Influence of Vocoder Frequency Bands on Perception of Malayalam Chimeric Sentences

Indu T S

Register Number: 14AUD004

This Dissertation is submitted as part fulfilment

for the Degree of Master of Science in Audiology

University of Mysore, Mysuru

ALL INDIA INSTITUTE OF SPEECH AND HEARING

MANASAGANGOTHRI, MYSURU-570 006

MAY 2016

CERTIFICATE

This is to certify that the dissertation entitled "**Influence of vocoder frequency bands on perception of Malayalam chimeric sentences**" is the bonafide work submitted in part fulfillment for the degree of Master of Science (Audiology) of the student (Registration No. 14AUD004). 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, 2016.

CERTIFICATE

This is to certify that the dissertation entitled "**Influence of vocoder frequency bands on perception of Malayalam chimeric 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.

Ms. Devi N

Mysuru

May, 2016.

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

DECLARATION

This is to certify that this dissertation entitled "**Influence of vocoder frequency bands on perception of Malayalam chimeric sentences**" is the result of my own study under the guidance of Ms. 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.

Mysore **Register No:14AUD004**

May , 2016

Acknowledgement

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, Ms Devi N, Lecturer in Audiology, All India institute of speech and hearing, Mysore 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'm. thank you Ma'm 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 kishore tanniru ,Former Lecturer in Audiology 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.

The study would not have come to a successful completion, without the help received from staff of electronics and library. I acknowledge and appreciate them for all their efforts.

Since from the very beginning of my career in speech and hearing my classmates Devika, Haritha, Merin, Rini and the new members to the gang Irfaneechi, Veena, Bincy and Jeena all have extended their support in a very special way and I gained a lot from them through their personal and scholarly interactions. I would like to acknowledge them and all my old pals from NISH, "Chillerzzz" for their well wishes.

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 – my Achan and Amma ,who supported me in every possible way to see the completion of this work.

A special thanks to you Appu for being my backbone in every situation. You have been my constant support ever since i started thinking about my study. Without your help i wouldn't have been even started this work. Thank you for being the kind of man who does all the things without a second thought, Thank you for being a man who doesn't require any kind of praise or gratitude for being a caring and trustworthy best friend to me.

Above all ,I owe it all to almighty god for granding me wisdom ,health and strength to undertake this task and enabling me to its completion.

Table of contents

List of tables

List of figures

Abstract

The temporal cues are play an important role for the perception of different languages. Literature reports that temporal fine structure cues are important for the perception of tonal languages were as envelope cues plays an important role for non-tonal languages. In indian context most of the languages are found to be non-tonal and this study attempts to investigate the effects of temporal cues on an Indian language 'Malayalam' using auditory chimera. The present study was carried out in two phases which includes preparation of stimuli and presentation of the same to normal hearing individuals respectively. 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 cue and increased perception for envelope cues. The same trend was also obtained for when only fine structure or envelope cues were presented.

Chapter 1

Introduction

The human auditory system has amazing capabilities in discriminating and understanding complex sounds. It passes the external sounds to the brain and also analyze them during the transmission. Humans show great sensitivity and take in sounds over the range of 20 Hz to 20 kHz. Sound is converted into vibrations after passing through the outer ear canal. Movements from the eardrum in response to the sound pressure are passed through ossicular chain to the cochlea. 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, 1996). Sensory receptors, known as hair cells, transduce the mechanical wave 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).

Speech perception in humans had been the subject of intensive research 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 (Clark, 2003). When these signals are analyzed in inner ear, resulting in a sequence of band pass signals, each one analogous to a location on 'Basilar Membrane'. Excitation patterns code the spectrum in formation of the acoustic stimulus, which is referred to as 'spectral' or 'place' information (Nuttal et al., 1981). It is assumed that spectral information plays a crucial task in speech recognition as many phonetic features are characterized by their frequency spectrum. Because of the frequency selectivity of inner ear, it acts as a bank of band-pass filters with each filter correlate to a particular position on the BM. The signal at the output of the 'Cochlear Filters' carries important temporal information as well. It can be viewed as a gradually changing envelope modulation superimposed on more fast oscillations or 'Temporal fine structure/fine structure (TFS or FS)' in the waveform (Moore & Sek, 2009). This temporal information is relayed to the afferent AN fibers through changes in the firing rate, which is linked to the signal envelope and the times between spikes, which reflects the TFS information (Young & Sachs, 1979; Young, 2008). The relative envelope magnitude across channels carries information that can be used by the system to identify the signal spectral shape and its slow short-term spectral changes. The TFS conveys cues regarding the 'Fundamental frequency' of the signal and concerning its 'short-term spectrum'. 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 necessery 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). Recognition in background noise, however, requires more frequency bands in ENV speech generation process (Qin & Oxenham, 2003; Stone & Moore, 2003). It has been concluded that E cues are enough to provide fine intelligibility in quiet, while the recognition performance is slightly degraded in fluctuating background noise. The robust speech identification in quiet from ENV cues from relatively small frequency bands is the reason that current cochlear implants provide E over a small number (Eight to sixteen) of electrodes (Wilson et al., 1991). On the other hand, TFS is associated with perception of pitch for both simple and compound tones as well as ‘sound localization’ (Smith, Delgutte & Oxenham,2002; Nelson et al., 2003; Plack & Oxenham, 2005; Qin & Oxenham, 2003, 2006; Stickney et al., 2005; Fullgrabe et al., 2006).

The slow changes in the amplitude of the speech signal (ENV) are the main cues used by the human auditory system to understand speech in quiet (Flanagan, 1980; Shannon, Zeng, Kamath, Wygonski & Ekelid,1995; Smith, Delgutte & Oxenham, 2002). The FS, which is the rapid changes in the speech signal, is generally linked to melody identification and pitch perception as well as sound localization (Qin & Oxenham, 2003, 2006; Nelson et al., 2003; Stickney et al., 2005; Fullgrabe et al., 2006).

To study the relative roles of E and FS cues in speech understanding, several experiments were performed where speech signals are processed to remove E or FS cues. The first attempt to separate the envelope from fine structure cue was done using peak clipping mechanism of speech signal (Licklider & Pollack, 1948) later by using Hilbert transform (Bracewell, 1986). Combining the features of different sounds can be used to check the perceptual importance of E and FS. This is achieved by making hybrid sounds known as auditory chimera. Smith, Delgutte & Oxenham (2002), elobrated a process to split FS from E where two acoustic waveforms are processed with a bank of band pass filters followed by the 'Hilbert transform' to create 'ENV-only and TFS-only' condition of the signals. In each band, the envelope of one waveform is multiplied by the FS of the other. The stimuli are then summed across frequency bands to construct the auditory chimaeras. 'Speech-speech chimaeras' are constructed when both waveforms contains speech signals. However, to make 'Speech-noise chimaeras', one waveform is the speech signal and the other is noise. Another approach is to compute the envelope in each frequency band using rectification and low-pass filtering (Shannon, Zeng, Kamath, Wygonski & Ekelid,1995). The benefit of this approach is that it produces a known, fixed maximal bandwidth of the ENV signals, determined by the cutoff frequency of the low-pass filter. However, rectification is a nonlinear process which introduces distortions that affect the quality of the extracted envelope.

General findings suggest that temporal mechanisms implicated in the perception of phonetic features are vital for language developement even in infants (Nazzi, Floccia & Bertonini, 1998; Nazzi et al., 2000, 2006). Research shows that in older individuals ,cochlear lesions looses the ability to encode and /or use fine structure cues while preserving the capacity to use envelop cues (Lorenzi et al., 2006). Reduced perception of envelop cues has been reported repeatedly in the case of 'central damage' to the auditory system and 'language acquisition disorders' (Lorenzi, Dumont & Fullgrade, 2000; Wable et al ., 2012.)

Need for the study

There are previous studies exploring the effect of envelope and fine structur' in normal hearing and hearing loss individuals especially in foreign languages like Mandarian Chinese and English (Heinz & Swaminathan,2009). Mandarian Chinese is a tonal language and it was reported that fine structure plays an important role in the perception (Xu & Pfingst, 2005), which is in contradiction with that of English. However, in Indian context most of the languages are non-tonal and hence a study is required to investigate if there could be any differences in cues required for the identification of Indian languages.

Aim

To investigate the influence of vocoder frequency bands on the perception of Malayalam chimeric sentences in individuals with normal hearing

Objective

- To study the influence of envelope cues on word identification in a sentence
- To study the influence of fine structure cues on word identification in a sentence
- To find out the number of frequency bands needed to clearly differentiate between envelope and fine structure of chimeric sentence

Hypothesis

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

- There are no significant effect of envelope on word recognition in Malayalam language.

- There are no significant effect of fine structure on word recognition in Malayalam language.
- There are no significant effect of number of frequency bands for the identification of envelope and fine structure

Chapter 2

Review

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 believed that spectral information plays a important function in language recognition as many phonetic features are characterized by their frequency spectrum. Because of the frequency selectivity of the cochlea, it behaves as a collection of band-pass filters with every filter corresponding to a exceptional place on the basilar membrane (Russel et al.,1986). The signal at the end product of the cochlear filters carries important temporal information as easily. It is viewed 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.,2005).

Auditory nerve (AN) fibers through changes in the firing rate, which is 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 TFS information is encrypted 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 usually assumed that 'TFS information is not used for frequencies above that boundary'.

Envelope or ENV information is significant for speech understanding and supports robust speech identification in silent when provided in as few as four frequency bands (Flanagan,1980; Shannon et al.,1995; Smith et al., 2002). Recognition in background noise, however, requires more frequency bands in the ENV speech generation process (Qin & Oxenham, 2003; Stone & Moore, 2003). Results reveal that E cues are enough to provide intelligibility in silent, while the identification performance is slightly degraded in 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 (Moore, 2003; Nelson et al., 2003; Qin & Oxenham, 2003, 2006; Smith et al., 2002; Stickney et al., 2005; Plack & Oxenham, 2005; Fullgrabe et al., 2006).

Smith et al., 2002; Xu and Pfingst, 2003; Zeng et al., 2005 have investigated the comparative roles of Speech-E and FS 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. In their experimentation, they processed syllables to create vocoders with variable amount of spectral and temporal cues. Spectral cues are altered by altering the number of channels in the vocoder processing, while temporal cues are altered by changing the cutoff frequency of the E extractor low-pass filter. The experiment tested both consonant and vowel recognition and proved that there was a tradeoff between the spectral and temporal cues in phoneme identification, where enhanced spectral cues can make up for reduced temporal ones and vice versa.

Nie et al., (2005) studied spectral and temporal cues in individuals with hearing impairment and normal hearing individuals. They altered the amount of spectral and temporal cues by changing the number of channels and pulse rate, respectively It was noticed that normal hearing subjects can create usage of TFS cues more than hearing impaired subjects and this was related to the reduced capability of hearing impaired persons to understand speech in fluctuating background sound (Moore & Skrodzka, 2002; Moore, 2003; Lorenzi et al., 2006; Moore et al., 2006; Hopkins & Moore, 2007; Hopkins et al., 2008). The results indicate that 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, 1986) 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, 2008b).

Consonant identification in nonsense processed vowel-consonant-vowel (VCV) stimuli was used in Lorenzi et al., (2006) to evaluate the contribution of TFS to speech identification. Normal and hearing-impaired 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. Their results show that normal hearing subjects show significant intelligibility for TFS cues, where up to 90% recognition is reported after some training. Moore (2008b) 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 et al., 2009), it has been shown that children with normal hearing aged 5 to 7 are able to make use of TFS cues. They reasoned out 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 in the work of Hopkins and 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. They desired 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. They establish 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 (2010) 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. They noticed that when there are more channels containing TFS cues, SRT were decreased, showing benefits from the introduced cues. They also noted some redundancy in TFS information as adding TFS in some channels does not always improve the threshold. They performed another experiment where they filtered 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. They processed the stimuli 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. They have 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. They concluded that both frequency selectivity and TFS cues are significant for the ability to listen in the dips. Gilbert and Lorenzi (2010) assessed the comparative use of ENV and TFS cues in reconstructing missing information in interrupted speech. In their subject field, they used 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. They generated four types of sentences; reference, partially empty, vocoded and partially vocoded. The resulting sentences were still understandable but the intelligibility significantly deteriorated after adding a silence gap. They showed that TFS cues have an 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; Rice, 1973; Logan,1977). 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 responses to chimeric speech. They have also computed '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).

Chapter 3

Method

The study aimed to determine the performance of normal hearing young adults aged 17 to 30 years on the identification speech using of temporal fine structure cues and temporal envelope cues inauditory chimeric Malayalam sentences.

The study was carried out in two phases:

Phase I: Preparation of the test stimuli

This involved selection of sentences in Malayalam and making auditory chimeric stimuli out of it and each chimeric stimulus was divided into different frequency bands also selection of participants.

Phase II: Administration of the test stimuli

Stage 1: Administration of the chimeric sentences for 5 frequency bands

Stage 2: Administration of the chimeric sentences for 8 frequency bands

Participants

- Participants of the study included 30 normal hearing adults, aged between 17 years to 30 years (Mean = 21.03, SD = 3.52) who were native speakers of Malayalam.

- 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).

- Ipsilateral and contralateral acoustic reflex thresholds within 100 dB HL at 0.5 kHz, 1 kHz and 2 kHz.

- Native speakers of Malayalam

Instrumentation

- Sony vaio laptop, core i3 processor loaded with the following softwares:

- Hilbert transform using MATLAB software [MATLAB 7.12.0 (R2011a)] Vocoder software (serious vocoder developed by Zerius Development Inc) for imposing the vocal effects of speech to another sound.

- Paradigm V2.5 software (available from http://www.paradigmexperiments.com/) for the presentation of chimeric stimuli.

- 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.

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).

Material

Seven Malayalam sentences having same syllabic count were selected from ‘Sentences in Malayalam’ list (Sreeraj & Kishore, 2012). The total number of syllables in each sentence is limited to eight-nine syllables and each word in sentences were having not more than three syllables. Speech –speech chimera was developed using the seven sentences which has been synthesized to form eight frequency bands out of which five frequency bands including one, four, six, thirteen, sixteen was used for the second phase – stage 1 of the study. Remaining three frequency bands including twenty-eight, thirty-two, and sixty-four was used in the Second phase – stage 2 of the study .

Procedure

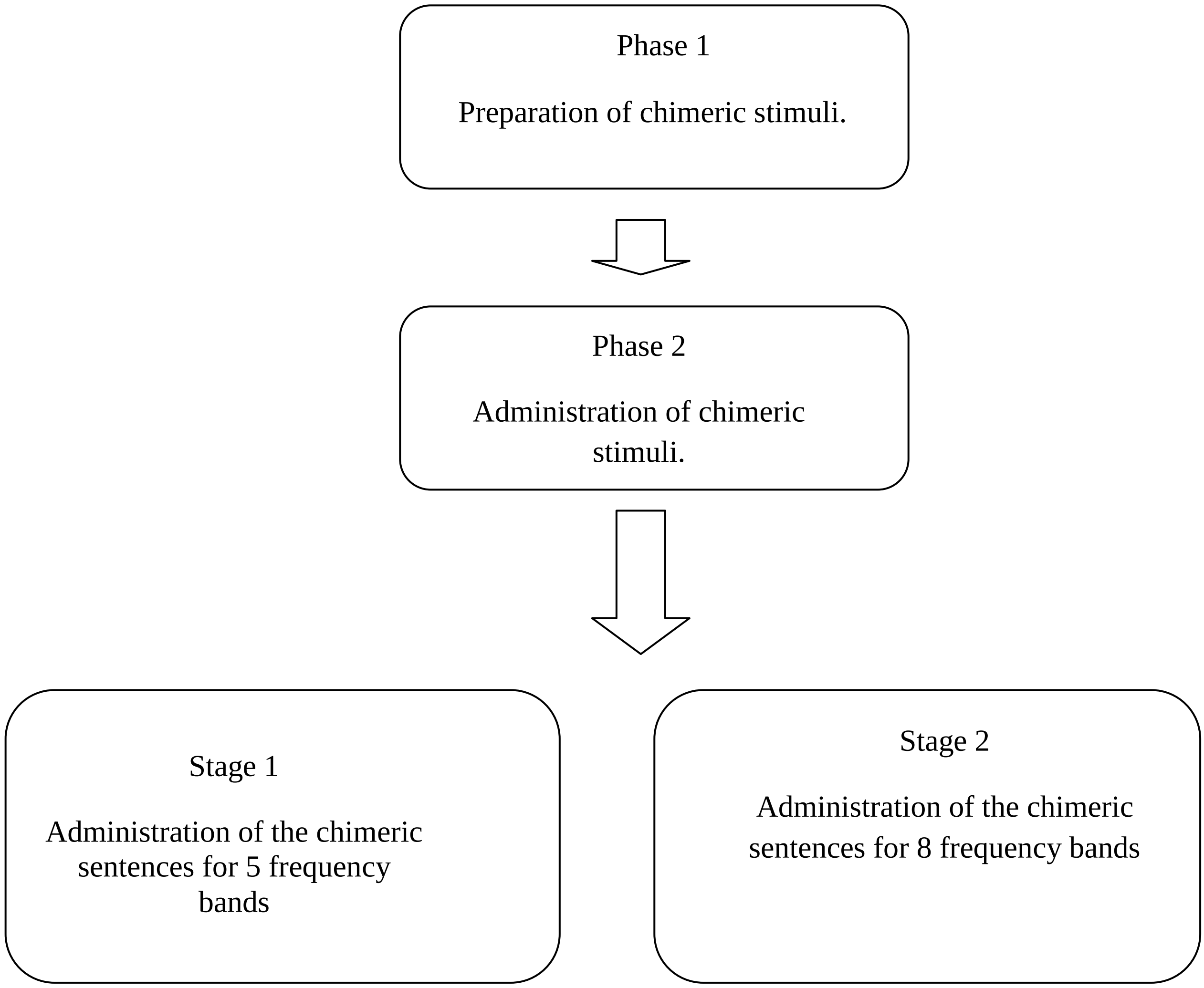

Figure 3.1. A flow chart of the different phases of the study

Phase 1: Preparation of stimuli and selection of participants

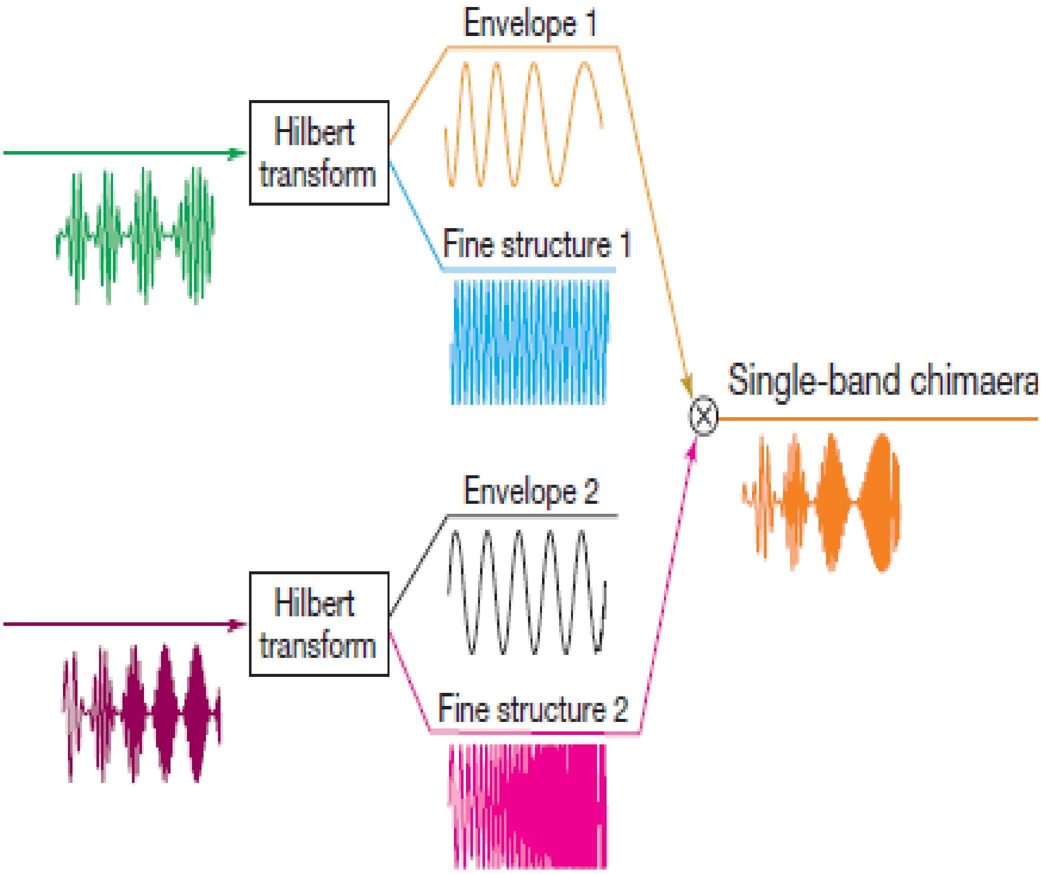

Figure 3.2. Diagrammatic representation of preparation of chimeric stimuli

The selected seven Malayalam 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, Popper & Fay,2008)

- 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 ∂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 (Liu, 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, E of sentence one is combined with fine structure of sentences two to make one chimeric sentence. Likewise cues were exchanged between all sentences and 30 chimeric sentences were made. Each sentence was divided into different frequency bands using vocoder software with 50% window overlap and was normalized. All the stimuli were generated using a sampling frequency of 44.1 kHz.

Phase 2: Administration of chimeric stimuli

Stage 1- Administration of the chimeric sentences for 5 frequency bandsAll chimeric stimuli having frequency bands of one, four, six, thirteen, and sixteen were routed via two channel clinical audiometer (GSI-61) using paradigm software. The output through the headphones was maintained at 65dBHL. For each participants practice trial using three chimeric stimuli was given prior to testing. Post practice trials, each participant were provided with seventy speech–speech chimeric sentences. Instruction provided includes

‘Listen carefully to each sentence and repeat back the words that were identifiable in each sentence’. Speech identification scores were considered based on the percentage of correct words identified from each auditory chimeric sentences.

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

Additionally chimeric sentences having three more frequency bands such as twenty eight, thirty two, and sixty four were provided two channel clinical audiometer (GSI -61). Also, speech – silence chimeric stimuli was created by providing envelope of the stimuli intact and fine structure cue is exchanged with silence having same duration and vice versa.In this phase also, speech identification scores were calculated based on the percentage of number of correct words identified from each auditory chimeric sentences.

Statistical analysis

To fulfill the objectives of the study, the data was tabulated and subjected to appropriate statistical analyses.

- ‘Statistical Package for Social Sciences (SPSS) software version 20 (SPSS Inc. Chicago)’ was used for descriptive statistics and parametric/non parametric tests.
    - Mean and standard deviation was measured for all the data obtained. Following this, Shapiro-Wilk test of normality was

administered. As indicated by the normality test, parametric/non parametric tests were used.

- Whenever main effects or interactions were significant, the post hoc test was done using pairwise comparisons applied for multiple comparisons.

Chapter 4

Results

The present work aimed at studying the influence of vocoder frequency bands on the perception of Malayalam chimeric sentences on normal hearing individuals.The data was collected initially on 30 participants across 5 frequencies band (1, 4, 6, 13 & 16) further 10 more participants were randomly selected within the 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 20. First descriptive statistics (mean, median and standard deviation) were reported for all the measurements. Following this, Shapiro-Wilk test of normality was administered. As indicated by the normality test, parametric/non parametric tests were used as the case may be. Whenever main effects or interactions were significant, the post hoc test was done using pairwise comparisons

The outcomes of the study are explained under the following:

1. Fine structure cue for speech identification
2. Envelope cue for speech identification.
3. Speech identification scores across different frequency bands

4.1. Fine structure and envelope cue for speech identification scores in speech-speech chimeric condition

The mean, median and standard deviation values of speech identification scores for fine structure cue across different frequency bands were calculated using the descriptive statistical analysis as indicated in table 4.1

Table 4.1.

Mean, Median and Standard Deviation of Speech Identification Scores for Fine Structure Cue Across Different Frequency Bands (1-16).

Fine structure Frequency bands	**Mean (%)**	**Std. Deviation**	**Median (%)**
FS1	18.43	10.09	17.50
FS4	21.30	10.59	21.00
FS6	19.73	9.69	19.00
FS13	24.70	12.08	23.00
FS16	24.00	10.97	24.50

Note: *FS: fine structure in speech-speech chimeric condition, 1-16: frequency bands).*

Repeated measure ANOVA across different frequency bands (1, 4, 6, 13, & 16) as within subject factor was carried out to examine the essence of fine structure cues in speech-speech chimeric condition . The results revealed that there was a significant main effect of number of frequency bands on fine structure cue identification [F (4,30) =8. 441, $p < 0.05$]. This can be inferred that as the number of frequency bands increased fine structure perception also increases . However no envelope perception was observed in frequency bands from one to sixteen bands. Hence the hypothesis, there are no significant effect of fine structure on word identification in Malayalam language, is rejected.

Wilcoxon signed rank test was done to compare the speech identification scores of fine structure cues in each frequency bands.

Table 4.2.

Pairwise comparison across 5 frequency bands.

Fine structure frequency bands	FS1	FS4	FS6	FS13	FS16
FS1	-	2.88*	-1.30	6.27*	5.57*
FS4		-	1.56	-3.40	-2.70
FS6			-	4.97*	-4.27
FS13				-	0.70
FS16					-

*Note: *p <0.05*

Results showed that for fine structure perception there was no difference found for frequency band one (FS1) and six (FS6). Fine structure of frequency band four (FS4) was not set up to be different from FS6, FS13, and FS16. FS6 found to be different from FS4, FS13 have significant difference with FS1 and FS6, also FS16 significantly different with FS1. The frequency bands were the fine structure perception reaches almost 50% closer is in FS13 and FS16. Lowest speech identification scores were obtained for FS1 and maximum scores were obtained for FS13.

Speech perception scores for fine structure were increasing as the number of frequency bands increases. In all the frequency band's, envelope perception was found to be nil, indicating for Malayalam language perception fine structure cues are playing the major role compared to envelope cues. Hence the hypothesis there are no significant effect of envelope on word identification in Malayalam language, is accepted.

Since the individuals were not able identify the envelope cues till 16 frequency bands chimeric stimuli with additional frequency bands (28,32 and 64) were administered on 10 individuals who were randomly selected from the participants inorder to see the effect of envelope.

4.2.Fine structure and envelope perception in speech-speech chimera for 8 frequency bands.

The mean, median and standard deviation values of speech identification scores for fine structure cue (Table 4.3) and envelope (Table 4.4) across different frequency bands were calculated using the descriptive statistical analysis as indicated.

Table 4.3.
Mean,Median and Standard deviation of fine structure speech identification score in different frequency band (1-64) for speech-speech chimera

Fine structure Frequency bands	Mean(%)	Std. Deviation	Median(%)
FS1	35.00	20.91	39.00
FS4	41.60	22.10	42.00
FS6	35.00	20.89	37.00
FS13	47.60	22.93	55.00
FS16	44.60	24.73	49.00
FS28	27.80	19.07	29.00
FS32	23.60	16.37	23.00
FS64	18.00	11.35	14.00

Note:FS-fine structure in speech-speech chimeric condition,1-64:frequency bands.

Table4.4.

Mean,Median and Standard Deviation of Envelope Speech Identification Score in Different Frequency Band (1-64) for Speech-Speech Chimera.

Envelope frequency bands	Mean(%)	Std. Deviation	Median(%)
E1	.00	.000	.00
E4	.00	.000	.00
E6	.00	.000	.00
E13	.00	.000	.00
E16	.00	.000	.00
E28	5.00	3.29	4.00
E32	4.80	2.52	4.00
E64	4.20	3.04	3.00

Note : E-envelope in speech –speech chimeric condition,1-64 :frequency bands.

From the above results it can be inferred that as the number of frequency bands increased fine structure perception was increasing (Table 4.3). The maximum speech identification scores for fine structure perception was obtained at FS13 and found to reduce as the number of frequency bands increased (Table 4.3) also envelope perception was found to start from 28^{th} frequency bands (Table 4.4).

A non parametric Friedman test of differences among repeated measures was conducted and rendered χ^2 (7) =18.087, p <0.05(FS) and χ^2(7) =60.640, p <0.05 (E). Further a Wilcoxon signed rank test was run for the comparison of speech identification scores of envelope and fine structure in different frequency bands.

Table 4.5.

Pair wise comparison of all the frequency bands for fine structure perception speech-speech chimera.

Fine structure Frequency bands	FS1	FS4	FS6	FS13	FS16	FS28	FS32	FS64
FS1	-	-2.44	-0.07	-2.25	-2.49*	-0.89	-0.97	-1.84
FS4		-	-2.72*	-0.87	0.00	-1.17	-1.84	-2.51
FS6			-	-2.62*	-2.51*	-0.53	-1.63	-1.68
FS13				-	-0.89*	-1.68	-1.99*	-2.29*
FS16					-	-1.27	-2.19*	-2.24*
FS28						-	-2.53*	-1.94
FS32							-	-1.48
FS64								-

*Note: *p <0.05*

Results of Table 4.5 showed that for fine structure perception there was significant difference found for frequency band one (FS1) and sixteen (FS16), FS 4 with FS6, FS6 with FS13 and FS16, FS13 with FS16 ,FS32 and FS64,FS16 with FS 32 and FS64 ,FS28 with FS32.

Table 4.6:

Pairwise comparison of all frequency bands for envelope perception in speech-speech chimera.

Envelope Frequency bands	E1	E4	E6	E13	E16	E28	E32	E64
E1						-2.82*	-2.76*	-2.69*
E4						-2.82*	-2.76*	-2.69*
E6						-2.82*	-2.76*	-2.69*
E13						-2.82*	-2.76*	-2.69*
E16						-2.82*	-2.76*	-2.6*
E28							-0.073	-0.54
E32								-0.57
E64								

*Note: *p <0.05*

Results of Table 4.6 indicates that there was significant difference for envelope perception between frequency bands 1, 4, 6, 13, 16 with 28 ,32 and 64. The null hypothesis is rejected for fine structure in all the frequencies, but accepted for envelope till 16^{th} frequency band.

4.3.Fine structure only condition (FSO)

The mean, median and standard deviation values were calculated using the descriptive statistical analysis.

Table4.7.

Mean,Median and Standard Deviation of Fine Structure Speech Identification Score in Different Frequency bands(1-64).

Fine structure Frequency bands	**Mean (%)**	**Std. Deviation**	**Median (%)**
FSO1	18.00	20.30	10.00
FSO4	19.00	20.24	12.50
FSO6	25.50	19.50	25.00
FSO13	30.00	21.47	25.00
FSO16	45.00	20.27	50.00
FSO28	10.00	4.37	9.00
FSO32	11.10	3.84	11.00
FSO64	11.10	3.84	11.00

Note: FSO –fine structure only ,1-64 number of frequency bands

From the above Table 4.7 it can be inferred that as the number of frequency bands increased fine structure perception was increasing. The maximum scores for fine structure perception was obtained at 16^{th} frequency bands. Also ,as the number of frequency bands are increased from 28 to 64 there found to have a reduction in the speech identification score for fine structure.

A non parametric Friedman test of differences among repeated measures was conducted and rendered χ^2 (7) =48.587, $p < 0.05$. Further a Wilcoxon signed rank test was run for the comparison of different frequency bands.

Table 4.8.

Pair wise comparison of all frequency bands for envelope perception in speech-speech chimera.

Fine structure Frequency bands	FSO1	FSO4	FS06	FSO13	FSO16	FSO28	FSO32	FSO64
FSO1		-1.41	-2.21*	-2.53*	-2.81*	-1.13	-.83	-.83
FSO4			-2.21*	-2.53*	-2.81*	-1.54*	-1.19*	-1.19*
FSO6				-1.60	-2.82*	-2.55*	-2.49*	-2.49*
FSO13					-2.53*	-2.66*	-2.67*	-2.67*
FSO16						-2.81*	-2.81*	-2.81*
FSO28							-2.03*	-2.03*
FSO32								0.00
FSO64								

*Note: *p <0.05*

Results of Table 4.8 indicates that speech identification scores of fine structure only condition in first frequency band (FSO1) is significantly different from FSO6, FSO13 and FSO 16. FSO4 is found significantly different from FSO6,FSO13,FSO16,FSO28,FS032 and FSO64.6th frequency band (FSO6) is significantly different from FSO16, FSO28, FSO32 and FSO64. Significant difference is also found to be present between FSO13 with FSO16, FSO28, FSO32 and FSO64. FSO16 is also found to significantly different from FSO28, FSO32 and FSO64. Results also indicates that FSO28 is also significantly different from FSO32 and FSO 64.

4.4.Envelope only condition (EO)

The mean, median and standard deviation values of speech identification scores for envelope only condition across different frequency bands were calculated using the descriptive statistical analysis as indicated in table 4.9.

Table 4.9.

Mean ,Median and Standard Deviation of Speech Identification Scores for Envelope Only Condition in Different Frequency Bands(1-64).

Envelope frequency bands	Mean(%)	Std. Deviation	Median(%)
EO1	.00	.00	.00
EO4	.00	.00	.00
EO6	.00	.00	.00
EO13	7.00	9.18	.00
EO16	8.50	11.06	.00
EO28	10.00	11.05	7.50
EO32	25.00	11.78	25.00
EO64	25.00	11.78	25.00

Note:EO : Envelope only,1-64:frequency bands.

Table 4.9 indicates that in envelope only condition (were fine structure cues were replaced by silence) the speech identification was started from 13th frequency band and found to be increasing with the number of frequency bands.

A non parametric Friedman test of differences among repeated measures was conducted and rendered χ^2(7) =48.587, p <0.05. A Wilcoxon signed rank test was run for the comparability of different frequency sets

Table 4.10.

Pair wise comparison of all the frequency bands for envelope perception in speech-speech chimera.

Envelope Frequency bands	EO1	EO4	EO6	EO13	EO16	EO28	EO32	EO64
EO1				-1.86	-1.89	-2.04*	-2.68*	-2.68*
EO4				-1.86	-1.89	-2.04*	-2.68*	-2.68*
EO6				-1.86	-1.89	-2.04*	-2.68*	-2.68*
EO13					-.53	-0.75	-2.25*	-2.25*
EO16						-0.37	-2.14*	-2.14*
EO28							-1.97*	-1.97*
EO32								0.00
EO64								

*Note: *p <0.05*

Results from Table 4.10 indicated that speech identification scores in envelope only condition for firest frequency band is significantly different from EO28, E032 and EO64.EO4 and EO6 are found to be significantly different from EO28, EO32 and EO64. EO13, EO16 and EO28 are significantly different from EO32 and EO64.

Table 4.1.

Comparison of envelope and fine structure perception across frequency bands.

Frequency bands	*Z*	*P*
E1 - FS1	-2.66	.008*
EO1 - FSO1	-2.68	.007*
E4 - FS4	-2.80	.005*
EO4 - FSO4	-2.67	.008*
E6 - FS6	-2.80	.005*
EO6 - FSO6	-2.67	.007*
E13 - FS13	-2.80	.005*
EO13 - FSO13	-2.43	.015*
E16 - FS16	-2.80	.005*
EO16 - FSO16	-2.67	.008*
E28 - FS28	-2.52	.012*
EO28 - FSO28	-.29	.766
E32 - FS32	-2.55	.011*
EO32 - FSO32	-2.31	.021*
E64 - FS64	-2.65	.008*
EO64 - FSO64	-2.31	.021*

Note :FS-fine structure in chimeric condition, E-envelope in chimeric condition, FSO-fine structure only condition, EO-envelope only condition,1-64: number of frequency bands. $p<0.05$.*

Results from Table 4.11 reveals that there are significant difference between envelope and fine structure perception in all frequency bands except for envelope only (E0) and fine structure only (FSO) condition in 28th frequency band.

Chapter 5

Discussion

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 centre 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) which has been used in the current study. When a channel signal is manipulated by the Hilbert envelope, ie 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 (Lorenzy, 2006) . There are different vocoder processing present, one among them intended to disrupt fine structure cues while preserving envelope cue is were the envelope of the channel signal is used to modulate a carrier signal that may be either a sinusoid or a narrow band noise or using a silent portion. Thus original fine structure is replaced by a synthetic fine structure. Speech processed in this manner are usually mentioned as envelope speech, explained in the study as envelope only condition, ie, removing the fine structure information while preserving envelope information. Infact the processed signal still contains fine structure information. Since the original fine structure is replaced by a ‘Synthetic fine structure’, it alter the neural response relative to that evoked by the original signal, provides reduced perception of envelope in all the condition.

Problem with estimating envelope and fine structure with Hilbert transform is that the instantaneous frequency of the fine structure correspons to its instantaneous rate of change of phase. When envelope of physical signal approaches zero there is a abrupt change in fine structure of signal, which is equivalent to a large jump in instantaneous frequency (Hopkins et al .,2010). More generally whenever envelope of the physical signal has a very low value, the instantaneous frequency of fine structure can show wild excursions, because of low level noise in the original channel signal. This noise is effectively amplified when the envelope is removed leading to large audible effects in the processed signal making it sound very noisy (Hopkins et al.,2010) which supports the finding of the present study for the poor perception of fine structure only condition.

Another 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. Results by Gilbert and Lorenzi, (2006); Irino and Patterson, (2001) suggest that these reconstruction did not play a major role in speech identification when number of bands are equal to or greater than 16 frequency bands, which supports the test findings in the current study, ie is, no envelope perception was found around 16th frequency bands for envelope only (EO) condition and also in speech- speech chimeric condition.

The study started with the hypothesis that the relative importance of envelope and fine structure for Malayalam language resembles that for English speech recognition or not. Results indicated that in Malayalam language the individuals mainly depends on the fine structure cues for speech perception compared to envelope. Our results qualitatively matches with the results of smith et al (2002),where it was observed that speech perception is better with fewer frequency bands when speech information contains only fine structure . Also when speech information contained only envelope, speech perception improved as the number of vocoder bands are increased (Smith et al., 2002). When silence is added, speech perception scores are reduced indicating the 'Speech-speech chimera is influenced by processed stimuli' (Smith et al., 2002). Poor perception on envelope in lower vocoder bands indicates that , 'The amount of envelope reconstruction is having a negligible significance in lower frequency bands' (Irno & Patterson., 1997). Our results are conflicting with the English speech-speech chimera results of Smith et al., (2002),where they found that envelope cues are dominating fine structure for speech perception when chimeric signal contains only speech information.

Chapter 5

Summary and conclusion

Literature reports that fine structure have an important role in the perception of tonal language (Mandarian chineasse) and envelope cues are important for the perception of English language. However, in Indian context most of the languages are non-tonal ,a study was required to investigate the differences in temporal cues required to for the perception of Indian languages. Hence the aim of the study was to investigate the influence of vocoder frequency bands on the perception of Malayalam chimeric sentences in individuals with normal hearing. The objectives aimed were, to study the influence of envelope and fine structure cues on word identification and also the number of frequency bands required to differentiate the cues for envelope and fine structure using chimeric sentences. The study started with the null hypothesis that is envelope and fine structure is not having any effect on word identification in Malayalam language. The study was carried out in two phases were phase 1 included the preparation of chimeric stimuli and phase 2 included the presentation of stimulus. Since with 16 frequency bands used in the stage 1 of phase 2, no envelope perception was obtained , three additional frequency bands were added in the stage 2 of phase 2. Also, to check the perception of envelope and fine structure perception two more condition were added as envelope only and fine structure only were either only envelope cue or fine structure cue was preserved. Descriptive stastistical analysis was carried out on SPSS software (version 20). The results revealed that fine structure cues are important for the perception of Malayalam language. Envelope perception is not found to happen for less number of frequency bands and as the number of frequency bands are increased ,improved envelope and reduced fine structure perception was observed. This is the 1^{st} study on envelope and fine structure cues using chimeric sentences in Malayalam language on normal hearing individuals . 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.

Clinical implications

The present study's results reveal that fine structure cues are important for the perception of Malayalam language. However current speech processing strategiesemployed in the hearing aids and cochlear implants do not efficiently use fine structure cues. The current findings of the study highlights the need to include methods such as encoding fine structure 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 fine structure cues for better perception of Malayalam language. In the literature it has been reported that individuals having cochlear lesion have reduced ability to use fine structure cues. Hence the current study highlights the need to focus on fine structure cues in intervention of individuals with cochlear lesion.

Future directions

1) Study of 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 disordered population and in individuals using hearing aids and cochlear implants.

References

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*(1-2), 123-131. doi:10.1016/s0378-5955(00)00265-3

Apoux, F., Yoho, S. E., Youngdahl, C. L., & Healy, E. W. (2013). Role and relative contribution of temporal envelope and fine structure cues in sentence recognition by normal-hearing listeners. *The Journal of the Acoustical Society of America*, *134*(3), 2205. doi:10.1121/1.4816413

Ardoint, M., & Lorenzi, C. (2010). Effects of lowpass and highpass filtering on the intelligibility of speech based on temporal fine structure or envelope cues. *Hearing Research*, *260*(1-2), 89-95. doi:10.1016/j.heares.2009.12.002

Bertoncini, J., Serniclaes, W., & Lorenzi, C. (2009). Discrimination of Speech Sounds Based Upon Temporal Envelope Versus Fine Structure Cues in 5- to 7-Year-Old Children. *J Speech Lang Hear Res*, *52*(3), 682. doi:10.1044/1092-4388(2008/07-0273)

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

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

Davidson, S. A., Gilkey, R. H., Colburn, H. S., & Carney, L. H. (2009). Diotic and dichotic detection with reproducible chimeric stimuli. *The Journal of the Acoustical Society of America*, *126*(4), 1889. doi:10.1121/1.3203996

Drennan, W. R., Won, J. H., Dasika, V. K., & Rubinstein, J. T. (2007). Effects of Temporal Fine Structure on the Lateralization of Speech and on Speech Understanding in Noise. *Jaro-journal of The Association for Research in Otolaryngology*. doi:10.1007/s10162-007-0074-y

Drullman, R. (1995). Temporal envelope and fine structure cues for speech intelligibility. *The Journal of the Acoustical Society of America*, *97*(1), 585. doi:10.1121/1.413112

Drullman, R. (1995). Temporal envelope and fine structure cues for speech intelligibility. *The Journal of the Acoustical Society of America*, *97*(1), 585. doi:10.1121/1.413112

Eaves, J. M., Kitterick, P. T., & Summerfield, A. Q. (2011). Benefit of temporal fine structure to speech perception in noise measured with controlled temporal envelopes. *The Journal of the Acoustical Society of America*, *130*(1), 501. doi:10.1121/1.3592237

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, 412–419.

Fogerty, D., & Humes, L. E. (2012). The role of vowel and consonant fundamental frequency, envelope, and temporal fine structure cues to the intelligibility of words and sentences.*The Journal of the Acoustical Society of America*, *131*(2), 1490. doi:10.1121/1.3676696

Fu, Q., Zeng, F., Shannon, R. V., & Soli, S. D. (1998). Importance of tonal envelope cues in Chinese speech recognition. *The Journal of the Acoustical Society of America*, *104*(1), 505-510. doi:10.1121/1.423251

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., & 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.

Grose, J. H., & Mamo, S. K. (2010). Processing of Temporal Fine Structure as a Function of Age. *Ear and Hearing*. doi:10.1097/AUD.0b013e3181e627e7

Grose, J. H., & Mamo, S. K. (2010). Processing of Temporal Fine Structure as a Function of Age. *Ear and Hearing*, *31*(6), 755-760. doi:10.1097/aud.0b013e3181e627e7

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

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. and Moore, B. C. J. (2010). The importance of temporal fine structure information in speech at differentspectral regions for normal-hearing and hearing-impaired subjects. *J Acoust Soc Am*, 127, 1595–1608.

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

Irino, T. and Patterson, R. D. (1997). A time domain, level dependent auditory filter: The gammachirp. *J Acoust Soc Am*, 101, 412–419.

Kale, S., & Heinz, M. G. (2010). Envelope Coding in Auditory Nerve Fibers Following Noise-Induced Hearing Loss. *Jaro-journal of The Association for Research in Otolaryngology*. doi:10.1007/s10162-010-0223-6

Kathryn Hopkins. (2003). The effect of compression speed on intelligibility: Simulated hearing-aid processing with and without original temporal fine structure information. *Acoustical Society of America*.

Kathryn, H. S & Hopkins, T. (2008). *The role of temporal ne structure information in the perception of complex sounds for normal-hearing and hearing-impaired subjects*(Unpublished doctoral dissertation). St Catharine's College.

Léger, A. C., Desloge, J. G., Braida, L. D., & Swaminathan, J. (2015). The role of recovered envelope cues in the identification of temporal-fine-structure speech for hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *137*(1), 505-508. doi:10.1121/1.4904540

Léger, A. C., Moore, B. C., & Lorenzi, C. (2012). Abnormal speech processing in frequency regions where absolute thresholds are normal for listeners with high-frequency hearing loss. *Hearing Research*, *294*(1-2), 95-103. doi:10.1016/j.heares.2012.10.002

Leger, A. C., Reed, C. M., Desloge, J. G., Swaminathan, J., & Braida, L. D. (2014). Speech intelligibility and masking release using temporal fine structure and recovered envelope cues for normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *135*(4), 2160-2161. doi:10.1121/1.4877010

Léger, A. C., Reed, C. M., Desloge, J. G., Swaminathan, J., & Braida, L. D. (2015). Consonant identification in noise using Hilbert-transform temporal fine-structure speech and recovered-envelope speech for listeners with normal and impaired hearing. *The Journal of the Acoustical Society of America*, *138*(1), 389-403. doi:10.1121/1.4922949

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*,106,2097-2103

Lorenzi, 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

Moon, I. J., & Hong, S. H. (2014). What Is Temporal Fine Structure and Why Is It Important.*Korean J Audiol*, *18*(1), 1. doi:10.7874/kja.2014.18.1.1

Moore, B. C. J. (2003). An introduction to the psychology of hearing. Academic Press San Diego, fifth edition.

Moore, B. C. J. (2008a). The choice of compression speed in hearing aids: Theoretical and practical considerations and the role of individual differences. *Trends in Amplification*,12, 103–112.

Moore, B. C. J. (2008b). 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. and Glasberg, B. R. (1981). Auditory filter shapes derived in simultaneous and forward masking. *J Acoust Soc Am*, 70, 1003–1014.

Moore, B. C. J. and Glasberg, B. R. (1981). Auditory filter shapes derived in simultaneous and forward masking. *J Acoust Soc Am*, 70, 1003–1014.

Moore, B. C. J. and Skrodzka, E. (2002). Detection of frequency modulation by hearingimpaired listeners: Effects of carrier frequency, modulation rate, and added amplitude modulation. *J Acoust Soc Am*, 111, 327–335.

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., Glasberg, B. R., Stoev, M., Füllgrabe, C., & Hopkins, K. (2012). The influence of age and high-frequency hearing loss on sensitivity to temporal fine structure at low frequencies (L). *The Journal of the Acoustical Society of America*, *131*(2), 1003. doi:10.1121/1.3672808

Moore, B., & Carlyon, R. (0). Perception of Pitch by People with Cochlear Hearing Loss and by Cochlear Implant Users. doi:10.1007/0-387-28958-5_7

Narne, V. K. (2013). Temporal Processing and Speech Perception in Noise by Listeners with Auditory Neuropathy. *PLoS ONE*, *8*(2), e55995. doi:10.1371/journal.pone.0055995

Nazzi, T., Floccia, C., & Bertoncini, J. (1998). Discrimination of pitch contours by neonates. *Infant Behavior & Development*. doi:10.1016/S0163-6383(98)90044-3

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., Atlas, L., and Rubinstein, J. (2008). Single sideband encoder for music coding in cochlear implants. In Proc Int Conference on Acoust, Speech, and Signal Processing (ICASSP), pages 4209–4212.

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.

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*. doi:10.1097/01.aud.0000179689.79868.06

R. Palmer, and R. Meddis, editors, Neurophysiological Bases of Auditory Perception, pages 429–438. Springer, New York.

Robert, H. P. & Bernhard, U. S. (2010.). Indications for temporal fine structure contribution to co-modulation masking release. *Acoustical Society of America*, 43.66.Dc, 43.66.Ts, 43.66.Mk, 43.66.Fe [RYL]. Retrieved from DOI: 10.1121/1.3500673

Rocheron, I., Lorenzi, C., & Dumont, A. (2002). Temporal envelope perception in dyslexic children. *Neuroreport*. doi:10.1097/00001756-200209160-00023

Rosen, S. (1992). Temporal Information in Speech: Acoustic, Auditory and Linguistic Aspects. *Philosophical Transactions of The Royal Society B: Biological Sciences*. doi:10.1098/rstb.1992.0070

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.

Schatzer, R., Krenmayr, A., Au, D. K., Kals, M., & Zierhofer, C. (2010). Temporal fine structure in cochlear implants: Preliminary speech perception results in Cantonese-speaking implant users. *Acta Oto-Laryngologica*, *130*(9), 1031-1039. doi:10.3109/00016481003591731

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.

Scott, S. K., & Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception. *Trends in Neurosciences*. doi:10.1016/S0166-2236(02)00037-1

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.

Shamma, S., & Lorenzi, C. (2013). On the balance of envelope and temporal fine structure in the encoding of speech in the early auditory system. *The Journal of the Acoustical Society of America*, *133*(5), 2818. doi:10.1121/1.4795783

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

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

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

Sreeraj, S. & Kishore, T. (2012). Development of sentence list in Telugu and Malayalam. (ongoing project in AIISH)

Tasell, D. J. (1987). Speech waveform envelope cues for consonant recognition. *Journal of The Acoustical Society of America*. doi:10.1121/1.395251

Turner, C. W. (1995). Use of temporal envelope cues in speech recognition by normal and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *97*(4), 2568. doi:10.1121/1.411911

Van der Horst, R., Leeuw, A. R., & Dreschler, W. A. (1999). Importance of temporal-envelope cues in consonant recognition. *The Journal of the Acoustical Society of America*, *105*(3), 1801. doi:10.1121/1.426718

von B´ek´esy, G. (1960). Experiments in hearing. McGraw-Hill, New York.

Wang, S., Dong, R., Liu, D., Wang, Y., Liu, B., Zhang, L., & Xu, L. (2015). The Role of Temporal Envelope and Fine Structure in Mandarin Lexical Tone Perception in Auditory Neuropathy Spectrum Disorder. *PLOS ONE*, *10*(6), e0129710. doi:10.1371/journal.pone.0129710

Wang, S., Mannell, R., Newall, P., & Han, D. (2010). Contribution of Spectral Cues to Mandarin Lexical Tone Recognition in Normal-Hearing and Hearing-Impaired Mandarin Chinese Speakers. *Ear and Hearing*, 1. doi:10.1097/aud.0b013e3181ec5c28

Wang, S., Xu, L., & Mannell, R. (2011). Relative Contributions of Temporal Envelope and Fine Structure Cues to Lexical Tone Recognition in Hearing-Impaired Listeners. *JARO*,*12*(6), 783-794. doi:10.1007/s10162-011-0285-0

Warren, J. D., Jennings, A. R., & Griffiths, T. D. (2005). Analysis of the spectral envelope of sounds by the human brain. *Neuroimage*. doi:10.1016/j.neuroimage.2004.10.031

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

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

APPENDIX – i

Hilbert transform used to separate envelope and fine structure

```
clc;clf;clear;
[a,Fs,nbit]=wavread('waveform');
Fs2=44100
sound(a)
subplot(3,1,1);
plot(a);
b=hilbert(a);
c=abs(b);
%sound(c);
c = 0.99*c/max(abs(c));
wavwrite(c,Fs2,'envelope');
%subplot(3,1,2);
figure;
plot(c);
d=360*angle(b)/(2*pi);
%sound(d)
d = 0.99*d/max(abs(d));
wavwrite(d,Fs2,'fine structure');
figure;
plot(d);
```

Appendix 2

Sentences in Malayalam	*IPA*
അവര് ഇളന്ീ് വാങ്ങി കുടിച്ച	aʊar ɪɭaɪː ʊaːŋɪ kʊʈɪccʊ
അയാള് വയലില് വാഴ ന്ട്ടച	ajaːɭ ʊajalil ʊaː□a aʈʈʊ
അേ കുഞ്ഞിന് രതാട്ടിലില്പ് ഉറക്കി	amma kʊʈʈɪjɜ ʈʈɪlɪl ʊrakkɪ
വവ്വാല്പ് മാങ്ങ തിന്നു	ʊaʊʊaː maːŋa tɪnnʊ
കുപ്പിയില്പ് മഷി ന്ിറച്ച	kʊppɪjɪl maʂɪ nɪraccu
കുട്ടി ന്ലല ഉടുപ്പ് ധ്രിച്ച	kʊːʈʈɪ nalla ʊʈʊppə arɪccʊ
മുക്കുവന് മീന് പ്ിടിച്ച	mʊkkʊʊan mɪːn ɪʈɪccʊ
പാപ്തത്തില് ക്ഷണം കഴിച്ച	Pa:trattil baʂaɳam ka□ɪtccʊ
അയാള് അവന് പുത്തകം രകാടുത്തു	ajaːɭ aʊanə pʊstakam kotuʈʈʊ
രപന് താരഴ വീണു	peːn a ta□ɜ vɪːɳə
അവള് കുരറ സമയം പ്പാര്ളിച്ച	aʊaɭ kʊrɜ sa ajam

	pra:ttɪccʊ
അേ അവരr ഇഷ്ടത്തിന് വഴങ്ങി	amma aʋantɜ ɪʂʈaɪnə va□aŋɪ
അവന് കത്ത്അയച്ച	aʋan kattə ajaccʊ
വ്ഷം രതാറും പ്പശ്ന്ം കൂടി	ʋarʂa m to: ʊm praʃnam kʊ:dɪ