INFLUENCE OF NUMBER OF FREQUENCY BANDS ON PERCEPTION OF KANNADA CHIMERIC WORDS AND SENTENCES

Naveen C P
Register Number: 15AUD017

This Dissertation is submitted as part fulfillment
for the Degree of Master of Science in Audiology
University of Mysuru, Mysuru

ALL INDIA INSTITUTE OF SPEECH AND HEARING
MANASAGANGOTHRI, MYSURU-570006
MAY 2017

Dedication

Dedicated to my Family, Friends and my Guide

CERTIFICATE

This is to certify that the dissertation entitled "**Influence of number of frequency bands on perception of Kannada chimeric words and sentences**" is the bonafide work submitted in part fulfillment for the degree of Master of Science (Audiology) of the student (Registration No. 15AUD017). This has been carried out under the guidance of a faculty of this institute and has not been submitted earlier to any other University for the award of any other Diploma or Degree.

Dr. S. R. Savithri

Director

All India Institute of Speech and Hearing,
Manasagangothri, Mysuru-570006

Mysuru,

May, 2017.

CERTIFICATE

This is to certify that the dissertation entitled "**Influence of number of frequency bands on perception of Kannada chimeric words and sentences**" has been prepared under my supervision and guidance. It is also certified that this has not been submitted earlier in any other University for the award of any Diploma or Degree.

Dr. Devi N

Mysuru
May, 2017.

Lecturer
Department of Audiology
All India Institute of Speech and Hearing,
Manasagangothri, Mysuru.

DECLARATION

This is to certify that this dissertation entitled “**Influence of number of frequency bands on perception of Kannada chimeric words and sentences**” is the result of my own study under the guidance of Dr. Devi N, Lecturer in Audiology Department of Audiology, All India Institute of Speech and Hearing, Mysuru, and has not submitted earlier in any other University for the award of any Diploma or Degree.

Mysuru **RegisterNo:15AUD017**

May, 2017

Acknowledgments

Completion of this Master's dissertation was possible with the support of several people. I would like to express my sincere gratitude to all of them. First of all, I am extremely grateful to my dissertation guide, Dr. Devi N, Lecturer in Audiology, All India institute of speech and hearing, Mysuru for her valuable guidance., scholarly inputs and constant encouragement I received throughout my research work. This work was possible only because of the unconditional support provided by Ma'am. thank you, Ma'am, for all your help and support.

I thank Dr. S R Savithri, Director AIISH, Mysuru for permitting me to do my study. I thank HOD, Dept of Audiology for providing academic support and facilities provided to carry out the research work at the institute.

Some faculty members of the institute have been very kind enough to extend their help at various phases of this study, whenever I approached them and I do hereby acknowledge all of them. I thank Mr. Nike Gnanateja for his valuable suggestions and concise comments on the study. Ms. Vasanthalakshmi, have extended her support in statistical data analysis and I thank her for her contributions.

I would like to thank my friends Ishu Mittal, Darshan, Madhu Sagar, Deepak, Nishanth, Shivakumar, Mahesh, Arun and Shrikanth

I owe a lot to my parents, who encouraged and helped me at every stage of my personal and academic life and longed to see this achievement come true. I am much indebted to my family who supported me in every possible way to see the completion of this work.

Abstract

The temporal cues play an important role in perception of different language. Previous investigations report that temporal fine structure cues are important for perception of tonal language and also it reports envelope as an major cue for speech identification. In Indian context, most of the languages are found to be non-tonal and this study attempts to investigate the effect of temporal cues on South Indian Kannada language using auditory chimera. The present study includes identification of chimeric words and chimeric sentences in normal adult participants with native language as Kannada. Investigations for finding out effect of fine structure and envelope was carried out across eight frequency bands and a comparison was made between identification of chimeric words and chimeric sentences. Results showed an improvement in chimeric word and chimeric sentence identification as number of frequency band increases and envelope cues were found to be dominated in identification of both chimeric word and chimeric sentences. Identification of chimeric sentences were found to be better compared to chimeric words.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

The human auditory system has the ability to hear sounds from 20 Hz to 20 kHz. Our auditory system has the abilities to understand the most complex sound signals. The human auditory system comprises of 3 components (outer ear, middle ear and inner ear). Acoustic energy received by the outer ear will be transferred to the middle ear and it gets converted into mechanical energy by the movements of tympanic membrane and middle ear ossicles. Further this mechanical energy will be transferred to the cochlea (inner ear). The cochlea behaves as a frequency analyzer with each place on the cochlea responding more favorably to a particular frequency known as the characteristic frequency (CF) (Bekesy, 1960). Sensory receptors, known as hair cells, transduce the mechanical energy into neural activity (spikes), which are elicited on Auditory Nerve (AN) fibers innervating those hair cells. The neural code in the Auditory Nerve fibers is then conveyed to higher nuclei in central auditory system for further analysis and processing. The auditory system performs several complex tasks such as sound localization, speech understanding, pitch and melody perception, which are usually required to function properly even in the existence of competing speech. Understanding the mechanism of operation of the auditory system requires good attention to the structure of the different sections in the auditory system as well as the interactions between the different parts (Billone & Ragnar, 1973).

Intensive research has been going on in the area of speech perception to identify the factors and mechanism by which humans understand speech in different listening conditions. When a sound is received by the cochlea, the frequency content of the signal is mapped into a pattern of excitation along the Basilar Membrane (Chen, Clark, & Jones, 2003). These signals are analyzed in inner ear and converted into a sequence of band pass signals, each one analogous to a location on ‘Basilar Membrane’. Excitation patterns code the spectrum information of the acoustic stimulus, which is referred to as ‘spectral’ or ‘place’ information (Nuttall, Brown, Masta, & Lawrence, 1981). It is assumed that spectral information plays an important role in speech recognition as many phonetic features are characterized by their frequency spectrum. Because basilar membrane acts as a bank of band-pass filters, it yields in the frequency selectivity. Analyzed signals contains two forms of information provided by the temporal analysis mechanism; fluctuations in envelop (E) and fluctuations in Temporal Fine Structure (TFS) (Moore & Sęk, 2009).

Temporal fine structure refers to rapidly fluctuating variations in the waveform amplitude close to the center frequency of the band, whereas envelop refers to slower modulation superimposed on the fine structure. In order to investigate the relative role of both envelop and fine structure, signals can be breakdown into their respective envelop and fine structure using specific algorithm. The first attempt to separate the envelope from a fine structure cue was done using peak clipping mechanism of speech signal (Licklider & Pollack, 1948) later by using Hilbert transform (Bracewell & Bracewell, 1986). Combining the features of different sounds can be used to study the perceptual relevance of envelop and fine structure. This can be achieved by making hybrid sounds known as auditory chimera. Construction of chimeric sentences are done by interchanging the envelope and fine structure of different sentences or words. These two characteristics of each sound helps to understand the relative importance of each in perceiving a sound. Perceptual studies on normal auditory system demonstrates that envelop cues are sufficient for speech perception in quiet, whereas fine structure is important in pitch perception, lexical tone perception and also speech perception in noisy conditions (Loizou, Dorman, & Tu, 1999; Smith et al., 2002). The TFS information is coded through the phase locking property of the AN fibers and it is known that phase locking is weak at high frequencies with almost a complete loss of synchrony for frequencies above 4-5 kHz in mammalian auditory systems (Palmer & Russell, 1986). Hence, it is commonly assumed that ‘TFS information is not used for frequencies above that limit’. It has been demonstrated in many experiments that envelope (ENV/E) clue is necessary for speech understanding and it provides robust speech recognition in silence even when provided in as few as four frequency bands (Flanagan,1980; Shannon, Zeng, Kamath, Wygonski & Ekelid,1995; Smith, Delgutte & Oxenham, 2002). Hence auditory chimeras provide a way to study the relative importance of envelop and fine structure in speech perception and pitch perception.

Smith et al (2002) found that in speech –speech chimeras with 4-16 frequency bands, the words represented in the envelop were identified correctly much more frequently than words represented in the fine structure were as in melody chimeras (developed by exchanging the envelop and fine structure of two different melodies) participants were almost always reported hearing the melody represented by fine structure, which concludes the importance of number of frequency bands needed to identify speech and melody.

Auditory abilities and speech perception are known to mature over the first 10–12 years (Hnath-Chisolm, Laipply, & Boothroyd, 1998; Siegenthaler, 1969) little work has been done to investigate the developmental time course of the ability to use E and TFS speech cues. The only study conducted with children (5–12 years) using noise vocoded speech stimuli demonstrated that the ability to use E cues in speech matures before the age of 7 years, and becomes similar to adults' capacities around the age of 10 years(Eisenberg, Shannon, Schaefer Martinez, Wygonski, & Boothroyd, 2000) (Eisenberg, Shannon, Schaefer Martinez, Wygonski, & Boothroyd, 2000).

The temporal auditory mechanisms involved in the perception of phonetic features are crucial for language acquisition even in infants (Nazzi, Iakimova, Bertoncini, Frédonie, & Alcantara, 2006; Nazzi, Jusczyk, & Johnson, 2000). Research indicates that in adults, cochlear lesions abolish the ability to encode and /or use fine structure cues while preserving the ability to encode and/or use envelop cues (Lorenzi, Gilbert, Carn, Garnier, & Moore, 2006). Impaired perception of envelop cues has been reported repeatedly in the case of central damage to the auditory system and specific language acquisition disorders (Lorenzi, Dumont & Fullgrade, 2000; Lorenzi, Wable et al., 2000). Present study aims to investigate the effect of envelop and fine structure cues across frequency bands on speech perception using language specific material.

Need for the study

Heinz and Swaminathan, (2009) explored the effect of envelop and fine structure in normal hearing individuals especially in foreign tonal and non-tonal languages (Mandarian Chinese & English). The results revealed that the Mandarian Chinese is a tonal language and it was reported that fine structure plays an important role in the perception (Xu & Pfingst, 2005). According to Smith et al, (2002) results revealed that the recognition of English speech (non-tonal language) was dominated by the temporal envelope, whereas recognition of Mandarian speech (tonal language) was dominated by the temporal fine structure. Many studies suggest that the temporal envelope plays an important role in identifying the non-tonal languages and temporal fine structure in identifying the tonal languages.

According to Indu and Devi, (2015) the temporal fine structure plays an important role in identifying lower frequency bands and temporal envelope plays an important role in identifying higher frequency bands for identification of words in Malayalam. However, most of the south Indian languages are non-tonal languages. Hence, there is a need to investigate whether the same trend follows in Kannada language also and to check if there is any difference in cues required for the identification of the chimeric words and chimeric sentences of Kannada language.

Aim

To investigate the influence of frequency bands on perception of Kannada chimeric words and chimeric sentences in individuals with normal hearing.

Objective

- To study the influence of envelop on chimeric word and chimeric sentence identification.
- To study the influence of fine structure on chimeric word and chimeric sentence identification.
- To find out the number of frequency bands needed to clearly differentiate between envelop and fine structure in chimeric words and chimeric sentence.
- To compare identification of chimeric words and chimeric sentences across the frequency bands.

Hypothesis

The following null hypotheses were framed for each main objective of the study. They were

- There are no significant effect of envelope on chimeric word and chimeric sentence identification in Kannada language.

- There is no significant effect of fine structure on chimeric word and chimeric sentence identification in Kannada language.
- There are no significant effect of number of frequency bands for the identification of envelope and fine structure cues.
- There are no significant effect of identification of chimeric words and chimeric sentences across the frequency bands.

Chapter 2

Review of literature

When a sound is received by the cochlea, the frequency content of the signal is mapped into a pattern of excitation along the basilar membrane. The excitation pattern code the spectrum information of the acoustic stimulus, which is concerned to as ‘Spectral’ or ‘Place’ information (Dallos & Fay, 1996). It is assumed that spectral information plays an important role in language recognition as many phonetic features are characterized by their frequency spectrum. Basilar membrane is a bank of band pass filter which has the property of frequency selectivity, with every filter corresponding to an exceptional place on the basilar membrane (Russel et al., 1986). The output signal from these band pass filters carries important temporal information as easily. It is noticed as a gradually varying envelope modulation superimposed on fast oscillations or fine structure (TFS) in the wave form. This temporal information is relayed to the sensory nerve (Smith et al., 2002; Xu & Pfingst, 2003; Zeng et al., 2004).

Auditory nerve (AN) fibers firing rate changes are associated to the signal envelope and the times between spikes, which reflects the TFS or temporal fine structure information (Young & Sachs, 1979; Young, 2008). The relative envelope magnitude across channels carries information that can be used in the auditory system to identify the signal spectral shape and its slow short-term spectral changes. The TFS conveys cues regarding the central frequency of the sound and about its short-term spectrum.

The auditory nerve (AN) fibers phase locking property helps to encode the TFS information. The phase locking property of the AN fibers gets weak at high frequencies and its absent for frequencies above 4-5 kHz for mammalians (Palmer & Russell, 1986). Hence, it is usually assumed that ‘TFS information is not used for frequencies above that boundary’. For robust speech identification envelope or ENV information is significant in silent situation even when provided in as few as four frequency bands (Flanagan,1980; Shannon et al.,1995; Smith et al., 2002). But in the presence of background noise more frequency bands are required for the ENV speech generation process (Qin & Oxenham, 2005; Stone & Moore, 2003). As the results reveal that envelope ENV cues are enough to provide intelligibility in silent, But the identification performance is slightly degraded in the presence of fluctuating noise. The robust speech identification in quiet from ENV cues from relatively small frequency bands is the reason that current cochlear implants provide ENV information over a small number (8 to 16) of electrodes (Wilson et al, 1991). On the other hand, TFS or temporal fine structure is associated with perception of pitch for both gross and complex feelings as well as sound localization (Nelson et al., 2003; Qin & Oxenham, 2003, 2006; Smith et al., 2002; Fullgrabe et al., 2006).

Smith et al., (2002); Xu and Pfingst, (2003); Zeng et al., (2005) had studied the comparative roles of Speech envelope (E) and temporal fine structure (TFS) cues in speech identification. The relative one particular cue while leaving the other intact. One means to accomplish this is through the role of noise or tone vocoders. Vocoded speech is generated by separating out a broadband signal into a number of frequency bands, pulling up the E from every set to modulate a noise or tone carrier and combining the resulting signals from all frequency bands. Of late, various works have aimed out a potential contribution of TFS cues in language perception.

Xu and Zheng (2007) examined the comparative assistance of spectral and temporal cues to phoneme recognition. The spectral cues were altered by altering the number of channels in the vocoder processing, and the temporal cues were altered by changing the cutoff frequency of the envelope (E) extractor low-pass filter. Study included tests for both vowel and consonant recognition. The result revealed that there was a tradeoff between the spectral and temporal cues in phoneme identification, where enhanced spectral cues can make up for reduced temporal cues and vice versa.

Nie et al., (2005), studied on individuals with normal hearing and hearing impairment for identification of spectral and temporal cues. Number of channels and pulse rate were changed in order to alter the amount of spectral and temporal cues. The results revealed that the TFS cues are more efficiently used by the normal hearing individuals than the hearing-impaired individuals. This is because of the reduced capability of hearing impaired persons to understand speech in fluctuating background sound (Moore & Sek, 2009; Lorenzi et al., 2006; Moore et al., 2006; Hopkins & Moore, 2007; Hopkins et al., 2009). This might be due to reduced phase locking ability in hearing impaired subjects. Alternatively, the reduced ability of hearing impaired subjects to profit from FS cues might be induced by reduced ability to decode the TFS case where it is indicated that this process involves cross-correlation of the yields of two spots on the basilar membrane (Loeb et al., 1983; Shamma, 1985). Lastly, the broader tuning of the auditory filters in hearing impaired persons (Glasberg & Moore, 1987) may have a substantial purpose in their miserable performance in understanding TFS information. This is imputable to the limited frequency selectivity of the cochlear filters which has difficulty in decoding the complex and rapidly variable TFS information (Moore, 2008).

Lorenzi et al. (2006) evaluated the contribution of TFS in speech identification, consonant identification in nonsense processed vowel-consonant-vowel (VCV) stimuli was used. The individuals with hearing impairment and normal subjects were tested with TFS-only stimuli generated from nonsense VCV words by separating out the original signal into 16 contiguous frequency bands, computing the E and FS in every band using the Hilbert transform, and combining the TFS signals from the different frequency bands to construct the last stimulus. Results revealed that normal hearing subjects show significant intelligibility for TFS cues, where up to 90% recognition is reported after some training. Moore (2008) explained the need for training to achieve high recognition scores by the possibility that the auditory system is not applied for processing TFS cues in isolation from ENV cues or that TFS cues in processed stimuli are distorted compared to intact speech and hence training is needed. In a similar experiment (Lorenzi, Debruille, Garnier, Fleuriot, & Moore, 2009), it has been shown that children with normal hearing aged 5 to 7 are able to make use of TFS cues. Results revealed that normal hearing children can use both ENV and TFS cues at the same level as adults, which means that tests for the sensitivity to TFS cues can be performed at this very youthful age for the early spotting of any potential problems in the TFS process.

A different approach to measure the ability of normal and hearing-impaired persons to benefit from FS has been embraced (Hopkins & Moore, 2007, 2009, 2010). Processing of TFS cues is assessed by measuring changes in the speech recognition threshold (SRT). SRT is the least hearing level for spoken communication at which an individual can recognize 50% of the speech material. Hopkins and Moore (2007) quantified the importance of TFS cues by varying the number of frequency channels containing TFS information with the ease of the channels being noise or tone vocoded to suppress any TFS information. To examine the theory that hearing-impaired subjects can create usage of TFS cues only at low frequencies. Hence, removing TFS from low channels would affect the performance while removing TFS from high channels should not hold much meaning. Their answers demonstrate that hearing impaired subjects have less power to constitute use of TFS cues at medium and high frequency when listening in a competing talker background.

Hopkins and Moore (2009) measured the SRTs in normal hearing subjects while varying the cutoff channel which is the frequency band below which the stimulus is left intact, while TFS information is murdered from all bands above it. It was established that the SRT declined significantly as the value of the cutoff channel increased, which suggests that TFS has an significant part in understanding speech in fluctuating background noise. Hopkins and Moore (2009) measured the SRTs for speech processed to contain varying amounts of FS cues. The speech signals were filtered using 30 1-ERBN filters and treated to keep ENV only information or left unprocessed to preserve both ENV and TFS cues. It was noticed that when there are more channels containing TFS cues, SRT were decreased, showing benefits from the introduced cues. Redundancy in TFS information as adding TFS in some channels does not always improve the threshold was also observed. Another experiment was carried out where filtering of the speech signal through 5 6-ERBN channels and brought forth a tone vocoded signal in four of the available five channels. The fifth channel was either absent or was unprocessed. Normal hearing subjects benefitted from the added TFS cues over a spacious range of frequency, while the benefit was less in hearing-impaired subjects.

Gnansia et al., (2009) studied the effects of spectral smearing and degradation of TFS cues on masking release, which is the capability to listen in the dips of the background noise. Stimuli was processed using a spectral smearing algorithm or a tone vocoder technique. The spectral smearing algorithm computes the short-term spectrum using fast Fourier transform, and then the spectrum is smeared by a divisor of two or four using a smearing matrix for 2-ERBN or 4-ERBN auditory filters. Tit was discovered that the fundamental frequency information was more degraded by the vocoder than the spectral smearing algorithm. Masking release was reduced more with the tone vocoder than spectral smearing. Conclusion of study showed that both frequency selectivity and TFS cues are significant for the ability to listen in the dips. Gilbert and Lorenzi (2006) assessed the comparative use of ENV and TFS cues in reconstructing missing information in interrupted speech. In their subject field, four types of sentences processed into 32 frequency bands and information in 21 bands were removed or processed so that the final stimuli have different amounts of ENV and TFS cues. Four types of sentences were used; reference, partially empty, vocoded and partially vocoded. The resulting sentences were still understandable but the intelligibility significantly deteriorated after adding a silence gap. TFS cues have a significant part in reconstructing the broken sentences. The TFS is not sufficient alone, but is practiced along with ENV to understand interrupted speech.

A significant concern regarding the results for TFS contribution to speech understanding is that these effects may be influenced by possible ENV cues in signals. These ENV cues may be due to inefficient signal processing techniques applied to separate TFS from ENV, which is not an easy job given that the TFS and ENV are not totally independent (Ghitza, 2001). Some other significant constituent is the possible recovery of ENV cues by the human auditory filters from a correctly processed signal having only TFS cues. For example, narrow-band filtering can recover the signal ENV from the fine-structure information (Voelcker, 1966). This is especially important in humans because of the sharp cochlear tuning (narrow filters), which facilitate the retrieval of the slow amplitude variations (ENV) from the TFS signal (Ghitza, 2001; Zeng et al., 2004; Heinz & Swaminathan, 2009). Gilbert and Lorenzi (2006), it is argued that recovery of ENV cues from TFS-only signals has minimal contribution to speech recognition when the vocoder analysis filters, which are utilized to generate the TFS-only stimulus, have bandwidth less than

4 ERBN. According to them, using 16 frequency channels should be sufficient to prevent the use of recovered ENV cues. Heinz and Swaminathan (2009), nevertheless, presented physiological evidence for the presence of recovered ENV in chinchilla AN response to chimeric speech. It was computed that ‘Neural cross-correlation coefficients’ to measure the similarity between ENV or TFS to quantify the similarity between ENV (or TFS) components in the spike train responses.

Sheft et al., (2008) presented different ways to reduce the fidelity of ENV reconstruction from TFS signals. The TFS signal can be filtered by an all-pass filter with a random phase response. This is founded on the assumption that ENV and the instantaneous phase are connected, so that processing the TFS signal to produce a mismatch with the original ENV signal will reduce the fidelity of ENV recovery (Schimmel & Atlas, 2005). The other method to reduce the chances of meaningful ENV recovery from TFS cues is to increase the number of analysis filters. When the bandwidth of the analysis filter is narrower than 4 times the normal auditory filter, some studies argued that the role of recovered ENV cues in speech perception is negligible (Gilbert & Lorenzi, 2006).

The last method proposed by Sheft et al., (2008) is to limit the bandwidth of the extracted TFS signal of the analysis filter bandwidth in order to degrade ENV reconstruction. The results of Sheft et al., (2008) show that TFS stimuli, processed to reduce chances of intelligibility from recovered ENV cues, were still highly intelligible (50%– 80% correct consonant identification).

Indu and Devi (2016) investigated speech identification through Malayalam chimeric sentences in normal adult individuals across 8 frequency bands. Results have shown that for Malayalam language perception, fine structure cues are important and envelope perception is not obtained for lesser number of frequency bands. As the number of frequency bands are increased, there was a reduced perception for temporal fine structure cues and increased perception for envelope cues. The same trend was obtained when only fine structure or envelope cues were presented.

Chapter 3

Method

The present study aimed to determine the performance of normal hearing young adults aged 18 to 30 years on the identification of speech using of temporal fine structure cues and temporal envelope cues in auditory chimeric Kannada words and sentences.

3.1 Participants

Participants of the study included 30 normal hearing adults, aged between 18 years to 30 years, who were native speakers of Kannada.

3.1.1 Inclusion criteria:

- Air conduction pure tone hearing thresholds less than or equal to 15 dB HL in both ears at octave frequencies from 250 Hz to 8000 Hz as measured from pure tone audiometry using modified Hughson-Westlake procedure (Carhart & Jerger, 1959).
- Normal middle ear functioning as indicated by ‘A’ type tympanogram (Margolis & Heller, 1987).
- Ipsi-lateral and contralateral acoustic reflex thresholds within 100 dB HL at 0.5 kHz, 1 kHz and 2 kHz.
- Native speakers of Kannada

3.2 Instrumentation

- Dell laptop, core i3 processor loaded with the following softwares: Hilbert transform using MATLAB software [MATLAB 7.12.0 (R2011a)].
- A calibrated 'clinical audiometer' (GSI 61) with TDH 39 earphones enclosed in MX-41/AR supra-aural ear cushions to estimate the air-conduction thresholds, SRT and SIS; and Radio Ear B-71 bone vibrator to estimate the bone conduction thresholds.
- A calibrated 'Grason-Stadler TympStar (version 2)' middle ear analyzer to evaluate the status of the middle ear.
- The audio output of the laptop was routed through a THD-39 head phone housed in MX-41AR supra aural cushions.

3.3 Environment

The tests, including routine audiological evaluations and presentation of chimeric sentences were administered in a sound treated double room where the noise levels are within permissible limits (ANSI S3.1-1999).

3.4 Material

Words for preparing chimeric list were selected from "Development of phonemically balanced word lists in Kannada for adults'' developed by Manjula, Geetha, Sharath and Antony (2012). First 16 lists containing 25 words in each list were selected. Total of 200 pairs of words were taken to prepare speech – speech chimera across eight frequency bands which includes one, four, six, thirteen, sixteen twenty-eight, thirty-two, and sixty-four.

Sentences for preparing chimeric list was selected from “Sentence identification test in Kannada” developed by Geetha et al. (2014). The sentences were selected such that the total number of syllables in each sentence is limited to eight-nine syllables and each word in sentences was not have more than three syllables. Total of eighty pairs of sentences were taken to prepare speech – speech chimera across eight frequency bands which includes one, four, six, thirteen, sixteen twenty-eight, thirty-two, and sixty-four.

3.5 Procedure

The study was carried out in two phases:

Phase I: Preparation of the test stimuli: This involved selection of words and sentences in Kannada and making auditory chimeric stimuli. Each chimeric stimulus was divided into different frequency bands.

Phase II: Administration of the test stimuli: It was carried out in two stages. Which are as follows.

- Stage 1: Administration of the chimeric words for 8 frequency bands.
- Stage 2: Administration of the chimeric sentences for 8 frequency bands.

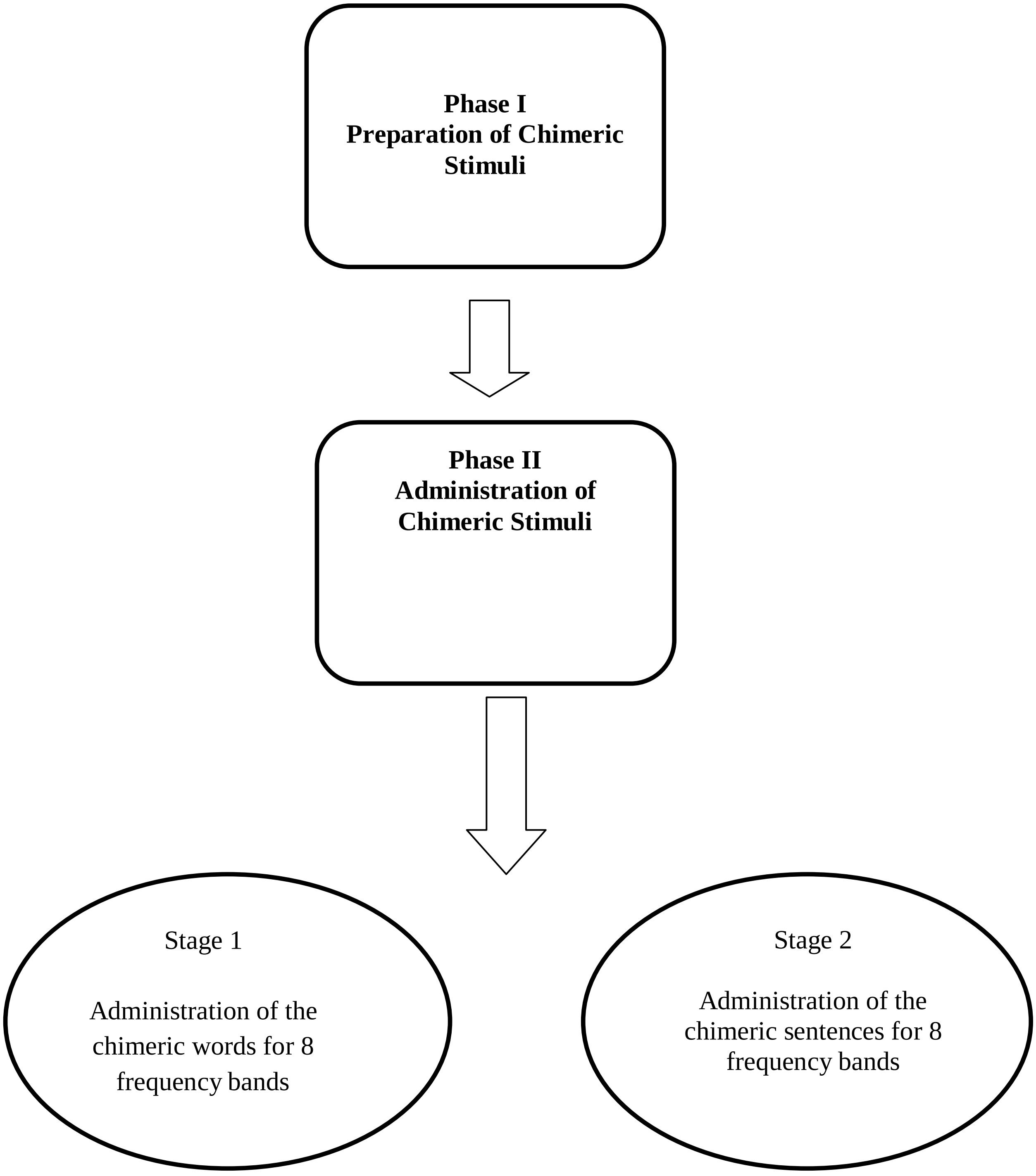

Figure 3.1: Flow chart depicting different phases of the study

3.5.1 Phase 1: Preparation of stimuli and selection of participants:

The selected eighty pairs of sentences were processed using Hilbert transform to extract the temporal cues such as envelope and fine structure. Hilbert transform is mainly used to derive envelope function or instantaneous amplitude of a signal. It mainly represents a filter without affecting the gain (Yost & Fay, 2007).

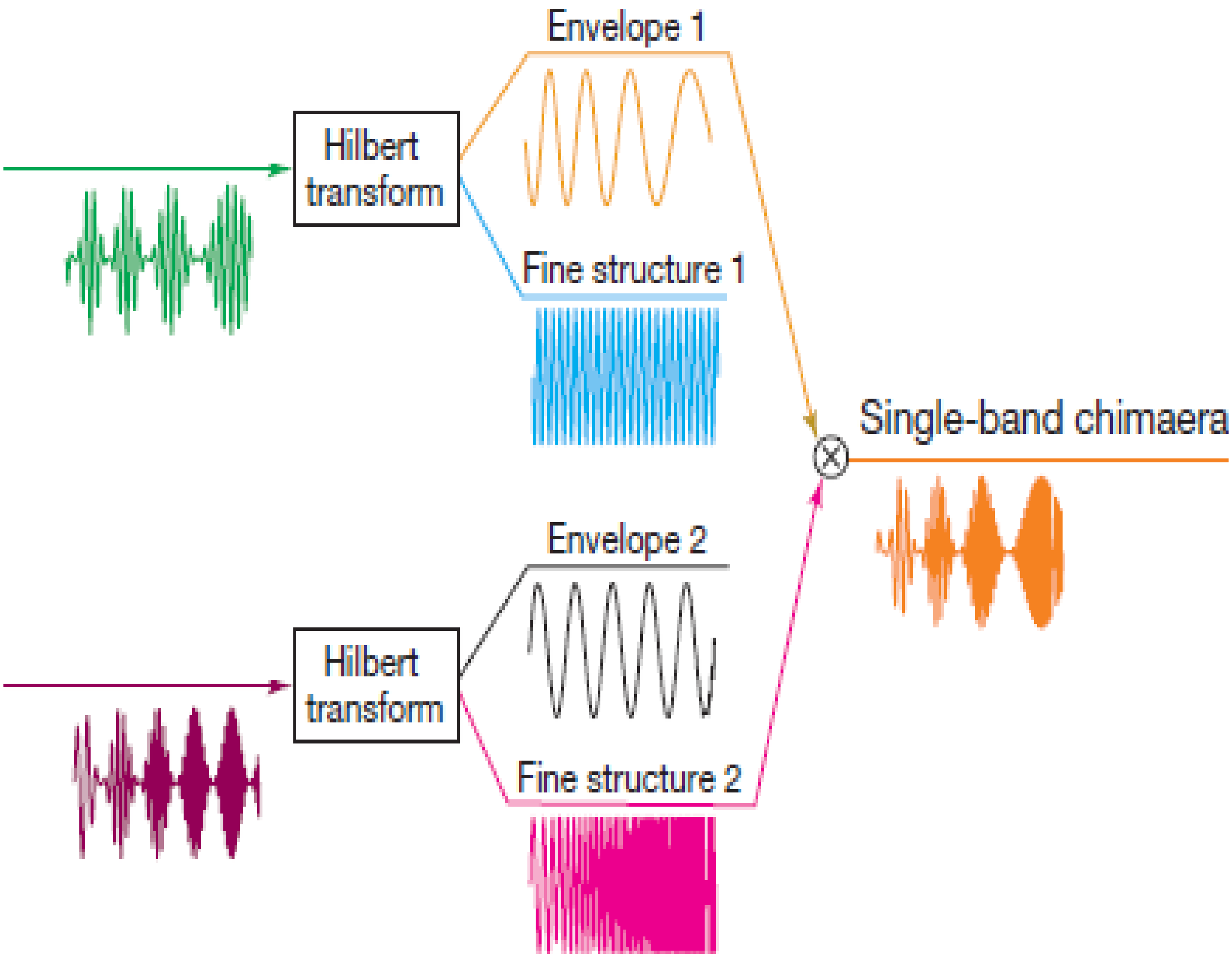

Figure 3.2. Diagrammatic representation of preparation of chimeric stimuli

Hilbert transform is computed in a few steps:

- First, calculate the Fourier transform of the given signal $x(t)$.
- Second, reject the negative frequencies.
- Finally, calculate the inverse Fourier transform, and the result will be a complex-valued signal where the real and the imaginary parts form a Hilbert-transform pair.

For example: When $x(t)$ is narrow-banded, $|z(t)|$ can be regarded as a slow-varying envelope of $x(t)$ while the phase derivative $\partial t\ [\tan{-1}(y/x)]$ is an instantaneous frequency. Thus, Hilbert transform can be interpreted as a way to symbolize a narrow-band signal in terms of amplitude and frequency modulation (Shi, Lee, Liu, Yang, & Wang, 2011). After obtaining envelope and fine structure for each sentence, these temporal cues were exchanged with each other in order to make speech-speech auditory chimeric sentences. For example, envelop of sentence one is combined with fine structure of sentences two to make one chimeric sentence. Likewise, cues were exchanged between all sentences and 80 chimeric sentences were made. Same procedure was followed to make speech-speech chimera for words.

3.5.2 Phase 2: Administration of chimeric stimuli: It was carried out in two stages which are explained in details below

3.5.2.1 Stage 1: Administration of the chimeric words for frequency bands. All chimeric stimuli having frequency bands of 1, 4, 6, 13, 16, 28, 32 and 64 were given through the headphone TDH-39 at the most comfortable level. For each participants practice trial using three chimeric stimuli was given prior to testing. Instruction provided include ‘Listen carefully to each word and repeat back the word that were identifiable’. Speech identification scores were considered based on the percentage of correct words identified from each auditory chimeric word.

3.5.2.2 Stage 2: Administration of the chimeric sentences for frequency bands. All chimeric stimuli having frequency bands of 1, 4, 6, 13, 16, 28, 32 and 64 were given through the headphone TDH-39 at the most comfortable level. For each participant practice trial using three chimeric stimuli was given prior to testing. Instruction provided include ‘Listen carefully to each word and repeat back the word that were identifiable’. Speech identification scores were considered based on the percentage of number of correct key words identified from each auditory chimeric sentence.

Chapter 4

Results

The present study aimed at determining the influence of number of frequency bands on the perception of Kannada chimeric words and chimeric sentences on normal hearing individuals. The data was collected on 30 participants for assessing the speech identification for fine structure and envelope cues across 8 frequency bands (1, 4, 6, 13, 16, 28, 32 & 64). The collected data was tabulated and subjected to statistical analysis using SPSS software version 21.

Results of Shapiro-Wilk test showed no normal distribution across the participants. So, non-parametric tests were carried out for further statistical analysis.

The outcomes of the study are explained under the following:

1. Identification of chimeric words and chimeric sentences across frequency bands
2. Comparison of chimeric word identification across frequency bands
3. Comparison of chimeric sentences identification across frequency bands
4. Comparison between chimeric words and chimeric sentences identification across frequency bands:

4.1 Identification of chimeric words and chimeric sentences across frequency bands:

The participants were presented with eight list of chimeric words representing different frequency bands containing 25 chimeric words and 10 chimeric sentences (4 key words in each). A score of 1 for each correct word identification was given. Whereas, in case of chimeric sentences each correct identification of key words was awarded with 1 score.

The mean raw score and standard deviation scores of speech identification for chimeric words and chimeric sentences across different frequency bands were calculated using the descriptive statistical analysis as indicated in table 4.1.

Table 4.1: Raw Mean and Standard Deviation of Chimeric words and Chimeric sentences across different frequency bands.

Stimulus	Mean	SD
W1	1.53	0.89
W4	3.36	2.12
W6	5.63	1.95
W13	14.83	2.53
W16	17.90	1.62
W28	20.96	1.32
W32	22.46	1.10
W64	23.43	0.89
S1	1.13	1.04
S4	5.66	3.94
S6	21.80	6.58
S13	29.66	5.07
S16	39.93	0.25
S28	40.00	0.00
S32	40.00	0.00
S64	40.00	0.00

Note: W = Chimeric word list, S = chimeric sentence list, 1-64 indicates number of frequency bands.

Above results of table 4.1 shows an improvement of both chimeric word and chimeric sentence identification as the number of frequency band increases. Further figure 4.1 and figure 4.2 shows percentage mean scores for chimeric words and chimeric sentence identification respectively.

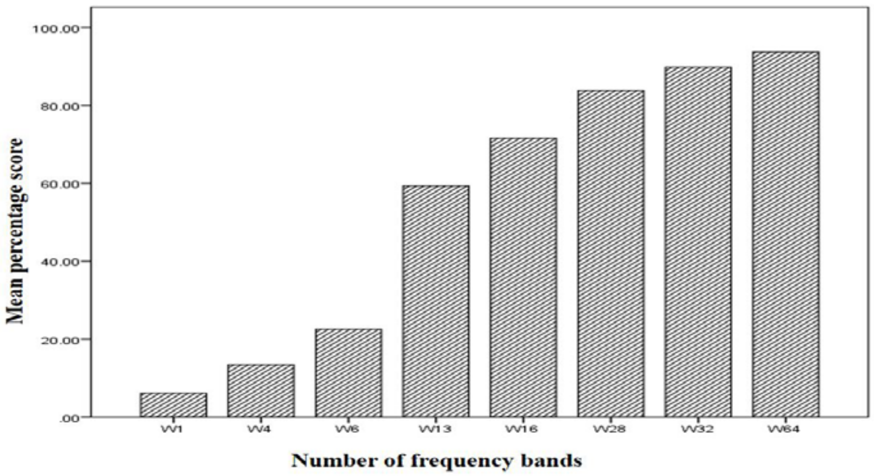

Figure 4.1: Mean percentage score of chimeric word identification across bands

Above figure 4.1 shows percentage score of chimeric word identification across the 8 frequency bands which reveals an improvement in word identification as frequency band increases from W1 to W64.

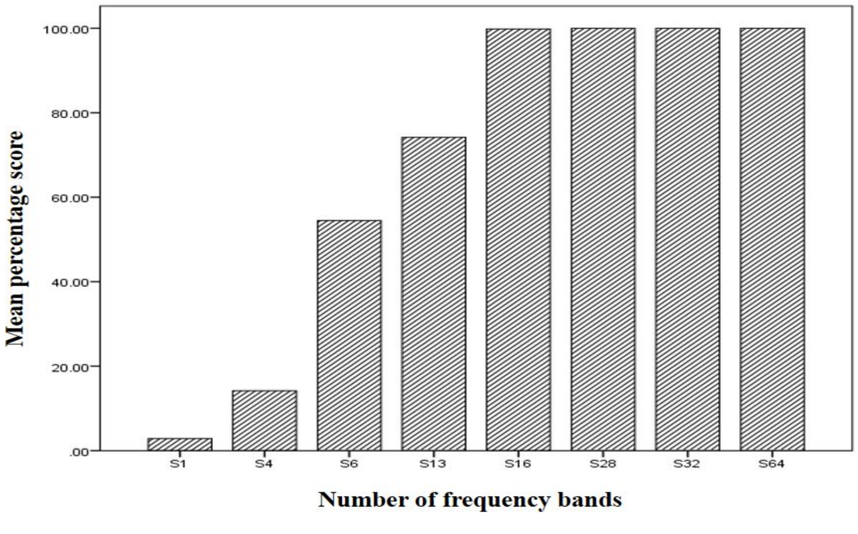

Figure 4.2: Mean percentage score of chimeric sentence identification across bands

Above figure 4.2 shows percentage score of chimeric sentence identification across the 8 frequency bands which shows an improvement in word identification as frequency band increases from S1 to S16 and above that improvement stabilizes till S64.

Further Friedman test was administered to check if there is any difference across bands for chimeric words and chimeric sentences. The results showed a significant difference among bands both in case of chimeric words and chimeric sentences as shown in table 4.2.

Table 4.2: Result of friedman test for **c**omparison of chimeric words and sentences across all frequency bands.

Stimuli	**$\chi 2$**
Chimeric words	200.64*
Chimeric sentences	206.67*

*Note: * indicates p <0.01*

Since there was a significant difference further pairwise comparison across different frequency bands was done separately for chimeric words and chimeric sentences.

4.2 Comparison of chimeric word identification across frequency bands:

Since the Friedman test revealed significant differences across bands, Wilcoxon signed rank test was done to find out pairwise comparison across different bands of chimeric words and results are shown in table 4.3.

Table 4.3: Pairwise comparison of frequency bands for identification of chimeric words.

	W1	W4	W6	W13	W16	W28	W32	W64
W1		3.64*	4.72*	4.81*	4.79*	4.81*	4.81*	4.85*
W4			4.20*	4.80*	4.79*	4.81*	4.79*	4.80*
W6				4.80*	4.79*	4.79*	4.81*	4.79*
W13					4.33*	4.81*	4.79*	4.80*
W16						4.68*	4.83*	4.81*
W28							3.46*	4.84*
W32								3.43*
W64								

*Note: *p < 0.01, W = Chimeric word list, 1- 64 indicates number of frequency bands*

Results from above table 4.3 revealed a significant difference for all pairwise comparison across the frequency bands.

4.3 Comparison of chimeric sentences identification across frequency bands: Since the Friedman test revealed significant differences across bands, Wilcoxon signed rank test was done to find out pairwise comparison across different bands of chimeric sentences and results are shown in table 4.4.

Table 4.4: Results of pairwise comparison of bands for sentence identification

	S1	S4	S6	S13	S16	S28	S32	S64
S1		4.55*	4.79*	4.79*	4.85*	4.85*	4.85*	4.85*
S4			4.79*	4.79*	4.80*	4.81*	4.81*	4.81*
S6				4.24*	4.79*	4.79*	4.79*	4.79*
S13					4.81*	4.82*	4.82*	4.82*
S16						1.41	1.41	1.41
S28							0.00	0.00
S32								0.00
S64								

*Note: *p < 0.01, S = Chimeric Sentence list, 1- 64 indicates number of frequency bands*

Results shows a significant difference across all pairwise comparison except for comparison of bands S28-S16, S32-S16, S64-S16, S32-S28, S64-S28 and S64-S32.

4.4 Comparison between chimeric words and chimeric sentences identification across frequency bands:

Results from table 4.1 shows mean values across frequency bands for chimeric sentences is higher as compared to chimeric words. Wilcoxon signed rank test was done to find out pairwise comparison across different bands of chimeric words and results are shown in table 4.5.

Table 4.5: Results of pairwise comparison of frequency bands for identification of chimeric words and sentence.

Pairs	**/Z/**	**p**
S1 - W1	3.62	0.00*
S4 - W4	0.62	0.53
S6 - W6	0.78	0.00*
S13 - W13	3.76	0.00*
S16 - W16	4.82	0.00*
S28 - W28	4.81	0.00*
S32 - W32	4.82	0.00*
S64 - W64	4.52	0.00*

*Note: * indicates p <0.01 W = Chimeric word list, S = Chimeric Sentence list, 1- 64 indicates number of frequency bands*

Above results of table 4.5 shows a significant difference across all pairwise comparison except for comparison of band S4 and W4.

4.5 Number of frequency bands needed to clearly differentiate between envelop and fine structure in chimeric words:

Participants were presented 200 chimeric words across 8 frequency bands (25 in each). Task was to identify the words and repeat them. Results of the responses for participants are presented in table 4.6.

Table 4.6: Representation of frequency (%) for identification of fine structure and envelope cues across frequency bands in chimeric words

Frequency band	**Fine Structure (frequency %)**	**Envelope (frequency %)**
W 1	90	10
W 4	10	90
W 6	20	80
W 13	0	100
W 16	0	100
W 28	0	100
W 32	0	100
W 64	0	100

Note: W = Chimeric word list, 1- 64 indicates number of frequency bands

Above table 4.6 results revealed that at lower frequency bands, identification of fine structure cues are prominent as the frequency band increases envelope cues dominates over fine structure cues in chimeric word identification.

4.6 Number of frequency bands needed to clearly differentiate between envelop and fine structure in chimeric sentences:

Participants were presented 80 chimeric sentences across 8 frequency bands (10 in each). All the sentences included 4 key words. Task was to identify the key words and repeat them. Results of the responses for participants are presented in table 4.7.

Table 4.7: Representation of frequency (%) for identification of fine structure and envelope cues across frequency bands in chimeric sentences.

Frequency band	**Fine Structure (frequency %))**	**Envelope (frequency %)**
S 1	93.34	6.66
S 4	0	100
S 6	0	100
S 13	0	100
S 16	0	100
S 28	0	100
S 32	0	100
S 64	0	100

Note: S = Chimeric Sentence list, 1- 64 indicates number of frequency bands

Above table 4.7 result revealed that at lower frequency bands, identification of fine structure cues are prominent as the frequency band increases envelope cues dominates over fine structure cues in chimeric sentence identification.

Chapter 5

Discussion

Sound can be mathematically factored into the product of a slowly varying envelope (also called modulation), and a rapidly-varying fine time structure (also known as carrier). Our goal is to find out which of the two factors (envelope or fine structure) is the most important for auditory perception in case of words and sentences. To do so, we synthesize novel stimuli which have the envelope of one sound and the fine structure of another sound. Those stimuli are called "auditory chimeras".

There are two common method of extracting envelope and fine structure of physical stimuli in each channel vocoders. One way is to ‘Rectify the channel waveform and then to low pass filter the rectified waveform with a cutoff frequency that is below the center frequency of the channel’. Here envelope of physical signal (Envelope of speech signal) is generally well behaved but the form of these types of envelope mainly depends on the cutoff frequency and slope of the low pass filter and the choice of these are usually arbitrary (Licklider & Pollack, 1948). Second common method of estimating envelope of physical signal is using Hilbert transform (Bracewell, 1986). The second method has been used in the current study. When a channel signal is manipulated by the Hilbert envelope, i.e. extracting the envelope cue using Hilbert transform leaving the signal without any envelope fluctuations, corresponding to fine structure of the physical signal, sometimes called as fine structure speech (Lorenzi, 2006).

Problem with this filtering using Hilbert transform is that, after filtering on the basilar membrane envelope cues are reintroduced which contains useful information (Ghitza., 2001; Zeng et al., 2004; Gilbert & Lorenzi,2006). This phenomenon is addressed as Envelope recovery or reconstruction. Which supports the test findings in the current study, i.e. Envelope perception was found to be improved from 4th band to 64th band for speech chimeric identification for both words and sentences.

Comparison of chimeric words and sentences shows better speech identification in case of sentences in all frequency bands except for band 1. Which can be justified by the contextual cues present in sentences. Same results were reported by Indu and Devi (2016), Better sentence identification of Malayalam chimeric sentences in case more numbers of frequency bands. The present study started with the hypothesis that the relative importance of envelope and fine structure for Kannada language resembles that for English speech recognition or not. Results indicated that in Kannada language the individuals mainly depends on the envelope cues for speech perception compared to fine structure. Whereas, Indu and Devi (2016) reported that fine structure cues dominates to identify Malayalam chimeric sentences. Our results qualitatively match with the results of Smith et al (2002), whereas, in present study it was found that participants were only able to identify fine structure cues in case fewer frequency bands. Also, when speech information contained more number of bands it was found that envelope cues were dominated in perception of speech in case of words and sentences (Smith et al., 2002). Poor perception on envelope in lower bands indicates that, “The amount of envelope reconstruction is having a negligible significance in lower frequency bands” (Irino & Patterson, 1997). Our results are in ease with the English speech-speech chimera results of Smith et al. (2002), where it was also found that envelope cues are dominating fine structure for speech perception. Also, the results show significant difference across frequency bands between words and sentence identification of chimeric stimuli. Hence, to achieve the identification of chimeric stimuli perception, both words and sentences lists would be preferred rather than individual.

In the present study comparison of chimeric words and sentences identification across frequency bands shows a significant difference across all the frequency bands except for comparison of S4 and W4. It could be because of inter-subject variability and less number of samples taken for the study.

Chapter 6

Summary and Conclusion

The fine structure cues of the speech have an important role in the perception of tonal language (Mandarin, Chinese) and envelope cues are important for the perception of English language. However, in Indian context most of the languages are non-tonal, only one study has been done in Malayalam language, before generalizing the results, a study was required to investigate the differences in temporal cues required for the perception of other Indian languages. Hence the aim of the study was to investigate the influence of number of frequency bands on the perception of Kannada chimeric words and chimeric sentences in individuals with normal hearing. The objectives aimed were, to study the influence of envelope and fine structure cues on chimeric word and chimeric sentence identification and also the number of frequency bands required to differentiate the cues for envelope and fine structure using chimeric words and chimeric sentences. The study started with the null hypothesis that envelope and fine structure does not have any effect on word identification in Kannada language. The study was carried out in two phases where phase 1 included the preparation of chimeric stimuli and phase 2 included the presentation of stimuli. Phase 2 was carried out in two stages i.e. administration of chimeric word and chimeric sentence identification separately. Descriptive statistical analysis was carried out on SPSS software (version 20). The results revealed that fine structure cues are important at lower frequency bands for the perception of Kannada language. Envelope perception was not present for lesser number of frequency bands and as the number of frequency bands increased, Perception through envelope was improved with reduced fine structure perception was observed. This is the first study on envelope and fine structure cues using chimeric word and chimeric sentences in Kannada language on normal hearing individuals.

Clinical implications

The results of the study could have some clinical implication in the modification of amplification devices. The current study was a preliminary attempt to investigate temporal cues on the language basis using auditory chimera especially in Indian languages. To gain in depth knowledge regarding the same, a detailed study across age groups, hearing and hearing-impaired individuals, in different languages to be carried out.

The present study's results reveal that envelope cues are important for the perception of Kannada language. However, current speech processing strategies employed in the hearing aids and cochlear implants do not efficiently use envelope cues. The current findings of the study highlight the need to include methods such as encoding envelope in the form of frequency modulation, or a race to spike algorithm, or using coherent demodulation in the single band encoder which preserves both envelope and fine structure. For hearing aids the strategies that can be adapted includes spatio-temporal pattern correlation or Neuro-compensator algorithm to provide envelope cues for better perception of Kannada language. In the literature, it has been reported that individuals having cochlear lesion have reduced ability to use envelope and fine structure cues. Hence, the current study highlights the need to focus on envelope in intervention of individuals with cochlear lesion.

Future directions

1) Can study the speech processing strategies in order to develop better algorithms, which may provide better fine structure representations.
2) Extend the work into different Indian languages both tonal and non-tonal.
3) Study using human auditory processing model to find out or estimate the recovered envelope in particular language.
4) Extend the population from normal adults to different hearing impaired population and in individuals using hearing aids and cochlear implants.

References

Bracewell, R. N., & Bracewell, R. N. (1986). *The Fourier transform and its applications* (Vol. 31999): McGraw-Hill New York.

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

Chen, B., Clark, G. M., & Jones, R. (2003). Evaluation of trajectories and contact pressures for the straight nucleus cochlear implant electrode array—a two-dimensional application of finite element analysis. *Medical engineering & physics, 25*(2), 141-147.

Dallos, P. and Fay, A. P. R. (1996). The Cochlea. Springer-Verlag, New York.

Eisenberg, L. S., Shannon, R. V., Schaefer Martinez, A., Wygonski, J., & Boothroyd, A. (2000). Speech recognition with reduced spectral cues as a function of age. *The Journal of the Acoustical Society of America, 107*(5), 2704-2710.

F¨ullgrabe, C., Berthommier, F., and Lorenzi, C. (2006). Masking release for consonant features in temporally fluctuating background noise. *Hear Res*, 211, 74–84.

Flanagan, J. L. (1980). Parametric coding of speech spectra. *J Acoust Soc Am*, 68, 412–41

Geetha, C., Kumar, K. S., Manjula, P., & Pavan, M. (2014). Development and standardisation of the sentence identification test in the kannada language. Journal of Hearing Science, 4(1).

Ghitza, O. (2001). On the upper cutoff frequency of the auditory critical-band envelope detectors in the context of speech perception. *The Journal of the Acoustical Society of America*, *110*(3), 1628-1640.

Gilbert, G., & Lorenzi, C. (2006). The ability of listeners to use recovered envelope cues from speech fine structure. *The Journal of the Acoustical Society of America*, *119*(4), 2438. doi:10.1121/1.2173522

Gilbert, G., Bergeras, I., Voillery, D., & Lorenzi, C. (2007). Effects of periodic interruptions on the intelligibility of speech based on temporal fine-structure or envelope cues. *The Journal of the Acoustical Society of America*, *122*(3), 1336. doi:10.1121/1.2756161

Gnansia, D., P´ean, V., Meyer, B., and Lorenzi, C. (2009). Effects of spectral smearing and temporal fine structure degradation on speech masking release. *J Acoust Soc Am*, 125,4023–4033.

Heinz, M. G., & Swaminathan, J. (2009). Quantifying Envelope and Fine-Structure Coding in Auditory Nerve Responses to Chimaeric Speech. *Jaro-journal of The Association for Research in Otolaryngology*. doi:10.1007/s10162-009-0169-8

Heinz, M., & Swaminathan, J. (2008). Neural cross-correlation metrics to quantify envelope and fine-structure coding in auditory-nerve responses. *Journal of The Acoustical Society of America*. doi:10.1121/1.2932776

Hnath-Chisolm, T. E., Laipply, E., & Boothroyd, A. (1998). Age-related changes on a children's test of sensory-level speech perception capacity. *Journal of Speech, Language, and Hearing Research, 41*(1), 94-106.

Hopkins, K. and Moore, B. C. J. (2007). Moderate cochlear hearing loss leads to a reduced ability to use temporal fine structure information. *J Acoust Soc Am*, 122, 1055–1068.

Hopkins, K. and Moore, B. C. J. (2009). The contribution of temporal fine structure to the intelligibility of speech in steady and modulated noise. *J Acoust Soc Am*, 125, 442–446.

Hopkins, K., King, A., & Moore, B. C. (2012). The effect of compression speed on intelligibility: Simulated hearing-aid processing with and without original temporal fine structure information. *The Journal of the Acoustical Society of America*, *132*(3), 1592. doi:10.1121/1.4742719

Hopkins, K., Moore, B. C., & Stone, M. A. (2008). Effects of moderate cochlear hearing loss on the ability to benefit from **t**emporal fine structure information in speech.*Journal of The Acoustical Society of America*. doi:10.1121/1.2824018

Ibrahim, R. A., & Bruce, I. C. (2002). Effects of Peripheral Tuning on the Auditory Nerve's Representation of Speech Envelope and Temporal Fine Structure Cues. doi:10.1007/978-1-4419-5686-6_40

Indu, T. S., & Devi,N. (2015). Influence of Vocoder Frequency Bands on Perception of Malayalam Chimeric Sentence. Unpublished Dissertation, University of Mysuru, Mysuru.

Irino, T., & Patterson, R. D. (1997). A time-domain, level-dependent auditory filter: The gammachirp. *The Journal of the Acoustical Society of America*, *101*(1), 412-419.

Licklider, J. C. R., & Pollack, I. (1948). Effects of differentiation, integration, and infinite peak clipping upon the intelligibility of speech. *The Journal of the Acoustical Society of America*, *20*(1), 42-51.

Liu, S., & Zeng, F. (2006). Temporal properties in clear speech perception. *The Journal of the Acoustical Society of America*, *120*(1), 424. doi:10.1121/1.2208427

Loeb, G. E., White, M. W., and Merzenich, M. M. (1983). Spatial cross correlation: A proposed mechanism for acoustic pitch perception. Biol Cybern, 47, 149–163.

Loizou,P.C., Dorman,M.,& Tu,Z(1999).on the number of channels neded to understand speech. *The journal of acoustical society of America*,106,2097-2103Lorenzi, C and Gilbert, G. (2006). The ability of listeners to use recovered envelope cues from speech fine structure. *J Acoust Soc Am*, 119, 2438–2444.

Lorenzi, C., Wable, J., Moroni, C., Derobert, C., Frachet, B., & Belin, C. (2000). Auditory temporal envelope processing in a patient with left-hemisphere damage.*Neurocase*. doi:10.1080/13554790008402773

Lorenzi, C., Debruille, L., Garnier, S., Fleuriot, P., & Moore, B. C. (2009). Abnormal processing of temporal fine structure in speech for frequencies where absolute thresholds are normal. *The Journal of the Acoustical Society of America, 125*(1), 27-30.

Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., & Moore, B. C. (2006). Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. *Proceedings of the National Academy of Sciences, 103*(49), 18866-18869.

Manjula P., Geetha C., Sharath K. S., & Antony J (2012). *Development of phonemically balanced word lists in Kannada for adults.* Developed in Department of Audiology. All India Institute of Speech and Hearing, Mysore.

Margolis, R. H., & Heller, J. W. (1987). Screening Tympanometry: Criteria for Medical Referral: Original Papers. *Audiology, 26*(4), 197-208.

Moore, B. C. J. (2008). The role of temporal fine structure processing in pitch perception masking, and speech perception for normal-hearing and hearing- impaired people. *J Assoc Res Otolaryngol*, 9, 399–406.

Moore, B. C. J., Glasberg, B. R., and Hopkins, K. (2006). Frequency discrimination of complex tones by hearing-impaired subjects: Evidence for loss of ability to use temporal fine structure. *Hear Res*, 222, 16–27.

Moore, D. R. (1987). Physiology of higher auditory system. Brit Med Bull, 43, 856–870.

Moore, B. C., & Ernst, S. M. (2012). Frequency difference limens at high frequencies: Evidence for a transition from a temporal to a place code. *The Journal of the Acoustical Society of America*, *132*(3), 1542. doi:10.1121/1.4739444

Moore, B. C., & Sęk, A. (2009). Sensitivity of the human auditory system to temporal fine structure at high frequencies. *The Journal of the Acoustical Society of America, 125*(5), 3186-3193.

Nazzi, T., Iakimova, G., Bertoncini, J., Frédonie, S., & Alcantara, C. (2006). Early segmentation of fluent speech by infants acquiring French: Emerging evidence for crosslinguistic differences. *Journal of Memory and Language, 54*(3), 283-299.

Nazzi, T., Jusczyk, P. W., & Johnson, E. K. (2000). Language discrimination by English-learning 5-month-olds: Effects of rhythm and familiarity. *Journal of Memory and Language, 43*(1), 1-19.

Nelson, P. B., Jin, S., Carney, A. E., & Nelson, D. A. (2003). Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners.Journal of The Acoustical Society of America. doi:10.1121/1.1531983.

Nie, K., Stickney, G., and Zeng, F. G. (2005). Encoding frequency modulation to improve cochlear implant performance in noise. IEEE Trans Biomed Eng, 52, 64–73.

Nuttall, A. L., Brown, M. C., Masta, R. I., & Lawrence, M. (1981). Inner hair cell responses to the velocity of basilar membrane motion in the guinea pig. *Brain research, 211*(1), 171-174.

Palmer, A. R. and Russell, I. J. (1986). Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner haircells. Hear Res, 24, 1–15.

Qin, M. K., & Oxenham, A. J. (2005). Effects of envelope-vocoder processing on F0 discrimination and concurrent-vowel identification. *Ear and Hearing, 26*(5), 451-460.

Russell, I. J., Cody, A. R., and Richardson, G. P. (1986). The responses of inner and outer hair cells in the basal turn of the guinea-pig cochlea and in the mouse cochlea grown in vitro. *Hear Res*, 22, 199–216.

Schimmel, S. and Atlas, L. (2005). Coherent envelope detection for modulation filtering of speech. *In Proc Int Conference on Acoust, Speech, and Signal Processing (ICASSP)*, volume 1, pages 221–224.

Shamma, S. A. (1985). Speech processing in the auditory system. I: The representation of speech sounds in the responses of the auditory nerve. *J Acoust Soc Am*, 78, 1612–1621.

Shannon, R. V., Zeng, F., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech Recognition with Primarily Temporal Cues. Science. doi:10.1126/science.270.5234.303

Sheft, S., Ardoint, M., & Lorenzi, C. (2008). Speech identification based on temporal fine structure cues. *The Journal of the Acoustical Society of America*, *124*(1), 562. doi:10.1121/1.2918540

Shi, J., Lee, W.-J., Liu, Y., Yang, Y., & Wang, P. (2011). *Short term wind power forecasting using Hilbert-Huang Transform and artificial neural network.* Paper presented at the Electric Utility Deregulation and Restructuring and Power Technologies (DRPT), 2011 4th International Conference on.

Siegenthaler, B. M. (1969). Maturation of auditory abilities in children. *International Audiology*, *8*(1), 59-71.

Sit, J. J., Simonson, A. M., Oxenham, A. J., Faltys, M. A., and Sarpeshkar, R. (2007). A low-power asynchronous interleaved sampling algorithm for cochlear implants that encodes envelope and phase information. IEEE Trans Biomed Eng, 54, 138–149.

Smith, Z. M., Delgutte, B., & Oxenham, A. J. (2002). Chimaeric sounds reveal dichotomies in auditory perception. *Nature*, *416*(6876), 87-90. doi:10.1038/416087a

Stone, M. A., & Moore, B. C. (2003). Tolerable hearing aid delays. III. Effects on speech production and perception of across-frequency variation in delay. *Ear and Hearing, 24*(2), 175-183.

Voelcker, H. B. (1966). Toward a unified theory of modulation part I: Phase-envelope relationships. *Proceedings of the IEEE, 54*(3), 340-353.

Von Békésy, G. (1960). Experiments in hearing (Vol. 8). E. G. Wever (Ed.). New York: McGraw-Hill.

Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K., and Rabinowitz W. M. (1991). Better speech recognition with cochlear implants. Nature, 352, 236–238.

Xu, L., & Pfingst, B. E. (2003). Relative importance of temporal envelope and fine structure in lexical-tone perception (L). *Journal of The Acoustical Society of America*. doi:10.1121/1.1623786

Yost, W. A., & Fay, R. R. (2007). *Auditory perception of sound sources* (Vol. 29): Springer Science & Business Media.

Young, E. D. (2008). Neural representation of spectral and temporal information in speech,volume 363.

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

Zeng, F., Nie, K., Liu, S., Stickney, G., Rio, E. D., Kong, Y., & Chen, H. (2004). On the dichotomy in auditory perception between temporal envelope and fine structure cues (L). *Journal of The Acoustical Society of America*. doi:10.1121/1.1777938