Perception of Speech with Spectral and Temporal Modifications by Individuals with Auditory Dys-synchrony

A Thesis

Submitted to the University of Mysore,
for the award of degree of
Doctor of Philosophy (Ph.D) in Speech & Hearing.

By

Vijaya Kumar Narne

Under the Guidance of

Dr. C.S. Vanaja

All India Institute of Speech and Hearing, Mysore - 6, India
May, 2008.

Certificate

This is to certify that the thesis entitled **"Perception of Speech with Spectral and Temporal Modifications by Individuals with Auditory Dys-synchrony"** submitted by Mr. Vijaya Kumar Narne for the degree of Ph.D. in Speech and Hearing, to the University of Mysore, Mysore, is the result of the work done by him at All India Institute of Speech and Hearing, Mysore, under my guidance. I further declare that the results of this work have not been previously submitted for any degree.

Place: Mysore
Date: 8/5/08

Prof. C.S. Vanaja
Guide
Professor in Audiology,
BVU School of ASLP,
Bharati Vidyapeeth University,
Pune.

Certificate

This is to certify that the thesis entitled **"Perception of Speech with Spectral and Temporal Modifications by Individuals with Auditory Dys-synchrony"** submitted by Mr. Vijaya Kumar Name for the degree of Doctor of Philosophy to the University of Mysore, Mysore was carried out at All India Institute of Speech and Hearing, Mysore.

Place: Mysore
Date: 26/05/08

Dr.Vijayalakshmi Basavaraj
Director
A.I.I.S.H
Mysore - 570 006

Declaration

I declare that the thesis entitled **"Perception of Speech with Spectral and Temporal Modifications by Individuals with Auditory Dys-synchrony"** which is submitted herewith for the award of the degree of Doctor of Philosophy (Speech and Hearing) at the University of Mysore, Mysore, is the result of work carried out by me at the All India Institute of Speech and Hearing, Mysore, under the guidance of Prof.C.S. Vanaja, Ph.D, Professor in Audiology, BVU School of ASLP, Pune.

I further declare that the results of this work have not been previously submitted for any degree.

Place: Mysore
Date: 18/5/08

Vijaya Kumar Narne
Candidate

ACKNOWLEDGEMENTS

First of all, I would like to express my deep sense of respect and gratitude towards my Guide Dr. C.S. Vanaja, who has been the guiding force behind this work. I am greatly indebted for her constant encouragement and invaluable advice in every aspect of my academic life. Thank you very much madam.

I respectfully acknowledge Dr. Vijayalakshmi Basavaraj, Director, AIISH, for all her support, and advice.

I would also like to thank Dr. M. Jayaram for the various opportunities he gave me and his support.

I extend my appreciation to Prof. Asha Yathiraj, for granting me the required support from the department and help for carrying out one of my studies.

My sincere thanks to Dr. Lorenzi, C., Dr. Apoux, F., Eddins and Jagadish N, for guiding me through the signal processing.

My heartfelt acknowledgements are due to Dr. Venkatesan, Prof. Savithri, Mr. Animesh Barman, Manjula, P., Ajith Kumar, U., Reddy S., and Moore, B.C.J for the productive discussions and advice.

My family - Amma, Nanna and my dear Brothers you have been my biggest and unstinting support at every little step of this journey. I couldn’t have done without you!!

My immense gratitude is due to all the participants of my study.

Warm thanks are also due to all those who have directly or indirectly helped me to make this thesis a reality.

Table of Contents

List of Tables

List of Figures

CHAPTER I

Introduction

Spoken language is one of the most common forms of communication that enables humans to convey information. For most of us perceiving speech is an effortless and overlooked task whereas, for many individuals with a hearing problem, communication is a difficult task as speech perception is affected. Different hearing disorders impair speech perception at different stages of processing in the auditory system and different rehabilitative strategies are required for management of these disorders.

Auditory dys-synchrony is a hearing disorder with unique perceptual consequences (Starr, Picton, Sininger, Hood, & Berlin, 1996; Starr et al, 2003; Zeng Oba, Grade, Sininger, & Starr, 1999). One of the main characteristics of auditory dys-synchrony is disrupted auditory nerve activity with normal or near normal cochlear amplification. It is reported that clinically disrupted auditory nerve activity is reflected by absent auditory brainstem response with preserved oto-acoustic emissions and/or cochlear microphonics (Hood, Berlin, Bordelon, & Rose, 2003; Starr et al., 1996). Psychophysical studies indicate that the consequence of disrupted auditory nerve activity is reflected as a significant impairment in temporal processing and difficulty in speech understanding that is disproportionate to the degree of hearing loss measured by pure-tone thresholds (Sininger, & Oba, 2001; Zeng et al., 1999; Zeng, & Liu, 2006).

A number of investigators have studied open-set speech perception in individuals with auditory dys-synchrony and have reported that the scores range from 0 to 100%

(Narne & Vanaja, 2008 b; Starr et al., 1996; Rance, 2005) but approximately 60 to 70 % of individuals have identification scores well below the scores estimated from their pure-tone thresholds (Narne & Vanaja, 2008 b; Sininger & Oba, 2001; Starr et al., 1996). Very few studies have investigated speech perception in noise and their results illustrate that the noise has more detrimental effect on speech perception than that observed for listeners with normal hearing and those with cochlear hearing loss (Rance et al., 2007; Zeng, & Liu, 2006). These results indicate that in contrast to individuals with cochlear pathology, speech perception abilities of individuals with auditory dys-synchrony depend on the extent of distortion of temporal cues at suprathreshold levels rather than audibility (Rance, Mc Kay, & Grayden, 2004; Zeng et al., 1999).

Temporal resolution has been measured in individuals with auditory dys-synchrony using two measures, gap detection (Starr et al., 1996; Zeng et al., 1999; Zeng, Kong, Michalewski, & Starr, 2005) and amplitude modulation detection at different modulation frequencies (Zeng et al., 1999; Rance et al., 2004; Zeng et al., 2005; Kumar, & Jayaram, 2005). Temporal modulation transfer function (TMTF) is defined as a measure of the modulation depth that is required to just detect the presence of modulation of sinusoidal amplitude modulated noise as a function of modulation frequency (Viemeister, 1979). Previous studies indicated that in individuals with normal hearing, sensitivity to modulation is relatively good at low modulation frequencies and becomes progressively poor at higher modulation frequencies (Viemeister, 1979; Bacon & Viemeister, 1985). The modulation detection threshold has been reported to be abnormally high in individuals with auditory neuropathy and there is a good correlation between the modulation threshold and speech perception scores in quiet (Kumar & Jayaram, 2005; Rance et al., 2004; Zeng et al., 2005; Zeng et al., 1999). Kumar and Jayaram (2005) found that the mean peak sensitivity of the modulation detection is also related to speech identification scores. The peak sensitivity was better in individuals who had speech identification scores of greater than 50% when compared to those who had identification scores of less than 20%.

Zeng et al. (1999) edited temporal modulations in speech signal to simulate processing of speech in individuals with auditory dys-synchrony and presented it to listeners with normal hearing. It was found that speech recognition abilities of individuals with normal hearing for the edited speech samples were similar to those seen in individuals with auditory dys-synchrony for normal speech. From these studies, it can be inferred that deficits in their ability to follow amplitude variations in speech signals probably contribute to the difficulty in understanding speech.

The findings observed in individuals with auditory dys-synchrony are in contrast to those observed in patients with cochlear hearing loss whose speech understanding difficulties are substantially accounted for reduced audibility, loudness recruitment and broadening of auditory filters (Moore, 2003). Individuals with cochlear hearing loss, gain significant benefit from hearing aids, which employ nonlinear compression. A number of investigators have demonstrated that conventional hearing aids are of not much use for individuals with auditory dys-synchrony (Berlin, Hood, Morlet, Rose, & Brashears, 2003; Berlin, Hood, Hurely, & Wen, 1996). Other management strategies adopted by individuals with auditory dys-synchrony include FM systems, cochlear implants, perceptual training, speech reading and cued speech (Kraus, 2001). Cochlear implant may be a viable option for some of the individuals with auditory neuropathy. Earlier studies have shown that, benefit from cochlear implants for individuals with auditory dys-synchrony is comparable to that of individuals with cochlear hearing loss (Peterson, et al., 2003; Shallop, Peterson, Facer, Fabry, & Driscoll, 2001; Trautwein, Sininger, & Nelson, 2000). However, recent investigators have shown that not all children with auditory dys-synchrony show comparable benefit with cochlear implants, a few children perform poorer than those with cochlear pathology (Gibson, & Sanli, 2007; Runge-Samuelson, Drake, & Wackym, 2008). Further, the degree of hearing loss may not justify the use of cochlear implant as an option of management for individuals with auditory dys-synchrony (Zeng & Liu, 2006). Therefore, there is a need to explore alternative strategies that are much less invasive and expensive than cochlear implants but may benefit individuals with auditory dys-synchrony.

Zeng and Liu (2006) reported that subjects with auditory dys-synchrony showed improved performance, both in quiet as well as in the presence of noise when clear speech is presented and this improvement is attributed to enhanced envelopes in the clear speech. Narne and Vanaja (2008 a) have shown that enhancing the envelope improved consonant identification in quiet. However, further studies are needed in this direction before making an attempt for online implementation of this technique in hearing aids.

The psycho-acoustical studies have also demonstrated impaired frequency discrimination in individuals with auditory neuropathy. It has been reported (Zeng et al., 1999; Rance et al., 2004) that frequency discrimination is poorer for low frequency sounds (500 Hz) and the performance improves as the frequency is increased, reaching near normal values at 4000 Hz. Consistent to these findings, it has been observed that individuals with auditory dys-synchrony show good identification for phonemes that lie in the high frequency range than those phonemes that lie in the low frequency range (Starr et al., 1991; Narne, & Vanaja, 2008 a). Zeng et al. (2005) hypothesized that, eliminating the low frequency content of speech signal or shifting the low frequency content to high frequencies may improve speech perception and reduce undesirable masking in individuals with auditory neuropathy.

Contrary to this hypothesis, shifting spectral information upward may lead to poor speech perception as shifting low frequency information to higher frequencies will bring a complete a change in frequency coding in the auditory system. It has been well established that low frequency signals are coded using phase locking and high frequency signals are coded using place coding (Moore, 1995). Also at higher frequencies, auditory filters get broader and broader leading to poor frequency resolution (Moore, 1995). Thus, it may become more difficult to understand speech when low frequency information is also shifted to high frequencies. However, usefulness of this option has not been investigated in individuals with auditory dys-synchrony.

1.1.Need for the Study

In the early 1990's it was believed that auditory dys-synchrony is a rare disorder. Growing literature on auditory dys-synchrony has suggested that it is not such an extremely rare disorder. The current literature has reported that the prevalence of auditory dys-synchrony varies from 11% to 0.5% in sensory-neural hearing loss (Kumar & Jayaram, 2006; Rance, et al., 1999; Tang, McPherson, Yuen, Wong, & Lee, 2004). The increasing incidence and prevalence of this problem, and the significant difficulty experienced by this population in daily communication warrant further studies on these populations.

1.1.1. Need for investigating temporal modulation transfer function

Speech understanding involves suprathreshold auditory processing of signals that are spectrally and temporally more complex than pure-tones. The reason for individuals with auditory dys-synchrony having difficulty in understanding speech in quiet and in the presence of noise conditions may be related to poor temporal processing (Narne, & Vanaja, 2008 b; Rance et al., 2007; Zeng et al., 1999). Temporal aspects of speech signal, such as temporal envelope (amplitude variation) and periodicity, have been shown to convey important information about syllable and phrase boundaries, as well as consonant identification (Price & Simon, 1984; Rosen, 1992, Van Tasell, Soli, Kirby & Widin, 1987). Recent studies have shown that the temporal envelope of speech is necessary and almost sufficient for correct speech recognition in quiet (Drullman, Festen, & Plomp, 1994 a, b; Narne, Manjula, & Vanaja, 2007; Shannon, Zeng, Kamath, Wygonisk, & Ekelid, 1995). These results suggest that poor speech understanding may be related in part to abnormal temporal processing. Understanding of speech is adversely affected when modulation depth of the temporal envelope is reduced by noise and/or reverberation (Houtgast, & Steenken, 1985; Houtgast, Steenken, & Plomp, 1980; Nabeleck & Robinson, 1982). Temporal envelope processing is measured by obtaining the sensitivity to sinusoidal amplitude-modulated noise as a function of the modulation frequency (Dau, Kollmeier, & Kohlrausch, 1997).

Previous investigators have assessed the TMTF in individuals with auditory dys-synchrony (Kumar & Jayaram, 2005; Rance et al., 2004; Zeng et al., 1999; Zeng et al., 2005). A majority of the investigators have assessed modulation threshold for the modulation frequencies of 2, 4, 8, 16, 32, 64 Hz, 128 Hz, 256 and 512 Hz (Kumar & Jayaram, 2005; Viemeister, 1979; Zeng et al., 1999). Results of some of the studies indicate that individuals with auditory dys-synchrony have varying degrees of impairment in modulation detection (Kumar & Jayaram, 2005; Zeng et al., 1999; Rance et al., 2004). They further demonstrated that there is a good correlation between peak sensitivity of modulation detection threshold and speech perception. Though investigators have studied the TMTF and their relation with speech perception in individuals with auditory dys-synchrony (Kumar & Jayaram, 2005; Rance et al., 2004; Zeng et al., 1999), a majority of these investigations have been carried out on a very small sample of subjects and not all the parameters of TMTF have been investigated. Rance et al. (2004) investigated modulation detection threshold only at three modulation frequencies (10 Hz, 50 Hz & 100 Hz). Kumar and Jayaram (2005) investigated at six modulation frequencies but they did not study the bandwidth. Studies have investigated the relation between peak sensitivity and speech identification scores, but the relation between bandwidth of TMTF and speech perception has not been investigated. Further studies are needed to replicate the results of TMTF on larger population and also to investigate the relation between speech perception and all the parameters of the TMTF.

1.1.2. Need to study identification of speech with temporal modification

Management of individuals with auditory dys-synchrony continues to be debatable (Berlin et al., 2003; Zeng, & Liu, 2006). As conventional amplification has achieved limited success, it is important to explore alternative strategies that improve speech perception in these individuals. The psycho-acoustical studies on individuals with auditory dys-synchrony have shown that impaired temporal resolution is one of the hallmark characteristic of auditory dys-synchrony (Zeng et al., 1999; Rance et al., 2004). The impaired temporal resolution worsens the ability to process amplitude fluctuations in the speech signal. Poor processing of temporal envelope (amplitude fluctuations) in speech signal impairs processing of syllable and phrase boundaries and further causes smearing of consonantal energy to the following vowel (Drullman et al., 1994 a, b) causing a profound impact on speech perception. Enhancing temporal modulations in speech signal in such a way that it enhances the consonantal energy over the vowel, may improve speech perception in these individuals.

Apoux, Tribut, Debruille, and Lorenzi (2004) applied an envelope enhancement scheme, which enhanced the amplitude of the consonantal portion of the signal while compressing the vowel portion. They observed a significant improvement in identification scores in the presence of background noise when the envelope-enhanced stimuli were presented to individuals with cochlear hearing loss. Research has indicated that increasing consonant vowel ratio improves consonant identification and speech perception in the presence of noise for individuals with hearing impairment (House, William, Hecker, & Kryter, 1965; Revoile, Holden-Pitt, Edward, Pickett, & Brandt, 1986).

A preliminary study carried out by Narne and Vanaja (2008 a) assessed the consonant identification in quiet when the modulations in the speech signal are enhanced. Results revealed a significant improvement in consonant identification when envelope was enhanced. Further studies need to be carried out to corroborate these findings. The usefulness of envelope enhancement in understanding speech in presence of noise has not been investigated. In addition, there are no studies investigating usefulness of envelope enhancement with meaningful stimuli both in quiet and in the presence of noise.

1.1.3. Need to study identification of speech with spectral modification

Proportional frequency compression is another approach that has improved speech perception in individuals with cochlear hearing loss whose hearing thresholds at high frequencies precludes the perception of any useful amplification (Turner, & Hurting, 2002; Reed, Hicks, Brida, & Drulach, 1983). Proportional frequency compression is a concept employed in hearing amplification to transpose high frequency information to low frequency region (McDermott, Dorkos, Dean, & Ching 1999; Turner, & Hurting, 2002). Turner and Hurting (2002) reported that, individuals with high frequency sloping hearing loss showed more improvement in speech recognition with proportional frequency compression than that observed with conventional amplification.

Pure-tone audiograms of subjects with auditory dys-synchrony show more hearing loss at low frequencies with relatively better or normal thresholds at high frequencies. Probably processing of low frequency signal is affected, as it requires phase locking responses, which are affected due to dys-synchronous firing of auditory nerve (Sininger, & Oba, 2001). Further, the energy that is audible in the low frequencies may not be useful, instead may cause unwanted masking (Zeng et al., 2005). Hence, Zeng et al. hypothesized that transferring the low frequency information to high frequencies by upward spectral shift may improve speech perception by individuals with auditory dys-synchrony. However, there is a dearth of research in this direction.

Individuals with auditory dys-synchrony may have deficits either in modulation detection threshold and/or in discrimination at low frequencies. Therefore, there is a need to investigate whether speech perception can be improved if speech is modified both temporally and spectrally.

1.2.Objectives of the Study

The present study was designed to investigate the following in individuals with auditory dys-synchrony and those with normal hearing:

1. Temporal modulation transfer function for sinusoidal amplitude modulated white noise
2. Speech identification ability in quiet and its relation with temporal modulation transfer function
3. Identification of speech with envelope enhancement, in quiet.
4. Speech identification ability in the presence of speech spectrum shaped noise and its relation with temporal modulation transfer function
5. Identification of speech with envelope enhancement, in the presence of speech spectrum shaped noise.
6. Identification of speech with upward spectral (frequency) shifts, in quiet.
7. Identification of speech with upward spectral shift along with envelope enhancement, in quiet.

CHAPTER 2

Review of Literature

Contents

Auditory dys-synchrony is a form of hearing disorder in which cochlear amplification is spared, but neural conduction in the auditory pathway is affected. Starr, Picton, Sininger, Hood and Berlin first defined it in 1996 and termed it as auditory neuropathy as eight of ten subjects investigated by them had peripheral neuropathies. Later, Berlin, Hood and Rose (2001) suggested the term auditory dys-synchrony as a more accurate indicator of the underlying condition as the exact site of dysfunction and the pathology is not known.

The exact prevalence of auditory dys-synchrony is still not established. The existing limited data suggest that prevalence rate varies from 11% to 0.5% (Kraus, Ozdamar, Stein, & Reed, 1984; Rance et al., 1999; Tang, McPherson, Yuen, Wong, & Lee, 2004). Kumar and Jayaram (2006) estimated the prevalence of auditory dys-synchrony in India by a retrospective study and reported that one in 183 persons with sensory-neural hearing loss has auditory dys-synchrony.

2.1. Audiological Characteristics of Individuals with Auditory Dys-synchrony

2.1.1. Pure-tone Audiometry

Audiological profile of individuals with auditory dys-synchrony is variable. The degree of hearing loss may range anywhere from normal hearing to profound hearing loss. Sininger and Oba (2001) reported that a majority of the subjects (82%) in their study showed bilateral symmetrical hearing loss with only 14 % showing asymmetrical hearing loss and about 4% had unilateral hearing loss.

Kumar and Jayaram (2006) conducted a survey of 61 participants with auditory dys-synchrony and observed that 26 of them had peaked audiogram, 11 participants showed reverse sloping audiogram, 11 had flat loss, and sloping hearing loss was observed in five participants. Audiogram could not be obtained in two children whose age was less than 2 years. Starr, Sininger and Pratt (2000) observed a flat hearing loss in 41%, reverse sloping in 29%, irregular saw-tooth pattern in 9%, 'U' shaped audiogram in 5%, a tent shaped (peaked) audiogram in 5% and high frequency sloping hearing loss in 11% of the individuals with auditory dys-synchrony.

The difference in results observed in different studies may be related to the age of the subjects and severity of the hearing loss. Rance et al. (1999) observed that audiometric configuration varied with the degree of hearing loss. Ears with normal or near normal hearing acuity showed equal sensitivities at all the frequencies. Individuals with mild to severe hearing loss had audiograms that showed poor hearing sensitivities in the low and mid frequencies, but better thresholds in the high frequencies.

The configuration of the audiogram can give an indication about the underlining pathophysiology in individuals with auditory dyssynchronny. According to the theories of pitch perception, frequencies above 4 kHz are coded by place of excitation on the basilar membrane whereas frequencies below 4 kHz are coded by temporal information (Moore, 1995). Reverse sloping audiogram observed in individuals with auditory dys-synchrony suggests the probability of impairment in processing phase locked temporal information (Sininger, & Oba, 2001).

2.1.2. Physiological and Electrophysiological Responses

It has been reported that acoustic reflexes are absent in a majority of the individuals with auditory dys-synchrony (Kumar, & Jayaram; 2006; Starr et al., 1996). Evaluations of 128 children by Berlin et al. (2005) revealed that acoustic reflexes were absent in 113 (88.3%) and elevated in nine (7 %) participants. Though acoustic reflexes were present at normal level (95 dB HL or below) in six (4.7 %) participants, it was present only at 500 Hz and 1000 Hz. Starr et al. (2000) also observed similar results in 50 individuals they examined.

Presence of robust evoked oto-acoustic emissions is a hallmark of the audiological profile of individuals with auditory dys-synchrony. Sininger and Oba (2001) reported that about 80% of the patients with auditory dys-synchrony showed robust Transient oto-acoustic emissions (TEOAE). Only 9% of the participants did not show TEOAEs even in the initial evaluation while the remaining 11 % showed presence of TEOAEs in initial evaluation, which disappeared overtime. Rance et al. (1999) found absence of TEOAEs in one-half of their participants however, cochlear microphonics were present in all the participants they studied. Similarly, Starr et al. (2000) have observed that TEOAE, disappeared over time in 20% of the participants they investigated whereas cochlear microphonics were consistently present over time in all the patients they studied. It has been reported that cochlear microphonics are robust and present for several milliseconds (Berlin et al., 1998; Deltenre et al., 1998; Starr et al., 2000).

Auditory brainstem responses (ABR) are either absent or severely abnormal in persons with auditory dys-synchrony. Starr et al. (2000) reported that 73% of their patients did not show any component of ABR regardless of the stimulus level. Though, 21% showed wave V and 6% showed wave III and V, the peaks were poorly defined with prolonged latency and reduced amplitude in a majority of the individuals. Other investigators have also reported similar results (Berlin et al., 1998; Sininger, & Oba, 2001). Berlin et al. (1998) observed that many patients with auditory dys-synchrony demonstrate long duration cochlear microphonics, which may be confused with early peaks of ABR. Therefore, it is recommended that while recording ABR, responses for the condensation and rarefaction clicks should be compared to rule out the possibility of misinterpreting cochlear microphonics as ABR. Presence of cochlear microphonics with absent/abnormal ABR is regarded as a gold standard for diagnosis of auditory dys-synchrony.

2.2. Patho-physiology of Auditory Dys-synchrony

Based on the electrophysiological and pathological findings, Starr, Picton and Kim (2001) have put forth two mechanisms to account for distorted auditory nerve activity in individuals with auditory dys-synchrony. There is impaired synchrony among the auditory nerve fibers and/or reduced neural input. Desynchronized neural discharge can occur due to demyelination and ion-channel dysfunction in the auditory nerve (Starr et al., 1998) and/or dysfunctional synaptic transmission between the inner hair cells and the auditory nerve (Fuchs, Glowatzki, & Moser, 2003). Loss of the neural input to the brain can occur due to inner hair cell loss (Harrison, 1998) and/or auditory nerve loss (Starr et al., 2003).

Figure Re.1 illustrates sensory nerve action potentials in individual nerve fibers and the resulting compound action potential when there is demyelination, axonal loss and both. Activation of normally functioning inner hair cells and auditory nerve, will normally lead to synchronized excitation of discharge of auditory nerve fibres connected to inner hair cells, leading to clearly averaged compound action potential (Figure Re.1 a) that is distinguished from spontaneous discharge. Alterations of the synchrony of discharge due to demyelination of auditory nerve fibres or impaired synaptic coupling between inner hair cells and auditory nerve dendrite result in changes of both amplitude and duration of the averaged nerve responses (Figure Re.1b). When the numbers of fibres are reduced due to axonal loss, the compound action potential is unchanged in form but markedly reduced in amplitude (Figure. Re.1c). A combination of reduced neural input and dys-synchrony (Figure. Re.1 d) will have profound effects on the formation of a detectable compound action potential.

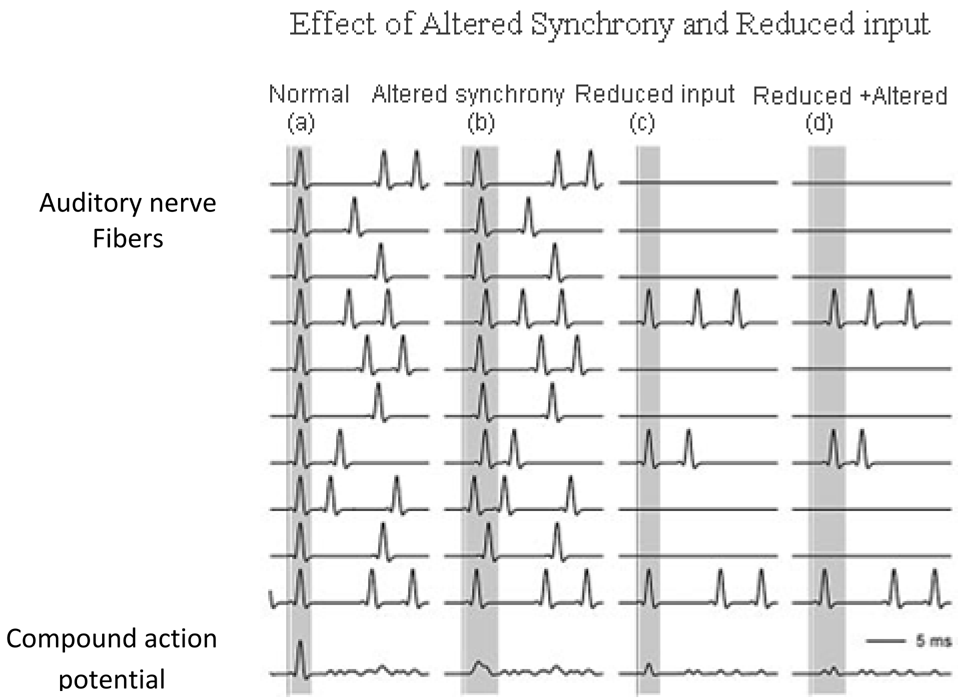


Figure Re.1. Action potential of individual fibers and resultant compound action potential. (From pathology and physiology of auditory neuropathy with a novel mutation in the MPZ gene (Tyr145⟶ Ser) (2003). Reprinted with permission from advanced accesses publications).

2.3. Speech Perception in Individuals with Auditory Dys-synchrony

2.3.1. Speech Perception in Quiet

Many researchers have investigated speech perception in individuals with auditory dys-synchrony. Difficulty in perceiving speech in listeners with auditory dys-synchrony has shown no correlation with the pure-tone sensitivity (Narne, & Vanaja, 2008 a; Sininger, & Oba, 2001; Starr et al., 2000), and has in most cases been significantly poorer than would have been expected for sensorineural losses of equivalent degree. Starr et al. (1996) estimated open-set word recognition scores for eight of the ten subjects with auditory neuropathy. It was observed that the word recognition scores ranged from 0 to 92%, and the speech identification scores of 12 of the 16 ears, were significantly lower than that predicted from the norms generated by Yellin, Jerger and Fifer (1989) for ears with sensorineural hearing loss. Similarly, Sininger and Oba (2001) assessed speech identification scores (CID W-22 lists) of 36 subjects with auditory dys-synchrony and reported that the scores for 25 subjects (69 %) were below the normative range reported by Yellin et al. (1989).

The data presented in these studies demonstrate that in many adults with auditory dys-synchrony, speech signal disruption is more severe than that observed in individuals with sensorineural hearing loss. However, not all of the adults with auditory dys-synchrony have reported to have poor speech understanding in quiet listening conditions. Twenty-five percent of the ears assessed by Starr et al. (1996) and thirty percent of the subjects evaluated by Sininger and Oba (2001) for example, showed speech perception scores within the normal range for sensorineural losses of equivalent degree.

Attempts have been made to analyze the errors made by individuals with auditory dys-synchrony while understanding speech. Starr et al. (1991) assessed word identification for one participant with auditory dys-synchrony and reported that the high frequency consonants (/s/, /sh/, /ch/) were perceived better than other consonants. Ramirez and Mann (2005) investigated the perception of consonants (p, t, k, b, d, g, k, m, n, r, w) in the vowel context /a/ in four participants with auditory dys-synchrony, ten dyslexia and fourteen normal hearing listeners. Glides and nasals were better perceived than stops consonants in individuals with auditory dys-synchrony and dyslexia. Similar error patterns were observed in individuals with auditory dys-synchrony and dyslexia.

A few studies have investigated the effect of auditory dys-synchrony on perception of different cues used for speech perception. Kraus et al. (2000) systematically examined fine-grained speech perception abilities in an adult with auditory dys-synchrony. They observed that the subject had good discrimination for speech sounds along a /ba-wa/ continuum where the speech sounds differ in their manner of articulation. However, the subject displayed very poor discrimination for speech sounds /da-ga/, that vary in place of articulation. Narne and Vanaja (2008 b) examined the perception of consonants /p, b, t, d, k, g, s, l, r, m, n, t̪, d̪, tʃ, ʃ/ in vowel /a/ context, in eight individuals with auditory dys-synchrony. It was noted that a majority of the participants had significant difficulty in perceiving place of articulation when compared to manner and voicing cues. Further, it was also noticed that, participants had more difficulty in perceiving stops and liquids when compared to fricatives, affricates and nasals.

It can be inferred from the results of the studies on speech perception that individuals with auditory dys-synchrony perform similar to normal hearing listeners on speech perception if the rate of change of temporal features is slow (eg. fricatives, affricates), but their speech processing mechanism breaks down while dealing with rapid spectro-temporal changes (Kraus et al., 2000). Speech events critical for the perception of stop consonants involves rapid temporal changes at the onset of the stimulus (for example burst and transition) and hence it is logical to predict that individuals with auditory dys-synchrony will have problems in speech perception.

2.3.2 Speech Perception in Noise

Speech perception difficulties in the presence of background noise are not unique to subjects with auditory dys-synchrony. Individuals with sensorineural loss also have difficulty in understanding speech in presence of competing signals (Moore, 2003). The effects of noise on speech perception are more severe in individuals with auditory dys-synchrony.

Rance et al. (2007) evaluated twelve children with auditory dys-synchrony, twenty children with cochlear hearing loss and twenty-five children with normal hearing on open-set word recognition scores for CNC words at three different S/N ratio conditions (0 dB, 5 dB, 10 dB and quiet). Results showed that children with normal hearing and those with cochlear hearing loss could maintain the performance at 70 % till +5 dB S/N ratio, whereas in individuals with auditory dys-synchrony the performance dropped down to 30 % at +5 dB S/N ratio. They further demonstrated that reduction in score observed depends on the speech identification scores in quiet. The performance dropped down to 40 % at 0 dB S/N ratio for the individuals having identification scores greater 60% in quiet condition, whereas for those with scores less than 60% in quiet, the performance dropped down to 20%.

Zeng and Liu (2006) have investigated the perception of four participants for Bamford-Kowal-Bench (BKB) sentence in presence of speech spectral shaped noise at

15 dB, 10 dB, 5 dB, and 0 dB S/N ratios. In a participant who had identification score of 90% in quiet, scores reduced to 40% at 10 dB S/N ratio and dropped down to 5% at 0 dB S/N ratio.

Kraus et al. (2000) presented findings of 24 years old subject, who was diagnosed as auditory dys-synchrony, with normal hearing thresholds and no concomitant medial history. On CUNY- sentence assessment subject obtained 100 % scores in quiet, but in presence of background noise at +3 dB S/N ratio the identification scores were 10%. Similarly, Shallop (2002) also reported of a auditory dys-synchrony subject with mild to moderate hearing loss whose identification scores for Hearing in Noise Test (HINT) sentences, was 100% in quiet, 25 % at +15 dB S/N ratio and 0 % at +12 dB S/N ratio. The mechanisms underlying these extreme perceptual difficulties in noise are unclear. They are however consistent with the findings of recent psychophysical studies that have shown excessive masking of pure tones in auditory dys-synchrony subjects by simultaneous noise, as well as noise bursts presented before and after the test signal (Kraus et al., 2000; Zeng et al., 2001; Zeng et al., 2005).

In summary, a majority of individuals with auditory dys-synchrony have shown severely disrupted speech perception. The proportion of participants that have trouble in understanding speech and factors that determine the amount of difficulty they experience is not completely determined. Speech understanding is impaired in a quiet condition and in the presence of noise for individuals with auditory dys-synchrony, but the amount of difficulty experienced in the presence of noise is more extreme than that observed in listeners with normal hearing and cochlear hearing loss.

2.4. Importance of Temporal Information in Speech

Speech is composed of spectral information, which changes overtime, sometimes slowly, often quickly (Kewley-Port, 1983). This dynamic property provides information essential for distinguishing among phonemes (van Wieringen & Pols, 1998). Apart from spectral variation, speech is also composed by amplitude changes overtime. Amplitude variations are as important as spectral variations (Drullman, Festen, & Plomp, 1994 a, b; Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995) as it provides information crucial for segmentation of speech syllabic and word level (Van Tasell, Soli, Kirby & Widin, 1987; Price & Simon, 1984; Rosen, 1992).

Plomp (1983) and Rosen (1992) developed a framework for describing the acoustic structure of speech, based on temporal properties. According to their framework, speech is composed of three main temporal features, envelope, fine structure and periodicity. Envelope or temporal envelope refers to the fluctuation in overall amplitude in 2-50 Hz range of speech signal and they convey linguistic information regarding manner of articulation, voicing and vowel identity. Fine structure cue has dominant fluctuations in the frequency range of 600 Hz to 10 kHz and conveys linguistic information regarding place of articulation, voicing, and voice quality. Periodicity, temporal fluctuations found between 50Hz-500Hz, provides linguistic information about voicing and manner.

2.4.1. Temporal Cues and Speech Perception in Quiet

The support for importance of temporal envelope cues in speech signal comes from two groups of studies. First group of studies conducted on listeners with normal hearing and those with hearing impairment has shown that accurate speech identification may be obtained by preserving low frequency amplitude modulations and severally degrading spectral information (Narne, Manjula, & Vanaja, 2007; Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995; Turner, Souza, & Forget, 1995; van Tasell, Soli, Kirby, Widin, 1987). The second group of studies examined speech perception when temporal envelope information was parametrically removed from the speech signal (Drullman et al., 1994 a, b). Results showed that speech recognition was unchanged as long as envelope information below 16 Hz were preserved. Further, it was noticed that removing the temporal envelopes from speech signal affected consonants more than vowels and within the consonants; stops were significantly degraded when compared to fricatives, affricatives, glides and nasals.

Studies have assessed contribution of temporal fine structure information to speech perception by completely removing the temporal envelope from the speech signal. Smith, Delgutte, and Oxenham, (2002) have reported that sentence identification is significantly impaired when only temporal fine structure information is provided. In addition, Zeng et al. (2004) made similar observation on sentence identification in normal hearing listeners. They suggested that fine structure information in speech is neither necessary nor sufficient for understanding speech in quiet (Zeng et al., 2004; Smith, Delgutte, & Oxenham, 2002). In contrast, other investigators have assessed the consonant perception in listeners with normal hearing with only temporal fine structure information and showed that perfect consonants identification could be achieved with only fine structure information (Gilbert, & Lorenzi 2006; Narne, Manjula, & Vanaja, 2007). The perfect consonant identification with only fine structure information may be due to preserved formant transitions.

2.4.2. Temporal Cues and Speech Perception in Noise

A number of investigators studied the cause for reduction in speech intelligibility in background steady state noise (Drullman, 1995 a, b; Houtgast, & Steeneken, 1985). They have demonstrated that one of the major reasons for reduced speech intelligibility in background noise, is in part because noise reduces the modulation depth of temporal envelope. In addition to reducing the modulation depth of temporal envelope, noise introduces spurious modulations, obscuring the relevant speech modulations. Drullman (1995 b) showed that at critical S/N ratio conditions listeners with normal hearing utilizes temporal fine structure cues for understanding speech. Other studies evaluating the importance of envelope and fine structure in speech perception supports the reports of Drullman (1995 b). It has been observed that listeners with normal hearing obtained 100 % scores in quiet when only envelope cues and 8 channels of spectral information (speech shaped noise) was provided, whereas in presence of noise, speech understanding reduced dramatically and was as low as 10 to 20 % when the S/N ratio was 0 dB (Dorman, Loizou, Tu, 1998; Stickney, Nie, Zeng, 2004). They attributed the reduced identification scores in adverse noise conditions to removal of fine structure information.

The temporal envelope information is very critical for understanding speech in quiet and in presence of noise. In addition to the envelope cues, fine structure also plays important role in understanding speech in critical S/N ratio conditions. The poor perception of speech in adverse listening conditions and competing talker background in listeners with cochlear hearing loss is attributed to impaired ability to process temporal fine structure information (Lorenzi, Gilbert, Carn, Garnier, & Moore, 2006; Narne et al., 2007).

2.5. Speech Perception and Psycho-acoustic Abilities

Examination of the perceptual importance of various spectro-temporal properties in speech is difficult as speech signal is dynamic and highly complex by nature and single properties of the speech signal cannot be isolated and manipulated independently (van Wieringer & Pols, 2006). In such conditions, non-speech stimuli such as tonal stimuli would be necessary to assess the importance of specific properties of speech signal. Assessing psychoacoustic abilities in hearing disorders may provide a logical explanation for perceptual difficulties.

A number of investigators have studied the psycho-acoustical abilities that underpin the speech perception in individuals with auditory dys-synchrony and those with normal hearing. Some of the psycho-acoustical abilities that are impaired and cause poor speech perception include temporal resolution and frequency discrimination (Kumar, & Jayaram, 2005; Rance et al., 2004; Zeng et al., 1999; Zeng et al., 2005).

2.5.1. Temporal Resolution and Speech Perception in Auditory Dys-Synchrony

a) Temporal resolution in auditory dys-synchrony

Temporal resolution is reported to be abnormal in a majority of the individuals with auditory dys-synchrony (Kumar & Jayaram, 2005; Rance et al., 2004; Zeng et al., 1999). One of the temporal resolution measures that has been extensively studied in individuals with auditory dys-synchrony is temporal modulation transfer function (TMTF). The TMTF is a measure of sensitivity to amplitude fluctuation over a range of modulation frequencies (Viemeister, 1979). Zeng et al. (1999) measured the TMTF in ten individuals with auditory dys-synchrony using three intervals, three alternative, forced choice methods and two down one up procedure employed to track the threshold with 70.7% accuracy. Results showed that the individuals with auditory dys-synchrony had a high peak sensitivity of - 8.7 dB with a bandwidth of 17 Hz whereas the peak sensitivity was -20.4 dB with a bandwidth of 247 Hz in individuals with normal hearing. Zeng et al. (2005) reported similar results in 16 individuals with auditory dys-synchrony. In contrast to the report of Zeng et al., earlier investigators reported of a bandwidth of only 50 Hz in listeners with normal hearing (Viemeister, 1979; Bacon & Viemeistor, 1985; Eddins, 1993). The difference in the bandwidth reported for listeners with normal hearing in the above studies may attributed to type of procedure employed for estimating threshold and the method with which 3 dB cut-off frequency was derived.

Kumar and Jayaram (2005) measured the TMTF in 10 individuals with auditory dys-synchrony using a discrimination task with two down one up procedure to track threshold in three dB steps. The results showed that average modulation detection threshold in individuals with auditory dys-synchrony was three times higher than listeners with normal hearing (-6.6 dB in auditory dys-synchrony and -17.7 dB in normal hearing subjects). Difference between normal hearing listeners and individuals with auditory dys-synchrony was pronounced for higher modulation frequencies. The threshold obtained by Kumar and Jayaram, in normal hearing listeners and those with auditory neuropathy was lower than that reported by earlier investigators (Viemeister, 1979; Zeng et al., 1999; Zeng, et al., 2005). The lower threshold obtained by Kumar and Jayaram could be because of the larger step size used when compared to that employed by earlier investigators. In addition, the procedure used by Kumar and Jayaram is different from that used by the earlier investigators.

The TMTF has also been studied in children with auditory dys-synchrony. Rance, McKay, & Grayden (2004) evaluated modulation detection threshold for fourteen children with auditory dys-synchrony, ten children with normal hearing and ten children with sensory-neural hearing loss. Modulation frequencies of 10 Hz, 50 Hz and 500 Hz were used and the thresholds were established using three interval three alternative forced choice, two-down one up procedure in three dB steps. Results showed that the modulation threshold was significantly lower at all the modulation frequencies in participants with auditory dys-synchrony (-8.7 dB at 10 Hz) when compared to those with normal hearing (-15.8 dB at 10 Hz) and sensory-neural hearing loss (-15.1 dB at 10 Hz). The discrepancy between the thresholds obtained by Rance et al. and the previous investigators may be due to the age of participants, and the step size used. Rance et al. have investigated only at three-modulation frequency and these modulation frequencies investigated have not been extensively studied by other investigators.

Other temporal resolution measurers that have been evaluated by Zeng et al. (2005) in individuals with auditory dys-synchrony are gap detection, backward masking, forward masking and simultaneous masking. They adopted an adaptive a three-interval three-alternative forced choice, two down one up procedure to track 70.7% correct response criterion. The gap detection thresholds were 2 to 3 ms at high sensation levels for individuals with normal hearing, whereas in those with auditory dys-synchrony, gap detection thresholds were 25 ms to 28 ms.

Temporal masking and simultaneous masking paradigm have shown that individuals with auditory dys-synchrony have difficulty in separating the sounds that occur in close succession and detecting signal in noise (Zeng et al., 2005). Individuals with auditory dys-synchrony showed 60% masking in forward masking paradigm, when the signal and the masker were separated by as much as 100 ms whereas normal hearing listeners showed only 15% masking with a signal delay of less than 20ms. Kraus et al. (2000) also reported exaggerated masking effect in one patient with auditory dys-synchrony who had near normal hearing thresholds.

Results of limited number of studies carried out on subjects with auditory dys-synchrony reveal a significant impairment in temporal resolution of individuals with auditory dys-synchrony, however the severity of impairment is not same in all the individuals. These studies have investigated only in a small number of subjects and not all the studies have investigated all the parameters of temporal modulation transfer function. Further studies are needed to replicate these results on a larger population.

b) Relation between temporal resolution and speech perception

It has been speculated that the difficulty in understanding speech in individuals with auditory dys-synchrony is related with abnormal temporal resolution (Kumar, & Jayaram, 2005; Zeng et al., 2005; Zeng et al., 1999; Rance et al., 2004). Speech perception difficulties experienced by participants with auditory dys-synchrony resemble those experienced by subjects with learning disability and elderly. Research has indicated a good correlation between speech perception abilities and temporal resolution in elderly listeners (Gordon-Salant, & Fitzgibbons, 1993) and individuals with learning disability (Tallal, & Stark, 1981). Hence, attempts have been made to correlate modulation detection threshold and speech perception in individuals with auditory dys-synchrony and the results of these investigations have indicated that the severity of abnormality is significantly correlated with speech understanding difficulties (Kumar, & Jayaram, 2005; Rance et al., 2004; Zeng et al., 1999). Kumar and Jayaram (2005) grouped ten participants with auditory dys-synchrony into two groups based on speech identifications and compared the peak modulation threshold for the two groups. It was noted that individuals with speech identification score of greater than 50% had lower threshold (-10.4 dB) and those with speech identification score of less than 50% had higher modulation detection threshold (-5.6 dB). Similarly, Rance et al. (2004) grouped children with auditory dys-synchrony into two groups based on speech identification scores and observed that individuals with speech identification scores poorer than 30 % had a peak sensitivity of -3.4 dB whereas the peak sensitivity was 14.3 dB in those with identification scores greater than 30%. These results suggest that speech perception is highly dependent on ability to detect amplitude variations in continuous signal.

Zeng et al. (2001) assessed the relation between speech identification scores and peak sensitivity in TMTF in ten participants. They obtained the following regression equation for predicting identification scores from the TMTF peak sensitivity, bandwidth and pure-tone thresholds:

$$\mathrm{SIS}\,\% = 49.5 - (0.99 \times \mathrm{PTA}) - \big(2.02 \times \mathrm{TMTF}\,(\mathrm{dB})\big) + (0.04\ \times \mathrm{TMTF}\,(\mathrm{Hz}))$$

This equation revealed a high correlation of 0.93 %. It can be noted from the equation that, prediction of identification is mainly dependent upon the TMTF threshold, but one can note that the pure-tone threshold do contribute minimally for identification scores. This suggests that once the speech signal is audible, impaired understanding of speech is accounted by impaired ability to follow amplitude variation in speech signal.

Furthermore, Zeng et al. (2001) edited the speech sample according to TMTF thresholds observed in individuals with auditory dys-synchrony and presented it to listeners with normal hearing. The errors made by normal hearing listeners for edited speech tokens were similar to those observed in individuals with auditory dys -synchrony for normal speech suggesting that processing of temporal envelope of the speech signal is impaired in these individuals. In addition, consonantal errors observed in listeners with auditory dys-synchrony (Narne &Vanaja, 2008; Kraus et al., 2000) were similar to that observed in normal hearing listeners when the temporal envelope of the speech signal was temporally smeared (Drullaman et al. (1994 a). This suggests that speech perception difficulties seen in individuals with auditory dys-synchrony were partly due to the inability in perceiving the amplitude variations in the speech signal.

Thus, a review of literature indicates that the disorder affects individual's ability to cope with the dynamic nature of speech signal. That is individual's ability to follow amplitude variations in speech signal is impaired, leading to smearing of signal in the temporal domain. The amount of impact on speech perception depends on degree of impairment in temporal processing.

2.5.2. Frequency Discrimination and Speech Perception in Auditory Dys-synchrony

a) *Frequency discrimination in auditory dys-synchrony*

Frequency discrimination is the ability to perceive changes in frequency over time. For steady state stimuli of 4 kHz and above, frequency discrimination depends primarily on place mechanisms based on spatial changes in the basilar membrane excitation pattern (Moore, 1973; Sek & Moore, 1995). In contrast, discrimination of frequency less than 4 kHz depends on the use of temporal information through neural phase locking (Moore, 1973; Sek & Moore, 1995).

Frequency discrimination ability in individuals with auditory dys-synchrony is yet to be investigated thoroughly, but the data that has been presented so far suggests extreme perceptual deficits. Starr et al. (1991) measured "just noticeable differences" (JNDs) for pairs of 500 ms tone-burst stimuli at octave frequencies from 250 Hz to 8 kHz in a 11 years old subject with auditory dys-synchrony. Frequency discrimination results in this case were consistently depressed, showing JNDs of approximately 4.5 times higher than those obtained from a group of five age-matched children across the frequency range tested.

Zeng et al. (2005) found impaired frequency discrimination ability in 12 individuals with auditory dys-synchrony. Frequency difference limens were obtained at octave frequencies (250 Hz – 8 kHz) using a three alternative forced-choice adaptive procedure. It was observed that the discrimination task was considerably poorer than normal hearing, particularly in the low-to-mid frequency range (≤ 2 kHz), but discrimination in the high frequency range appeared to be less impaired, approaching near normal range at 8 kHz test frequency. It can be inferred from this that the low frequency discrimination is partly due to the disrupted temporal resolution in listeners with auditory dys-synchrony (Moore, 1995; Zeng et al., 1999; Zeng et al., 2001; Zeng et al., 2005).

Rance et al. (2004) reported similar results for children with auditory dys-synchrony. It was observed that the mean difference limen for 4 kHz pure tones were 4.5 times the normal value, whereas the discrimination at 500 Hz was on an average 11 times poorer than that of the normally hearing cohort. They further measured frequency discrimination with frequency-modulated stimuli (modulation frequency 10 Hz) and observed that difference limen obtained for frequency modulated stimuli is much poor than that observed for pure-tone stimuli in participants with auditory dys-synchrony. It was inferred from these results that children with auditory dys-synchrony cannot utilise phase locking cues to the same extent as children with normally hearing or those with sensorineural hearing loss children (Rance et al., 2004).

b) Relation between frequency discrimination and speech perception

Speech identification scores have shown good correlation with frequency discrimination ability in subjects with auditory dys-synchrony. For example, in the children assessed by Rance et al. (2004) there was a strong correlation between open set-speech identification score and difference limen for frequency at all the test frequency assessed. The children with poor frequency discrimination ability typically presented with greater impairment in speech identification.

The inability to discriminate signal at low frequencies will impair the ability to perceive frequency difference in signal at low frequency. Further, greater degree of hearing loss at low frequency will cause an imperception of important speech cues in low frequencies. This impairs their ability to utilize low frequency information for speech perception. In addition, poor processing of frequency-modulated signal indicates that participants with auditory dys-synchrony are impaired in following the change in frequency of over time. This impaired ability to follow or discriminate frequency change will impairs perception of fast spectral-temporal changes in speech signal and reduces phoneme perception ability.

2.6. Management of Individuals with Auditory Dys-synchrony

Management of individuals with auditory dys-synchrony is a challenging task for clinical audiologists because of the uncertainty in etiology and pathophysiology. Generally, hearing aids are not of much benefit for individuals with auditory dys-synchrony. A number of researchers have demonstrated that, in adult and children with auditory dys-synchrony, acceptance of amplification is universally poor with reports ranging from little or no benefit to detrimental effects (Berlin, Hood, Hurely, & Wen, 1996; Berlin, Hood, Morlet, Rose, & Brashears, 2003; Starr et al., 1996). In contrast, other investigators have demonstrated that a small group of participants do benefit with amplification but the amount of benefit is not comparable with those observed in individuals with cochlear hearing loss (Cone-Wesson, Rance, & Sininger, 2001; Rance, Cone-Wesson, Wunderlich, & Dowell, 2002; Rance & Barker, 2008). Zeng et al. (1999) pointed that linear amplification used in conventional hearing aids do not change amplitude fluctuations of speech and the use of a nonlinear amplitude compression circuit reduces the fluctuations (Van Tasell, 1993) causing detrimental effects on speech perception. They further opined that, current hearing aid technology improves only speech awareness.

Cochlear implants have been found to be useful in some individuals with auditory dys-synchrony (Mason, De Michele, Stevens, Ruth, & Hashisaki, 2003; Paterson et al., 2003; Shallop, Peterson, Facer, Fabry, & Driscoll, 2001; Walteon, Gibson, Sanli, & Prelog, 2008). It has been reported that performance with cochlear implants was comparable in subjects with auditory dys-synchrony and those with sensory-neural hearing loss (Mason, et al., 2003; Paterson et al., 2003; Shallop et al., 2001). However, recent investigators have demonstrated that not all subjects with auditory dys-synchrony show benefit comparable to those with sensory-neural hearing loss (Rance, & Barker, 2008; Runge-Samuelson, Drake, & Wackym, 2008; Walteon, Gibson, Sanli, & Prelog, 2008).

Although cochlear implants improve speech perception in some individuals with auditory dys-synchrony, Rance and Barker, (2008) showed that the amount of benefit derived with cochlear implants was similar to that obtained with amplification. In addition, the degree of hearing loss may not justify the use of cochlear implant as an option of management in some of these individuals with auditory dys-synchrony. Though hearing aids may not show the expected benefits, as they do not address issues relating to processing of temporal information in speech, they still seem to be a better option of management because of their lower cost. Therefore, there is a need to develop algorithms in hearing aids, which enhances the temporal information in speech.

2.6.1. Alternative Approaches to Improve Speech Perception

a) Clear speech

Sentences spoken 'clearly' in presence of background noise are more intelligible than those spoken 'conversationally' for listeners with normal hearing, hearing-impairment, and learning disability (Liu, & Zeng, 2006; Picheny, Durlach, & Braida, 1985, 1986; Uchansky, Choi, Braida, Reed, & Durlach, 1996). Acoustical properties of clear speech include increase in spectral energy at mid as well as high frequencies, increased modulation depth (enhanced envelope cues) and increased consonant vowel ratio (Krause, & Braida, 2004; Liu, & Zeng, 2006; Picheny, et al., 1985, 1986). Recently, Liu and Zeng (2006) evaluated the properties of clear speech that improves speech perception in background noise for listeners with normal hearing. They reported that enhanced envelope cues in the clear speech improved speech perception at higher

S/N ratios (> 5 dB), but at lower S/N ratios (0 dB & -10 dB), it is the enhanced fine structure cues in clear speech that contributed for improved speech perception.

Zeng and Liu (2006) studied the advantage of clear speech over conversational speech in 13 individuals with auditory dys-synchrony. Out of 13 participants, seven participants were fitted with cochlear implants. Perception of sentences were assessed in quiet and in presence of speech spectrum shaped noise. Results demonstrated a significant benefit with clear speech over conversational speech in quiet and in presence of noise. Clear speech improved speech perception in all the listening conditions:
i) acoustical hearing ii) electrical hearing (cochlear implants) iii) acoustical and electrical hearing. They attributed the advantage observed for clear speech to enhanced temporal envelope cues.

Evidence for better encoding of clear speech than conversational speech has also come from animal studies. Cunningham, Nicol, King, Zecker and Kraus (2002) recorded aggregate neural responses from auditory midbrain, thalamus and auditory cortex of anesthetized guinea pigs to a syllable /ada/ in quiet and in noise. Results demonstrated that when the consonantal portion of speech signal was enhanced (amplifying burst), it resulted in better representation of consonants in auditory midbrain, thalamus and auditory cortex. These results suggest that enhancing envelope cues artificially in the speech signal would improve speech perception in quiet and in presence of noise.

b) Temporal Modification

i) Time scale modification of speech

Formant transitions, voice onset time (VOT), closer duration and burst are short acoustic vents that have been shown to be of importance in differentiating stop consonants (Pickett, 1999). These short acoustic events occur in the beginning of stop consonants and are brief. It has been shown that children with dyslexia, learning problems, specific language impairment and elderly listeners have difficulty in perceiving stop consonants due to poor processing of dynamic cues (Tallal & Piercy, 1975; Tallal et al., 1996). One of the techniques that has been advocated to overcome poor temporal processing abilities such as those observed in these clinical populations is time scale modification of the speech.

Tallal et al. (1996) and Nagarajan et al. (1998) modified the temporal cues in speech signal using time scale modification (doubling the duration) without altering spectral content and differential enhancement of fast modulations in the range of 3Hz to 30 Hz. The children with language impairment were trained for 4-weeks with modified speech stimuli and there was a significant improvement in speech discrimination and language proficiency following training.

It has been shown that individuals with auditory dys-synchrony have difficulty in processing short duration sounds (Zeng et al., 2005). Lengthening the short acoustic events may improve speech perception in individuals with auditory dys-synchrony. The evidence to this comes from study conducted by Kumar (2006) in which stop consonant identification were assessed with time scale modification (lengthening) of the VOT, burst and transition duration. He observed that the perception of place of articulation and voicing improved with lengthening the formant transition and lengthening of VOT showed a greater improvement for voicing perception than place of articulation. Combined modification of transition duration, burst duration and VOT did not bring about significant improvement in speech identification scores compared to lengthening of transition duration alone. These results suggest that lengthening VOT and transition duration significantly improved consonant perception. However, one should be cautious in using this approach as lengthening the duration may bring an unnatural speech quality and interferes in daily communication due to delay between receiving and producing speech.

ii) Envelope enhancement

A considerable range of speech enhancement techniques have been developed to address the problem of understanding speech in adverse listening conditions. These enhancement techniques can broadly be categorised into spectral subtraction and envelope enhancement /expansion (Clarkson, & Bahgat, 1991). The studies examining the importance of envelope enhancement techniques enhance the low frequency amplitude modulations (envelope) of the speech signal, which are the actual carriers of speech information. Enhancing the envelope of speech signal produces a spectrum in which greater than normal portion of spectrum energy is more narrowly focused around important features of speech signal (Rance 2005).

The envelope enhancement strategies were originally implemented to improve speech perception in noise for individual with cochlear hearing loss and normal hearing listeners (Apoux, Courzet, & Lorenzi, 2001; Apoux, Tribut, Debruille, & Lorenzi, 2004; Clarckson & Bahgat, 1991; Lorenzi, Berthommier, Apoux, & Bacri, 1999). Freyman & Narbonne (1996) evaluated the consonant perception in 50 normal hearing listeners in quiet and at different S/N ratios for unprocessed and envelope-enhanced signal. The envelope enhancement was done by "simple power law function" (squaring envelope). Results showed that envelope enhancement deteriorated the performance in quiet and in presence of noise. They further demonstrated that envelope expansion by simple power law function reduces consonant vowel ratio, and affect consonant perception.

Lorenzi et al. (1999) conducted experiments to study of effect of degrading, preserving and expanding the temporal envelope of the speech signal under conditions of greatly reduced spectral information. Spectral information was degraded to allow direct examination of modification of envelope on speech recognition. They assessed consonant recognition in vowel-consonant-vowel /aCa/ context. Degrading of envelope was performed by adding white noise at 0 dB S/N ratio. Envelope expansion was performed by rising the temporal envelope of the speech stimuli to the power 2. They showed that the expansion improved consonant recognition against background noise but introduced deleterious effects on consonant recognition in quiet. Apoux et al. (2001) reported similar results for consonant recognition in listeners with normal hearing and cochlear hearing loss. The discrepancy in the results obtained by Lorenzi et al. and Freyman, Nerbonne could be due to type of stimuli used. Freyman and Nerbonne used speech stimuli without spectral degradation whereas Lorenzi et al. and Apoux et al. used spectrally degraded stimuli.

Apoux et al. (2004) examined sentence identification in 24 elderly participants with cochlear hearing loss and 8 listeners with normal hearing. Sentence identification in stationary and fluctuating background noise was examined for speech processed to enhance envelope by simple power law function. Results showed that speech identification deteriorated for envelope-expanded sentences. They attributed the poor performance to the simple power law expansion scheme, which reduces the consonant-vowel ratio when overall RMS levels of expanded and control stimuli are equated. They opined that when a simple “power law expansion” is applied high input amplitudes envelope (corresponding mainly to the vocalic components of the word/sentence) are more amplified than low amplitude envelopes (corresponding mainly to the consonant components of word/sentence) leading to reduced consonant vowel ratio.

To overcome the disadvantage of simple power function, Apoux et al. (2004) have designed an envelope enhancement procedure called “compressed /expanded scheme”. This program was designed to enhance consonants more than vowels by expanding and compressing the low and high-level parts of temporal envelope respectively. Sentence identification was evaluated in quiet and in presence of background noise for listeners with normal hearing and those with hearing impairment. Results showed a significant improvement with this scheme over “power law expansion function” in both normal hearing and those with hearing impairment.

Narne and Vanaja (2008 a) investigated the ability to identify consonants /p, b, t, d, k, g, s, l, r, m, n, t̪, d̪, tʃ, ʃ/ in vowel /a/ context by individuals with auditory dys-synchrony when the envelope of the speech signal was enhanced by15 dB for different modulation bandwidths (3 to 10 Hz; 3 to 20 Hz; 3 to 30 Hz; 3 to 60 Hz). The results revealed that consonant identification improved in six individuals and only two individuals did not show any improvement. The amount of improvement was greater for 3-30 Hz bandwidth condition when compared to others. The detailed evaluation of voicing, manner, and place of articulation cues, revealed that manner cues were better perceived than voicing and place cues. Within the manner cues, fricatives and affricates were better perceived than stops and liquids. With envelope enhancement, perception of stops and affricates showed greater improvement than the other consonants. Further, it was noted that voicing perception did not improve with envelope enhancement.

Psycho-acoustical data has demonstrated that auditory dys-synchrony impairs perception of the amplitude variations in continuous signal and causing poor speech perception difficulties. (Kumar & Jayaram, 2005; Zeng et al., 1999; Zeng et al., 2005). Poor processing of temporal envelope (amplitude fluctuations) in speech signal impairs processing of syllable and phrase boundaries and further causes smearing of consonantal energy to the following vowel (Drullman et al., 1994 a). Probably signal-processing strategy that enhances the envelope in such a way that it enhances the consonant vowel ratio will improve speech perception. There is a dearth for studies investigating usefulness of envelope enhancement with meaningful stimuli and in presence of background noise.

c) *Spectral modification*

A vast majority of individuals with sensory-neural hearing loss have a greater loss of hearing in high frequencies than in the low frequencies. For this reason, the speech information in higher frequencies was shifted to the low frequencies to make high frequency speech information accessible. Although it is logical that it should improve speech perception, initial clinical reports did not show the expected results (Brida et a., 1979; Reed, Hicks, Brida, & Drulach, 1983; Turner & Hurtig, 1999). This could be because initially frequency shifting was implemented with either frequency transposition or vocoder techniques. In these techniques, a band of high frequencies information was shifted by a fixed number of Hertz's to lower frequencies using amplitude modulation technique or nonlinear distortions. This shifted band information was mixed with original low frequency signal. This form of frequency shifting will degrade the speech signal affecting speech perception (Brida et al., 1979).

Recent investigators have studied the speech perception with digital frequency shifting strategies. Reed et al. (1983; 1985) have studied speech perception with frequency shifting scheme, which did not modify the fundamental frequency of speech, but lowered the formant frequencies of the speech by a fixed factor. They observed that this type of frequency lowering scheme typically produced unnatural speech leading to poor speech perception. The reason for poor speech recognition may be because of alteration in the ratio of fundamental frequency and corresponding formants, which is an invariant cue for vowel and speech perception (Neary, 1989).

Recently, Turner and Hurtig (1999) investigated the usefulness of proportional frequency compression on speech recognition in listeners with hearing impairment. The proportional frequency compression technique compresses the frequency range to a specified frequency range and this technique keeps the ratio between formant frequencies and fundamental frequency constant. Results showed a significant improvement in speech perception for listeners with steeply sloping hearing loss. A number of other investigators also demonstrated a significant benefit with proportional frequency compression amplification in participants steeply sloping hearing loss participants (McDermott, Dorkos, Dean, & Ching 1999; Parent, Chmiel, and Jerger 1997).

Nagafuchi (1976) studied the effect of compressing frequency range and expanding frequency range in normal hearing listeners. They have shown that speech identification deteriorates with either compressing (shifting to low frequency) or expanding (shifting to higher frequency) frequency range of the speech signal. When frequency (spectral) information in speech was expanded by 150% (an octave shift) speech perception reduces by 35%. Scott and Assmann (2001) investigated the vowel and sentence perception with original and synthesized stimuli, where the formant frequencies and fundamental frequency was systematically varied either upward or downward. Upward shifts in the formant frequencies affect the performance of sentence and vowels similarly. Upward shift of formant frequencies reduces the performance by about 25% when formants are raised by a factor of 2 (equal to a shift of 500 Hz). Effect of upward spectral shift was less when the when upward shifts in the formant frequency are combined with increases in fundamental frequency. Thus a review of literature suggest that proportional frequency shift either upward or downward affect speech less significantly than shift in only the either fundamental frequency or formant frequencies.

Similarly, a number of other investigators studied the effect of shifting spectral information upward in a spectrally degraded speech where fine structure is replaced with noise. They have also demonstrated that spectral shifting can severely limit the speech recognition ability (Fu, & Shannon, 1999; Rosen et al., 1999). Investigators have also demonstrated that normal hearing listeners are able to adapt quickly to frequency-shifted speech and the time required for adaption to spectrally shifted speech depends on the degree of spectral shift (Fu, Shannon, & Galvin, 2002).

A majority of adult listeners with auditory dys-synchrony have low frequency hearing loss. In addition, frequency discrimination ability is also poor at low frequencies. The low frequency hearing affects ability to detect sounds in low frequencies, and even when they are able to detect, they may not be able to distinguish between different low frequency sounds due to poor frequency discrimination. Further, the low frequency sounds may increase upward spread of masking. Zeng et al. (2005) suggest that eliminating the low frequencies or shifting low frequency information to higher frequency may improve speech perception. There are no published reports on investigation of the effect of upward spectral shift on speech perception in individuals with auditory dys-synchrony.

CHAPTER 3

Method

Contents

The study was conducted in two phases to investigate the following objectives: (a) temporal modulation transfer function in participants with auditory dys-synchrony (b) identification of speech with and without envelope enhancement in individuals with auditory dys-synchrony and their relation with modulation detection threshold (c) identification of speech with upward spectral shift and identification of speech with both upward spectral shift and envelope enhancement in individuals normal hearing and those with auditory dys-synchrony.

In the first phase, a psycho-acoustical task was administered for estimating the temporal modulation transfer function (TMTF). Second phase, included three experiments, in the first experiment, speech identification ability was assessed for speech with and without envelope enhancement in quiet. Second experiment assessed the effect of speech spectrum shaped noise on identification of speech with and without envelope enhancement. In the third experiment identification of speech with upward spectral shift with and without envelope enhancement was investigated.

3.1. Participants

Two groups of individuals, Group I and Group II participated in the study. Group I included individuals with auditory dys-synchrony, and Group II included individuals with normal hearing.

3.1.1. Group I

40 individuals (16 males and 24 females) diagnosed as having auditory dys-synchrony based on the audiological evaluations were included in this group. Out of the 40 participants, speech identification in noise could be tested only in 25 individuals with auditory dys-synchrony. All the participants could not be tested due to time constraints and only those participants who were willing to stay for longer time or willing to attend another session for testing were included for this experiment. The age of the participants ranged from 12 to 39 years with a mean of 19 years. This age range was selected as it has been reported that psycho-acoustical abilities reach a plateau in normal subjects by 12 years of age (Lynne, Werner, & Gray, 1998). None of the participants reported of any history of external ear or middle ear problems. There was no history of usage of ototoxic drug and none of the participants had undergone any formal auditory training. The primary language of all the participants was Kannada, a Dravidian language spoken in a southern state of India.

All the participants had bilateral auditory dys-synchrony with symmetrical hearing loss. Figure M.1 shows the mean pure-tone threshold and standard deviation (error bars), for right ear and left ear at different octave frequencies. The hearing loss in terms of pure-tone average ranged from 11 dB HL to 53 dB HL with a mean of 36.4 dB HL for the right ear and 35.9 dB HL for the left ear.

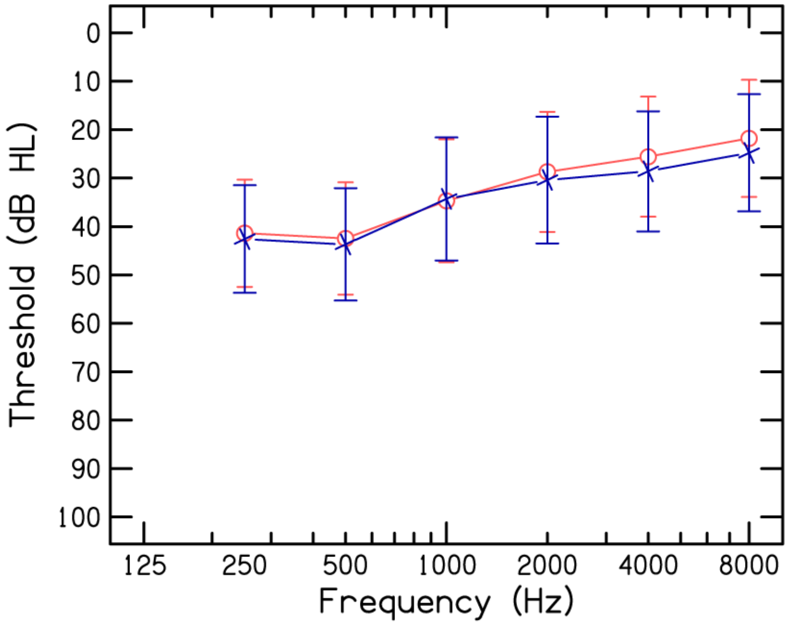


Figure M.1. Mean pure-tone threshold and standard deviation (error bars) for the right (red line) and left ear (blue line) of Group I.

The audiograms were classified into different configurations based on the classification described by Hall and Mueller (1992). The configuration of the audiogram was rising (reverse slope) in 33.3 % of the participants, peaked (peak at 2000 Hz) in 30 %, flat in 26.6 %, irregular saw tooth pattern in 6.6 % and 'U' shaped in 3.3 %. All the individuals had robust transient evoked otoacoustic emissions with absent middle ear acoustic reflexes (both ipsilateral and contralateral) and auditory brainstem responses. All the participants had undergone hearing aid evaluation, as a part of regular clinical evaluation, with linear gain behind-the-ear hearing aid adjusted to comply with National Acoustic Laboratory prescription (Byrne & Dillon, 1986) for gain and frequency response. None of the participants showed an improvement in speech recognition scores with a conventional linear gain hearing aid.

3.1.2. Group II

30 participants (12 males and 18 female) with normal hearing, in the age range of 12 to 40 years with a mean age of 20 years constituted the normal group. The primary language of all the participants was Kannada. It was ascertained from a structured interview that none of the normal hearing subjects selected for the study had difficulty in understanding speech in daily listening conditions, and that they did not have any history of neurologic or otologic disorder. Participants in the normal group had pure-tone thresholds <15 dB HL (ISO, 389) at octave frequencies between 250 Hz to 8 kHz and speech identification score of greater than 90 % at 40 dB SL (ref: average hearing thresholds at 500 Hz, 1000 Hz and 2000 Hz). Immittance evaluation and recording of auditory brainstem responses and transient evoked otoacoustic emissions revealed normal findings in all the participants.

3.2. Instrumentation

The following instruments were used for the study:

1. Orbiter OB-922 (Madsen Electronics, Denmark), two channel diagnostic audiometer calibrated as per ISO 389 with supra aural head phones (Telephonics TDH-39), bone vibrator (Radio ear B-71) and speakers (C 115 Martin Audio) to assess the pure-tone threshold and to administer the experiments in Phase I as well as Phase II.
2. GSI-Tympstar (Grason-Stadler Inc., USA), middle ear analyzer calibrated as per ANSI, 1987 to administer tympanometry and reflexometry.

3. ILO 292 (Otodynamics Inc., UK), an oto-acoustic emission analyzer to record transient click evoked oto-acoustic emissions.
4. Smart EP Version 2.12 C (Intelligent Hearing systems, USA) with AgCl electrodes and ER-3A insert earphones for recording ABR.
5. A computer with MATLAB-7 (Language of Technical computing, USA) software for generating stimuli for Phase I of the study and processing speech stimuli for temporal and spectral modification.
6. A computer with sound card (Sigma Tel High definition audio) and Adobe audition software (Version 2), for playing the stimuli.

3.3. Test Environment

All the tests were conducted in an air conditioned, double room suite where the ambient noise levels were within permissible limits (ANSI, 1999).

3.4. Procedure for Selection of Participants

3.4.1. Pure-tone Audiometry and Speech Audiometry

Air conduction and bone conduction pure-tone thresholds were determined using modified version of Hughson and Westlake procedure (Charhart, & Jerger, 1959). Paired words in Kannada were used to check the speech recognition threshold. Monosyllables (Mayadevi, 1974) were used to assess open set speech identification abilities.

3.4.2. Tympanometry and Reflexometry

Tympanometry and reflexometry was done using a calibrated immittance meter. Tympanograms were obtained for a 226 Hz probe tone. Ipsilateral and contralateral acoustic reflex threshold was measured at 500 Hz, 1000 Hz, 2000 Hz and 4000 Hz frequencies.

3.4.3. Auditory Brainstem Responses

ABR was recorded for 100 µsec clicks presented through Etymotic research ER-3A insert earphones at a maximum level of 90 dB nHL and a repetition rate of 11.1/sec . The recorded EEG was passed through a band pass filter with cut-off frequencies of 100 Hz to 3000 Hz and amplified one lakh times. The time window for averaging and recording was set to 12 msec and 2000 samples were averaged at each intensity. Responses were recorded twice to ensure reproducibility of the waveforms.

3.4.4. Oto-acoustic Emissions

Transient evoked otoacoustic emissions were recorded using nonlinear clicks presented at 80 dB ± 5 dB peSPL. Emissions in response to 260 stimuli were averaged over a time window of 20 msec. The presence of normal OAE in post stimulus period of 2.5 msec to 20 msec was determined by overall amplitude of at least 5 dB SPL with a wave reproducibility of 75%.

3.4.5. Otological Evaluation and Neurological Evaluation

All the subjects, before being selected for the study underwent an ototlogical examination which was done to rule out any external or middle ear problem. Similarly, all the participants were subjected to a neurological examination by a qualified neurologist for ruling out any space-occupying lesion.

3.5. Data Collection

3.5.1. Phase-I: Temporal Modulation Transfer Function

a) Test Stimuli

Two stimuli, un-modulated white noise and sinusoidal amplitude modulated white noise, of 500 ms duration with a ramp of 20 ms were used for this experiment. Figure-M.2 shows a representation of the stimuli used in the study. The stimuli were generated using a 16-bit digital to analog converter with a sampling frequency of 44.1 kHz and low pass filtered with a cut off frequency of 20 kHz. The modulated signal was derived by multiplying the white noise by a dc-shifted sine wave. The depth of the modulation was controlled by varying the amplitude of the modulating sine wave. The stimuli were developed based on the procedure described by Lorenzi et al. (2000). Equation (1) shows the expression describing the sinusoidal amplitude modulated stimuli.

$$m(t) = c \times [1 + m(sin2\pi\, fmt)] \times \mathrm{n(t)} \qquad (1)$$

In the Equation (1), m is the modulation depth (0<m<1), *fm* is the modulation frequency (2, 4, 8, 16, 32, 64, 128, 256, 512), and n (t) is the white noise. The term c is a multiplicative compensation term (Viemester, 1979) set such that the overall power is same in modulated and un-modulated stimuli. The expression used for c is given in Equation (2).

$$c = \left[1 + \frac{\mathrm{m}^2}{2}\right]^{-0.5} \qquad (2)$$

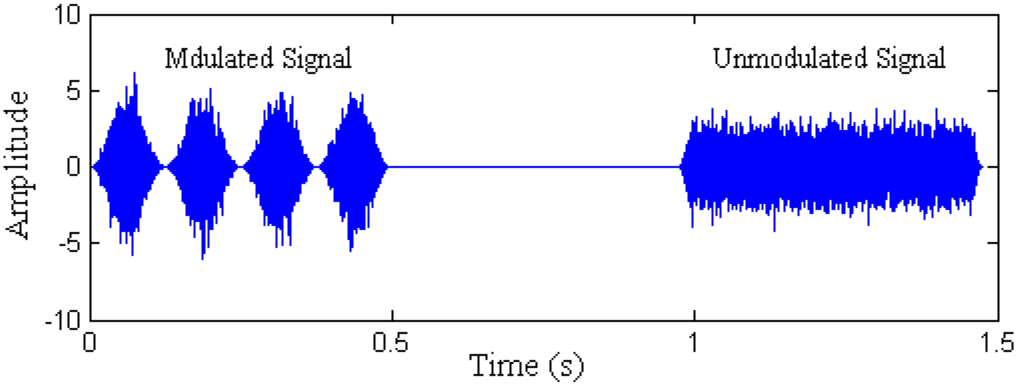


Figure M.2. Waveform of sinusoidal amplitude modulated and un-modulated white noise.

b) *Procedure*

The stimuli were presented at 40 dB SL (Re: Speech recognition threshold). The stimuli were played in a computer and routed through an audiometer (OB-922). The participants received the signal from a loudspeaker of the audiometer, kept at a distance of one meter at 0° azimuth. The participant's task was to discriminate between amplitude-modulated noise and un-modulated noise.

On each trail, un-modulated and modulated stimuli were successively presented with an inter stimulus interval of 500 ms. Modulation thresholds were obtained using adaptive two down one up procedure. Modulation depth was converted into decibels

($20\times\log_{10}$ (m), where m refers depth of modulation) and the step size and threshold were represented in decibels. The step size was initially 4 dB and reduced to 2 dB after two reversals. This procedure provides an estimate of the value of amplitude modulation necessary for 70.7% estimate correct responses (Levitt, 1971). The mean of eight reversals in a block of 14 were taken as threshold. The worst threshold that could be measured was 0 dB, and corresponded to a modulation depth of one (100% modulated noise).

The average power of a modulated noise was higher, by $m^2/2$, than un-modulated noise. This difference in the average power or energy could have introduced potential artifact for noise modulated at higher frequencies, where m is larger. To remove this potential artifact, the stimuli presented in discrimination task were adjusted to have same expected power. In addition, the level of presentation was randomized over a range of 10 dB for each modulation frequency.

3.5.2. Phase-2: Assessment of Speech Identification

c) Test Stimuli

The speech stimuli used in the present study were taken from bi-syllabic wordlists in Kannada developed by Yathiraj and Vijayalakshmi (2005). This test contains four different word lists of equal difficulty, each containing 25 bi-syllabic words, which are phonemically balanced. The words, spoken in a conversational style by a female native speaker of Kannada, were digitally recorded in an acoustically treated room on a data acquisition system with 16-bit analog to digital converter at a sampling frequency of 44.1 kHz.

i) Envelope enhancement

The envelope enhancement technique was designed to increase the amplitude of modulation depth in different frequency bands of speech signal. The amplitude modulations were enhanced in such a way that, it enhances low amplitude envelopes and compresses the high amplitude envelopes. The enhancement procedure for the present study was adopted from the procedure used by Apoux et al. (2004) and the signal was compressed and expanded using MATLAB-7.

The signal X (t) was divided into 4 bands using band pass filters (3rd order Butterworth filter) with cut-off frequencies of 150-550 Hz; 550-1550 Hz; 1550-3550 Hz; and 3550-8000 Hz. The temporal envelopes E (t) was extracted from each band by full-wave rectification and low pass filtering (3rd order Butterworth filter) with a cut-off frequency of 32 Hz. The extracted envelope was either left intact or raised to the power K, with K ranging from a highly expansive value ($K_{max} = 4$) to a highly compressive value ($K_{min} = 0.3$) as a function of the instantaneous envelope amplitude value (E_i). More precisely, exponent k was computed via a decreasing exponential function of the instantaneous envelope amplitude value (E_i), and was set such that: (i) maximum expansion ($K_{max} =4$) was applied to the lowest envelope amplitude value (E_{min}) and (ii) maximum compression ($K_{min} =0.3$) was applied to the highest envelope amplitude value. The expression used for K is given in Equation (3).

$$k_i = e^{-\frac{(E_i - E_{min})}{\tau}}(\mathrm{K}_{\max} - \mathrm{K}_{\min}) + \mathrm{K}_{\min} \qquad (3)$$

In this equation, τ was a constant (0.5 for each word) and E_{min} , the minimum envelope amplitude value, was computed over the whole signal duration within the band. A correction factor was then obtained for all the bands by computing the ratio of the expanded and the original envelope for each sample. The correction factor obtained was then multiplied with the original band pass signal at each corresponding point in time, and finally the resulting bands were added at the output and low pass filtered (3rd order Butterworth filter) with a cut-off frequency of 8000 Hz. The RMS amplitude of the expanded signals was then equated to that of the unprocessed signals. All the eight lists were processed using this scheme. Figure M.3 shows an example of the waveform and envelope of a signal with and without envelope enhancement. The envelope for unprocessed (blue line) and envelope-enhanced signal (red line) is represented from four bands. It can be observed from the figure that low amplitude envelope of the signal (consonantal portion) is more enhanced than high amplitude envelope (vowel portion).

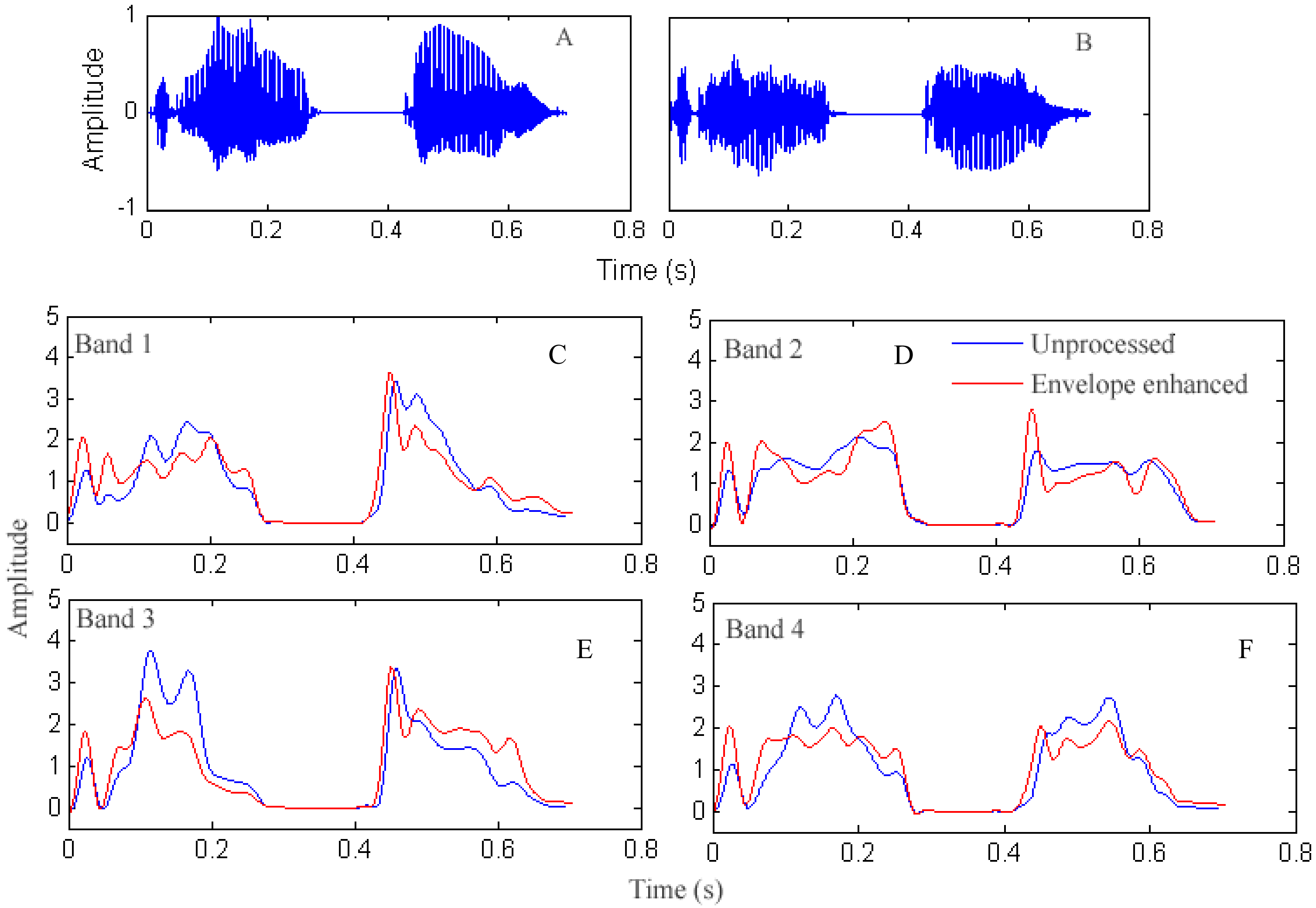


Figure M.3.Waveform of unprocessed signal (Panel A), envelope enhanced signal (Panel B) for the word /rai t̪a/ and envelope of the unprocessed and envelope enhanced signal in four different bands (Panels C, D, E and F).

ii) Speech in Noise

The words were individually mixed with a speech-spectrum-shaped noise at a specific signal to noise ratios (S/N ratio). Three S/N ratios, 0 dB, 5 dB and 10 dB were used. The speech spectrum shaped noise was generated by randomizing the phase of Fourier spectrum of concatenated words of original signals. The noise was added to both unprocessed and envelope enhanced speech based on RMS level of the signal. For envelope enhanced speech, noise was added after envelope enhancement. Figure M.4 shows the wave form of unprocessed signal with and without addition of spectrum shaped noise. Totally eight lists were available for assessing speech perception in noise.

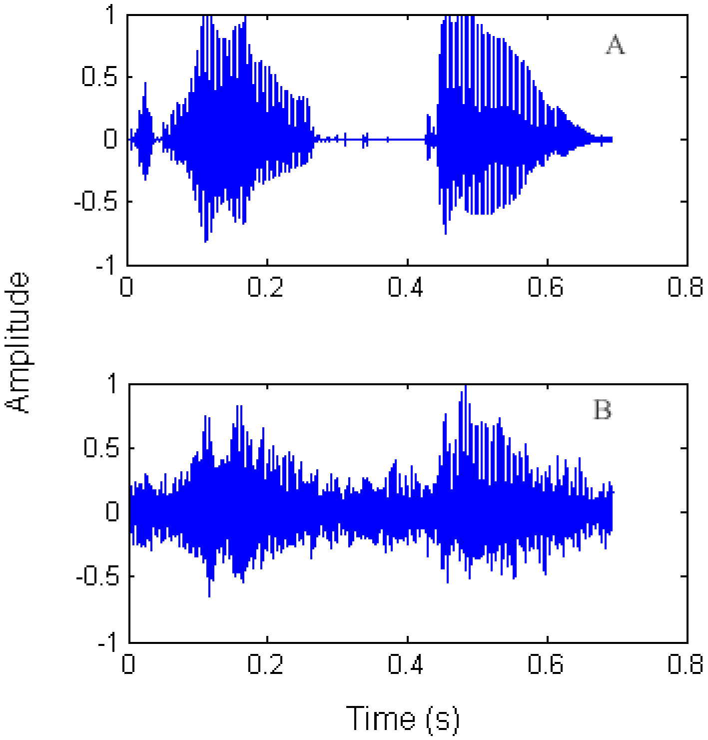


Figure M.4. Waveform of signal without noise (Panel A) and signal with noise at 0 dB S/ N ratio (Panel B) for word /rai t̪a/.

iii) Upward spectral shift

The upward spectral shift implemented in this study shifted the frequency range linearly upward above the knee-point frequency (500 Hz). The original frequency range 0-8 kHz was shifted linearly upward above 500 Hz. This processor shifts all the formant frequencies by the same amount in the frequency and this program was implemented in MATLAB -7.

For the upward spectral shift, speech signal was initially passed through a pre-emphasis filter (Butterworth -6 dB/octave below 1200 Hz). The output of the pre-emphasis filter was passed through Gamma tone filter. The Gamma tone filter was used to divide speech signal into 24 frequency bands based on the ERB scale (Patterson, 1989) to extract the envelope and carrier of the speech signal (Mc Cook 1993). The extracted carrier of the speech signal was shifted above 500 Hz and the shifted carrier was then multiplied with extracted envelope of the speech signal. Finally, the output of each band was summed together. The RMS (root mean square) energy of the output signal was made equal to the original speech signal. This speech-processing algorithm was implemented in the MATLAB -7(Math Works Inc). From Figure M.5 (B), it can be observed that information from 0 Hz was shifted above 500 Hz with preserved formant frequency ratio, which indicates that formant frequencies are proportionally shifted upward. There was no difference in the envelope of spectrally shifted speech and unprocessed speech.

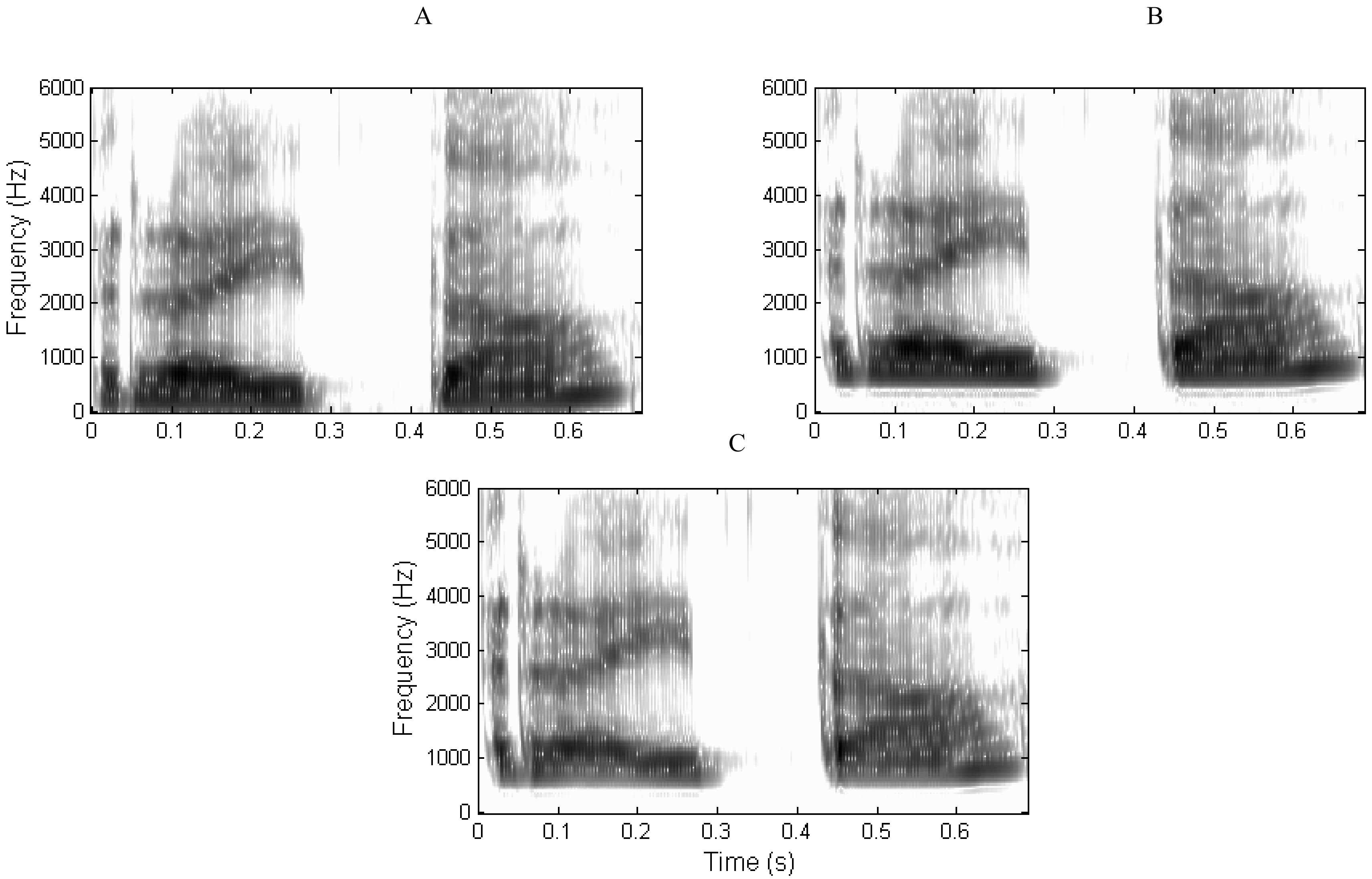


Figure M.5. Spectrogram of unprocessed speech (Panel A), upward spectral shifted speech (Panel B), and upward spectral shifted along with envelope enhanced speech (Panel C) for the word /rai t̪a/.

iv) Upward spectral shift along with envelope enhancement

Speech material was first spectrally shifted upward using the procedure described in the section 3.5.2 (*iii*) and then envelope was enhanced according to procedure described in section 3.5.2 (*i*). Figure M.5 (C) shows an example of spectrogram of a word with upward spectrally shift along with envelope enhancement.

d) Procedure

The speech stimuli were played manually in a PC at sampling frequency of 44.1 kHz and routed to a calibrated (ISO 389) diagnostic audiometer (Madson OB-922 with speaker). The participants received the signal from the loudspeaker of the audiometer kept at a distance of one meter at 0° azimuth. The presentation level of the stimuli was 40 dB SL (re: Speech Recognition Threshold). Each participant received all the lists. The order of presentation of lists was randomized across the participants. Participants had to repeat the speech token heard by them. The speech recognition scores were calculated by counting the number of words correctly repeated.

The data collected were tabulated and analyses were carried to investigate the objectives of the present study. The results obtained are discussed in the next chapter.

CHAPTER 4

Results

Contents

- 4.1. Analysis
  - 4.1.1. Phase I
  - 4.1.2. Phase II
- 4.2. Results of Temporal Modulation Transfer Function
- 4.3. Results of Speech Identification Assessment
  - 4.3.1. Speech Identification in Quiet
    - *a) Speech identification of unprocessed speech*
    - *b) Identification of speech with envelope enhancement*
  - 4.3.2. Speech Identification in Noise
  - 4.3.3. Identification of Upward Spectrally Shifted Speech with and without Envelope Enhancement

The objectives of the present study were to investigate (a) temporal modulation transfer function in individuals with auditory dys-synchrony (b) identification of speech with and without envelope enhancement in individuals with auditory dys-synchrony and their relation with modulation detection threshold (c) identification of speech with upward spectral shift and identification of speech with both upward spectral shift and envelope enhancement in individuals with auditory dys-synchrony. The data from two phases of study were tabulated and subjected to suitable statistical analyses. Statistical Package for the Social Sciences (SPSS) (version-15) and MATLAB -7 (Math works, UAS) was used to analyse the data.

4.1. **Analysis**

4.1.1. Phase I

The following analyses were carried out on the TMTF data obtained from the first phase of the study:

a) To quantify the temporal modulation detection threshold data, a first order low pass Butterworth filter was fitted from which peak sensitivity and band width measures were obtained using the Equation (1):

$$\mathrm{Y} = -10 \times \log_{10}\left[\frac{X0}{1+(\frac{f}{fc})^2}\right] \qquad (1)$$

where y is the modulation index (m) in dB (-10 log m), f is the modulation frequency in Hz, -10 log (X_0) is the peak sensitivity or gain in dB, fc is the

3 dB cutoff frequency or band width in Hz. The fitted function was adapted from Zeng et al. (2005). The curve fit program was executed using MATLAB-7.

b) Mixed design Analysis of Variance (ANOVA) was performed for within group and between group comparisons of modulation detection thresholds of the two groups. As there was a significant interaction, Independent sample‘t’ test was performed to compare modulation detection thresholds of the two groups for each modulation frequency.
c) Independent sample‘t’ test was performed to compare peak sensitivity and bandwidth of the two groups.

4.1.2. Phase II

The speech identification scores obtained for speech with temporal and/or spectral modifications obtained during the second phase of the study were converted into rationalized arc sine (rau) transform units (Studebaker, 1985) to stabilize the error variance for inferential statistics. The following analyses were carried out twice, once for raw scores and once for “rau” scores of the data obtained in the second phase of the study, but only results for “rau” scores are given.

a) Independent sample ‘t’ test was performed to compare identification scores in the two groups.
b) Paired sample ‘t’ test was performed to compare identification scores between unprocessed and envelope enhanced stimuli in listeners with normal hearing and auditory dys-synchrony.

c) Mixed design ANOVA was performed for within-group and between-group comparisons of speech identification scores in quiet for unmodified and temporally modified speech in subgroups of auditory dys-synchrony. As there was an interaction effect this was followed by one-way ANOVA.

d) Pearson product moment correlation analysis was performed to assess the correlation of speech identification scores with peak sensitivity, bandwidth of TMTF and pure-tone threshold.

e) Non-linear regression analysis was carried out to predict speech identification scores from peak sensitivity.

f) Mixed design ANOVA was performed for within group comparison and between conditions (0dB, +5dB and +10dB SNR) of speech identification score in noise for unprocessed speech and speech with temporal modification. As there was an interaction effect this was followed by one-way ANOVA.

g) Pearson product moment correlation analysis was carried out between speech identification scores in noise and peak sensitivity of TMTF.

h) Mixed design ANOVA, was performed for within-group and between-group comparisons of speech identification scores in quiet for unprocessed, spectrally modified and spectrally as well as temporally modified speech.

4.2. **Results of Temporal Modulation Transfer Function**

Temporal modulation transfer function, a graph representing modulation threshold as a function of modulation frequency (Viemeister, 1979) was plotted for subjects with normal hearing as well as those with auditory dys-synchrony. Figure R.1 shows the mean sinusoidal amplitude modulation detection threshold, with error bars representing the standard deviation (SD), for individuals with auditory dys-synchrony and those with normal hearing. Sensitivity to modulation was expressed as 20× log m, on the ordinate, and modulation frequency on the abscissa. It can be noted from the Figure R.1 that the sensitivity for modulation detection is poorer in individuals with auditory dys-synchrony when compared to that of normal hearing. The standard deviation was also larger in subjects with auditory dys-synchrony indicating greater inter subject variability.

The temporal modulation transfer function for listeners with normal hearing as well as those with auditory dys-synchrony mimics the shape of a low pass filter. In normal hearing listeners, modulation detection thresholds were around -20 dB for low modulation rates (fm =2-16 Hz), and worsened by 3-4 dB per octave when the modulation frequency was increased above 16 Hz. In subjects with auditory dys-synchrony, the thresholds were poorer than that obtained for subjects with normal hearing for all the modulation frequencies. It was around -10 dB for low modulation frequencies (2 to 8 Hz) and worsened further with increase in modulation frequencies. The average threshold dropped by 5-7 dB/octave as the modulation frequency was increased from 16 Hz to 64 Hz and worsened at the rate of 2-3 dB/octave for modulation frequencies above 64 Hz. A majority of the participants (25 subjects) with auditory dys-synchrony could not even detect a modulation depth of 0 dB when the modulation frequency was 128 Hz or higher.

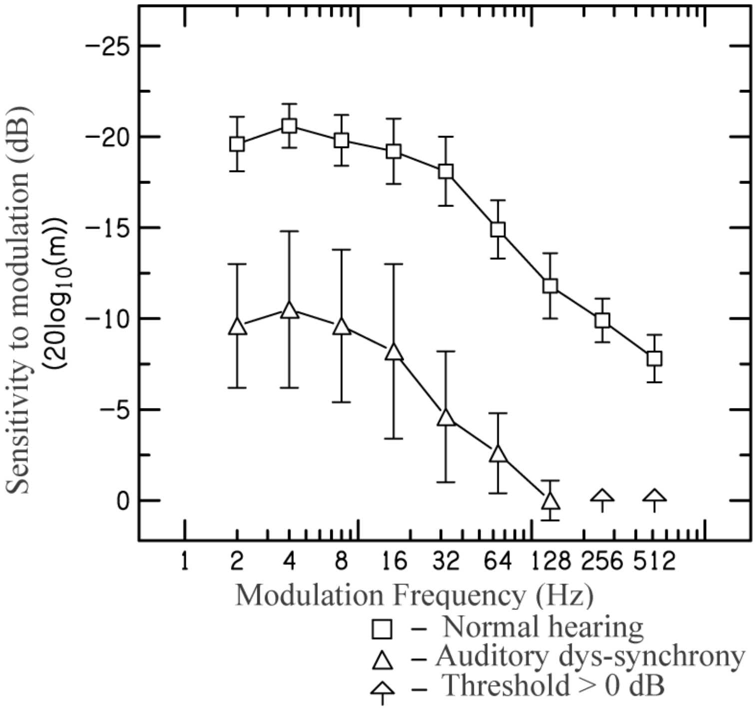


Figure R.1. Mean sensitivity to modulation as a function of modulation frequency in individuals with normal hearing and those with auditory dys-synchrony.

A Mixed ANVOA (for repeated measures), with modulation frequency (8 levels) as within subject factor and group (2 levels) as between subject factor, was performed to assess the effect of group and modulation frequency on modulation detection threshold. Results revealed a significant main effect of modulation frequency ($F_{(8,520)} = 475.8$, $p < 0.01$) and group ($F_{(1, 65)} = 441.76$, $p < 0.01$). There was a significant interaction between the effect of group and modulation frequencies ($F_{(8, 520)} = 43.8$, $p < 0.01$) indicating that the difference in sensitivity, between individuals with normal hearing and those with auditory dys-synchrony was not same for all the modulation frequencies. It can be observed from Figure R.1 that the difference between the two groups was lesser for low modulation frequencies and larger at higher modulation frequencies. The Bonferroni pairwise analysis indicated that sensitivity at 2 Hz was not significantly different from that of 8 Hz ($p=0.65$). Sensitivity at all the other modulation frequencies was significantly different ($p < 0.01$) from one another.

To assess whether the difference in mean threshold between the groups reach significance at all modulation frequencies, a separate Independent sample'tʼ test was performed. Analysis revealed a significant difference between the two groups at all the modulation frequencies. Table R.1 provides'tʼ test value and level of significance for all the modulation frequencies.

Table R.1

't' value and level of significance for difference in modulation detection threshold at different modulation frequencies

Modulation Frequency (in Hz)	**2**	**4**	**8**	**16**	**32**	**64**	**128**	**256**	**512**
't' value	-17.6*	-14.7*	13.7*	13.3*	14.7*	17.4*	17.2*	15.5*	14.6*

**: p< 0.01level*

The data of listeners with normal hearing and those with auditory dys-synchrony were modeled as a first order low pass filter similar to that carried out by Zeng et al. (2005). From the fitted function, 3 dB bandwidth (*fc*) and peak sensitivity were derived.

For individuals with normal hearing a typical low pass pattern was observed, with an average peak sensitivity of -20.6 dB (SD=1.1) and *f*c of 54.5 Hz (SD = 10.5). Individuals with auditory dys-synchrony also showed a low pass pattern, but the average peak sensitivity was -9.9 dB (SD = 4.3) and *f*c was 27.1 Hz (SD = 12.5). Independent sample'‘t’ test showed that the mean difference was statistically significant for both peak sensitivity (t= -11.6, $p<0.01$) and *f*c (t= -9.9, $p<0.001$). The RMS error for the fitted function across the subjects ranged from 0.8 to 2.2 dB for individuals with normal hearing and 0.8 to 3.4 dB for individuals with auditory dys-synchrony.

The modulation detection thresholds of individuals with auditory dys-synchrony were variable. Therefore, they were grouped based on peak sensitivity into three subgroups, namely, mild, moderate, and severe. Table R.2 provides the mean and range of peak sensitivity and bandwidth of each group. In individuals with auditory dys-synchrony two participants had peak sensitivity within normal range. The mean modulation detection threshold for different subgroups at each modulation frequency is depicted in Figure R.2.

The TMTF data was analysed to investigate if there is a significant difference in the bandwidth obtained for different subgroups of auditory dys-synchrony. One-way ANOVA revealed a significant main effect of severity of auditory dys-synchrony on bandwidth ($F_{(2,\ 1074)} = 10.05$, $p<0.01$). Bonferroni post-hoc analysis revealed that the mean bandwidth of mild group was significantly ($p<0.05$) different from that of severe group. There was no significant difference among the bandwidths of other subgroups.

Table R.2

Mean, range of peak sensitivity and bandwidth of different subgroups

Subgroups	N	Mean peak sensitivity (dB)	Bandwidth (Hz)
Normal	30	- 20.0 (-20 to -18)	54.5 (42 to 64)
Mild	15	- 13.5 (-17 to -13)	36.1 (30 to 42)
Moderate	14	- 8.5 (-12 to -8)	28.3 (20 to 35)
Severe	9	- 4.0 (-7 to 0)	18.3 (14 to 24)

Note: Values in the parenthesis indicate range

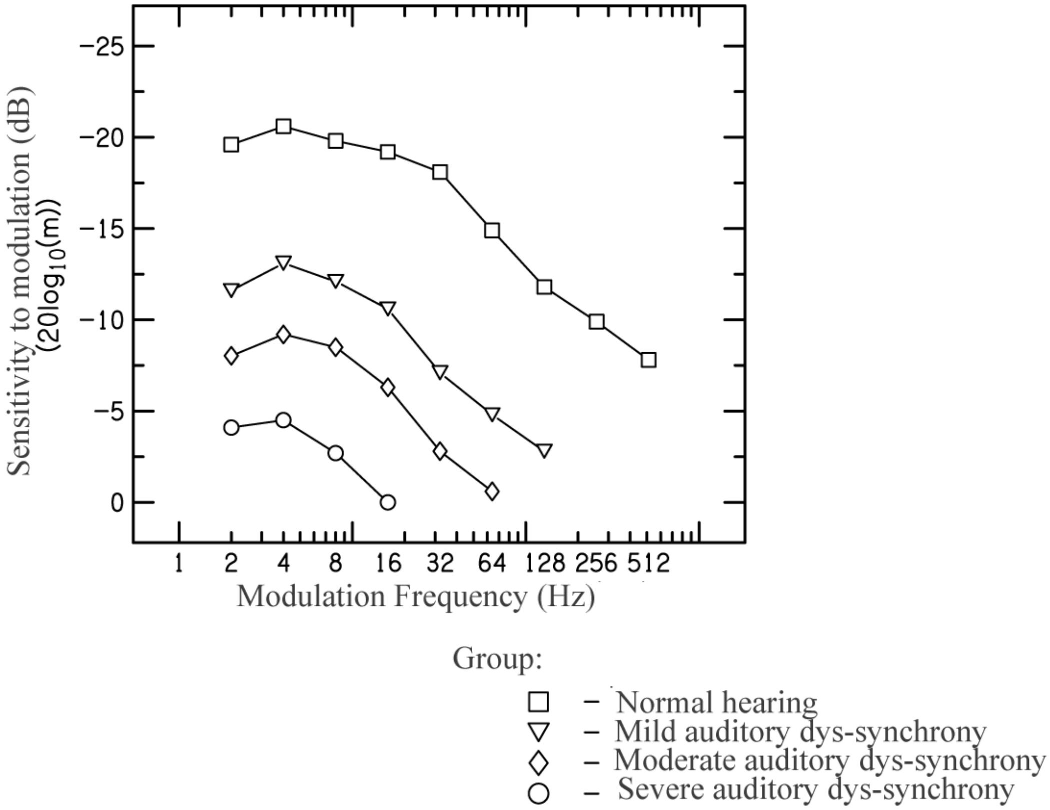


Figure R.2. Mean sensitivity to modulation as a function of modulation frequency for different subgroups.

Pearson product moment correlation was examined to assess the relationship between better ear pure-tone average with the modulation detection threshold for different modulation frequencies, peak sensitivity and bandwidth of TMTF. Pure-tone average showed no significant correlation with peak sensitivity (r= 0.3, p = 0.65) and bandwidth (r= 0.063, p = 0.686). As shown in Table R.3, the pure-tone average did not correlate with modulation detection threshold at any of the modulation frequencies.

Table R.3

Correlation Coefficient ('r') between better ear pure-tone threshold and modulation detection threshold at different modulation frequencies

Modulation Frequency (in Hz)	**2**	**4**	**8**	**16**	**32**	**64**	**128**	**256**	**512**
Correlation Coefficient 'r'	0.22	0.29	0.26	0.27	0.28	0.32	0.14	0.058	0.09

Note: *None of values were significant*

4.3. **Results of Speech Identification Assessment**

4.3.1. Speech Identification in Quiet

a) Speech identification for unprocessed speech

The speech identification scores of subjects with normal hearing ranged from 92 % to 100 % with a mean of 96 % for the unprocessed stimuli whereas the scores of individuals with auditory dys-synchrony ranged from 0% to 100% with a mean of 52.24% and a standard deviation of 34%. Independent sample 't' test revealed a significant difference between the mean speech identification scores of the two groups (t = 5.77, p<0.01).

Pearson product moment correlation analysis revealed no significant correlation between identification scores in quiet and pure-tone average (r = 0.47, p = 0.62). The correlation analysis was carried out between the better ear pure-tone thresholds at each octave frequency with identification score. Analysis revealed no significant correlation with threshold at any of the octave frequencies. Table R.4 provides correlation coefficient values at each of the octave frequencies.

Table R.4

Correlation Coefficient ('r') between speech identification scores for unprocessed and pure-tone threshold at octave frequencies

Frequency (Hz)	**250**	**500**	**1000**	**2000**	**4000**	**8000**
Correlation Coefficient 'r'	-0.32	-0.22	-0.35	-0.32	-0.38	-0.25

Note: None of the values were significant

Pearson product moment correlation was carried out to assess the relationship between speech identification scores, peak sensitivity and bandwidth of TMTF. The correlation analysis was carried out to assess the relation between identification scores and modulation detection threshold at all the modulation frequencies. Speech identification scores showed a significant high negative correlation with peak sensitivity (r = - 0.94, p<0.01) and a significant high positive correlation with bandwidth (r= 0.63, p<0.01). As shown in Table R.5, there was a significant negative correlation with threshold of TMTF for all the modulation frequencies though the correlation value decreased with increase in modulation frequency.

Table R.5

Correlation Coefficient ('r') between speech identification scores for unprocessed speech and modulation detection threshold at different modulation frequencies

Modulation Frequency (Hz)	**2**	**4**	**8**	**16**	**32**	**64**	**128**	**256**	**512**
Correlation Coefficient 'r'	-.78*	-.86*	-.86*	-.82*	-.83*	-.75*	-.64**	-.54**	-.44**

*: $p<0.01$ level, **: $p< 0.05$ level;

An attempt was made to predict the speech identification scores from peak sensitivity. A linear regression was drawn to predict the identification scores from peak sensitivity. The regression line was fitted (raw scores) in a scatter plot as shown in the Figure R.3 (Panel A) and Equation 2 was obtained to predict speech identification scores from peak sensitivity.

$$\text{Speech identification score} = (1.78 \times \text{peak senstivity}) - 4.5 \qquad (2)$$

The linear regression showed a high r^2 (0.877). From the regression equation it can be noted that the gradient is close to one and the constant is less, which indicates that speech identification scores can be predicted from peak sensitivity using linear regression equation. However, a plot of residuals depicted Figure R.3 (Panel B) displays a systematic variation, which indicates that the data fits poorly in a linear model.

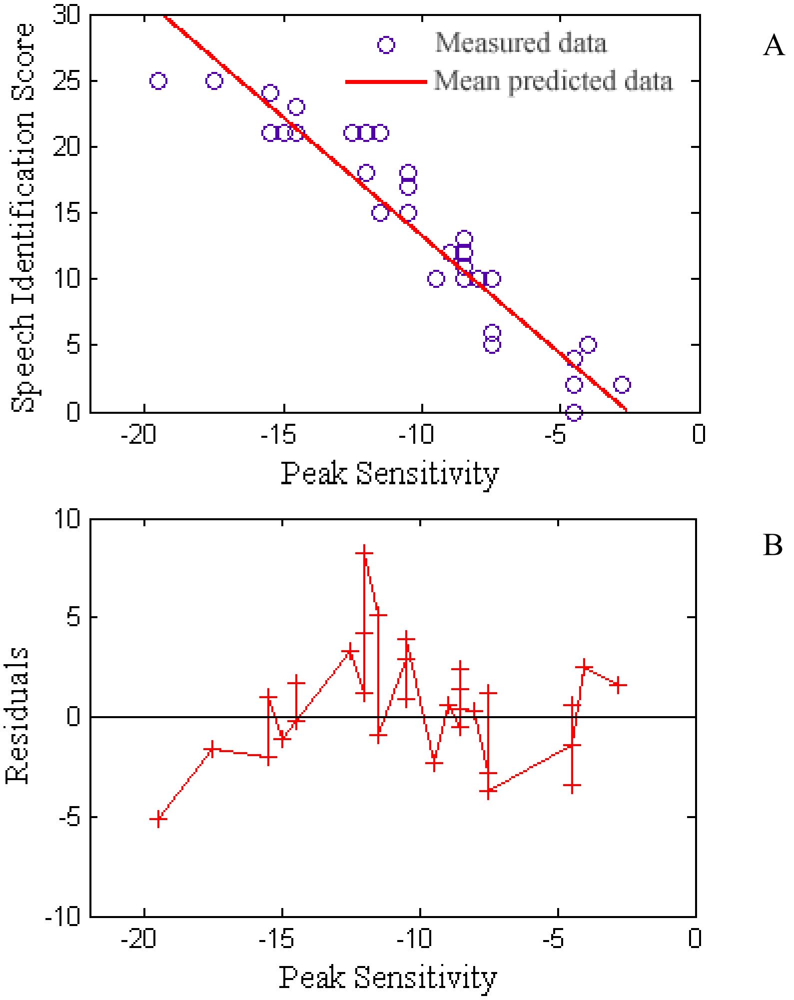


Figure R.3. Measured and predicted speech identification scores from linear regression (Panel A) and a plot of residuals (Panel B).

Analysis using MATLAB curve fit tool revealed that a nonlinear regression, using a three parameter sigmoid function gives best fit to the data. Equation 3 shows the function used to fit the data.

$$y = \frac{a}{1+e^{\left(\frac{x-c}{b}\right)}} \tag{3}$$

In the formula /*y*/ represents the predicted speech identification score, /*x*/ represents the peak sensitivity and *a, b*, and *c* are constants (a= 23.98, b=-0.53, and c= 9.08) . Figure R.4 (Panel A), shows the fitted function of the relationship between speech identification scores, and peak sensitivity of TMTF. The solid lines in the figure shows the mean predicted scores and dashed line represents 95% confidence intervals. The residuals in the Figure R.4 (Panel B) do not show any trend, indicating that the sigmoid function yields a good fit to the data. The r^2 value of the model was 0.96 and the standard error of the estimate was 2.9.

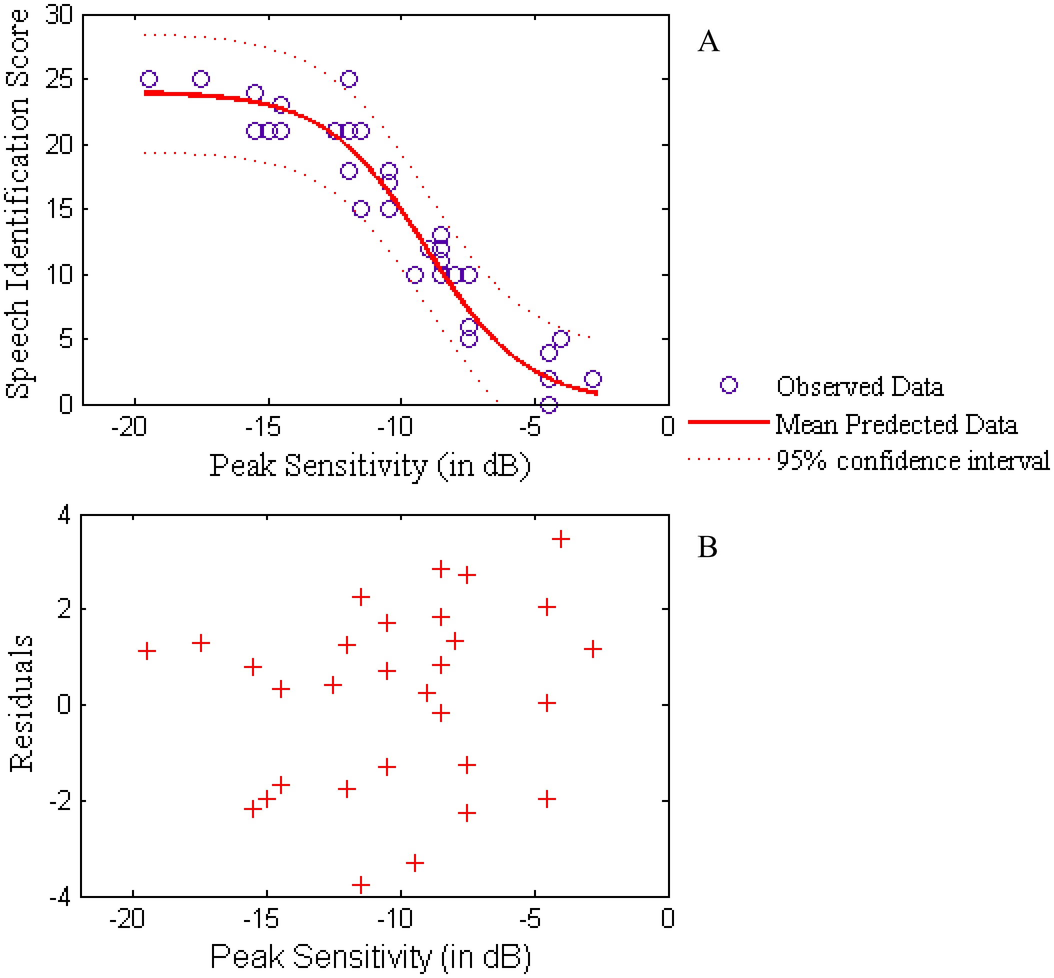


Figure R.4. Measured and predicted speech identification scores from nonlinear regression (Panel A) and a plot of residuals (Panel B).

b) Identification of speech with envelope enhancement

All the individuals in the normal hearing group obtained a speech identification score of 100% for the envelope enhanced stimuli. The identification scores for individuals with auditory dys-synchrony ranged from 0% to 100% with a mean of 68.24 % and a SD of 35.2%. Independent sample‘t’ test revealed a significant difference between the mean speech identification scores of the two groups (t = 3.77, p < 0.01). For subjects with auditory dys-synchrony, the mean speech identification score for the envelope enhanced stimuli was 14.4 % higher than the mean score for the unprocessed stimuli and Paired Sample t-test revealed that this difference was statistically significant ($t= -6.5$, $p < 0.01$).

Two individuals with auditory dys-synchrony, whose peak sensitivity was within normal limits had identification scores greater than 90% for both unprocessed and envelope enhanced speech and were not included for further analysis of the subgroups. To evaluate the benefit derived by envelope-enhanced speech in individuals with different subgroups, a mixed ANOVA (for repeated measures) was performed, with subgroups (4 levels) as between subject factor and stimuli (2 levels) as within subject factor. The analysis showed a significant main effect of stimuli ($F_{(1, 37)} = 70.9$ $p<0.01$) and subgroup ($F_{(3, 36)} = 69.5$, $p<0.01$). The Bonferroni Post hoc analysis indicated a significant difference among all the subgroups ($p<0.01$). The analysis also showed a significant interaction between subgroups and stimuli ($F_{(1, 37)} = 7.43$ $p<0.05$).

As there was an interaction between subgroups and stimuli, a separate one-way ANOVA was carried out to assess the significant difference among subgroups for speech identification scores for unprocessed and envelope enhanced speech. Results revealed a significant effect of subgroups on identification scores for both unprocessed speech ($F_{(3,215.9)} = 109.8$, $p<0.01$) and envelope enhanced speech ($F_{(3, 36.6)} =28.3$, $p<0.01$). Bonferroni post hoc analysis for unprocessed speech revealed that the difference in mean scores reached significance ($p< 0.01$) for all the subgroups. For envelope enhanced speech Bonferroni post hoc analysis revealed a significant ($p< 0.01$) difference among all the subgroups, only mean scores of individuals in normal hearing group did not differ significantly from individuals in the mild group.

Figure R.5 shows the mean speech identification scores with standard deviation as error bars for unprocessed and envelope enhanced speech for different subgroups. The mean improvement for participants with moderate group was 23%, which is higher than that observed for mild (15.2 %) and severe (7.5 %) group. Variability in improvement was higher for participants in the severe group when compared to those in mild and moderate group.

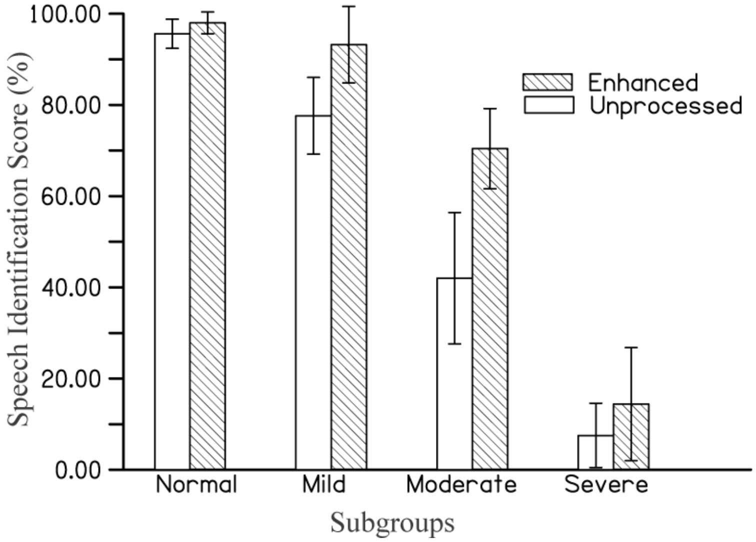


Figure R.5. Mean and standard deviation (error bar) of speech identification scores across different subgroups.

Pearson product-moment correlation revealed a high negative correlation between speech identification scores and peak sensitivity ($r = -0.87, p < 0.01$). There was a high positive correlation between speech identification scores and bandwidth ($r= 0.78, p < 0.05$). That is individuals with good speech identification scores also had lower peak sensitivity and higher bandwidth. The modulation detection thresholds at each of the modulation frequencies were compared with the speech identification score by computing a Pearson's Product Moment Correlation between the two factors. Analysis revealed a negative correlation between modulation detection threshold and speech identification scores. Table R.6 shows the 'r' value for all the modulation frequencies.

Table R.6

Correlation Coefficient ('r') between speech identification scores with envelope enhancement and modulation detection threshold at different modulation frequencies

Modulation Frequency (Hz)	**2**	**4**	**8**	**16**	**32**	**64**	**128**	**256**	**512**
Correlation Coefficient 'r'	-.84*	-.84*	-.86*	-.78*	-.77*	-.66*	-.49*	-.36**	-.28

***: $p < 0.05$ level *: $p < 0.01$ level;*

4.3.2. Speech Identification in Noise

Speech perception in noise was assessed for 25 individuals with normal hearing and 25 participants with auditory dys-synchrony at three S/N ratios (0 dB, +5 dB, +10dB) for unprocessed and envelope enhanced speech. Table R.7 presents the mean and standard deviation of speech identification scores of both the groups, for unprocessed, and envelope enhanced speech, at three different S/N ratios. It can be discerned from the table R.7 that as the S/N ratio decreased, there was a reduction in speech identification scores in both the groups, but this reduction was greater for individuals with auditory dys-synchrony when compared to individuals with normal hearing. Envelope enhancement improved speech identifications scores for both the groups. Individuals with normal hearing showed greater improvement at 5 dB S/N ratio and 0 dB S/N ratios. Participants with auditory dys-synchrony showed improved identification ability for envelope-enhanced stimuli at all the S/N ratios but maximum improvement was observed at +10 dB S/N ratio.

To evaluate the effect of envelope enhancement in quiet and at three S/N ratios in the two groups, Mixed ANOVA, for repeated measures, was carried out with S/N ratio (4 levels) and stimuli (2 levels) as within subject factors and group (2 levels) as between subject factor. The analysis showed a significant main effect of S/N ratio ($F_{(3,126)} = 117.5$, $p < 0.01$), stimuli ($F_{(1, 42)} = 89.3$, $p < 0.01$) and group ($F_{(3, 47)} = 298.1$, $p < 0.001$). Two way interaction analysis revealed that the effect of group had a significant interaction with S/N ratio ($F_{(3,126)} = 24.3$, $p < 0.01$) and stimuli ($F_{(1, 42)} = 7.9$, $p<0.05$), but there was no significant interaction between the effect of S/N ratio and stimuli ($F_{(3,126)} = 1.9$, $p = 1.29$). Three way interaction was also non-significant among the effects of group, S/N ratio, and stimuli ($F_{(3, 126)} = 26.7$, $p < 0.01$). Bonferroni Pairwise comparison showed a significant difference in speech identification scores across different S/N ratio conditions.

Table R.7

Mean and standard deviation of speech identification scores (in percentage) for unprocessed and envelope enhanced speech at different S/N ratios

Group	Stimuli	S/N ratio			
		Quiet	10 dB	5 dB	0 dB
Auditory Dys-synchrony	Unprocessed speech	57.2 (33.2)	33.6 (18.4)	21.6 (9.6)	13.6 (5.6)
	Envelope Enhanced speech	74.8 (31.4)	50.4 (25.4)	30.2 (18.4)	19.0 (11.2)
Normal Hearing	Unprocessed speech	98.4 (2.16)	94.4 (3.9)	86.2 (4.4)	62.2 (9.2)
	Envelope Enhanced speech	100.0 (0)	96.0 (3.4)	92.4 (5.2)	88.5 (6.4)

Note: Value in the parenthesis indicate standard deviation

A separate one-way ANOVA was administered to assess the effect of S/N ratio on speech identification scores for unprocessed and envelope enhanced speech in participants of both the groups. Results revealed that in normal hearing listeners, there is a significant effect of S/N ratio on speech identification scores for both unprocessed speech ($F_{(3,215.9)} = 109.8$, $p<0.01$) and envelope enhanced speech ($F_{(3, 36.6)} = 28.3$, $p<0.01$). Bonferroni post hoc analysis for unprocessed speech showed that the mean scores in quiet condition is significantly different from those obtained in 5 dB and 0 dB S/N ratio conditions ($p<0.01$), but it did not differ significantly from scores obtained in 10 dB S/N ratio condition. For envelope enhanced speech, scores obtained in 0 dB S/N ratio was significantly different from those obtained in other S/N ratio conditions, but there was no significant difference among the other S/N ratio conditions.

Similar results were obtained in listeners with auditory dys-synchrony. There was a main effect of S/N ratio on scores obtained for both unprocessed ($F_{(3,689.4)} = 26.5$, $p<0.01$) and envelope enhanced speech ($F_{(3, 77.2)} =16.1$, $p<0.01$). Bonferroni post hoc analysis for unprocessed speech revealed that the mean scores in quiet condition was significantly different from those obtained in the presence of noise ($p<0.01$), but there was no significant difference among the scores obtained for various S/N ratio conditions. For envelope enhanced speech, mean identification scores were significantly different from those obtained in noise conditions. Among the different noise conditions significant difference in the mean scores was observed only between 10 dB S/N ratio and 0 dB S/N ratio conditions.

Inspection of the individual data showed that in some of the subjects in whom the score in quiet was already poor, the actual deterioration in performance due to the presence of noise could not be measured due to floor effect. Therefore, to minimize this effect, data obtained at different S/N ratios were further analyzed to see if there is a significant difference among subgroups of auditory dys-synchrony. Among the 25 participants with auditory dys-synchrony there were nine individuals in mild group, nine individuals in the moderate group and seven individuals in the severe group.

Mixed ANOVA (for repeated measures) was performed to assess the effect of S/N ratio on speech identification scores for speech presented with and without envelope enhancement for subgroups. The analysis was performed with S/N ratio (4 levels) and Stimuli (2 levels) as within subject factors and subgroups (4 levels) as between subject factor. Analysis revealed a significant main effect of S/N ratio ($F_{(3,123)} = 547.1$, $p < 0.01$), stimuli ($F_{(1, 41)} =156.1$, $p < 0.01$) and subgroups ($F_{(3, 41)} = 524.7$, $p < 0.01$). There was a significant two way interaction of subgroups with S/N ratio ($F_{(9,123)} =66.4$,

$p<0.01$) and stimuli ($F_{(3, 41)} = 11.5$, $p<0.01$). There was no significant interaction between S/N ratio and stimuli ($F_{(3,123)} = 2.47$, $p < 0.01$). There was a significant three way interaction among S/N ratio, subgroup and stimuli ($F_{(9,123)} = 22.1$, $p < 0.01$). Bonferroni post-hoc analysis revealed a significant difference among the subgroups and Bonferroni pair wise comparison revealed a significant difference among the scores in all S/N ratio conditions.

A one-way ANOVA was performed to assess the significant difference in speech identification scores across different S/N ratio conditions for different subgroups. A separate analysis was carried out for unprocessed and envelope enhanced speech. As shown in Table R.8 there was a significant main effect of subgroup at all S/N ratio conditions for both unprocessed and envelope enhanced speech.

Table R.8

F ratio and significance values of the difference in mean values at different S/N ratios for unprocessed and envelope enhanced speech

Stimuli	**S/N Ratio**	**F ratio**
Unprocessed Speech	Quiet	200.0*
	10 dB S/N	308.1*
	5 dB S/N	477.9*
	0 dB S/N	245.8*
Envelope Enhanced Speech	Quiet	93.1*
	10 dB S/N	109.1*
	5 dB S/N	208.3*
	0 dB	313.6*

**: p< 0.01level*

Bonferroni post-hoc analysis revealed that there is a significant difference among different subgroups for unprocessed speech at all S/N ratios ($p<0.01$), except at 0 dB

S/N ratio, where the mean identification scores of individuals in moderate group did not differ significantly from those in severe group. For the envelope-enhanced stimuli, Bonferroni post-hoc analysis showed that mean scores were significantly different among the subgroups at all S/N ratios. Figure R.6 provides mean and standard deviations of identification scores at different S/N ratios for all the subgroups.

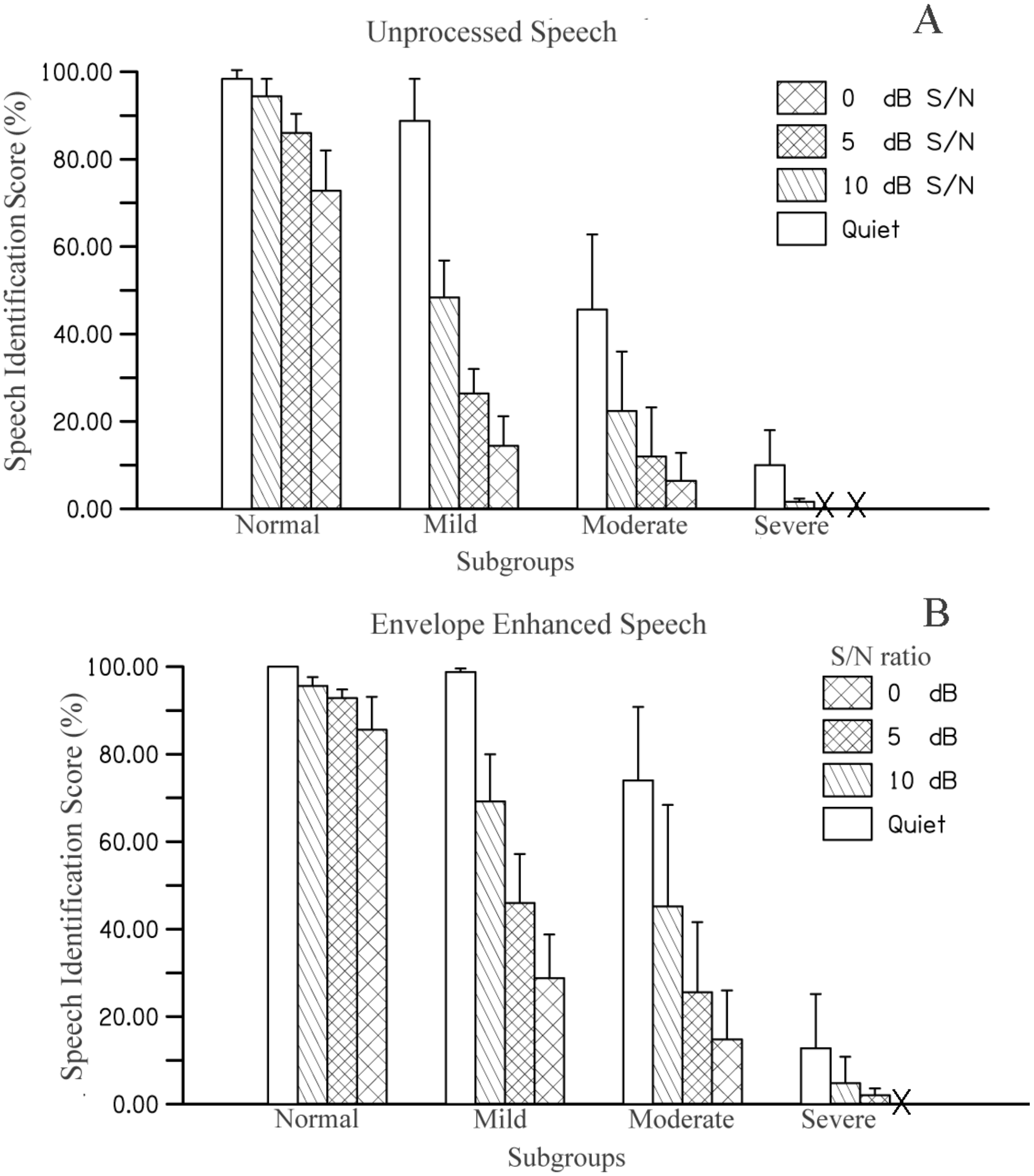


Figure R.6. Mean and standard deviation of identification scores for unprocessed speech (panel A) and envelope enhanced speech (panel B) at different S/N ratio conditions.

Pearson product moment correlation analysis was carried out to investigate the relationship between speech identification scores in noise at three S/N ratios, and peak sensitivity. Speech identification scores showed a significant high negative correlation with peak sensitivity at all the S/N ratios. The results indicated that individuals with good speech identification scores in noise also had lower peak sensitivity. The correlation coefficient 'r' value along with level of significance is shown in Table R.9.

Table R.9

Correlation Coefficient ('r') between speech identification scores at different S/N ratios and Peak sensitivity

S/N ratio	**Quiet**	**10 dB**	**5 dB**	**0 dB**
Correlation Coefficient 'r'	-0.91**	-0.81**	-0.75**	-0.61*

**: $p< 0.05$ level **: $p < 0.01$ level;*

4.3.3. Identification of upward spectrally shifted speech with and without envelope enhancement

The speech identification scores of individuals in the normal hearing group ranged from 60 to 92% with a mean of 75 % for both upward spectrally shifted speech and upward spectrally shifted speech with envelope enhancement. Identification scores for spectrally modified speech in individuals with auditory dys-synchrony varied from 0 to 40 % with a mean identification score of 9% for upward spectrally shifted speech and 10.8% for spectrally shifted speech with envelope enhancement. It was observed from the data that among the individuals with auditory dys-synchrony, only 12 subjects had speech identification scores above chance factor and 28 individuals had a score of 0 %. The mean and standard deviation for unprocessed, spectrally shifted speech and speech spectrally shifted, along with envelope enhancement is given the Figure R.7.

To examine the effect of upward spectral shift and envelope enhancement with upward spectral shift of speech in individuals with auditory dys-synchrony and those with normal hearing, a mixed ANOVA (for repeated measure) was performed on the raw scores, with stimuli (3 levels) as within subject factor and group (2 levels) as between subject factor. Results revealed a significant main effect of stimuli ($F_{(1, 87)} = 537.7$, $p<0.01$) and group ($F_{(1, 38)} =124.3$ $p<0.01$). Interaction analysis showed a significant interaction between group and stimuli ($F_{(1, 87)} = 537.7$, $p<0.01$). Bonferroni pairwise analysis revealed that the scores for unprocessed speech was significantly different ($p<0.01$) from those of spectrally shifted speech as well as envelope enhanced with spectrally shifted speech but there was no significant difference ($p>0.05$) between the scores for speech with upward spectral shift and upward spectral shift along with envelope enhanced speech.

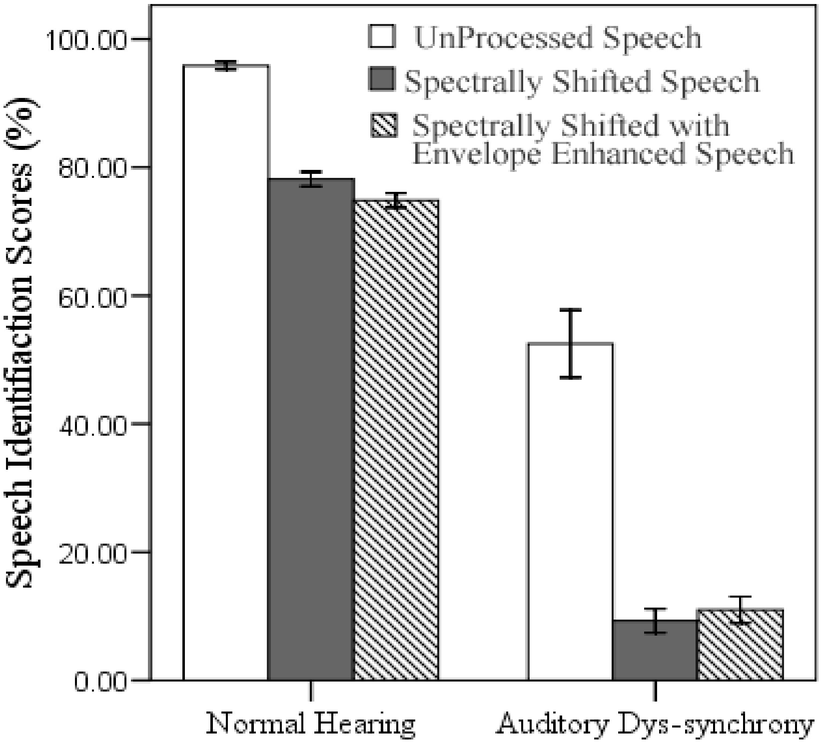


Figure R.7. Mean and standard deviation of identification scores for three different stimuli.

To summarize, the following results were observed in the present study:

a) The sensitivity to modulation was significantly lower in auditory dys-synchrony group, compared to normal hearing group. The threshold of modulation detection was higher for auditory dys-synchrony compared to those with normal hearing. The bandwidth of the TMTF in individuals with auditory dys-synchrony was one-half of that obtained in normal hearing listeners.

b) Identification scores were significantly poorer in individuals with auditory dys-synchrony when compared to those of normal hearing listeners. Identification of speech in individuals with auditory dys-synchrony correlated with the sensitivity to modulation.

c) Envelope enhancement improved speech identification in individuals with auditory dys-synchrony, but the improvement was higher for individuals with moderate degree when compared to those with severe and mild degree of impairment.

d) The effect of noise on speech identification was greater for individuals with auditory dys-synchrony when compared to those with normal hearing listeners.

e) Envelope enhancement improved speech identification scores, in presence of speech spectrum shaped noise for individuals with auditory dys-synchrony. The mean improvement was higher for 10 dB S/N ratio condition when compared to 0 dB and 5 dB S/N ratio.

f) Upward spectral shift or upward spectral shift with envelope enhancement did not improve identification scores in individuals with auditory dys-synchrony.

Chapter 5

Discussion

Contents

5.1 Temporal Modulation Transfer Function

Investigation of modulation detection thresholds for sinusoidal amplitude modulated white noise revealed the following results:

a) Listeners with normal hearing and those with auditory dys-synchrony were most sensitive to slow fluctuations and the sensitivity to modulations decreased with increase in the fluctuation rate.

b) The average peak sensitivity and bandwidth (*f*c) in listeners with auditory dys-synchrony were about one-half of that observed for listeners with normal hearing.

The results of the present study are consistent with those reported in the literature for individuals with normal hearing (Eddin, 1993; Lorenzi et al., 2002; Viemiestar, 1979) and those with auditory dys-synchrony (Kumar & Jayaram, 2005; Zeng, Kong, Michalewski, & Starr, 2005; Zeng, Oba, Grade, Sininger, & Starr, 1999). It was observed that the modulation detection threshold in participants with auditory dys-synchrony was two times higher when compared to that of normal hearing listeners. The difference in modulation detection threshold, between individuals with normal hearing and those with auditory dys-synchrony increased with increase in frequency. Kumar and Jayaram (2005) have also reported similar trend in individuals with auditory dys-synchrony. These results suggest that individuals with auditory dys-synchrony have greater impairment in processing faster modulation when compared to processing of slow modulations. The average peak sensitivity reported in individuals with auditory dys-synchrony in the present study is higher than that observed by previous investigators (Kumar, & Jayaram 2005; Zeng et al., 2005). This difference in the mean data for peak sensitivity may be due to variations in the procedure employed, step size used, number of subjects investigated and the severity of the disorder.

High modulation detection threshold observed in listeners with auditory dys-synchrony is unlikely to be because of reduced audibility and age. Correlation analysis in the present study demonstrated no relationship between pure-tone threshold and modulation threshold. Also, Viemeister, (1979) has demonstrated that sensitivity to modulation is level independent. Hall and Grose (1994) have reported that modulation sensitivity reaches adult values by 9 years of age and the age of the participants of the present study was 12 years and above. Therefore, age may not have been a contributing factor for the poor TMTF threshold. The impaired sensitivity to modulation in individuals with auditory dys-synchrony may be related to the underlying neural pathology.

The abnormalities in the auditory pathway that lead to the profile of auditory dys-synchrony are not clear, as it is difficult to determine the exact mechanism by which temporal cues are disrupted in the affected subjects. Results of electrophysiological tests indicate two neuro-physiological manifestations dys-synchronized spike discharge (Waxman, 1977) and/or reduced spike count (Starr et al., 2003). The dys-synchronies and/or reduced spike can lead to smearing of the internal representation of aggregate average neural response for physical stimuli (Starr et al., 2003; Zeng et al., 2005). This smeared neural representation causes difficulty in discriminating the small variations in the amplitude of sinusoidal amplitude modulated stimuli from un-modulated stimuli. So they require larger modulation depths, that is larger variation in amplitude to differentiate them from the un-modulated stimuli (Rance, 2005; Zeng et al., 2005).

In the auditory system, higher modulation frequencies are processed at the level of auditory nerve and brainstem, whereas lower modulation frequencies are processed mainly in the thalamus and auditory cortex (Giraud et al., 2000). Synchronous response for temporal coding is important for processing of the signal at the level of auditory nerve and brainstem while efficient processing can occur with less synchronous firings at higher levels (Frisina, 2001). This could have lead to greater problem in processing high rate of modulations in subjects with auditory dys-synchrony.

The results of the present study suggest that the modulation detection thresholds for different modulation frequencies (TMTF function) are affected in individuals with auditory dys-synchrony and there is variability in the degree of impairment. The results suggest that probably TMTF can be used as a potential tool for assessing the degree of abnormality in individuals with auditory dys-synchrony.

5.2. Speech Identification

In the second phase of the study speech identification ability was assessed for unprocessed and envelope enhanced speech in quiet and in the presence of noise. Speech identification ability was also assessed for speech with upward spectral shift and speech with both upward spectral shift and envelope enhancement. The results of the second phase can be summarized as follows:

a) Speech identification scores were significantly lower in individuals with auditory dys-synchrony when compared to listeners with normal hearing. In a majority of the participants, speech identification scores were poorer than that would be expected from their pure-tone threshold.

b) Noise had more detrimental effect on speech perception in individuals with auditory dys-synchrony when compared to listeners with normal hearing.

c) Both speech identification scores in quiet and noise showed a significant correlation with modulation detection threshold and bandwidth.

d) Envelope enhancement improved identification scores in individuals with auditory dys-synchrony, both in quiet and in presence of background noise. The improvement was greater for individuals with mild and moderate impairment in temporal processing. Individuals with severe impairment in temporal processing did not show any improvement with envelope enhancement.

e) Upward spectral shift or upward spectral shift with envelope enhancement did not improve speech identification ability, rather significantly degraded speech perception.

5.2.1 Speech Identification in Quiet

a) Speech identification of unprocessed speech

Perception of speech in individuals with auditory dys-synchrony was significantly lower when compared to normal hearing listeners. Further, the correlation analysis revealed that speech identification scores for individuals with auditory dys-synchrony were disproportionate to their pure-tone threshold. This is best illustrated by comparing the speech identification scores in individuals with auditory dys-synchrony to those expected by degree of hearing loss for patients with cochlear hearing loss (Vanaja and Jayaram, 2003). In the present study, in 68 % of individuals, the speech identification scores were lower when compared to those reported by Vanaja and Jayaram for ears with sensorineural hearing loss. Results of the present study are supported by the earlier findings (Narne & Vanaja 2008 b; Sininger & Oba, 2001; Starr et al., 1996). These results suggest that factors other than pure-tone thresholds impair the speech understanding capability in these individuals.

Modulation detection threshold and bandwidth were abnormally low in many of the participants with auditory dys-synchrony. Speech identification scores showed a good correlation with peak sensitivity and bandwidth. Previous investigators (Kumar and Jayaram, 2005; Rance, Mc Kay, Grayden, 2004; Zeng et al., 1999) reported similar results for peak sensitivity. But, there are no reports assessing the relationship between speech identification scores and bandwidth. Based on the peak sensitivity, individuals with auditory dys-synchrony were grouped into three subgroups, namely, mild, moderate and severe. Two participants who had normal peak sensitivity had speech identification scores of greater than 90%. It was observed that participants with mild impairment demonstrated speech identification scores greater than 65% whereas those with moderate impairment had identification scores in the range of 30% to 60%. Speech identification scores of participants with severe impairment were less than 8%. These results are similar to with those reported by Zeng et al. (1999). In an earlier investigation in individuals with auditory dys-synchrony, Rance et al. (2004) found that the mean peak sensitivity of the modulation detection was -3.4 dB for individuals with speech identification scores of less than 30% and the mean peak sensitivity of individuals with speech identification scores of greater than 30% was -14.3 dB.

Further, the results of the present study revealed a nonlinear relationship between speech identification scores and peak sensitivity. The speech identification scores showed a linear increase when the sensitivity to modulation increased from -5 dB to -18 dB. The speech identification scores reached 0% when the modulation threshold was less than -5 dB and the scores showed a saturation when the modulation sensitivity was greater than -18 dB. In contrast to this, earlier investigators have shown a linear relationship (Rance et al., 2004; Zeng et al., 2001) between modulation sensitivity and identification scores. Probably the number of subjects studied and range of difficulties experienced by those individuals contributed for the difference in results. Inspection of the individual data in the previous studies reveals that a majority of the participants had moderate impairment. If the data of only those with moderate degree of impairment was analyzed in the present study probably a linear relationship would have been observed.

The results of the present study and the previous investigations suggest that impaired ability to follow the amplitude variations (envelope) in speech signal probably account for a substantial part of the difficulty experienced by individuals with auditory dys-synchrony. Slow amplitude variations (envelope) of speech signal are associated with changes in syllabic and phonetic-segment constituents (Rosen 1992). Therefore, impairment in the ability to follow amplitude variations can impair the processing of syllabic and phonetic segments. This leads to smearing of the segmentation cues in speech signal, and this has a profound impact on speech intelligibility (Drullman, Festen, & Plomp, 1994a 1994b). Thus it can be hypothesized that assessing the temporal modulation transfer function (peak sensitivity and bandwidth) provide information about perception of the temporal envelope cues in speech signal.

Results of studies simulating perceptual difficulties experienced by individuals with auditory dys-synchrony (Zeng et al., 1999) support this hypothesis. Zeng et al. (1999) observed that when the speech signal was temporally distorted by filtering and reducing modulation depth in the speech signal, the perception of speech by normal hearing listeners was similar to that observed in listeners with auditory dys-synchrony. Artificially decreasing modulation depth of natural speech, generally causes low amplitude portion of the speech signal to be amplified and high amplitude regions to be attenuated, because, these amplitude variations in each band gets smaller and eventually becomes stationery sound (Drullman et al., 1994a). This supports our hypothesis that the difficulty in understanding speech experienced by individuals with auditory dys-synchrony is related to temporal resolution.

b) *Identification of speech with envelope enhancement*

Results of the present study unequivocally demonstrated a significant benefit from envelope enhancement. Envelope enhanced speech improved speech identification by 14.4% (13.6 %) over unprocessed speech in individuals with auditory dys-synchrony. Out of the 40 participants, 26 participants showed a significant improvement and only eight participants did not show any improvement with envelope enhancement. Six participants had identification scores of 90 to 100 % in quiet and hence improvement was not observed due to ceiling effect. Among the participants, those with moderate impairment in temporal resolution showed significantly greater improvement when compared to those with mild and severe degree.

Improvement in identification scores obtained by individuals with moderate degree is in close agreement with those reported by Narne and Vanaja (2008 a). Although investigation by Narne and Vanaja did not include assessment of the modulation detection threshold, inspection of speech identification scores indicates that a majority of the participants had moderate degree of impairment. The improvement observed for clear speech in individuals with auditory dys-synchrony (Zeng & Liu, 2006) is comparable to that observed in the present study. Thus, the results of the present study along with the previous literature shows that enhancing the amplitude variations (envelope) of the speech signal improves speech perception in individuals with auditory dys-synchrony.

Perceptual difficulties experienced by individuals with auditory dys-synchrony, is in part due to the impairment in perceiving the amplitude variations in continuous signal. Reduced ability to process the envelope of speech signal impairs the perception of segmental cues and salient cues for consonantal perception by blurring/smearing the consonant-vowel distinction (Drullman et al., 1994 a). As shown in Figure M.2, enhancing the envelope by compression/expansion scheme enhanced the segmental cues and consonantal portion of the signal. It can be reasoned that enhancing the segmental and consonantal portion of signal, may have improved their ability to processes more salient cues and they less likely to be subjected to forward or backward masking by vowels which have high amplitude. Furthermore, several researchers have shown that increasing consonant vowel ratio and their distinction is one of the contributing factors for word intelligibility (Freyman & Newrbonne, 1989; House et al., 1965). But, participants with severe impairment in temporal processing showed small improvement. This could be because the amount of envelope enhancement employed in the present investigation was not being sufficient to enhance speech perception in these individuals.

Absence of auditory brain stem responses and abnormal cortical potentials in individuals with auditory dys-synchrony revealed poor representation of auditory signals in brainstem and cortical structures (Krause et al., 2000; Narne & Vanaja, 2008 b; Starr et al., 1996). It has been reported that enhancing the envelope by differential emphasis may improve representation of signals in cortical structures (Cunningham, Nicol, King, Zecker, & Kraus, 2002; Lorenzi, Micheyl, & Berthommier, 1995; Tallal et al., 1996). Enhancing the envelope may have improved representation of signals in cortical structure in turn improving speech perception.

To summarize, speech identification is significantly impaired in most individuals with auditory dys-synchrony when compared with normal hearing individuals. This may be a reflection of impaired ability to follow amplitude variations. Enhancing the envelope (amplitude variations) of the speech signal improved identification scores in a majority of the individuals with auditory dys-synchrony. The results of the present study suggest that employing envelope enhancement strategies in hearing instruments may provide a significant benefit to many individuals with auditory dys-synchrony.

5.2.2. Speech identification in noise

a) Speech identification of unprocessed speech

In the present study, speech identification scores in the presence of noise was more severely affected in individuals with auditory dys-synchrony when compared to listeners with normal hearing. Effect of noise on speech identification in individuals with auditory dys-synchrony was more for those with severe impairment when compared to those with mild impairment. Overall, these data appear similar to observations made in the earlier studies (Kraus et al., 2000; Starr et al., 1996; Zeng, & Liu, 2006) and are consistent with the subjective complaint reported by participants with auditory dys-synchrony.

Speech intelligibility reduces in the presence of background noise, partly because the noise reduces the modulations of speech envelope (Houtgast & Steeneken, 1985). In addition, the decline in intelligibility may also result from distortion of temporal fine structure and introduction of spurious envelope modulations as these modulations obscure or mask the modulations pattern of speech, and obliterates some of the cues for identification (Drullman 1995 a, b; Noordhoek & Drullman, 1997). Drullman (1995) reported that fine structure of speech (temporal dips) play an important role in speech perception in presence of background noise in individuals with normal hearing. This was supported by other studies evaluating the importance of envelope and fine structure in speech perception. It has been reported that subjects with normal hearing obtained 100 % scores in quiet when only envelope cues and 8 channels of spectral information (speech shaped noise) was provided, whereas in presence of noise, speech understanding reduced dramatically and was as low as 10 to 20 % when the S/N ratio was 0 dB (Stickney, Nie, & Zeng, 2004; Dorman, Loizou, & Tu, 1998). These investigators have attributed the reduced identification scores in adverse noise conditions to removal of fine structure information.

The exact mechanism underlining excessive detrimental effect of noise in individuals with auditory dys-synchrony is unclear, although it may be related to temporal processing impairment. One plausible explanation is that, individuals with auditory dys-synchrony are impaired in extracting envelope from speech signal even in quiet. Adding noise to the speech signal may exaggerate their problem in perceiving the envelope of speech (amplitude variations) because of reduction in modulation depth and addition of spurious modulations (Drullman 1995 a, b; Noordhoek & Drullman, 1997). This explanation would explicate severe degradation in speech intelligibility in the presence of background noise for participants with moderate and severe impairment.

In the present study, identification scores for individuals with mild impairment reduced by 35 to 40% in presence of noise. These scores are similar to the scores obtained by earlier investigators (Drullman, 1995b; Stickney, Nie and Zeng, 2004) in subjects with normal hearing for speech signals without fine structure information. These results suggest that probably the participants who obtained 85% to 100 % scores in quiet were able to get envelope cues but were unable to extract fine structure cues. Physiological studies have demonstrated that fine structure is represented by phase locking responses of nerve spikes (Joris & Yin, 1992). It is possible that participants with auditory dys-synchrony were unable to extract fine structure cues from speech signal as phase locking is disrupted (Buss, Hall, & Grose, 2004) and hence had acute difficulty in understanding speech in adverse noise conditions.

To summarize, speech perception in the presence of noise was more affected for individuals with auditory dys-synchrony when compared to those with normal hearing. Dramatic effect of noise on speech perception in individuals with auditory dys-synchrony may, be due to introduction of spurious envelope modulations by noise. Probably these modulations mask the modulations pattern of speech and obscures relevant speech modulations. In addition, temporal processing impairment may affect the ability to utilize any fine structure information from speech.

a) Identification of speech with envelope enhancement

The data of the present study demonstrated a significant advantage of envelope enhancement regardless of the listening conditions in individuals with auditory dys-synchrony. Improved performance in presence of noise was comparable to improvement observed with clear speech in noise for four participants with auditory dys-synchrony (Zeng & Liu, 2006). Among the participants with auditory dys-synchrony, those with mild degree of impairment showed greater improvement when compared to moderate impairment at all S/N ratios, whereas those with severe impairment did not show any improvement with envelope enhancement.

Amount of benefit derived from envelope enhancement in noise depends upon the ability to utilize the enhanced envelopes of speech signal. The ability to detect amplitude variations was significantly impaired in participants with auditory dys-synchrony. The results indicate that participants with mild and moderate impairment in temporal resolution would benefit more from envelope enhancement compared to participants with severe impairment in temporal resolution. But at 0 dB S/N ratio even participants with mild and moderate impairment in temporal resolution showed only 15 % to 30% identification scores indicating that the envelope enhancement does not improve communication in adverse listening conditions. Similar results have been observed in subjects with cochlear pathology. Apoux et al. (2004) found that individuals with cochlear hearing loss show a minimal improvement when only 16 Hz envelope is enhanced but a significant improvement is observed in identification scores when 0 to 256 Hz envelope is enhanced. In adverse noise conditions normal hearing listeners extract speech information from fine structure (Drullman, 1995; Liu, & Zeng, 2006) but individuals with auditory dys-synchrony are poor in extracting information from troughs or fine structure (Buss, Hall, & Grose, 2004). In the present study, only 2-32 Hz region of envelope was enhanced and probably enhancing only the slow modulations did not improve speech recognition in adverse listening conditions. It is not known whether enhancing wider modulation frequency range would have improved speech perception in adverse listening conditions. Further studies need to be carried out to investigate the usefulness of enhancing the wider range of modulation frequencies.

Thus, the results of the present study consistently demonstrate that speech perception was significantly affected in participants with auditory dys-synchrony, both in quiet and in presence of noise. Impairment of speech perception substantially depends upon the degree of impairment in temporal resolution. Further, results unequivocally demonstrate that speech perception was significantly improved in both quiet and in the presence of noise for those who are mild and moderate impairment in temporal resolution.

5.2.3. Identification of upward spectrally shifted speech with and without envelope enhancement

Spectral modification degraded the speech understanding in quiet. It was observed that 70 % of participants had identification scores of 0% and other 30% of subjects had only 10 to 15 % identification scores. Even upward spectral shift with envelope enhancement showed similar results. Normal hearing listeners also showed reduced identification scores with upward spectral shift and upward spectral shift with envelope enhancement. Scott and Assmann (2001) reported similar results in normal hearing individuals when low frequency spectral information was shifted to high frequencies in the speech signal.

The reduction in identification scores for speech with upward spectral shift may not be due to any artifacts introduced on the envelope of the signal. The deterministic effect of upward spectral shifting on speech understanding is probably because, shifting frequency information to higher frequencies will bring a complete change in frequency coding in the auditory system. It has been well established that low frequency signals are coded using phase locking and high frequency signals are coded using place coding (Moore, 1995). Also at higher frequencies, auditory filters get broader and broader leading to poor frequency resolution (Moore, 1995).

Spectral modification technique probably introduced frequency mismatch causing deterioration in performance. Studies have shown that with training, listeners with normal hearing were able to adapt to spectrally shifted speech and eventually compensate for deficits caused by frequency mismatch (Rosen et al., 1999). No training was given in the present study. A long-term training may be needed for the participants adapt to new processing strategy.

To conclude, speech identification is significantly impaired in a majority of the individuals with auditory dys-synchrony. This is probably a reflection of diminished temporal processing capabilities in individuals with auditory dys-synchrony. Enhancing the envelope of the speech signal improved identification scores of individuals with auditory dys-synchrony. The results of the present study suggest that utilizing envelope enhancement strategies in hearing instruments will probably provide a significant benefit to many individuals with auditory dys-synchrony. The upward spectral shifting of speech did not improve speech identification ability. A long-term training program may help subjects adapt to the new processed sound.

CHAPTER 6

Summary and Conclusions

Auditory dys-synchrony is a hearing disorder with unique perceptual consequences caused due to disrupted neural activity (Starr, Picton, Sininger, Hood, & Berlin, 1996). The disrupted neural activity seen in individuals with auditory dys-synchrony results in significant impairment in temporal processing and difficulty in speech understanding, that is disproportionate to the degree of hearing loss measured by pure-tone thresholds (Sininger & Oba, 2001; Zeng Oba, Grade, Sininger, & Starr, 1999). Psycho-acoustical assessment in individuals with auditory dys-synchrony showed a significant impairment in the ability to detect amplitude variations in continuous signal and this modulation threshold has good correlation has good correlation with speech identification scores. However, very few studies have investigated the modulation detection ability and their relation with speech identification scores.

The management of auditory dys-synchrony has been a challenging task for audiologists as the conventional amplification has shown limited success in these individuals (Berlin, Hood, Morlet, Rose, & Brashears, 2003). New signal processing strategies need to be developed based on their psychoacoustic data to help them understand speech in quiet as well as in the presence of noise. Studies have indicated that amplitude modulations are crucial for understanding speech in quiet and noise. Enhancing these amplitude modulations in speech may improve speech perception in individuals with auditory dys-synchrony. Research has indicated that envelope enhancement improved speech identification in participants with cochlear hearing loss (Apoux et al., 2004) and learning disability (Tallal et al., 1996). Data from psycho-acoustical studies has also shown poor discrimination of signals in low frequency. The energy in the low frequencies may not be useful and furthermore may cause unwanted masking. Zeng et al., (2005) suggested that transferring low frequency information to high frequencies may improve speech perception in individuals with auditory dys-synchrony. However, there are no published reports in this direction.

Objectives of the study

The present study was designed to investigate the following in individuals with auditory dys-synchrony and those with normal hearing:

1. Temporal modulation transfer function for sinusoidal amplitude modulated white noise.
2. Speech identification ability in quiet and its relation with temporal modulation transfer function.
3. Identification of speech with envelope enhancement, in quiet.
4. Speech identification ability in the presence of speech spectrum shaped noise and its relation with temporal modulation transfer function.
5. Identification of speech with envelope enhancement, in the presence of speech spectrum shaped noise.
6. Identification of speech with upward spectral (frequency) shifts, in quiet.
7. Identification of speech with upward spectral shift along with envelope enhancement, in quiet.

Forty individuals with auditory dys-synchrony and thirty individuals with normal hearing in the age range of 12 to 39 years participated in the study. The study was conducted in two phases. In the first phase, psycho-acoustical experiment was conducted to measure temporal resolution. This was measured using modulation detection task at different modulation frequencies. Modulation detection threshold was assessed through discrimination task, in which participant had to discriminate between an un-modulated and amplitude modulated noise. Modulation detection threshold was estimated for noise modulated at different modulation frequencies.

The second phase of the study involved three experiments. In the first experiment, speech identification ability was assessed in quiet for bi-syllabic words with and without envelope enhancement. The envelope enhancement was carried out using a program developed in MATLAB, where the speech signal was processed by a compression/expansion scheme. Second experiment involved assessing speech identification ability for bi-syllabic words with and without envelope enhancement presented in the presence of speech spectral shaped noise.

In the third experiment, identification of speech with upward spectral shift as well as speech with upward spectral shift along with envelope enhancement was assessed. For upward spectral shift the information in the frequency range of 0 to 8000 Hz was linearly shifted above 500 Hz. This signal processing strategy was implemented in MATLAB.

The data collected in two phases of the study were tabulated and appropriate statistical analyses were conducted. Analysis of the data warranted the following conclusions:

1. The temporal modulation transfer function resembles a simple low pass filter, in both listeners with normal hearing and those with auditory dys-synchrony.
2. The modulation detection threshold in individuals with auditory dys-synchrony is significantly higher than that observed in listeners with normal hearing.
3. The peak sensitivity and bandwidth of TMTF are also reduced in individuals with auditory dys-synchrony.
4. Understanding of speech is significantly poorer in individuals with auditory dys-synchrony when compared to listeners with normal hearing.
5. Identification scores for unprocessed speech in quiet show a significant high correlation with the modulation detection threshold, peak sensitivity and bandwidth. Regression analysis further showed that identification scores can be predicted from peak sensitivity.
6. In individuals with auditory dys-synchrony, identification scores for unprocessed speech reduce more dramatically in presence of speech spectrum shaped noise when compared to normal hearing listeners.
7. Enhancing the envelope of speech improved speech identification scores in individuals with auditory dys-synchrony. Individuals with moderate degree of

impairment show greater improvement in comparison to those with severe degree of impairment in temporal resolution.

8. Speech identification scores in the presence of noise also improve when the envelope of the speech is enhanced. The improvement for individuals with mild degree of impairment in temporal resolution is greater than that observed for moderate and severe degree of impairment.
9. In subjects with auditory dys-synchrony, speech understanding deteriorates with upward spectral shift and upward spectral shift with envelope enhancement.

Implications of the Study

1. Present study provides norms for modulation detection threshold at different modulation frequencies. These norms can be used as a reference against which individual with suspected auditory perceptual deficits can be compared to make a decision regarding temporal processing abilities.
2. The results suggests that temporal modulation transfer function can be used as a tool to assess the degree of impairment in understanding speech in individuals with auditory dys-synchrony.
3. The improved identification scores with envelope enhancement in quiet and the presence of noise are particularly important in the context of implementing the envelope enhancement strategies in hearing devices for the benefit of individuals with auditory dys-synchrony.

CHAPTER 6

References

American National Standards Institute (1987). Specifications for instruments to measure aural acoustic impedance and admittance (aural acoustic immittance). ANSI S3.39-1987-R1996. New York: American National Standards Institute.

American National Standards Institute (1999). Maximum permissible ambient noise for audiometric rooms. ANSI S3.1-1999. New York: American National Standards Institute.

Apoux, F., Crouzet, O., & Lorenzi, C. (2001). Temporal envelope expansion of speech in noise for normal-hearing and hearing-impaired listeners: effects on identification performance and response times. *Hearing Research, 153,* 123-131.

Apoux, F., Tribut, N., Debruille, X., & Lorenzi. C. (2004). Identification of envelope expanded sentences in normal-hearing and hearing-impaired *listeners. Hearing Research, 189, 13-24.*

Bacon, S.P & Viemeister, N.F. (1985). Temporal modulation transfer functions in normal-hearing and hearing-impaired subjects. *Audiology* 24, 117–134.

Berlin, C. I. (1999). Auditory Neuropathy: Using OAEs and ABRs from screening to management. Seminars in Hearing, 20, 307-315.

Berlin, C. I., Hood, L. J., Hurely, A., & Wen, H. (1996). Hearing aids: Only for hearing impaired patients with abnormal otoacoustic emissions. In C.I. Berlin (Ed.), *Hair cells and hearing aids* (pp 99-111). San Diego: Singular publishing group.

Berlin, C. I., Hood, L. J., Morlet, T., Rose, K., & Brashears, S. (2003). Auditory Neuropathy/Dys-synchrony: Diagnosis and Management. *Mental Retardation and Developmental Disabilities Research Reviews, 9*, 225-231.

Berlin, C. I., Hood, L. J., Morlet, T., Wilensky, D., John, P., Montgomery, E., Thibodaudx, M. (2005). Absent or Elevated Middle Ear Muscle Reflexes in the Presence of Normal Otoacoustic Emissions: A Universal Finding in 136 Cases of Auditory Neuropathy/Dys-synchrony. *Journal of American Academy of Audiology, 16,* 550-557.

Berlin, C. I, Morlet, T., & Hood, L. J. (2003). Auditory neuropathy/dyssynchrony: Its diagnosis and management. *The Pediatric Clinics of North America, 50, 331-340.*

Berlin, C.I., Bordelon, J., St John, P., Wilensky, D., Hurleym, A., Kluka, E., & Hood, L.J. (1998). Reversing click polarity may uncover auditory neuropathy in infants. *Ear and Hearing, 19,* 37-47.

Berlin, C.I., Hood, L., & Rose, K. (2001). On renaming auditory neuropathy as auditory dys-synchrony. *Audiology Today, 13,* 15-17.

Braida, L. D., Durlach, N. I., Lippman, R. P., Hicks, B. L., Rabinowitz, W. M., & Reed, C. M. (1978). Hearing Aids—A review of past research on linear amplification, amplitude compression, and frequency lowering. ASHA Monographs Number 19, American Speech-Language-Hearing Association, Rockville, MD.

Buss, E., Hall, J. W. III., & Grose, J. H. (2004). Temporal Fine-Structure Cues to Speech and Pure Tone Modulation in Observers with Sensorineural Hearing Loss. Ear and Hearing*, 25,* 242-250.

Byrne, D., & Dillon, H. (1986). The National Acoustics Laboratories (N.A.L.) new procedure for selecting the gain and frequency response of a hearing aid*. Ear and Hearing, 7,* 257–265.

Carhart, R., & Jerger.J. (1959). Preferred method for clinical determination of pure tone thresholds. Journal of Speech and Hearing Disorders, 24, 330-345.

Clarckson, P. M., & Bahgat, S. F. (1991). Envelope expansion methods for speech enhancement. *Journal of the Acoustical Society of America, 89*, 1378-1382.

Cone-Wesson, B., Rance, G., & Sininger, Y. (2001). Amplification and rehabilitation strategies for patients with auditory neuropathy. In Y. Sininger, A. Starr (Eds.), *Auditory neuropathy: A new perspective on hearing disorders* (pp.233-248). Canada: Singular-Thomson learning.

Cunningham, J., Nicol, T., King, C., Zecker, S.G., & Kraus, N. (2002). Effects of noise and cue enhancement on neural responses to speech in auditory midbrain, thalamus and cortex. *Hearing research, 169*, 97-111.

Dau, T., Kollmeier, B., & Kohlrausch, A. (1997). Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carrier. *Journal of Acoustical society of America, 102*, 2892-2905.

Davis, H., & Hirsh, S. K. (1979). A slow brainstem response for low frequency audiometry. *Audiology, 18,* 445-465.

Deltenre, P., Mansbach, A. L., Bozet, C., Christiaens, F., Barthelemy, P., Paulissen, D. & Renglet, T. (1999). Auditory neuropathy with preserved cochlear microphonics and secondary loss of otoacoustic emissions. *Audiology 38,* 187-95.

Dorman, M., Loizou, P., & Tu, Z. (1998). The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processor with 6–20 channels. *Journal of the Acoustical Society of America, 104*, 3583– 3585.

Drullman, R. (1995 a). Temporal envelope and fine structure cues for speech intelligibility. *Journal of the Acoustical Society of America, 97,* 585–592.

Drullman, R. (1995 b). Speech intelligibility in noise: Relative contribution of speech elements above and below the noise level. *Journal of the Acoustical Society of America, 98,* 1796-1798.

Drullman, R., Festen , J.M., & Plomp, R. (1994a). Effect of temporal envelope smearing on speech reception. *Journal of the Acoustical Society of America, 95,* 1053-1064.

Drullman, R., Festen, J. M., & Plomp, R. (1994b). Effect of reducing slow temporal modulations on speech reception. *Journal of the Acoustical Society of America, 95,* 2670–2680.

Drullman, R., Festen, J. M., & Houtgast, T. (1996). Effect of temporal modulation reduction on spectral contrasts in speech. *Journal of the Acoustical Society of America, 99*, 2358–2364.

Eddins, D. A. (1993). Amplitude modulation detection of narrowband noise: effect of absolute bandwidth and frequency region. *Journal of Acoustical Society of America, 93*, 470-479.

Fuchs, P. A., Glowatzki, E., & Moser, T. (2003). The afferent synapse of cochlear hair cells. *Current Opinions in Neurobiology, 13,* 452-458.

Fu, Q-J., Shannon, R. V. (1999b). Recognition of spectrally degraded and frequency shifted vowels in acoustic and electric hearing. *Journal of the Acoustical Society of America, 105,* 1889-1900.

Fu, Q-J., Shannon, R. V., & Galvin, J.J III. (2002). Perceptual learning following changes in the frequency-to-electrode assignment with the Nucleus-22 cochlear implants. *Journal of Acoustical society of America, 112*, 1664-1674.

Gibson, W.P., & Sanli, H. (2007). Auditory neuropathy: an update. *Ear and Hearing, 28 (2 supplement), 102S-106S.*

Gilbert, G., & Lorenzi, C. (2006). The ability of listeners to use recovered envelope cues from speech fine structure. *Journal of the Acoustical Society of America, 119,* 2438–2444

Giraud, A., Lorenzi, C., Ashburner, J., Wable, J., Johnsrude, I., Frackowiak, R., & Kleinschmidt, A. (2000). Representation of the Temporal Envelope of Sounds in the Human Brain. *Journal of Neurophysiology, 84,* 1588-1598.

Gordon-Salant, S. (1987). Effects of acoustic modifications on consonant recognition by elderly hearing-impaired subjects. *Journal of the Acoustical Society of America*, *81*, 1199–1202.

Gordon-Salant, S., & Fitzgibbons, P. J. (1993). Temporal factors and speech recognition performance in young and elderly listeners. *Journal of Speech Language and Hearing Research, 31,* 1276-1285.

Hallpike, C. S., Harriman, D. G., and Wells, C. E. (1980). A case of afferent neuropathy and deafness. *Journal of Laryngology and Otology, 94,* 945-964.

Harrison, R. V. (1998). An animal model of auditory neuropathy. *Ear and Hearing, 19,* 355-361.

Hood, L. J., Berlin, C. I., Bordelon, J. and Rose, K. (2003). "Patients with auditory neuropathy/dys-synchrony lack efferent suppression of transient evoked otoacoustic emissions." *Journal of American Academy of Audiology, 14,* 302-313.

House, A. S., William, C. E., Hecker, M. H. L., & Kryter, K. D. (1965). Articulation-testing methods: Consonantal differentiation with a closed-response set. *Journal of Acoustical Society of America*, ***37***, 158–166.

Houtgast, T., & Steeneken, H. J. M. (1985). A review of the MFT concept in room acoustics and its use for estimating speech intelligibility in auditoria. *Journal of the Acoustical Society of America, 77,* 1069-1077.

International Standards Organization (1981). Acoustics-reference zero for calibration of audiometric equipment, part7: Reference equivalent threshold for under free-field and diffuse listening conditions. ISO 398-7. Geneva, Switzerland: International Standards Organization.

International Standards Organization (1994). Acoustics-reference zero for calibration of audiometric equipment, part2: Reference equivalent threshold sound pressure level for pure-tones and bone vibrator. ISO 398-3. Geneva, Switzerland: International Standards Organization.

International Standards Organization (1998). Acoustics-reference zero for calibration of audiometric equipment, part1: Reference equivalent threshold sound pressure level for pure-tones and supra aural head phones. ISO 398-1-1998. Geneva, Switzerland: International Standards Organization.

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

Kennedy, E., Levitt, H., Neuman, C. A., & Weiss, M. (1998). Consonant–vowel intensity ratios for maximizing consonant recognition by hearing-impaired listeners. *Journal of Acoustical society of America, 103,* 1098-1114.

Kewley-Port, D. (1983). Time-varying features as correlates of place of articulation in stop consonants. *Journal of Acoustical society of America, 73,322-335.*

Kraus, N. (2001). Auditory neuropathy: An historical and current perspective. In Y. Sininger, & A. Starr (Eds.), *Auditory neuropathy: A new perspective on hearing disorder* (pp. 1-14). Canada: Singular publishing group.

Kraus, N., Bradlow, M. A., Cunningham, C. J., King, C. D., Koch, D. B., Nicol, T. G., Mcgee, T. J., Stein, L. K., & Wright, B. A. (2000). Consequences of neural asynchrony: A case of auditory neuropathy. *Journal of the Association for Research in Otolaryngology, 01, 33-45.*

Kumar, A. U. (2006). Perception of some temporal parameter of speech in individuals with auditory dys-synchrony. Unpublished doctoral thesis, University of Mysore, Mysore.

Kumar, A. U., & Jayaram, M. (2005). Auditory processing in individuals with auditory neuropathy. *Behavioral and Brain Functions,* ***1:****21.*

Kumar, A. U., & Jayaram, M. (2006). Prevalence and audiological characteristics in individuals with auditory neuropathy/auditory dys-synchrony. *International Journal of Audiology, 45,* 360-366.

Liu, S., & Zeng, F. G. (2006). Temporal properties in clear speech perception. *Journal of Acoustical society of America, 120*, 424-432.

Levitt, H. (1971). Transformed up–down methods in psychoacoustics. *Journal of Acoustical Society of America, 49*, 467–477.

Lorenzi, C., Berthommier, F., Apoux, F., & Bacri, N. (1999). Effect of envelope *expansion on speech recognition. Hearing Research, 136,* 131-138.

Lorenzi, C., Wable, J., Moroni, C., Derobert, C., Frachet & Belin, C. (2000). Auditory temporal envelope processing in a patient with left-hemisphere damage. *Neurocase, 6,* 231-244.

Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., Moore, B. C. J. (2007). Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. *Proceedings of the National Academy of Sciences of United States of America, 103,* 18866–18869.

Lynne, A., Werner, & Gray, L. (1998). Behavioral studies of hearing development. In E. Rubel, A. Popper & R.R. Fay (Eds.), *Development of auditory system* (pp.15-47). NY: Springer.

Mason, J. G., De Michele, A., Stevens, C., Ruth, R. A., & Hashisaki, G.T. (2003). Cochlear implantation in patients with auditory neuropathy of varied etiologies. *Laryngoscope, 113,* 45-49.

Mayadevi. (1974). *Development and standardization of a common speech discrimination test for Indians*. An unpublished dissertation, University of Mysore, Mysore.

McDermott, H. J., Dorkos, V. P., Dean, M. R., & Ching, T. Y. (1999). Improvements in speech perception with use of the AVR TranSonic frequency-transposing hearing aid. *Journal of Speech, Language, and Hearing Research, 42*, 1323–1335.

Moore, B. C. J. (1973). "Frequency difference limens for short-duration tones," *Journal of Acoustical Society of America*, 54, 610-619.

Moore, B. C. J. (1995). Speech perception in people with cochlear damage. *In Perceptual consequences of cochlear damage.* (pp147–172). Oxford: Oxford University Press.

Moore, B. C. J. (1996). Perceptual consequences of cochlear hearing loss and their implications for the design of hearing aids. *Ear and Hearing, 17,* 133–160.

Moore, B. C. J. (2003). Speech processing for hearing impaired: successes, failure, and implications for speech mechanism. *Speech communication, 41*, 81-91.

Nabelek, A. K., & Robinson, P. K. (1982). Monaural and binaural speech perception in reverberation for listeners of various ages. *Journal of Acoustical Society of America, 71,* 1242-1248.

Nagafuchi, M. (1976). Intelligibility of distorted speech sounds shifted in frequency and time in normal children. *Audiology, 15*, 326-337.

Nagarajan, S., Wang, X., Merzenich, M.M., Schreiner, C.E., Johnston, P., Jenkins, W.M., Miller, S., Tallal, P. (1998). Speech modification algorithms used for training language learning-impaired children. *IEEE Transactions on Rehabilitation Engineering, 6,* 257-268.

Narne, V. K, Manjula, P & Vanaja.C. S. (2007). Speech identification with temporal envelope and fine structure cues in listeners with cochlear hearing loss. *Proceeding of Frontiers research in Speech and Music,* 130-134.

Narne, V. K., & Vanaja, C. S. (2008a). Effect of envelope enhancement on speech perception in individuals with auditory neuropathy. *Ear and Hearing, 29*, 45-53.

Narne, V. K., & Vanaja, C. S. (2008b). Speech Perception and cortical potentials in individuals with auditory neuropathy. *Behavioral and Brain Functions* 4:15.

Neary, T. (1989). Static, dynamic, and relational properties in vowel perception. *Journal of Acoustical Society of America. 85,* 2088–2113.

Parent, T. B., Chmiel, R., & Jerger, J. (1997). Comparison of performance with frequency transposition hearing aids and conventional hearing aids. *Journal of the American Academy of Audiology, 8*, 355–365.

Peterson, A., Shallop, J., Driscoll, C., Breneman, A., Babb, J., Stoeckel, R., Fabry, L. (2003). Outcomes of cochlear implantation in children with auditory neuropathy. *Journal of American Academy of Audiology, 14*, 188-201.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (1985). Speaking clearly for the hard of hearing. I: Intelligibility differences between clear and conversational speech. *Journal of Speech Language Hearing Research,* 28, 96–103.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (1986). Speaking clearly for the hard of hearing II. Acoustic characteristics of clear and conversational speech. *Journal of Speech Language Hearing Research, 29*, 434–446.

Pickett, J.M. (1999). *The acoustics of speech communication*. Needham Heights, MA: Allyn & Bacon.

Plomp, R. (1983). The negative effects of amplitude compression in multichannel hearing aid in light of the modulation-transfer function. *Journal of Acoustical Society of America, 83,* 2322-2327.

Price, P. J., & Simon, H .J. (1984). Perception of temporal differences in speech by “normal hearing” adults: Effects of age and intensity. *Journal of Acoustical society of America, 76*, 405-410.

Ramirez, J., & Mann, V. (2005). Using auditory-visual speech to probe the basis of noise-impaired consonant-vowel perception in dyslexia and auditory neuropathy. *Journal of Acoustical Society of America, 118,* 1122-1133.

Rance, G. (2005). Auditory neuropathy/dys-synchrony and its perceptual consequences. *Trends in amplification, 9,* 1-43.

Rance, G., & Barker, E. J. (2008). Speech perception in children with auditory neuropathy/dyssynchrony managed with either hearing aids or cochlear implants. *Otology neurotology, 29,* 179-182.

Rance, G., Barker, E., Mok, M., Dowell, R., Ricon, A., & Garratt, R. (2007). Speech perception in noise for children with auditory neuropathy /dys-synchrony type hearing loss. *Ear and Hearing, 28*, 351-360.

Rance, G., Beer, D. E., Cone-Wesson, B., Shepherd, R. K., Dowell, R. C., King, A.M., Rickards, F.W., & Clark, G.M. (1999). Clinical findings for a group of infants and young children with auditory neuropathy. *Ear and Hearing, 20,* 238-252.

Rance, G., Cone-Wesson, B., Wunderlich, J., & Dowell, R. (2002). Speech perception and cortical event related potentials in children with auditory neuropathy. *Ear and Hearing, 23,* 239-253.

Rance, G., McKay, C., Grayden, D. (2004). Perceptual characterization of children with auditory neuropathy. *Ear and Hearing, 25*, 34-46.

Reed, C. M., Hicks, B. L., Braida, L. D., and Durlach, N. I. (**1983**). ‘‘Discrimination of speech processed by low-pass filtering and pitch-invariant frequency lowering,’’ *Journal of Acoustical Society of America, 74*, 409–419.

Reed, C. M., Schultz, K. I., Braida, L. D., and Durlach, N. I. (**1985**). Discrimination and identification of frequency-lowered speech in listeners with high-frequency hearing impairment. *Journal of Acoustical Society of America. 78*, 2139-2141.

Revoile, S. G., Holden-Pitt, L., Edward, D., Pickett, J. M., & Brandt, F. (1987). Speech-cue enhancement for the hearing impaired: amplification of burst/murmur cues for improved perception of final stop voicing. *Journal of rehabilitation research and development, 24,* 207-216.

Rosen, S. (1992). Temporal information in speech: acoustic, auditory and linguistic aspects. *Philosophical Transactions: Biological Sciences, 336*, 367-373.

Rosen, S., Faulkner, A., & Wilkinson, L. (1999). Adaptation by normal listeners to upward spectral shifts of speech: implications for cochlear implants. *Journal of the Acoustical Society of America, 106*, 3629-3636.

Runge-Samuelson, C. L., Drake, S., & Wackym, A. P. (2008). Quantitative analysis of electrically evoked auditory brainstem responses in implanted children with auditory neuropathy/dys-synchrony. *Otology and Neurology, 29*,174-178.

Scott, J. M., & Assmann, P. F. (2001). Intelligibility of frequency-shifted speech. Poster presentation at the 141st Meeting of the Acoustical Society of America, June 4, Chicago.

Sek. A., & Moore, B. C. J. (1995). Frequency discrimination as a function of frequency, measured in several ways. *Journal of Acoustical Society of America,97,* 2479-2486.

Shallop, J. K. (2002). Auditory neuropathy/auditory dys-synchrony in adults and children. *Seminars in Hearing, 23*, 215-223.

Shallop, J. K., Peterson, A., Facer, G. W., Fabry, L. B., & Driscoll, C. L. (2001). Cochlear implants in five cases of auditory neuropathy: Postoperative findings and progress. *Laryngoscope, 111*, 555-562.

Shannon, R.V., Zeng, F. G., Kamath, V., Wygonisk, J., & Ekelid, M. (1995). Speech recognition with preliminarily temporal cues. *Science, 270*, 303-304.

Sininger, Y., & Oba, S. (2001). Patients with auditory neuropathy: who are they and what can they hear? In Y. Sininger, A. Starr, *Auditory neuropathy: A new perspective on hearing disorders* (pp. 15-35). Canada: Singular-Thomson learning.

Smith, Z. M., Delgutte, B., & Oxenham, A. J. (2002). Chimaeric sounds reveal dichotomies in auditory perception. Nature, 416, 87–90.

Starr, A., McPherson, D., Patterson, J., Don, M., Luxford, W., Shannon, R., Sininger, Y., Tonakawa, L., & Waring, M. (1991). Absence of both auditory evoked potential and auditory percepts dependent on timing cues. *Brain, 114,* 1157-1180.

Starr, A., Michalewski, H. J., Zeng, F. G., Brooks, S. F., Linthicum, F., Kim, C.S., Winnier, D., & Keats, B. (2003). Pathology and physiology of auditory neuropathy with a novel mutation in the MPZ gene. *Brain, 126*, 1604-1619.

Starr, A., Picton, T. W., & Kim, R. (2001). Pathophysiology of auditory neuropathy. In: Y. Sininger, & A. Starr (Eds.), *Auditory neuropathy: A new perspective on hearing disorder* (pp. 67-82). Canada: Singular publishing group.

Starr, A., Picton, T. W., Sininger, Y., Hood, L., & Berlin, C. I. (1996). Auditory neuropathy. *Brain, 119,* 741-753.

Starr, A., Sininger, Y. S., Nguyen, T., Michaelewski, H. J., Oba, S. & Abdala, C. (2001). Cochlear receptor (microphonic, summating potentials, and otoacoustic emissions) and auditory pathway (auditory brainstem potentials) activity in auditory neuropathy. *Ear and Hearing, 22,* 91-99.

Starr, A., Sininger, Y. S., & Praat (2000). Varieties of Auditory neuropathy. *Journal of Basic Clinical Physiology and Pharmocology, 11*, 215-229.

Starr, A., Sininger, Y. S., Winter, M., Derebery. M. J., Oba. S., & Michalewski, H. J. (1998). Transient deafness due to temperature-sensitive auditory neuropathy. *Ear and Hearing, 19,* 169-179.

Stickney, G. S., Zeng, F. G., Litovsky, R., & Assmann, P. F. (2004). Cochlear implant speech recognition with speech maskers. *Journal of Acoustical society of America 116,* 1081–1091.

Studebaker, G. A. (1985). "A "rationalized" arcsine transform". *Journal of Speech Language and Hearing Research, 28*, 455-462.

Tallal, P., Miller, S. T., Bedi, G., Byma, G., Wang, X., Nagarajan, S., Schreiner, C., Jenkins, W., et al. (1996). Language comprehension in language learning impaired children improved with acoustically modified speech. *Science, 271,* 81-84.

Tallal, P., & Stark, R. E. (1981). Speech acoustic-cue discrimination abilities of normally developing and language impaired children. *Journal of Acoustical Society of America, 69*, 568-574.

Tang, T. P., Mcpherson, B., Yuen, K. C., Wong, L. L., & Lee, J. S. (2004). Auditory neuropathy/auditory dys-synchrony in schoolchildren with hearing loss: Frequency of occurrence. *International Journal of Pediatric Otolaryngology,* 168, 175-183.

Trautwein,P. G., Sininger, Y. S., & Nelson, R. (2000). Cochlear implantation of auditory neuropathy. *Journal of American Academy of Audiology, 11*, 309-315.

Turner, C. W., & Hurtig, R. R. (2002). Proportional frequency compression of speech for listeners with sensorineural hearing loss. *Journal of the Acoustical society of America, 106,* 877-886.

Turner, C. W., Souza, P. E., & Forget, L. N. (1995). Use of temporal envelope cues in speech perception by normal hearing and hearing impaired listeners. *Journal of Acoustical Society of America, 97*, 2568-2576.

Uchansky, R. M., Choi, S., Braida, L. D., Reed, C. M., & Durlach, N. I. (1996). Speaking clearly for the hard of hearing IV: Further studies of role of speaking rate. *Journal of Speech and Hearing Research, 39,* 494-509.

Van Tasell, D. J., Soli, S. D., Kirby, V. M., & Widin, G. P. (1987). Speech waveform envelope cues for consonant recognition. *Journal of Acoustical society of America, 82,* 1152-1161.

Van Tasell, D. J. (1993). Hearing loss, speech, and hearing aids. *Journal of Speech and Hearing Research, 36*, 228-224.

van Wieringen, A., & Pols, L.C.W. (2006). Perception of highly dynamic properties in speech. In S. Greenberg, & W.A. Ainsworth (Eds) *Listening to Speech: An auditory perspective* (pp. 21-38). London: Lawrence Erlbaum Associates, Publishers.

Vanaja, C. S., & Jayaram, M. (2003). Sensitivity and specificity of audiological tests in differential diagnosis of auditory disorders. Unpublished project, All India Institute of Speech and Hearing, Mysore, India.

Viemeister, N. F. (1979). Temporal modulation transfer functions based upon modulation thresholds. *Journal of Acoustical society of America 66*, 1364–1380.

Waxman, S. G. (1977). Conduction in myelinated, unmyelinated, and demyelinated fibers. *Archives of Neurology, 34,* 585-589.

Yathiraj, A., & Vijayalakshmi, C.S. (2005). Auditory memory test. A test developed at the Department of Audiology, AIISH, Mysore.

Yellin, M. W., Jerger, J., &Fifer, R. C. (1989). Norms for disproportionate loss in speech intelligibility. *Ear and Hearing, 10,* 231-234.

Zeng, F. G., Oba, S., Grade, S., Sininger, Y., & Starr, A. (1999). Temporal and speech processing deficits in auditory neuropathy. *Neuro Report, 10*, 3429-3435.

Zeng, F. G., Oba, S., Garde, S., Sininger, Y., & Starr, A. (2001). Psychoacoustics and speech perception in auditory neuropathy. In: Y. Sininger, & A. Starr (Eds.), *Auditory neuropathy: A new perspective on hearing disorder* (pp. 141-164). Canada: Singular publishing group.

Zeng, F. G., Kong, Y. Y., Michalewski, H. J., & Starr, A. (2005). Perceptual consequences of disrupted auditory nerve activity. *Journal of Neurohysiol, 93,* 3050-3063.

Zeng, F. G., Nie, K., Liu, S., Stickney, G., Del Rio, E., Kong, Y. Y., & Chen, H. (2004). "On the dichotomy in auditory perception between temporal envelope and fine structure cues," *Journal of Acoustical Society of America. 116,* 1351– 1354.

Zeng, F. G., & Liu, S. (2006). Speech perception in individuals with auditory neuropathy. *Journal of Speech Language and Hearing Research, 49,* 367-380.

Effect of Envelope Enhancement on Speech Perception in Individuals with Auditory Neuropathy

Vijaya Kumar Name and C. S. Vanaja

Objectives: To investigate the ability to identify consonant-vowel (CV) stimuli by individuals with auditory neuropathy (AN) when the envelope of the speech signal is enhanced byl5 dB for different modulation bandwidths (3 to 10 Hz; 3 to 20 Hz; 3 to 30 Hz; 3 to 60 Hz).

Design: Eight individuals with auditory neuropathy whose pure-tone thresholds ranged from 30 to 70 dB HL participated in the present study. Speech material consisted of five lists of CV, one unprocessed and four with envelope enhancement. The magnitude of envelope enhancement was 15 dB and the bandwidths used were 3 to 10 Hz, 3 to 20 Hz, 3 to 30 Hz, and 3 to 60 Hz.

Results: Speech identification scores improved when the envelope of the speech was enhanced. The improvement was greater for the broader bandwidth (3 to 30 Hz) conditions when compared with the smaller bandwidth (3 to 10 Hz and 3 to 20 Hz) conditions. In the unprocessed condition, manner of articulation was transmitted better than voicing and place of articulation. In the envelope-enhanced conditions, cues for manner and place of articulation were transmitted better than voicing.

Conclusions: Envelope enhancement using digital techniques improves speech perception by individuals with auditory neuropathy, and it may be a viable option for the rehabilitation of these individuals. However, the magnitude of envelope enhancement of speech required for maximal improvement of speech perception is yet to be determined.

(Ear & Hearing 2008;29;4&-53)

Auditory neuropathy (AN) is a hearing disorder in which cochlear amplification is spared but processing of the auditory signal at eighth nerve and brain stem level is abnormal. By clinical definition, individuals with AN have normal otoacoustic emissions (OAEs) and/or cochlear microphonics (CM), but exhibit absent or severely abnormal auditory brain stem responses (ABR) and middle ear reflex without any space-occupying lesion (Starr, Picton, Sininger, Hood, & Berlin 1996). Hearing sensitivity may range from normal hearing to profound hearing impairment (Sininger, & Oba, 2001). A majority of the individuals with AN have low-frequency hearing loss with disproportionately poor speech recognition scores for the degree of hearing loss (Starr et al., 1996; Zeng, Oba, Garde, Sininger, & Starr, 1999).

Psycho-acoustical studies in individuals with AN have shown that they have poor temporal resolution (Starr et al., 1996; Zeng et al., 1999). A majority of the investigators (Kumar & Jayaram, 2005; Ranee, McKay, & Grayden, 2004; Starr et al., 1996; Zeng et al., 1999) have investigated temporal resolution in these individuals using the Temporal Modulation Transfer Function (TMTF). Normal-hearing listeners have low threshold for slow modulations and threshold increases as the modulation rate is increased in the TMTF task (Viemeister, 1979). Individuals with AN demonstrate impaired sensitivity to both slow and fast temporal modulations but processing of the faster modulations is affected more than those of the slower modulations (Kumar, & Jayaram, 2005; Ranee, McKay, & Grayden 2004; Zeng et al., 1999). A good correlation has been reported between the TMTF threshold and speech perception scores in individuals with AN (Kumar, & Jayaram, 2005; Ranee, McKay, & Grayden 2004; Zeng et al., 1999).

Zeng et al. (1999) edited temporal envelope modulations of speech to simulate processing of speech in individuals with AN and presented it to listeners with normal hearing. It was found that speech recognition abilities of individuals with normal hearing for the edited speech samples were similar to those seen in individuals with AN for normal speech. From these studies it can be inferred that the poor speech identification in individuals with AN is mainly due to reduced ability to follow the envelope (amplitude modulations) of the speech signal.

Rehabilitative measures for individuals with AN are an open question for audiologists. A number of investigators have reported that hearing aids are of limited use for these participants (Berlin, Morlet, & Hood, 2003; Ranee, Cone-Wesson, Wunderlich, & Dowell, 2002; Starr et al., 1996). The current hearing aid technology does not enhance the temporal envelope of the signal to compensate for temporal processing deficits in individuals with AN. Furthermore, conventional hearing aids reduce the amplitude fluctuations when a nonlinear amplitude-compression circuit is used (van Tassel, 1993) and this may result in

All India Institute of Speech and Hearing, Manasagangotri, Mysore, Karnataka, India.

0196/0202/08/2901-0045/0 • Ear & Hearing • Copyright © 2008 by Lippincott Williams & Wilkins • Printed in the U.S.A.

TABLE 1. Demographic and audiological profile of participants

Subject no.	Age/Sex	Pure-tone average in dB HL (right ear)	Pure-tone Average in dB HL (left ear)	Speech identification scores in sound filed	ABR	OAE	Acoustic reflex	Peripheral neuropathy
1	15 yr/M	51.1	56.1	70%	Absent	Present	Absent	No
2	11 yr/M	53.3	60.0	30%	Absent	Present	Absent	No
3	28 yr/F	63.6	70.0	0%	Absent	Present	Absent	No
4	10 yr/M	45.6	33.3	40%	Absent	Present	Absent	No
5	15 yr/F	80.0	60	0%	Absent	Present	Absent	No
6	18 yr/F	**60.0**	77.1	25%	Absent	Present	Absent	No
7	18 yr/F	**41.6**	43.3	15%	Absent	Present	Absent	No
8	17 yr/F	45	53.3	35%	Absent	Present	Absent	No

deterioration of performance in these individuals (Ranee, Cone-Wesson, Wunderlich, &, Dowell, 2002; Starr et al., 1996). Probably enhancement of amplitude fluctuations may improve speech perception in individuals with AN.

The speech signal has prominent low-frequency amplitude modulations, which are termed as envelope of the speech signal, and these envelopes provide important cues for consonant recognition (Rosen, 1992). The investigations (Drullman, Festen, & Plomp, 1994; Rosen, 1992; Shannon, Zeng, Kamath, Wyognisk, & Ekelid 1995) examining the importance of envelope in the speech signal have shown that the envelope cues in the frequency region of 2 to 50 Hz are sufficient for perception of manner of articulation and voicing, whereas perception of place of articulation requires envelope as well as spectral cues. Poor speech perception in noise by individuals with normal hearing and cochlear hearing loss is mainly attributed to the degradation caused by noise in processing the low-frequency modulations of the speech signal (Houtgast, & Steenken, 1985). It has been observed that individuals with normal hearing, cochlear hearing loss, and learning impairment show improved speech perception in noise when the envelope cues in speech signal are enhanced (Apoux, Tribut, Debruille, & Lorenzi, 2004; Lorenzi, Berthommier, Apoux, & Bacri 1999; Tallal et al., 1996). As poor speech perception by individuals with AN is mainly attributed to the reduced sensitivity to follow envelope cues in speech, we hypothesized that enhancing the envelope of the speech signal will improve the speech perception of individuals with AN. The present study was designed to examine the improvement in identification of monosyllable by individuals with AN when the depth of modulation of the speech envelope is enhanced byl5 dB for different modulation bandwidths.

METHODS

Participants

Eight participants (three male and five female), who had been diagnosed as AN, participated in the present study. Table 1 shows the demographic and audiological profile of these participants. The age of the participants ranged from 12 to 30 yr with a mean age of 19 yr. All the participants were native speakers of Kannada, a Dravidian language spoken in a southern state of India. Pure-tone average of these individuals ranged from 30 to 70 dB HL. Open-set speech recognition in quiet ranged from 0 to 70% correct with a median score of 30%, much lower than that expected by individuals with moderate pure-tone hearing loss (Vanaja & Jayaram, 2003, Reference Note 1). Transient evoked otoacoustic emissions were robust in all the participants. Middle ear acoustic reflexes (both ipsilateral and contralateral) and auditory brain stem responses were absent in all the participants. None of the participants showed improvement in speech recognition scores with a conventional linear gain hearing aid. Peripheral neuropathy was ruled out by neurological evaluation.

Speech Test Material

Speech material consisted of 15 consonant vowel (CV) /p, b, t, d, k, g, s, 1, r, m, n, t, d, /, f/. A 25-yr-old male native speaker of Kannada uttered these syllables in isolation. The stimuli were recorded digitally on a data acquisition system using at 44.1-kHz sampling frequency and a 16-bit analogue-to-digital converter.

PRAAT software (Version 4.4.12, developed in 2005, by the Institute of Phonetic Science, University of Amsterdam, Netherlands) was used for enhancing the envelope of the speech signal by a magnitude of 15 dB for different envelope modulation bandwidths. The bandwidths used were 3 to 10

TABLE 2. Classification of consonants by phonetic features

	pa	ba	ta	da	ta	**da**	**ka**	**ga**	**la**	**ra**	**ma**	**na**	**sa**	**Ja**	tfa
Voicing	0	1	0	**1**	0	**1**	**0**	**1**	**1**	**1**	**1**	**1**	**0**	**0**	**0**
Place	**1**	**1**	3	3	2	**2**	**5**	**5**	**3**	**3**	**1**	**3**	**3**	**4**	**4**
Manner	1	**1**	**1**	1	1	**1**	**1**	**1**	**5**	**6**	**3**	**3**	**2**	**2**	**4**

Voicing coding: Voiced = 1, voiceless = 0. Place of articulation coding: 1 -= labile, 2 = inter dental, 3 = alveolar, 4 - palatal, 5 - velar. Manner articulation coding: 1 •= plosive, 2 ~ fricative, 3 - nasal, 4 - affhcatrve, 5 = lateral. 6 = retroflex.

Hz, 3 to 20'Hz, 3 to 30 Hz, and 3 to 60 Hz. These frequencies were selected since, previous research has shown that normal hearing individuals require envelope modulation of 20 Hz or greater for correct identification of syllables and the performance of individuals with cochlear hearing loss improve with enhanced modulations for larger bandwidths (Apoux, Tribut, Debruille, & Lorenzi, 2004). The procedure used in the software for envelope enhancement is given in Appendix I.

In total, speech material consisted of five lists of CV, one unprocessed and four with envelope enhancement. All the lists were normalized for intensity using Audiolab software (version 2, Voice and Speech Systems, Bangalore).

Procedure

The participants were seated in an acoustically treated room where the ambient noise levels were within permissible limits (ANSI, 1999). The speech stimuli were played in a PC at 44.1-kHz sampling rate and routed to a calibrated (ANSI, 1996) diagnostic audiometer (MA-53). Participants received the signal from the loudspeaker of the audiometer kept at a distance of one meter at 0° azimuth. The presentation level of the stimulus was 40 dB SL (re: Speech Recognition Threshold). The participants were instructed to repeat the speech tokens heard by them.

Speech identification for both unprocessed and processed CV syllables was tested for all the participants. Each list, unprocessed and envelope-enhanced for different bandwidths, was presented 10 times to each participant. The order of the lists and presentation of CV syllables were randomized across participants and trials.

Data Analysis

The percentage of correct identification of consonants in different conditions was determined and Wilcoxon signed-rank test was performed to test the significance of difference between the conditions. Friedman's test was administered to assess significance of difference between trials. All the statistical analysis was computed using a statistical software, SPSS (version 10).

Results were also analyzed for information transmitted for each phonetic feature in different conditions. The consonants were classified based on the three phonetic features (i.e., Voicing, Place and Manner of articulation). These 15 consonants were further classified based on the different manners of articulation. Table 2, lists the 15 consonants and their classification with regard to three phonetic features. The recognition data of all the consonants were pooled across participants and a single confusion matrix was created for each condition. Confusion matrices of all the five conditions were then analyzed to determine the percentage of conditional information transmitted using SINFA (Sequential Information Analysis). SINFA was performed using the software Feature Information Xfer (FIX) (developed by University College of London, Department of Linguistics), which uses the procedure described by Wang & Bilger (1973). In SINFA procedure, features go through a number of iterations. In the preset study there were three iterations. In the first iteration of SINFA information transmitted for each feature is calculated. In the subsequent iterations, the feature with the highest percentage of information transmitted in the previous iteration is held constant and remains partialed out. Thus SINFA helps to estimate the redundancy of specific feature in their contribution in perception of consonants.

RESULTS

From Figure 1, it can be observed that the different envelope-enhanced conditions, used in the present study, improved the ability to identify con-

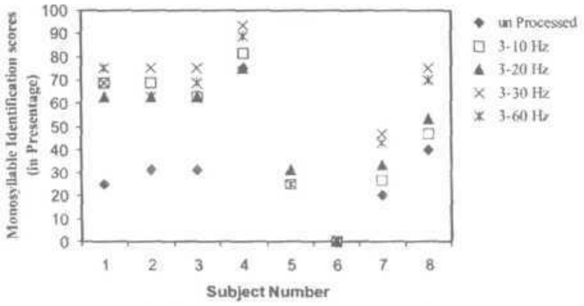


Fig. 1. Monosyllable identification scores (in percentage) in different test conditions.

TABLE 3. Mean (SD) of consonant identification scores and total information transmitted for different envelope-enhanced conditions

Conditions	Mean (SD)	Total transferred information (bits)
Unprocessed	35.2(18.6)	1.94
3 to 10 Hz	54.2(21.9)	2.71
3 to 20 Hz	54.3(16.3)	2.78
3 to 30 Hz	65.4(17.5)	3.01
3 to 60 Hz '	58.9(19.1)	3.03

sonants in six individuals with AN. Among the two individuals (subject 5 and subject 6) with no benefit in the enhanced conditions, one subject (subject 6) had zero percent scores in all the test conditions and his data was not considered for statistical analysis.

Friedman's test revealed that there was no significant difference ($p > 0.05$) in the scores obtained across trials. Therefore, further analysis was carried out for the data pooled from all the trials. The mean percentage of correct identification of consonants and the mean improvement for different envelope-enhanced conditions are shown in Table 3. It can be observed from the table that the improvement was greater when envelope bandwidth of 3 to 30 Hz was enhanced and there was lesser improvement for 3 to 10 Hz and 3 to 20 Hz conditions.

A Wilcoxon signed-rank test provided Z values which were assessed statistically. The scores in the unprocessed condition were significantly lower than those in the envelope-enhanced conditions irrespective of the bandwidth ($p < 0.01$). Among the envelope-enhanced conditions, scores for the broader bandwidth conditions were significantly different from those in the smaller bandwidth conditions ($p < 0.05$), except for the difference between the 3 to 10 Hz and 3 to 20 Hz conditions, which was not large enough to reach significance.

Information Analysis

Sequential information analysis (Wang & Bilger, 1973) was applied for each of the experimental conditions to assess the amount of information transfer from stimulus to response for a set of the most relevant phonetic features. The maximum information in bits that can be transmitted for the 15 stimuli of the present study is 3.9. It can be observed from Table 3 that the total information transmitted (bits) was increased in all the envelope-enhanced conditions when compared with the unprocessed condition. The information transmitted in the broader bandwidth (3 to 60 Hz) conditions was greater than those transmitted in the smaller bandwidth (3 to 10 Hz) conditions.

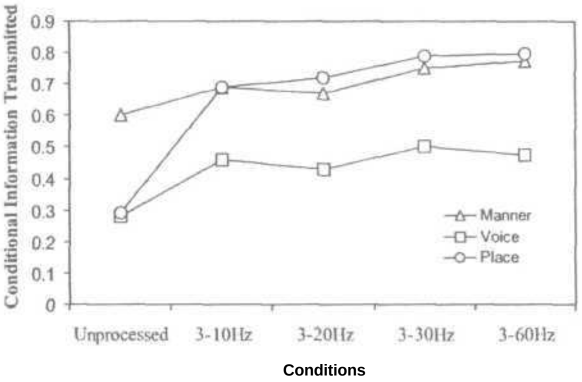


Fig. 2. Information transmitted for voicing, place, and manner in different conditions.

The percentage of conditional information transmitted for a feature is defined as conditional information transmitted (bits) divided by the input information for each feature. "Zero" indicates no transmission of that particular feature and a value of 1 indicates maximum transmission of the information. Figure 2 shows the percentage of conditional information transmitted for the feature that was identified and held constant in the SINFA procedure for all the five test conditions. In the unprocessed condition, manner of articulation conveyed greater information relative to the place of articulation and voicing. In the envelope-enhanced conditions, both place and manner of articulation conveyed more information than voicing. The information carried by the place and manner was more in the broader bandwidth conditions when compared with smaller bandwidth conditions. The information conveyed by voicing was the least in all the five conditions.

SINFA procedure was extended to assess the conditional information transmitted for different manners of articulation. Figure 3 shows the conditional information transmitted for the different conditions as a function of manner of articulation. It can be observed from Figure 3 that in the unprocessed condition nasals and fricatives conveyed greater amount of information relative to affricatives, plo-

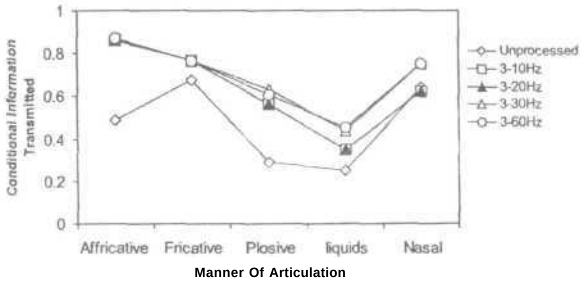


Fig. 3. Information transmitted for different manner of articulation in different conditions.

sives, and liquids. Information contributed by affricatives, fricatives, and nasals was more compared with liquids, irrespective of the envelope-enhanced conditions. From the Figure 3, it can be observed that, except liquids and nasals, other manner of articulation was processed better in all the envelope-enhanced conditions than in the unprocessed condition. Perception of liquids and nasals improved only in broader bandwidth conditions.

DISCUSSION

In the present study, envelope enhancement improved identification of consonants in a majority of the individuals with AN. A majority of the subjects had more difficulty in perceiving place of articulation. Kraus et al. (1996) observed difficulty in perceiving place of articulation in individuals with learning impairment. They postulated that the impaired ability to discriminate rapid spectro-temporal transitions for stop consonants in individuals with learning impairment is due to the poor temporal processing by cortical mechanisms (Kraus et al., 1996), recent findings also suggests a subcortical origin (brain stem) for these deficits (Cunningham, Nicol, & Kraus, 2000; Wible, Nicol, & Kraus, 2004). It can be hypothesized that in individuals with AN, dyssynchronous firing in the auditory nerve impairs the processing of rapidly varying signals at the subcortical level, which in turn impairs the encoding of rapid spectro-temporal transitions at the cortical level. Enhancing the envelope of speech signal by 15 dB results in rapid onset of the stimulus, which in turn increases the number of auditory nerve fibers that are synchronously activated (Lorenzi, Micheyl, & Berthommier, 1995). This probably resulted in better representation of consonants in the six participants who showed improvement in the envelope-enhanced conditions. It has been reported in the literature that synchronous firing of the auditory nerve fibers results in better representation of rapidly varying acoustical signals in the lower levels of the auditory system, which will be effectively integrated by the cortical structures for consonant perception (Cunningham, Nicol, King, Zecker, & Kraus, 2002; Merzenich, Scheriner, Jenkins, & Wang, 1995; Tallal et al., 1996). However, two participants did not benefit from the envelope enhancement used in the present study. Starr et al. (2003) have hypothesized that AN may be caused, because of reduced spike discharge, dyssynchronous firing and/or combination of above two. Probably two participants who showed no improvement had reduced number of fibers in the auditory nerve.

Comparison of scores in the four envelope-enhanced bandwidth conditions showed that there was greater improvement in the broader bandwidth condition when compared with the smaller bandwidth conditions. It was observed that this difference was mainly due to improvement in identification of plosives and liquids in the broader bandwidth conditions. Enhancing the envelope in broader-band width enhances the more salient cues such as formant transition and burst which in turn probably improved identification of consonants.

Perception of Manner of Articulation

Informational analyses for different features of consonants show that, in the unprocessed condition, manner information is processed more efficiently when compared with voicing and place information.

TABLE 4. Consonantal confusion for CV syllables in the unprocessed condition

							Response							-	
Stimulus	pa	ba	ta	da	ta	da	ka	ga	la	ra	ma	na	sa	**Ja**	tfa
pa	47	7	2			2	2								
ba	23	35				2									
ta	18		30	2	10										
da	9	16	5	20		10									
ta	16		4	7	18	15									
da	17	10	1	7	4	21									
ka	9	5	18	3	2	3	20								
ga	10		2	1	2	15	12	18							
la	5	5							33		10	7			
ra	25	5						3	5	20	1	1			
ma	5										40	10	3	2	
na	5										25	30			
sa	8												52		
Ja	3												4	53	
tfa	1	1	17								5		12		24

Shaded area indicates within-category substations among the consonants.

Table 4 shows the confusion matrix for unprocessed condition. It can be observed from the confusion matrix that a majority of the errors were of within category substitutions. Information analysis revealed that in the unprocessed condition, information processed was lesser for liquids (/I/ and /r/) and plosives (/p/ /b/ /k/, /g/ /t/ /d/ /t/ /d/) when compared with nasals (/m/) and fricatives (/s/ and /t//). It is probably because consonantal information for fricatives and nasals is carried by slow envelope (<8 Hz) of the speech signal (Drullman, Festen, & Plomp, 1994a) and the slow envelope of the speech signal is processed better than faster envelope in individuals with AN (Zeng et al., 1999). Two individuals who did not benefit from the envelope- enhancement in the present study perceived all the consonants as nasals and fricatives. Zeng et al. (1999) have reported that individuals who are severally impaired are able to perceive only <14 Hz envelopes in the speech signal. Perception of envelope of the speech signal in this frequency range is sufficient to perceive the fricatives and the nasals (Drullman, Festen, & Plomp 1994). Probably these individuals had reduced number of fibers and were able to perceive only < 14 Hz envelopes in the speech signal.

The poor perception of liquids /l/ and /r/ in the unprocessed condition is probably because, acoustically they are characterized by low-frequency formants and their transitions have little high frequency energy. Liquids are differentiated from nasals and other consonants by formant onset frequency and transitions (Johnson, 2003; Kent, & Read, 1995). Because differential sensitivity for low-frequency signals is poor in individuals with AN (Ranee, McKay, &, Grayden, 2004), they probably perceived nasals (/m/) and plosives (/p/) instead of liquids. Even in the envelope-enhanced conditions (3 to 30 Hz and 3 to 60 Hz) there was very little improvement as important cues that differentiate liquids from other consonants lie in the low-frequency region. This suggests that additional spectro-temporal modifications of speech are required for improving consonant perception.

It can be observed from the confusion matrix that affricate /tJ/ was perceived as either fricative /s/ or plosive /t/ in the unprocessed condition (refer to the Table 4). Because individuals with AN are poor at processing faster envelopes of the speech signal, perception of rise time of the fricative noise that differentiates affricates from the fricatives and stop consonants (Kent, & Read, 1995) was affected (Rosen, 1992; Drullman, Festen, & Plomp, 1994). Envelope-enhanced conditions enabled improved perception of rise times of the fricative noise, which in turn improved the perception of affricates (/tf/).

Perception of Place of Articulation

Place of articulation errors were mainly observed for stop consonants. In the unprocessed condition, a majority of the participants perceived alveolar as bilabials and velar as alveolar stop consonants (refer to Table 4). Similar place errors were also observed in a subject with AN reported by Kraus et al. (2000). Burst amplitude and formant transition are important cues for perception of place of articulation of stop consonants (Lahiri, Gewirth, & Blumstein, 1984; Ohde & Stevens, 1983). Drullman, Festen, & Plomp (1994) reported that in individuals with normal hearing, place errors were observed only when the faster envelopes of the speech signal are reduced and they related it to a reduction of changes in the distribution of spectral energy from burst onset to vowel onset. Individuals with AN require more depth of modulation to process faster envelopes in the speech signal. This would have precluded the perception of salient cues (formant transition, burst amplitude) in the unprocessed speech and lead to poor perception of place cue in the stop consonants. Similarly, even the nasal alveolar /n/ was perceived as the nasal bilabial /m/ probably because place perception depends on the second formant transition (Kent & Read, 1995), whose spectral energy lies in low frequency. As individual with AN are poor at processing low-frequency signals, perception of second formant transition is precluded leading to place errors.

From the information analysis, it was observed that in the all the envelope-enhanced conditions, there was improvement in perception of plosives. Probably enhancing the faster modulations of the speech signal helped AN participants to processes salient cues for consonant identification (formant transition and burst amplitude). Furthermore, the spectrogram of the enhanced signal (refer to Figure 1) shows more energy concentration in the formant frequency region. This supports the notion that envelope enhancement improves the perception of formant frequencies, which is one of the salient cues for speech perception. However, place perception for nasals did not improve even in broader bandwidth conditions. It has been reported that individuals with AN have poor differential sensitivity at low frequencies (Ranee et al., 2004; Zeng et al., 1999). Probably enhancing the envelope of the speech signal did not improve perception of second formant transition in nasals.

Perception of Voicing

Perception of voicing mainly depends on the voice onset time (VOT), and Fl (first formant) transition (Summerfield & Haggard, 1977). In the present study, individuals with AN perceived voiced sound as unvoiced. The poor perception of voicing may be due to low-frequency hearing loss or because of difficulty in detecting short duration VOT (Zeng, Kong, Michalewski, & Starr, 2005). Another factor that may have contributed is poor processing of the low-frequency signals due to dyssynchronous firing, (Ranee, McKay, & Grayden, 2004), which affects the perception of Fl onset and transition. In the present study, voicing perception is improved to negligible extent, because envelope enhancement did not increase the amplitude of the voicing bar. In the envelope-enhanced conditions there was a slight improvement for 3- to 30-Hz bandwidth conditions. This improvement may be attributed to the improved perception of first formant onset and transition.

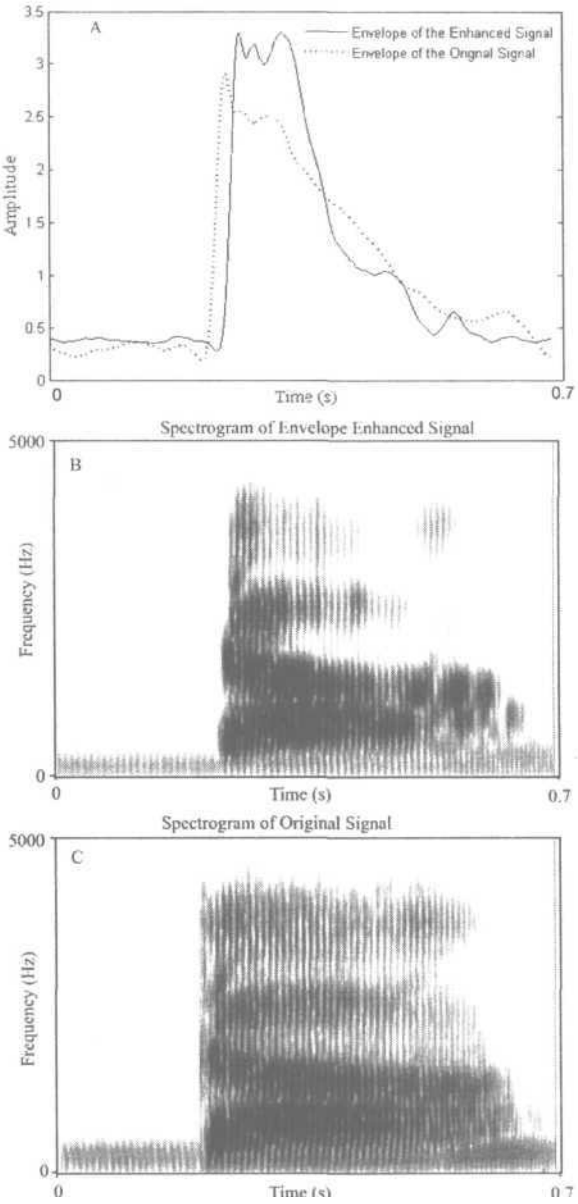


Fig. 4. Envelope and spectrogram of CV syllable /da/ with and without envelope enhancement. A, Envelope of the original and enhanced signal. B, The spectrogram of the original signal and envelope-enhanced signal are presented in right panel and left panels, respectively.

CONCLUSION

Envelope enhancement using digital techniques improves speech perception in individuals with AN, and it is a viable option for the rehabilitation of these individuals. However, the magnitude of envelope enhancement of speech required for maximal improvement of speech preparation is yet to be ascertained. Also, all of the individuals with AN did not show improved speech perception with envelope enhancement and envelope enhancement did not improve the consonant identification to 100% in any of these participants. Hence, other signal enhancement techniques need to be investigated for improving in speech perception by individuals with AN.

APPENDDC-I

Envelope Enhancement

The software incorporates overlap and add (OLA) procedure for speech processing. The speech signal is passed through stop band (300 Hz and 8000 Hz) (hamming) frequency domain filter after FFT. The spectral values at frequencies between 400 and 7900 Hz are set to zero. The spectral values from 200 to 400 Hz and from 7900 to 8100 Hz are multiplied by a raised sine, so as to give a smooth transition without ringing in the time domain. Finally, a backward Fourier transformation is done to obtain the stop band signal.

The processing is carried out on the spectrum in the 300 Hz to 8000 Hz range. The spectrum is divided into different critical bands using hamming window. Each frequency band is one Bark wide (based on bark scale), with 100 Hz overlap. Each critical band is converted to a pass-band filtered sound by means of the backward Fourier transform. The pass band filtered sound is subjected to FFT and envelope (intensity modulations) is obtained from the absolute value of FFT. The envelope within each band is then passed through a frequency domain filter using hamming window to extract frequency range of interest. The filtered band is converted into power. Based on the power and the maximum enhancement, a factor required for envelope expansion is calculated in dB scale and the filtered sound is multiplied by this factor. The manipulated band pass signals are added and a backward Fourier transformation is done to obtain the manipulated signal. The manipulated signal is finally added to the stop band signal to get the envelope-enhanced signal.

A representative sample of the envelope enhancement carried out in the present study is shown in Figure 4. It shows the spectrogram, envelope of a /CV/ syllable /da/ with and without envelope enhancement.

It can be'observed from the figure that the amplitude of the envelope of the enhanced signal is more when compared with that of the original signal. The spectrogram of the signal shows greater energy concentration in the formant frequency region for envelope-enhanced signal.

ACKNOWLEDGMENTS

We wish to thank our Director, All India Institute of Speech and Hearing, for allowing us to conduct the study. We are grateful to Editor Dr. Sasha John and the reviewers for their valuable and constructive comments. We extend our sincere thanks to Dr. Rupa Chandrasheker for her assistance in improving the manuscript. Finally we thank all our participants for their patient cooperation.

Address for correspondence: Mr. Vijaya Kumar Narne, Joiner Research Fellow, Department of Audiology, All India Institute of Speech and Hearing, Mysore, Manasagangotri, Mysore-570006, Karnataka, India.

Received April 9, 2006; accepted April 6, 2007.

REFERENCES

American National Standards Institute. (1996). Specification for audiometers. ANSI S3.6-1996. New York, NY: American National Standards Institute.

American National Standards Institute. (1999). Maximum permissible ambient noise for audiometric rooms. ANSI S3.1-1999. New York, NY: American National Standards Institute.

Apoux, F., Tribut, N., Debruille, X., Lorenzi, C. (2004). Identification of envelope-expanded sentences in normal-hearing and hearing-impaired listeners. *Hearing Research, 189,* 13-24.

Berlin, C. I., Morlet, T., Hood, L. (2003). Auditory neuropathy/dyssynchrony: Its diagnosis and management. *Pediatric Clinics of North America, 50,* 331-330.

Cunningham, J., Nicol, T., King, C, Zecker, G. S., Kraus, N. (2002). Effects of noise and cue enhancement on neural responses to speech in auditory midbrain, thalamus, and cortex. *Hearing Research, 169,* 97-111.

Cunningham, J., Nicol, T., Kraus, N. (2000). Neurophysiologic representation of clear speech in noise. *Journal of the Association for Research in Otolaryngology, 01,* 33-45.

Drullman, R., Festen, J. M., Plomp, R. (1994). Effect of temporal envelope smearing on speech reception. *Journal of the Acoustical Society of America, 95,* 1053-1064.

Houtgast, T., Steeneken, H. J. M. (1985). A review of the MFT concept in room acoustics and its use for estimating speech intelligibility in auditoria. *Journal of the Acoustical Society of America, 77,* 1069-1077.

Johnson, K. (2003). *Acoustic and auditory phonetics.* Oxford: Blackwell Publishing.

Kent, R. D., Read, C. (1995). *The acoustic analysis of speech.* Delhi, India: Laxman Chandarya.

Kraus, N., Bradlow, M. A., Cunningham, C. J., King, C. D., Koch, D. B., Nicol, T. G., Mcgee, T. J., Stein, L. K., Wright, B. A. (2000). Consequences of neural asynchrony: A case of auditory neuropathy. *Journal of the Association for Research in Otolaryngology, 01,* 33-45.

Kraus, N., McGee, T. J., Carrell, T. D., Zecker, S. G., Nicol, T. G., Koch, D. B. (1996). Auditory neurophysiologic responses and discrimination deficits in children with learning problems. *Science, 273,* 971-973.

Kumar, A. U., Jayaram, M. (2005). Auditory processing in individuals with auditory neuropathy. *Behavioral Brain Function, 21,* 1-8.

Lahiri, A., Gewirth, L., Blumstein, S. E. (1984). A reconsideration of acoustic invariance for place of articulation in diffuse stop consonants: Evidence from a cross language study, *Journal of the Acoustical Society of America, 76,* 391—404.

Lorenzi, C, Berthommier, F., Apoux, F., Bacri, N. (1999). Effect of envelope expansion on speech recognition. *Hearing Research, 136,* 131-138.

Lorenzi, C, Micheyl, C, Berthommier, F. (1995). Neuronal correlates of perceptual amplitude-modulation detection. *Hearing Research, 90,* 219-224.

Merzenich, M., Schreiner, C, Jenkins, W. M., Wang, X. (1993). Neural mechanisms underlining temporal interaction, segmentation, and input sequence representation: some implications of origin of learning. *Annals of the New York Academy of Sciences, 682,* 1-22.

Ohde, R. N., Stevens, K. N. (1983). Effect of burst amplitude on perception of stop consonant place of articulation. *Journal of the Acoustical Society of America, 74,* 706-714.

Ranee, G., Cone-Wesson, B., Wunderlich, J., Dowell, R. C. (2002). Speech perception and cortical event related potentials in children with auditory neuropathy. *Ear and Hearing, 23,* 239-253.

Ranee, G., McKay, C, Grayden, D. (2004). Perceptual characterization of children with auditory neuropathy. *Ear and Hearing, 25,* 34-^6.

Rapin, I., Gravel, J. (2003). 'Auditory neuropathy": Physiologic and pathologic evidence calls for more diagnostic specificity. *International Journal of Pediatric Otorhinolaryngology, 67,* 707-728.

Rosen, S. (1992). Temporal information in speech: acoustic, auditory, and linguistic aspects: Philosophical transactions. *Biological Sciences, 336,* 367-373.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonisk, J., Ekelid, M. (1995). Speech recognition with preliminarily temporal cues. *Science, 270,* 303-304.

Sininger, Y., Oba, S. (2001). Patients with auditory neuropathy: who are they and what can they hear? In Y. Sininger, A. Starr, Auditory neuropathy: A new perspective on hearing disorders (pp. 15-35). San Diego: Singular Publishing.

Starr, A., Michalewski, H. J., Zeng, F. G., FujiKawa-Brooks, S., Lithicum, F., Kim, C. S., Winner, D., Keats, B. (2003). Pathology and physiology of auditory neuropathy with novel mutation in the MPZ gene (tyrl45->ser). *Brain, 126,* 1604-1619.

Starr, A., Picton, T. W., Sininger, S., Hood, L. J., Berlin, C. I. (1996). Auditory neuropathy. *Brain, 119,* 741-753.

Summerfield, Q., Haggard, M. (1977). On dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. *Journal of the Acoustical Society of America, 62,* 435-448.

Tallal, P., Miller, S. L., Bedi, G., Byma, G., Wang, X., Nagarajan, S. S., Schreiner, C, Jenkins, M. W., Merzenich, M. M. (1996). Language comprehension in learning impaired children improved with acoustically modified speech. *Science, 271,* 82-84.

van Tassel, D. J. (1993). Hearing loss, speech and hearing aids. *Journal of Speech and Hearing Research, 36,* 228-244.

Viemeister, N. F. (1979). Temporal modulation transfers function based on modulation threshold. *Journal of the Acoustical Society of America, 66,* 1364-1380.

Wang, M. D., Bilger, R. C. (1973). Consonant confusions in noise: A study of perceptual features. *Journal of the Acoustical Society of America, 54,* 1248-1266.

Wible, B., Nicol, T., Kraus, N. (2004). Atypical brain stem representation of onset and formant structure of speech sounds in children with language-based learning problems. *Biological psychology, 67,* 299-317.

Zeng, F. G., Oba, S., Grade, S., Sininger, Y., Starr, A. (1999). Temporal and speech processing deficits in auditory neuropathy. *NeuroReport, 10,* 3429-3435.

Zeng, F. G., Kong, Y. Y., Michalewski, H. J., Starr, A. (2005). Perceptual consequences of disrupted auditory nerve activity. *Journal of Neurophysiology, 93,* 3050-3063.

REFERENCE NOTE

Vanaja C. S., Jayaram M. (2003). Sensitivity and specificity of audiological tests in differential diagnosis of auditory disorders. Unpublished project reported, submitted to AIISH, Mysore.

Behavioral and Brain Functions

BioMed Central

Research

Open Access

Speech identification and cortical potentials in individuals with auditory neuropathy

Vijaya kumar Name*[1] and CS Vanaja[2]

Address: 'Department of Audiology, All India Institute of Speech and Hearing, Mysore, Kamataka, 570006, India and [2]Bharati Vidyapeeth University School of Audiology and Speech Language Pathology, Dhankawadi, Pune, Maharashtra, 410021, India

Email: Vijaya kumar Name* - vijaynamel@yahoo.co.in; CS Vanaja - csvanaja@rediffmail.com

* Corresponding author

Published: 3 I March 2008

Behavioral and Brain *Functions* 2008. 4:15 doi: 10.1186/1744-9081-4-15

Received: 11 November 2007

Accepted: 31 March 2008

This article is available from: http://www.behavioralandbratnfunctions.eom/content/4/l/l5

Abstract

Background: Present study investigated the relationship between speech identification scores in quiet and parameters of cortical potentials (latency of PI, N I, and P2; and amplitude of NI /P2) in individuals with auditory neuropathy.

Methods: Ten individuals with auditory neuropathy (five males and five females) and ten individuals with normal hearing in the age range of 12 to 39 yr participated in the study. Speech identification ability was assessed for bi-syllabic words and cortical potentials were recorded for click stimuli.

Results: Results revealed that in individuals with auditory neuropathy, speech identification scores were significantly poorer than that of individuals with normal hearing. Individuals with auditory neuropathy were further classified into two groups, Good Performers and Poor Performers based on their speech identification scores. It was observed that the mean amplitude of NI/P2 of Poor Performers was significantly lower than that of Good Performers and those with normal hearing. There was no significant effect of group on the latency of the peaks. Speech identification scores showed a good correlation with the amplitude of cortical potentials (NI /P2 complex) but did not show a significant correlation with the latency of cortical potentials.

Conclusion: Results of the present study suggests that measuring the cortical potentials may offer a means for predicting perceptual skills in individuals with auditory neuropathy.

Background

Auditory neuropathy is one of the hearing disorders in which cochlear amplification is normal but neural transmission in afferent pathway is disordered. The integrity of cochlear function in this population is provided by the presence of evoked oto-acoustic emissions and/or cochlear microphonics (CM), and the abnormal neural transmission or dys-synchrony is indicated by the absence of auditory brainstem responses and middle ear muscle reflexes. Although the audiological findings in auditory neuropathy are suggestive of a retro-cochlear pathology, the exact site of pathology and patho-physiological mechanism leading to auditory neuropathy is not known. Two physiological explanations proposed for the neurophysiological manifestations observed include dys-synchronized spikes discharge and/or reduced spike of the auditory nerves [1,2]

Hearing sensitivity in individuals with auditory neuropathy may range from normal hearing to profound hearing impairment [3]. A majority of individuals with auditory neuropathy have low frequency hearing loss with disproportionately poor speech recognition scores for the degree of hearing loss (3|. Speech identification ability in individuals with auditory neuropathy varies considerably among patients but approximately 60 to 70% of individuals have identification scores well below the estimated identification scores from their pure-tone thresholds

Starr et al. (5) attempted to record click evoked cortical potentials (PI, Nl and P2) in four often adults subjects with auditory neuropathy. The responses to supra-threshold click stimuli were recordable in three of four subjects. They further observed that the subject with absent cortical potentials had poorer speech identification score than other three subjects. Kraus et al. [6] subsequently presented a case report showing cortical evoked potentials in a teenager with auditory neuropathy, whose identification score in quiet was 100%, whereas in adverse conditions, the identification scores were very poor. As the conical potentials were normal in this client, they hypothesized that speech perception in quiet was not significantly affected by poor synchronization at the brainstem level if synchronization is preserved at the cortical level. Results of some of the investigations carried out later support this hypothesis. Ranee et al. [7] observed better speech identification scores in children with auditory neuropathy who had normal cortical potentials when compared to those with abnormal cortical potentials. Vanaja and Manjula (8) reported that individuals who have higher amplitude in cortical potentials had better speech identification scores and also benefitted more with a hearing aid than those with lesser amplitude.

Thus, limited information available in literature shows that auditory neuropathy individuals having poor identification scores in quiet have abnormal or absent cortical potentials suggesting that integrity of processing at cortical level is important for speech understanding. The present study was undertaken to study the relationship between speech perception ability in quiet and parameters of cortical potentials in individuals with auditory neuropathy.

Methods

I. Participants

Ten individuals with auditory neuropathy and ten individuals with normal hearing participated in the present study. Out of ten individuals with normal hearing, five were males and five were females with ages ranging from 12 to 39 yr with a mean of 22 yr. The individuals with normal hearing had pure-tone sensitivity of less than 15 dB HL at octave frequencies from 250 Hz to 8000 Hz. These individuals were volunteers from local college and schools.

Participants with auditory neuropathy were recruited from the clients registered at the Audiology clinic of the All India Institute of Speech and Hearing, Mysore, India. Table 1 shows the audiological profile of the ten participants (5 males and 5 females) with auditory neuropathy. The age of the participants ranged from 12 to 39 yr with a mean of 20.7 yr. The pure-tone average (average of pure tone thresholds at 500, 1000, 2000, 4000 and 8000 Hz) ranged from 10 to 48 dB HL. A majority of the participants had symmetrical hearing loss in both the ears. The audiometric configuration was rising pattern in a majority of the participants. All the participants had present TEOAEs and absent middle ear acoustic reflexes (both ipsilateral and contralateral) and the auditory brainstem responses. None of participants had any family history or any other medical complications. All the participants were native speakers of Kannada, a Dravidian language spoken in a southern state of India.

Table I: Audiological profile of individuals with auditory neuropathy

S.No	Age/Sex	Pure-tone Average (dB HL) Right ear	Pure-tone Average (dB HL) Left ear	ABR in both ears	OAE in both ears	Acoustic reflex in both ears
ANI	I2ys/M	26.00	31.00	Absent	Present	Absent
AN2	20 ys/F	31.00	34.00	Absent	Present	Absent
AN3	15 ys/F	30.00	36.00	Absent	Present	Absent
AN4	39 ys/F	33.00	39.00	Absent	Present	Absent
ANS	I2ys/M	44.00	43.00	Absent	Present	Absent
AN6	24 ys/M	31.00	38.00	Absent	Present	Absent
AN7	27 yr/F	42.00	31.00	Absent	Present	Absent
AN8	20 yr/M	48.00	46.00	Absent	Present	Absent
AN9	I8yrs/M	19.00	10.00	Absent	Present	Absent
ANI0	20 yrs/M	43.00	**39.00**	Absent	Present	Absent

II. Data *Collection*

a. Assessment *of speech identification ability*

Stimuli

Speech Identification Test in Kannada developed by Vandana [9) was used to assess open set speech identification abilities. This test consists of 50 bi-syllabic meaningful words of Kannada. Validity and reliability of this test has been established on native speakers of Kannada [9].

Procedure

The participants listened to speech tokens individually in a double-walled, acoustically treated room where the ambient levels were within permissible limits [10]. The speech stimuli were presented through supra-aural headphones (TDH - 39) of a calibrated [11] diagnostic audiometer (Madson OB-922). The stimuli were presented at 40 dB SL (re: Speech Recognition Threshold) monaurally and the participants were asked to repeat the speech token. The speech recognition scores were calculated by counting the number of words correctly repeated.

b. *Cortical evoked potentials*

The participants were seated comfortably in a reclining chair and the cortical evoked potentials were acquired using the Intelligent Hearing Smart EP system. The responses were picked up from a disc electrode placed on the midline site, Cz, with reference to an electrode placed on the ipsilateral mastoid. The common electrode was placed at Fpz. It was ensured that the impedance at each electrode site was less than 5 k ohms and the inter-electrode impedance was less than 2 k ohms. The participants were instructed not to pay attention to the stimuli while recording.

The cortical potentials were recorded for each ear separately with click stimuli presented through insert-earphones (ER-3A) at a repetition rate of 1.1/sec at 80 dB nHL. Stimulus level used to elicit the cortical waveforms were supra-threshold for all participants. The EEG acquired was amplified 50,000 times and digitally filtered using a band pass filter of 1-30 Hz. The EEG was epoched using a window of 550 ms, including a 50 ms pre-stimulus baseline. Epochs greater than 45 uV were rejected. The EEG responses for 200 stimuli were averaged. The latency of P1, N1, P2, N2, and the amplitude of N1/P2 were measured. The amplitude of N1/P2 was measured with peak-to-peak.

Recordings were repeated twice to check for replicability. Only those peaks, which were replicable, were considered as a response. Three experienced audiologists independently analyzed the waveforms to identify and mark the peaks in cortical potentials. It was considered as a response only if all the three audiologists identified the cortical potentials at the same latency.

Results

Speech identification ability

Speech identification scores in individuals with normal hearing ranged from 95% to 100% with a mean of 96% in both eras whereas in individuals with auditory neuropathy identification scores in both eras ranged from 0 to 90% with a mean of 42.1% in the right ear and 41.2% in the left ear. Among the individuals with auditory neuropathy, AN-3 had 0% identification in both eras. Figure 1 shows the individual data for speech identification scores in individuals with auditory neuropathy. Paired sample "t" test revealed no significant difference (auditory neuropathy: t = 0.1, p = 0.88; Normal: t = 0.05, p = 0.9) between the ears for identification scores in both groups. Hence, the data from the two ears were merged for further statistical analysis.

The mean speech identification scores for subjects in the normal hearing group was 96% with a standard deviation of 2.5% whereas the mean scores of individuals with auditory neuropathy was 42% with a standard deviation of 25.4%. An Independent Sample't' test revealed a significant difference between the mean speech identification scores of the two groups (t = 5.77, p < 0.01).

Pearson product-moment correlation was performed between behavioral threshold and speech identification scores in individuals with auditory neuropathy. Figure 2 shows the scatter plot between pure-tone average and speech identification scores. Pearson correlation coeffi-

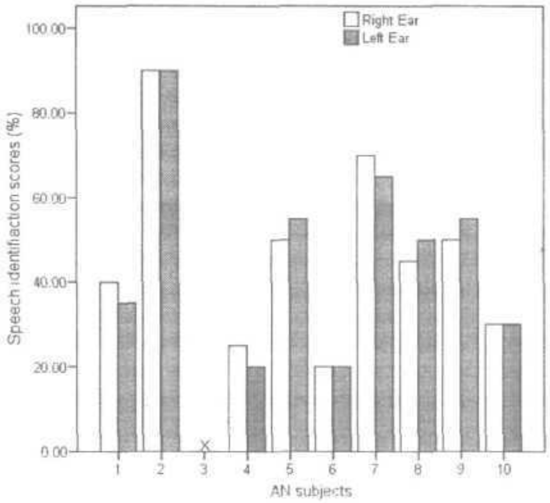


Figure 1
Speech identification scores of the individuals with auditory neuropathy. In the figure, X indicates of 0% identification scores.

cient revealed that there was no significant correlation between speech identification scores and pure-tone average in individuals with auditory neuropathy (r = -0.37, p = 0.6).

Cortical potentials

Cortical evoked potentials were present and symmetrical in alt the individuals with normal hearing. Cortical potentials were present and symmetrical in all the individuals with auditory neuropathy, except one participant (AN3). The responses were absent in a 15 year old participant with a pure-tone average of 30 dB HL. Therefore, the age and threshold cannot be the contributing factors for the absence of responses in this participant.

Paired Sample" t" test was performed to compare between two ears for latency of cortical potentials (PI, Nl, P2 and N2) and amplitude of N1/P2. The results revealed no significant difference between the two ears. For further analysis, data of right ear and left ear were combined. The mean and standard deviation of latencies of PI, Nl, P2, N2 in individuals with normal hearing and those with auditory neuropathy are presented in Table 2. From the table it can be noted that the latencies in subjects with auditory neuropathy were delayed by 20 - 50 ms for PI, 40-80 ms for Nl and 30-80 ms for P2 when compared to individuals with normal hearing.

Independent sample "t'testwas performed independently for latency of cortical potentials (PI, Nl, P2 and N2) and amplitude (N1/P2). Results revealed a statistically significant difference between the latencies of PI, Nl, and P2 peaks in individuals with normal hearing and those with auditory neuropathy but there was no significant difference for the latency of N2 peak. The "t" value and the level of significance are also shown in Table *2.* The N1/P2 amplitude of the participants with auditory neuropathy did not differ significantly from that of normal hearing individuals. However, the mean values of the amplitude for the participants with auditory neuropathy were slightly lower and the variability was greater when compared to those observed in normal subjects.

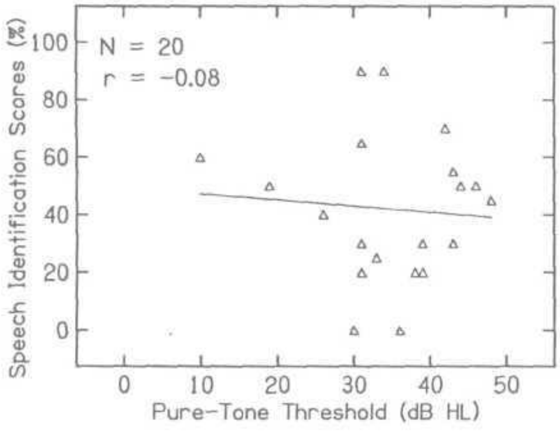


Figure 2
Relationship between the pure-tone threshold and identification scores of individuals with auditory neuropathy.

As there was more variability in measures of individuals with AN, the data of the participants with auditory neuropathy were further divided into two groups based on their speech identification scores. Group I included "Good Performers" whose speech identification score was more than 50% and Group II included "Poor Performers" whose speech identification score was less than or equal to 50%. The mean and standard deviation of latency and amplitude (N1/P2) cortical potentials for the two groups are presented in Table 3. It can be noted that the amplitude of Poor Performers was lower than that of Good Performers. Results of Kruskal Wallis test revealed that there is a significant effect of group on the amplitude ($p < 0.01$) of N1/P2 peak. Mann-Whitney test was performed to assess the paired comparison between the groups. Results revealed that the mean amplitude of Poor Performers was significantly lower than that of Good Performers ($p < 0.01$) and normal hearing subjects ($p < 0.01$). However, mean amplitude of Good Performers was not significantly different from that of normal hearing subjects ($p > 0.01$).

Kruskal Wallis test performed to study the effect of group on the latency of cortical potentials revealed that there was a significant effect of group on the latency ($p < 0.01$) for all the components except for N2. Mann-Whitney test was performed to assess the paired comparison between the groups. Results revealed that both in Good Performers and Poor Performers, the mean latency for all the peaks except N2 differed significantly from that of normal hearing subjects but there was no significant difference for latency for all the peaks between Good Performers and Poor Performers ($p < 0.01$).

Pearson product-moment correlation was carried out to study the correlation of the peak latency of PI, Nl, P2, N2 and the amplitude of N1/P2 with the behavioral thresholds (pure-tone average) and speech identification scores. It can be observed from Table 4 that the latency of cortical potentials did not show a significant correlation with the pure tone average or with speech identification scores. However, the amplitude of N1/P2 showed a significant correlation with speech identification scores. Relation between N1/P2 amplitude and speech identification scores is depicted in the scatter plot along with regression curve in Figure 3.

Table 2: Mean, SD, and "t" value of latencies and amplitude cortical potentials in individuals with normal hearing and auditory neuropathy

Participants	Latencies (m sec)				Amplitude NI/P2(MV)
	PI	**NI**	**P2**	**N2**	
AN	**76 (20)**	**124(31)**	**185(43)**	**243(50)**	**4.4(2.4)**
Normal i	**50(8.1)**	**85(9)**	**142(12)**	**218(13)**	**6.2(1.3)**
"t" Value	**4.1***	**2.8***	**3.05***	**1.6**	**-0.82**
"p" Value	**0.001**	**0.001**	**0.001**	0.057	**0.05**

* Significant at $p < 0.01$

Inspection of individuals data revealed that two participants had normal latencies with reduced amplitudes and their speech identification scores were poor (AN4, AN6), whereas two participants who had normal latencies with good amplitude (AN2 and AN7) had better speech identification scores. Five subjects who had prolonged latencies with normal amplitude also showed good speech identification scores. Figure 4 shows the waveforms of individuals with auditory neuropathy and those with normal hearing.

Discussion

Speech *identification in individuals with auditory neuropathy*

In the present study, speech identification scores in individuals with auditory neuropathy were significantly lower than that observed for participants with normal hearing. Further, the correlation analysis revealed that speech identification scores for individuals with auditory neuropathy were disproportionate to their pure-tone threshold. This is best illustrated by comparing the speech identification scores in individuals with auditory neuropathy to those expected by degree of hearing loss for patients with cochlear hearing loss [12]. In the present study, in 72% of individuals, the speech identification scores were lower when compared to those reported by Vanaja and (ayaram [12| for ears with sensorineural hearing loss. Sininger and Oba |3j observed that speech identification scores for 69% of their patients with auditory neuropathy were lower than that reported for patients with cochlear pathology by Yellin et al. **(13].** These results suggest that speech identification scores do not depend upon the pure-tone thresholds in individuals with auditory neuropathy. Other factors impair the speech understanding capability in these individuals.

One of the possible contributors for their poor speech identification score is disrupted neural synchrony, which impairs the listener's ability to processes the dynamic nature of speech signals. It has been reported that disrupted neural synchrony impairs the ability to use envelope cues in speech and also impair the ability to perceive rapid change of spectral shapes in the speech stimuli

Cortical potentials in individuals with auditory neuropathy

Latencies of cortical potentials in individuals with auditory neuropathy were significantly prolonged when compared to normal hearing listeners. Though not statistically significant, the mean amplitude of the cortical potentials was lower than that observed in participants with normal hearing and the variability was high in individuals with auditory neuropadiy. Latencies and amplitude variations in individuals with auditory neuropathy may not be due to increased pure-tone threshold, as there was no correlation between pure-tone thresholds and cortical potentials (latency and amplitude), suggesting that the latency and amplitude of cortical potentials were not affected by the hearing thresholds of the participants in the present study. Oates, Kurtzberg, andSatpells (15) reported that the latencies and amplitude of P1/N1/P2 were not significantly affected in subjects with cochlear hearing loss of less than moderate degree. Cortical potentials were absent in AN3 who had pure-tone threshold of 30 dB HL whereas it was

Table *3:* Mean and SD of latencies and amplitude cortical potentials for the two groups of auditory neuropathy and individuals with normal hearing

Participants	Latency (msec)				Amplitude
	PI	NI	P2	**N2**	NI/P2(uV)
Normal	**50(8.1)**	85(9)	142(12)	**218(13)**	6.2(1.3)
Good Performers	**84(16.8)**	**133(30.8)**	**186(53.8)**	**227(26.2)**	6.0(1.5)
Poor Performers	**78(20.2)**	**125(27.6)**	**184(27.4)**	**231(18.2)**	2.6(1.3)

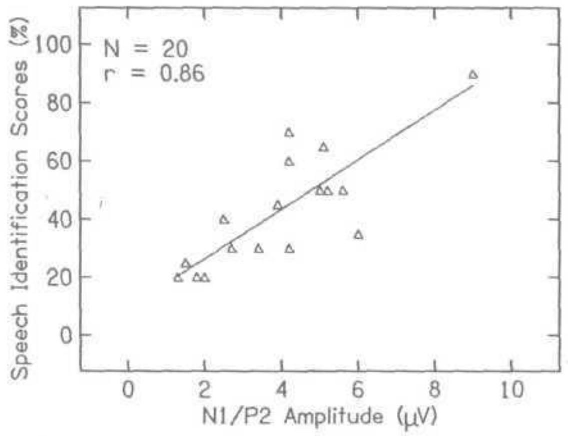


Figure 3
Scatter plot between speech identification scores and amplitude of cortical potentials in individuals with auditory neuropathy.

present in AN 4 whose pure-tone average of 48 dB HL. This further, supports the notion that cortical potentials did not depend upon the pure-tone average in the present study.

Cortical potentials mature and attain adult latency and morphology by the age of 9 years and there will not be any significant changes in latency until age of 50 yr [16). As the age range of the participants in the present study varied from 12 to 39 yr, the latency variations observed may not be due to maturational changes. It was hypothesized that probably the severity of the neural dys-synchrony rather than the hearing loss contributed for the variability in the cortical evoked responses observed in the present study.

An interesting observation in the present study was that, some of individuals with auditory neuropathy had abnormal latencies with normal amplitude, whereas some had normal latencies with abnormal amplitude in cortical potentials. Similar results have also been reported in the literature (5,8,17]. No clear-cut explanation can be provided for the variability observed in latencies of cortical potentials. The variability in latencies observed across individuals with auditory neuropathy in the present study may have been related to the underlining patho-physiology. That is prolonged latencies may be due to the dys-synchronous firing 118-20,14) whereas normal latencies may be due to the reduced numbers of fibers [1,14]. The magnitude of reduction in amplitude in either of patho-physiology depends upon the severity of the condition [14,21). Further investigation correlating cortical potentials with neurological findings need to be carried out to confirm this.

Relation between speech identification scores and cortical potentials

In the present study individuals who had cortical potentials with better amplitude had better speech identification scores than those with absent/abnormal amplitude in cortical potentials. Participant AN3 in the present study had absent cortical potentials and very poor speech identification scores. Similar results have been reported by earlier investigators [7,5,17]. These results suggest that it is possible to have good speech perception in quiet if the cortical responses are present even if the brainstem responses are abnormal. Good synchronization at the auditory nerve and brainstem level does not appear to be essential for understanding speech in quiet situations [6]. Results of physiological studies indicate that brainstem neurons process the fast modulations of the complex signals, whereas auditory cortex processes the slowly varying the amplitude modulations of the complex signal (22|, which plays an important role in auditory communication [23).

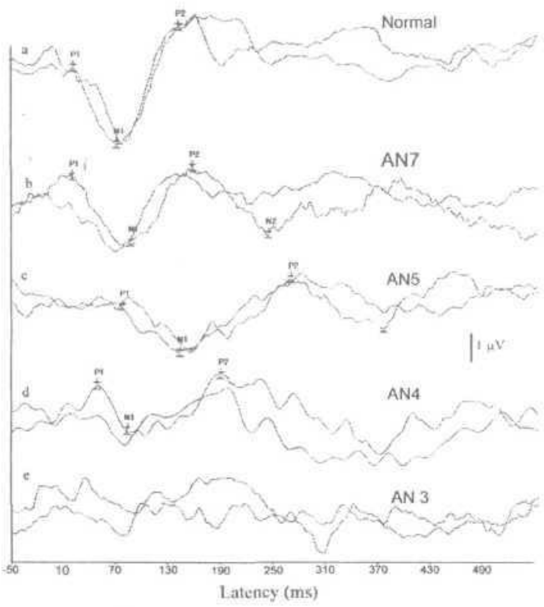
There was also a high positive correlation between speech identification and the amplitude of N1/P2. That is, individuals with better speech identification scores showed greater N1/P2 amplitude than those with poorer speech identification scores, However, no correlation was observed between latencies of cortical potentials and speech identification scores. Similar findings were observed in adults [8] and children[7] with auditory neuropathy using hearing aids. They reported that the cortical

Table 4: Correlation coefficients (r) of behavioral thresholds with cortical potentials and word recognition scores with cortical potentials

	"r" value				
	-atency *(ms)*				NI/P2
	PI	NI	P2	**N2**	amplitude (nV)
Behavioral Threshold	-0.46	**-0.25**	**-0.2**	-0.3	-0.16
Speech Identification score	0.45	**0.52**	**0.3**	-0.1	0.86*

v .

* Significant at p < 0.05



250 310
Latency (ms)

Figure 4

Representative samples of cortical potentials recorded from the two groups, a) Wave forms of individuals with normal hearing, b) Wave forms with normal latencies and normal amplitude obtained from an individual with auditory neuropathy, c) Wave forms with prolonged latencies obtained from an individual with auditory neuropathy, d) Wave forms with normal latency and reduced amplitude obtained from an individuals with auditory neuropathy e) No response obtained from an individual with auditory neuropathy.

potentials were present in individuals with auditory neuropathy who had good speech identification scores and these individuals also benefited from a hearing aid. Thus, it can be concluded that the speech recognition scores probably depend on the severity of the disorder rather than the underlining patho-physiology.

Conclusion

Results of the present study support the previously reported findings that speech perception ability cannot be reliably estimated from behavioral pure-tone audiogram in individuals with auditory neuropathy. Cortical potential testing may, however, offer a means of predicting speech understanding ability in individuals with auditory neuropathy. The presence and amplitude of cortical potentials showed a significant correlation with open set speech perception abilities. The absence of cortical potentials indicates extremely poor speech perception abilities. If these results are replicated in a larger group of individuals with auditory neuropathy, the procedure can be used to obtain important information regarding severity and management options for these participants.

Authors' contributions

VKN was involved in designing the study, data collection, analysis, interpretation of results and preparing the manuscript. CSV was involved in designing the study, analysis interpretation of results, and preparing the manuscript.

Acknowledgements

We thank Dr. Vijayalakshmi Bsasavaraj, Director, All India Institute of Speech and Hearing, for permitting us to conduct the study. We express our sincere thanks to all the participants for their patient cooperation.

References

1. Starr A, Michalewski HJ, Zeng FG, FujiKawa-Brooks S, Lithicum F. Kim CS, Winner D. Keats B: **Pathology and physiology of AN with novel mutation in the MPZ gene (tyrl45->ser).** *Brain* 2003, 1**26:1604-1619.**
2. Zeng FG, Kong YY, Michalewski HJ. Starr A: **Perceptual consequences of disrupted auditory nerve activity.** *J Neurohysiol* 2005, **93(6):3O5O-3O63.**
3. Sininger **Y,** Oba S: **Patients with AN: who are they and what can they hear?** In *A new perspective on hearing disorders* Edited by: Sininger Y. Starr A. Sandigo: Singular-Thomson learning; 2001:15-35.
4. Zeng FG, Oba S, Grade S, Sininger Y, Starr **A: Temporal and speech processing deficits in AN.** Neuroreport **1999, 10:3429-3435.**
5. Starr A, Picton TW, Sininger S, Hood LJ, Berlin CI: **Auditory Neuropathy.** *Brain* **1996, 119:741-753.**
6. Kraus N, Bradlow MA, Cunningham CJ. King CD, Koch DB, Nicol TG, Mcgee TJ, Stein LK, Wright BA: **Consequences of neural asynchrony: A case of AN.** *] Assoc Res Otohiyngol* 2000, **01:33-45.**
7. Ranee G, Cone-Wession B, Shepherd RK, Dowell RC, King AM, Rickards FW, Clark GM: **Speech perception and cortical event related potentials in children with AN.** *Ear Hear* **2002. 25:34-46.**
8. Vanaja CS, Manjula P: **LLR as a measure of benefit derived from hearing devices with auditory dys-synchrony.** In first *conference* on *Auditory Neuropathy* Edited by: Shivashanker N. Shashikala HR. Bangalore: Department of Speech pathology and Audiology, National Institute of Mental Health and Neuro sciences; 2004:136-146.
9. Vandana : **Speech identification test in Kannada.** Unpublished independent project. University of Mysore, Mysore, India; 1998.
10. American National Standards Institute: **Maximum permissible ambient noise for audiometric rooms. ANSI S3.1 -1999.** New York: American National Standards Institute.
11. American National Standards Institute: **Specification for audiometers. ANSI S3.6-I996.** New York: American National Standards Institute.
12. Vanaja CS, Jayaram M: **Sensitivity and specificity of audiological tests.** Unpublished Project Report, All India Institute of Speech and Hearing, Mysore, India; 2003.
13. Yellin MW, Jerger J. Fifer RC: **Norms for disproportionate loss in speech intelligibility.** *Ear Hear* 1989, 10:231-234.
14. Ranee G: **Auditory Neuropathy/Dys-synchrony and it's Perceptual Consequences.** *Trends Amplif* 2005, 9:1 -43.
15. Oates P, Kurtzberg D, Stapells DR: **Effect of sensoryneural hearing loss on cortical event-related potentials behavioral measures of sound processing.** *Ear Hear* 2002. 23:399-415.
16. Hall JW: **Effect of nonpathologic subject characteristics.** Edited by: Allyn & Bacon. Handbook of auditory evoked responses; 1992:70-103.
17. Satya-Murthi S, WolpawJR. Cacace AT. Scharffer CA: Late auditory **evoked potentials can occur without brain stem potentials.** Oectroencepha/ogr *Oin Neurophysiol* 1983, 56:304-307.
18. Hendler T, Squires NK, Moore JK, Coyle PK: **Auditory evoked potentials in multiple sclerosis: correlation with magnetic resonance imaging.** *] Basic* Gin *Physiol Pharmacol* 1996, 7:245-278.

19. Japaridze G, Shakarishvili R, Kevanishvili Z: **Auditory brainstem, middle-latency, and slow cortical responses in multiple sclerosis.** *Acta Neurol Scand* 2002, **106**:47-53.
20. Mc Donald WI, Sears TA: **The effect of experimental demyelination on conduction in the central nerves system.** *Brain* 1970, **91**:583-595.
21. Rapin I, Gravel J: **"Auditory neuropathy": Physiologic and pathologic evidence calls for more diagnostic specificity.** *Int J Pediatr Otorhinolaryngol* 2003, **67**:707-728.
22. Giraud AL, Lorenzi C, Ashburner J, Wable J, Johnsrude I, Frackowiak R, Kleinschmidt R: **Representation of the Temporal Envelope of Sounds in the Human Brain.** *J Neurophysiol* 2000, **84**:1588-1598.
23. Drullman R, Festen JM, Plomp R: **Effect of temporal envelope smearing on speech reception.** *J Acoust Soc Am* 1994, **95**:1053-1064.