21.998	170.151	.000
Error(Simulator)	Sphericity Assumed	1.939	45	.043		
	Greenhouse-Geisser	1.939	31.177	.062		
	Huynh-Feldt	1.939	36.283	.053		
	Lower-bound	1.939	15.000	.129		
Algorithm * Simulator	Sphericity Assumed	.634	9	.070	3.131	.002
	Greenhouse-Geisser	.634	5.352	.119	3.131	.011
	Huynh-Feldt	.634	8.667	.073	3.131	.002
	Lower-bound	.634	1.000	.634	3.131	.097
Error(Algorithm*Simulator)	Sphericity Assumed	3.039	135	.023		
	Greenhouse-Geisser	3.039	80.275	.038		
	Huynh-Feldt	3.039	130.012	.023		
	Lower-bound	3.039	15.000	.203		

Tests of Within-Subjects Contrasts

Measure:MEASURE_1

Source		Simulator	Type III Sum of Squares	df	Mean Square	F	Sig.
Algorithm	Level 2 vs. Level 1		.070	1	.070	5.512	.033
	Level 3 vs. Level 1		.046	1	.046	3.914	.067
	Level 4 vs. Level 1		.277	1	.277	18.168	.001
Error(Algorithm)	Level 2 vs. Level 1		.191	15	.013		
	Level 3 vs. Level 1		.177	15	.012		

	Level 4 vs. Level 1		.229	15	.015		
Simulator		Level 2 vs. Level 1	1.023	1	1.023	32.969	.000
		Level 3 vs. Level 1	6.135	1	6.135	555.830	.000
		Level 4 vs. Level 1	8.706	1	8.706	568.731	.000
Error(Simulator)		Level 2 vs. Level 1	.465	15	.031		
		Level 3 vs. Level 1	.166	15	.011		
		Level 4 vs. Level 1	.230	15	.015		
Algorithm * Simulator	Level 2 vs. Level 1	Level 2 vs. Level 1	.134	1	.134	.887	.361
		Level 3 vs. Level 1	.003	1	.003	.028	.869
		Level 4 vs. Level 1	.048	1	.048	1.054	.321
	Level 3 vs. Level 1	Level 2 vs. Level 1	.430	1	.430	5.386	.035
		Level 3 vs. Level 1	.311	1	.311	3.425	.084
		Level 4 vs. Level 1	.110	1	.110	2.291	.151
	Level 4 vs. Level 1	Level 2 vs. Level 1	.765	1	.765	6.643	.021
		Level 3 vs. Level 1	.011	1	.011	.105	.750
		Level 4 vs. Level 1	.004	1	.004	.150	.704
Error(Algorithm*Simulator)	Level 2 vs. Level 1	Level 2 vs. Level 1	2.270	15	.151		

		Level 3 vs. Level 1	1.847	15	.123		
		Level 4 vs. Level 1	.688	15	.046		
	Level 3 vs. Level 1	Level 2 vs. Level 1	1.197	15	.080		
		Level 3 vs. Level 1	1.362	15	.091		
		Level 4 vs. Level 1	.719	15	.048		
	Level 4 vs. Level 1	Level 2 vs. Level 1	1.727	15	.115		
		Level 3 vs. Level 1	1.560	15	.104		
		Level 4 vs. Level 1	.408	15	.027		

Tests of Between-Subjects Effects

Measure:MEASURE_1

Transformed Variable:Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	5.828	1	5.828	722.984	.000
Error	.121	15	.008		

Experiment I Paired Samples SPSS Output

T-Test

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	EE	.6333	16	.14207	.03552
	TE_EE	.6995	16	.11622	.02906
Pair 2	EE	.6333	16	.14207	.03552
	TE	.5797	16	.08397	.02099
Pair 3	EE	.6333	16	.14207	.03552
	Unprocessed	.5017	16	.10234	.02558
Pair 4	TE_EE	.6995	16	.11622	.02906
	TE	.5797	16	.08397	.02099
Pair 5	TE_EE	.6995	16	.11622	.02906
	Unprocessed	.5017	16	.10234	.02558
Pair 6	TE	.5797	16	.08397	.02099
	Unprocessed	.5017	16	.10234	.02558

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	EE & TE_EE	16	.636	.008
Pair 2	EE & TE	16	.648	.007
Pair 3	EE & Unprocessed	16	.529	.035
Pair 4	TE_EE & TE	16	.466	.069
Pair 5	TE_EE & Unprocessed	16	.527	.036
Pair 6	TE & Unprocessed	16	.250	.351

Paired Samples Test

		Paired Differences							
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
Pair 1	EE - TE_EE	-.06615	.11270	.02818	-.12620	-.00609	-2.348	15	.033
Pair 2	EE - TE	.05368	.10853	.02713	-.00415	.11151	1.978	15	.067
Pair 3	EE - Unprocessed	.13168	.12357	.03089	.06583	.19753	4.262	15	.001
Pair 4	TE_EE - TE	.11983	.10703	.02676	.06280	.17686	4.478	15	.000
Pair 5	TE_EE - Unprocessed	.19783	.10696	.02674	.14084	.25482	7.398	15	.000
Pair 6	TE - Unprocessed	.07800	.11505	.02876	.01670	.13930	2.712	15	.016

Experiment II Repeated Measures Output

General Linear Model

Within-Subjects Factors

Measure:MEASURE_1

Algorithm	Simulator	Dependent Variable
1	1	EE_Profound
	2	EE_Severe
	3	EE_Moderate
	4	EE_Mild
2	1	TE_EE_Profound
	2	TE_EE_Severe
	3	TE_EE_Moderate
	4	TE_EE_Mild
3	1	TE_Profound
	2	TE_Severe
	3	TE_Moderate
	4	TE_Mild
4	1	Unprocessed_Profound
	2	Unprocessed_Severe
	3	Unprocessed_Moderate
	4	Unprocessed_Mild
5	1	TE_SE_EE_Profound
	2	TE_SE_EE_Severe
	3	TE_SE_EE_Moderate
	4	TE_SE_EE_Mild

Multivariate Tests[c]

Effect		Value	F	Hypothesis df	Error df	Sig.
Algorithm	Pillai's Trace	.903	18.672[a]	4.000	8.000	.000
	Wilks' Lambda	.097	18.672[a]	4.000	8.000	.000
	Hotelling's Trace	9.336	18.672[a]	4.000	8.000	.000
	Roy's Largest Root	9.336	18.672[a]	4.000	8.000	.000
Simulator	Pillai's Trace	.983	178.138[a]	3.000	9.000	.000
	Wilks' Lambda	.017	178.138[a]	3.000	9.000	.000
	Hotelling's Trace	59.379	178.138[a]	3.000	9.000	.000
	Roy's Largest Root	59.379	178.138[a]	3.000	9.000	.000
Algorithm * Simulator	Pillai's Trace	.[b]	.	.	.	.
	Wilks' Lambda	.[b]	.	.	.	.
	Hotelling's Trace	.[b]	.	.	.	.
	Roy's Largest Root	.[b]	.	.	.	.

a. Exact statistic

b. Cannot produce multivariate test statistics because of insufficient residual degrees of freedom.

c. Design: Intercept

Within Subjects Design: Algorithm + Simulator + Algorithm * Simulator

Mauchly's Test of Sphericity[b]

Measure:MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[a]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Algorithm	.398	8.666	9	.476	.753	1.000	.250
Simulator	.512	6.508	5	.262	.698	.865	.333
Algorithm * Simulator	.000	.	77	.	.457	.963	.083

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

b. Design: Intercept

Within Subjects Design: Algorithm + Simulator + Algorithm * Simulator

Tests of Within-Subjects Effects

Measure:MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Algorithm	Sphericity Assumed	1.581	4	.395	16.883	.000
	Greenhouse-Geisser	1.581	3.011	.525	16.883	.000
	Huynh-Feldt	1.581	4.000	.395	16.883	.000
	Lower-bound	1.581	1.000	1.581	16.883	.002
Error(Algorithm)	Sphericity Assumed	1.030	44	.023		
	Greenhouse-Geisser	1.030	33.122	.031		
	Huynh-Feldt	1.030	44.000	.023		
	Lower-bound	1.030	11.000	.094		
Simulator	Sphericity Assumed	19.407	3	6.469	180.459	.000
	Greenhouse-Geisser	19.407	2.093	9.272	180.459	.000
	Huynh-Feldt	19.407	2.595	7.477	180.459	.000
	Lower-bound	19.407	1.000	19.407	180.459	.000
Error(Simulator)	Sphericity Assumed	1.183	33	.036		
	Greenhouse-Geisser	1.183	23.024	.051		
	Huynh-Feldt	1.183	28.550	.041		
	Lower-bound	1.183	11.000	.108		
Algorithm * Simulator	Sphericity Assumed	.768	12	.064	2.803	.002
	Greenhouse-Geisser	.768	5.481	.140	2.803	.021
	Huynh-Feldt	.768	11.554	.066	2.803	.002
	Lower-bound	.768	1.000	.768	2.803	.122
Error(Algorithm*Simulator)	Sphericity Assumed	3.015	132	.023		
	Greenhouse-Geisser	3.015	60.289	.050		
	Huynh-Feldt	3.015	127.097	.024		
	Lower-bound	3.015	11.000	.274		

Tests of Within-Subjects Contrasts

Measure:MEASURE_1

Source		Simulator	Type III Sum of Squares	df	Mean Square	F	Sig.
Algorithm	Level 1 vs. Level 5		.158	1	.158	11.338	.006
	Level 2 vs. Level 5		.015	1	.015	3.585	.085
	Level 3 vs. Level 5		.236	1	.236	24.832	.000
	Level 4 vs. Level 5		.636	1	.636	78.462	.000
Error(Algorithm)	Level 1 vs. Level 5		.153	11	.014		
	Level 2 vs. Level 5		.045	11	.004		
	Level 3 vs. Level 5		.104	11	.009		
	Level 4 vs. Level 5		.089	11	.008		
Simulator		Level 1 vs. Level 4	6.477	1	6.477	565.916	.000
		Level 2 vs. Level 4	2.036	1	2.036	95.314	.000
		Level 3 vs. Level 4	.168	1	.168	27.491	.000
Error(Simulator)		Level 1 vs. Level 4	.126	11	.011		
		Level 2 vs. Level 4	.235	11	.021		
		Level 3 vs. Level 4	.067	11	.006		
Algorithm * Simulator	Level 1 vs. Level 5	Level 1 vs. Level 4	.280	1	.280	3.170	.103
		Level 2 vs. Level 4	.212	1	.212	3.790	.078
		Level 3 vs. Level 4	.419	1	.419	7.431	.020
	Level 2 vs. Level 5	Level 1 vs. Level 4	.186	1	.186	12.736	.004
		Level 2 vs. Level 4	.226	1	.226	2.703	.128
		Level 3 vs. Level 4	.045	1	.045	1.611	.230
	Level 3 vs. Level 5	Level 1 vs. Level 4	.615	1	.615	9.057	.012
		Level 2 vs. Level 4	1.194	1	1.194	19.162	.001
		Level 3 vs. Level 4	.044	1	.044	2.078	.177

	Level 4 vs. Level 5	Level 1 vs. Level 4	.320	1	.320	2.671	.130
		Level 2 vs. Level 4	1.481	1	1.481	21.092	.001
		Level 3 vs. Level 4	.274	1	.274	6.006	.032
Error(Algorithm*Simulator)	Level 1 vs. Level 5	Level 1 vs. Level 4	.972	11	.088		
		Level 2 vs. Level 4	.615	11	.056		
		Level 3 vs. Level 4	.620	11	.056		
	Level 2 vs. Level 5	Level 1 vs. Level 4	.161	11	.015		
		Level 2 vs. Level 4	.918	11	.083		
		Level 3 vs. Level 4	.308	11	.028		
	Level 3 vs. Level 5	Level 1 vs. Level 4	.747	11	.068		
		Level 2 vs. Level 4	.685	11	.062		
		Level 3 vs. Level 4	.234	11	.021		
	Level 4 vs. Level 5	Level 1 vs. Level 4	1.319	11	.120		
		Level 2 vs. Level 4	.772	11	.070		
		Level 3 vs. Level 4	.503	11	.046		

Tests of Between-Subjects Effects

Measure:MEASURE_1

Transformed Variable:Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	4.812	1	4.812	650.326	.000
Error	.081	11	.007		

Experiment II Paired Samples SPSS Output

T-Test

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	TE_EE_Profound	.2702	12	.18892	.05454
	TE_SE_EE_Profound	.3490	12	.25069	.07237
Pair 2	TE_EE_Severe	.6343	12	.29588	.08541
	TE_SE_EE_Severe	.7257	12	.11889	.03432
Pair 3	TE_EE_Moderate	.9085	12	.13811	.03987
	TE_SE_EE_Moderate	.9241	12	.06691	.01931
Pair 4	TE_EE_Mild	.9960	12	.00938	.00271
	TE_SE_EE_Mild	.9502	12	.04760	.01374

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	TE_EE_Profound & TE_SE_EE_Profound	12	.948	.000
Pair 2	TE_EE_Severe & TE_SE_EE_Severe	12	.364	.245
Pair 3	TE_EE_Moderate & TE_SE_EE_Moderate	12	-.129	.691
Pair 4	TE_EE_Mild & TE_SE_EE_Mild	12	-.252	.430

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	TE_EE_Profound - TE_SE_EE_Profound	-.07883	.09369	.02704	-.13835	-.01930	-2.915	11	.014
Pair 2	TE_EE_Severe - TE_SE_EE_Severe	-.09137	.27580	.07962	-.26661	.08386	-1.148	11	.275
Pair 3	TE_EE_Moderate - TE_SE_EE_Moderate	-.01562	.16102	.04648	-.11793	.08668	-.336	11	.743
Pair 4	TE_EE_Mild - TE_SE_EE_Mild	.04574	.05078	.01466	.01347	.07800	3.120	11	.010