“DEVELOPMENT OF SPEECH PROCESSING ALGORITHM TO IMPROVE PERCEPTION OF HEARING IMPAIRED”

A

THESIS SUBMITTED

TO

SAVITRIBAI PHULE PUNE UNIVERSITY

FOR AWARD OF DEGREE OF

DOCTOR OF PHILOSOPHY

IN THE FACULTY OF

ELECTRONICS AND TELECOMMUNICATION ENGINEERING

SUBMITTED BY

JAYANT JANARDAN CHOPADE

[Ref. No. PGS/630]

Under the Guidance of

Dr. NITEENKUMAR PUNJABRAO FUTANE

DEPARTMENT OF

ELECTRONICS AND TELECOMMUNICATION ENGINEERING

MATOSHRI COLLEGE OF ENGINEERING AND RESEARCH CENTER, NASHIK

JULY, 2016

<u>Certificate of the Guide</u>

Certified that the work incorporated in the thesis **"Development of Speech Processing Algorithm to Improve Perception of Hearing Impaired"**, Submitted by **Mr. Jayant Janardan Chopade** was carried out by the candidate under my supervision and guidance. Such material as has been obtained from other sources has been duly acknowledged in the thesis.

Dr. Niteenkumar Punjabrao Futane

Research Guide
Dept. of E&TC,
Government College of Engineering.
and Research,
Avasari, Dist-Pune

Date:

Place:

<u>Declaration by the Candidate</u>

I declare that the thesis entitled **“Development of Speech Processing Algorithm to Improve Perception of Hearing Impaired”**, submitted by me for the degree of **Doctor of Philosophy** is the record of work carried out by me during the period from **15/3/2013** to **11/1/2016** under the guidance of **Dr. Niteenkumar Punjabrao Futane** and has not formed the basis for the award of any degree, diploma, associate ship, fellowship, titles in this or any other University or other institution of higher learning.

I further declare that the material obtained from other sources has been duly acknowledged and credited in the thesis.

Date:

Place:

Mr. Jayant Janardan Chopade

Research Scholar

Matoshri College of Engineering

and Research Center, Nasik

ACKNOWLEDGEMENT

First and foremost I would like to give all the credit to the Almighty for imparting me with enough strength to overcome the difficulties with ease.

I would like to express my profound gratitude to my mentor and guide ***Dr. Niteenkumar Punjabrao Futane*** whose support helped me to achieve this endeavor. He has been a constant source of motivation and helped me to channelize the research in proper direction.

I would like to thank from the bottom of my heart, ***Dr. G. K. Kharate***, Principal, Matoshri College of Engineering and Research Center for providing me the timely guidance, key inputs and the administrative support. I am also thankful to ***Dr. V. H. Patil***, Head, Computer Engineering, ***Dr. S. D. Pable,*** Ph.D Coordinator, ***Prof. D. D. Dighe,*** Head, E&TC Department, ***Prof. N. L. Bhale***, Head, I.T Department and all staff members of E&TC Department for extension of required infrastructure and lab facilities whenever required.

I sincerely appreciate valuable support and encouragement from ***Dr. V. J. Gond***, ***Dr. S. T. Gandhe*** and ***Dr. R. S. Patil***. I would also like to thank management of A.V.E.W. Trust Shatabdi Institute of Engineering and Research, Nashik.

I express my heartfelt gratitude to ***Dr. A. S. Indorewala*** & ***Dr. D. K. Chopade*** for giving me access to their esteemed hospitals. I would also thank audiologist ***Mr. Ramlal Singh*** and ***Mrs. Krupa Madam*** for locating the hearing impaired subjects and making their audiogram available. I sincerely thank to all people who actively participated in the listening test.

I am also thankful to ***Er. Anand Gharpure***, ***Er. Umesh Gite***, ***Prof. R. K. Agrawal***, ***Prof. M. P. Joshi***, ***Prof. V. N. Shah*** and ***Prof. S. D. Ratnakar*** for their timely support and consistent encouragement

I am greatly indebted to my mother Late ***Mrs. Chabbu J. Chopade*** and my father ***Mr. Janardan K. Chopade*** for always showering their blessings and best wishes which have driven me so far and have been key to my success; today and always.

Last, but not the least I would like to thank my wife ***Mrs. Meena*** and son ***Ajinkya*** who stood by me through thick and thin in the entire situation. I thank them for their patience, understanding and love.

Jayant Janardan Chopade

ABSTRACT

Life expectancy and rise in noise levels have increased the chances of hearing loss, which demand the development of sophisticated hearing aids based on improved algorithm. Basically hearing loss is classified into two categories viz. conductive loss and sensorineural loss. Sensorineural loss characteristics are frequency dependent shifts in hearing thresholds, compression in dynamic range of hearing, degradation of temporal resolution with an increase in temporal masking and damaged frequency selectivity with increase in spectral masking. Dominant possible reason for sensorineural loss is, spread spectral masking along the cochlear partitions. Features like duration, voicing and place features are not well perceived by persons with sensorineural hearing impairment, as they are cued by temporal and spectral masking.

It is observed that the increased masking effect is reduced when the signal is split into critical bands based on the complementary spectra and the obtained alternate bands are presented to both ears. It helps to improve the perception ability of hearing impaired subjects. The objective of this research is to develop speech processing algorithm so as to improve the perception of sensorineural hearing impaired subjects. Here, the speech is split into two signals with complementary spectra by using comb filters and optimized wavelet based filters with distinct basis functions. The test material used for experimental evaluation consists of fifteen nonsense syllables with English consonants in vowel-consonant-vowel context.

The experimental evaluation and implementation was carried out in two phases. In phase one, an algorithm based on spectral splitting scheme with 512-coefficient linear phase comb filter was developed and implemented in real time on Spartan6 XC6SLX45 CSG324 FPGA and an on-board audio codec (LM4550). For the evaluation of this algorithm, listening tests were carried out on seven sensorineural hearing impaired subjects having different degree of hearing loss (referred as Experiment I). Results show the improvement of response times, recognition scores, and transmission of consonantal features particularly voicing, manner and place for all subjects, indicating reduction in the effects of spectral masking.

In the second phase of research, wavelet based algorithms with three distinct basis functions were developed and filters were implemented using software based offline (MATLAB programming) and real time based hardware approach. For

evaluation of these algorithms, three experiments were conducted. In Experiment II, hardware based real time approach was used and listening tests were carried out on five normal people with simulated hearing loss (different SNR conditions). Results indicated reduction of load on speech perception process and improvement in speech perception by normal people under noisy environment. It was also observed that the relative improvements in recognition score due to processing were more for higher levels of masking noise. In Experiment III (software based offline) and Experiment IV (hardware based real time), eight sensorineural hearing impaired subjects participated for listening tests. The algorithms were found helpful in improving quality of speech, recognition scores, response times and information transmission of consonantal features particularly for frication, continuance, manner and place features.

Listening test results show improvement in speech quality, response time and recognition score for wavelet based algorithms over comb filter based algorithm. Latency and PSNR were analyzed and improvements in results were observed as compared to comb filter based algorithm. It may be concluded that, speech perception can be improved with dichotic processing schemes for people using binaural hearing aids.

The implementation of optimized wavelet filters with different basis function, utilize fraction of the resources available on FPGA with scope for implementation of other processing blocks of the hearing aid.

TABLE OF CONTENTS

APPENDICES

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ACE	Advanced Combinational Encoder
ADC	Analog-To-Digital Converter
Avg	Average
bior	Bi-orthogonal Wavelet
BM	Basilar membrane
BTE	Behind The Ear
BWT	Bionic Wavelet Transform
CB	Critical Bandwidth
CV	Consonant-Vowel
CVC	Consonant-Vowel- Consonant
DAC	Digital-To-Analog Converter
dB	Decibel
db	Daubechies Wavelet
DL	Difference lemen
DSP	Digital Signal Processing
DWT	Discrete Wavelet Transform
ERB	Equivalent Rectangular Bandwidth
FIR	Finite Impulse Response
FPGA	Field Programmable Gate Array
GUI	Graphical User Interface
HDL	Hardware Design Logic
HL	Hearing Level
HTL	Hearing Threshold Level
Hz	Hertz
ITC	In The Canal
ITE	In The Ear
JND	Just noticeable difference
LDL	Loudness discomfort level
MOS	Mean opinion score
PCM	Pulse Coded Modulation
PISO	Parallel in serial out
QMF	Quadrature Mirror Filters

RD	Relative Decrease
RI	Relative Improvement
RTL	Register Transfer level
SIPO	Serial in parallel out
SNR	Signal-To-Noise Ratio
SPL	Sound Pressure Level
sym	Symlet Wavelet
VC	Vowel-Consonant
VCV	Vowel-Consonant-Vowel
VHDL	VHSIC Hardware Description Language
VOT	Voice-Onset-Time
WP	Wavelet Packets
WPT	Wavelet Packets Transform

LIST OF SYMBOLS

ρ	Fluid Density
u	Flow to the velocity
F_{oval}	Force on oval window
F_{drum}	Force on drum
P_{oval}	Pressure on oval
P_{drum}	Pressure on drum
$\phi(t)$	Scaling function
$\psi(t)$	Wavelet function
Z_{Air}	Air impedance
Z_{fluid}	Fluid impedance
t_{av}	Sound intensity over time
B	Fluid viscosity
C	Speed of light
$c(j, k)$	Discrete scaling coefficients
$d(j, k)$	Discrete wavelet coefficients
F	Frequency
F1	First formant
F2	Second formant
F3	Third formant
H	Basilar membrane displacement
$H(e^{jw})$	Frequency response
$H(z)$	Transfer Function
$I(s)$	Information measure of stimulus *s*, in bits
$I(s; t)$	Information transmitted from stimulus *s* to response *t* in bits
$I(t)$	Information measure of response *t*, in bits
I^i	Intensity of incident waves
$I_{rel}(s; t)$	Information transmitted relative to stimulus information
I^t	Intensity of Transmitted waves
J	Relative current
K	Number of coefficients
K(x)	Stiffness of BM

L	Filter length
$p(s_i)$	Probability of response s_i
$p(s_i;t_j)$	Probability of stimulus response pair (s_i,t_j)
$p(t_j)$	Probability of stimulus t_j
Q	Constant relative bandwidth
R	Reflection Coefficient
T	Transmission Coefficient
λ	Wavelength

Chapter 1
INTRODUCTION

1.1 Overview

The researcher forecasts that number of people suffering from hearing loss will continue to increase, as the age of population increases. The problem in hearing leads to poor speech communication. Both, the old age of the people and the increased noise levels have contributed to the chances of hearing loss. Hence there is an immense need of improved hearing aids for speech perception.

On the basis of the predicted numbers of affected people over the globe, hearing impairment will be one of the serious health issues to be faced in recent future. The magnitude of the problem is very high on the basis of data collected by the National Center for Health Statistics and the U.S. Bureau of the Census. As per prediction, by the mid of 21st century there will be 36 Million people suffering from hearing impairment [1]. In developing countries like India majority of the people suffering from hearing impairment are using the conventional hearing aids which provide the amplification of speech. Many such aids are provided with filtering and amplitude compression. As the hearing loss is independent of frequency, these aids are not designed to improve the speech perception for the person suffering from sensorineural impairment. The available solutions in the market for hearing loss, urge to find a better solution between less effective conventional hearing aids and expensive surgical methods.

There are three sections of ear namely the outer ear, middle ear and the inner ear. As per the region affected, hearing impaired people can be classified. Depending upon the location of damage in the auditory system, hearing impaired people get categorized in Conductive and Sensorineural hearing impairment [2]. Conductive impairment occurs due to damage in the outer ear, ear drum or middle ear. It blocks the transmission of the sound to the inner ear. It also generates attenuation of the stimulus and results in an increase in both the hearing threshold levels and uncomfortable loudness levels. In many cases, medical and surgical treatments are effective in conductive hearing impairment [2], [3]. The conventional hearing aids provide the amplification of speech without paying attention to frequencies. Many such aids are provided with filtering and amplitude compression. They are usually used in addressing the problems of elevated thresholds. These aids are not designed to improve the qualitative speech presentation for the person suffering from sensorineural hearing impairment.

Sensorineural hearing impairment takes place due to damage to the transduction mechanism of the inner ear. Sensorineural hearing loss can be divided into two categories- when there is damage to the inner ear (cochlea) or when there is damage to the nerve pathways from inner ear (retrocochlear) to brain. The damage to inner ear (cochlear hearing impairment) is associated with loss of cochlear hair cells. [2].

The sensorineural loss is characterized by an increase in threshold of hearing, compression in dynamic range of hearing, degradation of temporal resolution with increase in temporal masking and degraded frequency selectivity with an increase in spectral masking [4], [7].

One of the possible reasons for the malfunction in sensorineural loss is due to spread of spectral masking along the cochlear partitions. Masking is a phenomenon in which perception of one sound is obscured by the presence of another. More specifically, when any person tries to listen two sounds simultaneously or concurrently, the threshold of hearing for other increases by the presence of first one. Simultaneous sounds cause frequency masking, where a lower-frequency sound generally masks a higher-frequency sound; Sounds delayed with respect to one another can cause temporal masking of one or both sounds. Masking is also termed as the inability of the human perceptual mechanism to perceive and extract a signal component. This inability is due to the presence of stronger component, which shares temporal or spectral location [7].

The masking effect explains and quantifies frequency selectivity. Frequency resolution or frequency selectivity occurs when two sounds of two different frequencies (pitches) are played at the same time and two separate sounds are often heard rather than a single combination tone. It is believed to take place on account of filtering inside the cochlea, also denoted as critical bandwidth. The complex sound is separated into distinct frequency components which generate a peak in the pattern of vibration at certain location on the cilia within the basilar membrane of the cochlea. The above components are individually coded on the auditory nerve that relays sound information to the brain. This individual coding takes place if and only if there is an enough distinction in spectral domain otherwise they are coded at the same place and are heard as single sound. The auditory filters are the ones that separate one sound from another. The auditory filters are broader than normal in increased spectral masking. Cues like voice-onset time, burst duration and formant transition important for consonants identification get masked by the following or preceding vowel segment which results in degraded speech perception.

Researchers have developed different techniques to minimize sensorineural loss viz. Compression technique, frequency transposition technique and speech processing using dichotic presentation. The problem of acoustic amplification can be dealt with by making use of compression techniques in hearing aids. It consists of those weak signals that are audible to the intense signals and reach upto uncomfortable level owing to the loudness recruitment in sensorineural hearing impairment.

In these techniques the amplification decreases with intensity, such that the wide dynamic range of input signal acquires compression in a smaller dynamic range of the output [8]. The frequency transposition technique deals with energy transposition from high frequency region to low frequency region. In case of sensorineural hearing impairment, to reduce the effect of increased spectral and temporal masking various speech enhancing schemes are based on the principle of alteration of energy of certain frequency components. Binaural hearing is the process of listening with both the ears. Presentation of the same signal to both the ears is diotic presentation, while presentation of two different signals to the ears is termed as dichotic presentation. The benefit of binaural dichotic mode of hearing to enhance the speech perception for people with the sensorineural hearing impairment is reported by many researchers. The splitting of information in speech signal for presenting signals to the ears, in some sort of complimentary fashion, to provide the relaxation for sensory cells of the basilar membrane, may help in reducing the effect of increased masking and thereby improve the speech reception in cases of bilateral sensorineural hearing impairment [1], [2].

It is revealed from the detailed literature review that the reduction in the spectral masking is better solution to improve the perception of hearing impaired. The core researchers from the relevant domain also asserted that the problems due to the spectral masking are not completely solved by using band pass filter banks and comb filters, except they are useful only for moderate deafness. Therefore, there is an urge to propose an algorithm to improve the perception of hearing impaired, that can be helpful for moderate to severe deafness. Further it is also essential to realize the prototype with efficacy. In our research, algorithms were developed for better perception of sensorineural hearing impairment and a prototype was also implemented. It was carried out in two phases – in first phase, a real time experiment (Experiment I) was implemented using comb filters and listening tests were carried out on hearing impaired people. The details of it are discussed in chapter 4. In second phase, three experiments were performed and their details are discussed in chapter 5. Experiment II was implemented using wavelet filters and listening tests were carried out on normal people with simulated hearing loss. Experiment III was software based offline implementation of wavelet filters and listening tests were carried out on hearing impaired subjects, while Experiment IV was hardware based real time implementation of wavelet filters and listening tests were performed on hearing impaired subjects.

1.2 Motivation

CHABA (Committee on Hearing, Bioacoustics and Biomechanics), National research council, Washington (1991) has presented statistics of percentage of hearing impaired for every decade and percentage of impairment according to age group, from which the committee has concluded that in future, the number of hearing impairment cases will grow continuously due to adverse environmental effect caused by unavoidable human activities. Hence the most effective aid for the vast majority of hearing impaired persons is and will remain in future the electro acoustic hearing aids [1].

By god's grace mankind is endowed with two important faculties namely hearing and speech. A great service can be offered to human race by restoring the above faculties to those who lack it.

By considering the need of society there is great necessity for further research in this area; specifically, sensorineural hearing loss, which is one of the most challenging problems in medicine and a large variety of hearing impairments fall under this category as it deals with most delicate part of human ear that is cochlea which is located at inner ear. The problem of sensorineural hearing loss can be better solved by developing an algorithm used in hearing aid, which can be used by hearing impaired patient instead of surgical treatment. This work would be of great help for those people who suffer from hearing impairment. The topic has been selected with the perspective of affected people so that the cost effective solution can be developed.

1.3 Objectives of Thesis

The prime objective is Development of Speech Processing Algorithm to Improve Perception of Hearing Impaired and to meet the said objective, proposed work includes-

- To study Physiology of Hearing
- To understand the mathematical model of human auditory system
- To study the problems associated with speech perception of sensorineural hearing impaired
- To study the different techniques proposed by the researchers to resolve the problems of speech perception for hearing impaired
- To understand the limitations/shortcomings of existing techniques
- To develop and propose the speech processing techniques to overcome the difficulties faced by hearing impaired persons
- To implement and test the proposed techniques and validate the results

1.4 Organization of the Thesis

This thesis is arranged in six chapters. Chapter 1 gives an introduction and briefly describes importance of sensorineural hearing loss and its characteristics like loudness recruitment, spectral masking and temporal masking. These characteristics are explained in context of the research. It also describes motivation, objectives and gives an outline of thesis.

Chapter 2 discusses literature on development of techniques and technologies for signal processing schemes supporting the hearing impaired people. A brief review of real time implementation along with speech material required for listening test is conferred. Findings of literature review are also highlighted.

Chapter 3 describes the physiology of hearing with sensorineural hearing impairment and its effects on speech perception. It also describes the mathematical model of human auditory system and a brief overview of proposed algorithm is discussed.

Chapter 4 deals with design and implementation of comb filters based on auditory critical bandwidth. The results obtained from experimental evaluation of the algorithm in hardware based real-time processing are also presented and discussed.

Chapter 5 deals with development of wavelet based filter algorithms and its implementation. Listening tests for overall evaluation of developed algorithms are presented. The results of software based offline and hardware based real-time processing are also discussed.

Chapter 6 provides a summary of the investigations, conclusions drawn from the results, and suggestions for further work.

Appendices provide additional information about the research. Appendix A describes performance assessment techniques and provides information about pure tone audiogram of sensorineural hearing impaired subjects. The description of hardware and software used in the research is discussed in Appendix B. Appendix C deals with the test instruction to normal and hearing impaired people.

Chapter 2

LITERATURE SURVEY

In sensorineural hearing impairment, frequency resolution is reduced in the peripheral auditory system. It is due to the spread of masking of frequency component by adjacent components, which degrades speech perception. So the input speech signal is the critical factor for the hearing aids. This chapter highlights development of techniques and technologies for signal processing schemes supporting the hearing impaired people.

We have thoroughly studied literature published by various researchers and summarized the major contributions. The different major techniques developed by researchers are compression technique, frequency transposition and speech processing using dichotic presentation to improve the perception of sensorineural hearing impairment. For the above said techniques, the experimental evaluation method along with speech materials proposed by different researchers has also been reviewed.

2.1 Techniques Used

2.1.1 Compression Technique

In this technique, by amplifying weaker signals to stronger one, we can compress wide dynamic range of the input signals to smaller dynamic range. Various compression techniques reported by researchers are Multiband compression systems, compression limiting and syllabic compression. [1], [2], [8]

In Multiband Compression, there is provision for different amount of compression in two or more frequency bands. So it’s more helpful for hearing impaired as it is frequency dependent. This method will create relatively weak high-frequency components in speech [9].

Some of the studies that use compression technique are reviewed in the following paragraphs.

A two channel amplitude compressor scheme has been proposed by Villchur E. (1973). Its frequency dependent compression ratio is adjusted to compensate the recruitment of the individual subject. This processing scheme improves speech recognition, both in quiet and in the presence of competing speech introduced before processing [9].

CHABA (1991) reported the improvements in conventional electro-acoustic hearing aids, cochlear stimulators and optical stimulating devices. It also presented statistics of percentage of hearing impaired of each decade and percentage of impairment according to age groups. Committee also concluded that, the most effective aid for the vast majority of hearing impaired persons is and will remain for future is the electro-acoustic hearing aids. The committee also highlights on major types of available hearing aids, different factors affecting the design, selection and use of conventional hearing aids. The committee has concluded that the long term automatic gain control and compression limiting were more helpful while describing three types of compression amplification [1].

With the objective to avoid the spectral flattening associated with Multi band compression and to compensate for narrow dynamic range, Asano et al. (1991) reported a digital hearing aid. It discussed that input signal was segmented into 8 ms window and each block was processed with an FIR filter with frequency-gain characteristic determined by the short-time magnitude spectrum of the segment and loudness compensation functions (relation between the loudness for normal listeners and that for the impaired listener). Spectral values in 6 octave bands were used for determining the frequency response and filter coefficients were calculated using frequency sampling technique. The scheme was evaluated by conducting tests on thirteen sensorineural hearing impaired subjects with monosyllabic speech as test material. Higher scores compared to linear amplification were reported [10].

Magotra, N et al. (1995) reported that speech amplitude compression controls the overall gain of a speech amplification system. In this, the dynamic range of the acoustic environment is mapped to the controlled dynamic range of the hearing impaired subject. The compression of amplitude is achieved by applying a gain of less than 1 to a signal whenever its power exceeds a predetermined threshold. In multi band compression method, 2–3 bands are sufficient to achieve adequate compensation for recruitment, since the compression reduces the spectral contrasts in complex signals [11].

An epochal method where the critical band was compressed along the frequency axis to adjust the shape of the auditory filters of hearing impaired persons has been reported by Yasu, K. et al. (2004). The Fourier transform based compression approach got better quality and intelligibility of speech signal for hearing impaired people. The compression rates ranges between 20 to 40% subject to shape of specific auditory filter [12].

2.1.2 Frequency Lowering Techniques

There is greater loss in high frequency range in sensorineural hearing impairment. This technique transposes energy from high frequency region to low frequency region. In cases of sensorineural hearing impairment, various speech enhancing schemes have tried to reduce the effect of increased spectral and temporal masking. To reduce the effect of spectral masking schemes are based on the alteration of energy of certain frequency components, while to reduce the effect of temporal masking schemes used clear speech for their evaluation [8], [13].

Reed et al. 1983 studied consonant discriminability by normal hearing. This uses frequency lowering with either uniform or non-uniform compression of the frequency axis. They concluded that the performance for 2500 Hz bandwidth was found superior than 1250 Hz bandwidth. For each bandwidth, compression schemes give better performance at high frequencies than lower frequencies. Lowering the frequencies was superior for fricatives but inferior for nasals and semivowels [14].

The sinusoidal model for frequency transposition was the base for algorithm as suggested by Munoz et al. (1999). The spectrum used is separated in two parts- high and low frequencies- by 3500 Hz central frequency. This threshold value was determined by choosing a value in which processing appeared to be curbed for most vowels, semivowels and nasals and triggered for affricates and fricatives. In each part the summation of peaks indicates energy [15].

The sinusoidal modeling is possible as amplitudes, phases and frequencies can be altered independently. The critical band responses of the human auditory system were used as a base to study the effect of wavelet based noise reduction systems. It also links nonlinear thresholding techniques and multi resolution analysis to wavelet coefficients [16].

The efficacy of frequency transposition techniques on speech perception was also studied by Iman M.S.EI Danasoury et al. (2013). They showed significant improvement in the audibility of high-frequency consonant. Audiological evaluation included aided sound field hearing thresholds in dBHL, Arabic speech perception tests, consonant recognition tests and Widex Infant Listening Skill Inventory (WILSI).The study was carried out on ten children with severe to profound sensorineural hearing loss with an average age of 8 years. The application of frequency transposition technique provided much better response to warble tones in the high frequencies [17].

Mohammed Alnahwi et al. (2015) reviewed comparison of different hearing instruments based on frequency compression and frequency transposition techniques They concluded that Frequency Compression can help to improve the recognition of consonants, monosyllabic words and sentences in noisy environment, whereas FT plays important role to detect fricative feature that finally leads to improved discrimination of consonants and does not show adverse effect on vowel recognition. Only one study has compared both FT and FC using various speech tests such as the vowel-consonant-vowel test, the nonsense syllable test, etc., to evaluate the performance of different hearing instruments. They also revealed that an organized review should be done to compare these techniques with same subjects and same schemes [18].

2.1.3 Speech Processing using Dichotic Presentation

Binaural hearing is listening with two ears while monaural hearing is listening with one ear. In Binaural hearing, there are two types, one is diotic, in which same signal is given to both the ears and second is dichotic, in which both the ears are fed with two different signals. It has been proved that binaural hearing is beneficial over the monaural hearing [19].

For people with moderate bilateral sensorineural hearing loss, the effects of masking can be reduced by using binaural dichotic presentation. In this scheme, speech signal to be presented gets split in complementary fashion and fed to both ears separately. This helps to relax the sensory cells of the basilar membrane, which reduces the effect of increased masking and improves the speech perception [1].

A review on comb filters and wavelet filters for speech processing that makes use of dichotic presentation is discussed below:

2.1.3.1 Use of Comb Filters in Speech Processing

Lyregaard (1982) developed an algorithm for complementary comb filter to split the speech signal. The analog delayed signal was used to add and subtract from the original signal to obtain the signals for binaural dichotic presentation. The comb filters with different bandwidths were obtained with variations in the delay. The listening tests showed no significant improvement dichotic over diotic presentation [20].

Lunner et al. (1991) investigated usage of eight channel digital filter bank for hearing aid. Filters used were complementary interpolated linear phase FIR filters of 700Hz constant bandwidth. As per the subjects hearing threshold the frequency response of filter was fine tuned. The signals were split in to odd bands and even bands and presented to left and right ear respectively. Digital signal processor TI/TMS320C25 was used to implement the comb filter. The listening tests were conducted on three subjects with five word-sentences speech material was used in presence of background noise. Test results showed an improvement of 2dB in signal-to-noise ratio for 50% correct recognition score due to dichotic presentation [21].

D. S. Chaudhari et al. (1998) implemented a comb filter based on 18 critical bands, corresponding to the auditory pattern described by [71] are about 15-17% of the center frequency for 1 to 5 KHz range. This scheme was implemented in real time using two DSP processors (TI/TMS320C50). For real time processing, two 128-coefficients FIR filters (comb filters) were realized, with filters designed by frequency sampling technique, to have sharp transition between pass and stop bands. For dichotic presentation, speech signal filtered by the alternate bands were presented to both ears. This scheme resulted in improving response time, recognition scores, speech quality and transmission of consonantal features [4], [5]. In 2008, information transmission analyses with speech material having VCV and CV contexts were studied. Results showed maximum contribution by the place features in improving the recognition score indicating reduced effect of spectral masking [7].

Jangamashetti, D. S.et al. (2000) have suggested the use of trapezoidal fading function in reducing the occurrence of high frequency components and thereby further increasing the speech intelligibility. The authors concluded that the inter-aural switching with trapezoidal transition helped in improving the perception by reducing the effect of temporal masking [22]. In 2008, researchers conducted listening tests on six normal hearing people with addition of broadband noise to speech material viz. phonetically balanced words, nonsense vowel –consonant-vowel syllables and modified rhyme words, keeping constant SNR on short time basis. Results indicated degradation in speech perception with decrease in SNR for all type speech materials used. An increase in response time was also observed due to decrease in SNR that indicated an increased load on the speech perception process [23].

Cheeran, A.N. et al. (2002) suggested binaural dichotic presentation using critical bandwidth (based on spectral splitting with perceptually balanced 256-coefficient linear phase comb filters) helps in reducing the effect of spectral masking for persons with moderate bilateral sensorineural hearing impairment [24]. Further, in

2003, the scheme was evaluated by conducting listening tests on five normal people with simulated hearing loss. Twelve vowel-consonant-vowel syllables were used as speech material. Significant improvements lead to further tests on five hearing-impaired subjects with moderate bilateral sensorineural hearing loss. Results of these listening tests showed decrease in response time, improvement in recognition scores and transmission of consonantal features, mainly duration and place. The frequency response of a pair of filters was attuned within a 6dB range and was cascaded with comb filters so as to partially compensate for frequency dependent hearing threshold shifts. This combination of filters additionally improved the results, particularly for subjects having relatively uniform loss [25].

Two methods to split the spectral information was studied by Loizou, C. et al. (2003) First method dealt with presenting odd indexed channel to one ear and even indexed channel to another, while second method dealt with presenting inferior frequency channels to one ear and sophisticated frequency channel to another ear. The dichotically presented sentences in the odd-even frequency condition were identified more accurately than when presented in the low-high frequency condition [13].

Kulkarni, P. N.et al. (2008) analyzed the optimization of the comb filters with respect to the number of bands and bandwidth. Authors explored and experimented with constant bandwidth comb filters, whose number of bands varied from two to twenty, critical band based comb filter and 1/3rd octave band based comb filter. Of these three types of comb filters, it had been proved that 1/3rd octave band and critical band based comb filters are superior to constant bandwidth filters [26]. Further in 2012, usefulness of binaural dichotic presentation using 513 coefficients linear phase FIR filters (comb filters) with complementary magnitude responses, based on fixed bandwidth of eighteen bands and auditory critical bandwidth were studied. Listening tests were conducted on eleven people with moderate bilateral sensorineural hearing loss and six normal-hearing subjects with simulated hearing loss in quiet environment. It involved source direction identification and consonant recognition. The results indicated better performance using comb filter based on auditory critical bandwidth with significant increase in recognition score (average 22%) and decrease in response time with no significant effect on broadband source direction identification [29].

2.1.3.2 Use of Wavelets in Speech Processing

The frequency analysis of human’s cochlea has constant Q or constant relative bandwidth that makes it a fine wavelet analysis organ. The time-scale analysis of wavelets and properties of cochlea coincide with each other making it possible to use in various speech signal processing schemes. The simulation parameters of cochlear implants are generated by features extracted using wavelet based method from speech signal. This method can provide tonal patterns for some languages and envelope signals in each band. Results indicate that speech analysis using wavelets is capable of executing process of speech signal for cochlear implants [30].

Recruitment of loudness has been compensated by wavelet based multiband dynamic range compression as cultivated by Drake, L. A., et al. (1993). The algorithm associates standard compression with intensity-level dependent gain calculation making its performance and complexity comparable to conventional procedures. Further, the methods established could be applied to more adaptive wavelet packets and local cosine basis that prototype the input speech signal more faithfully [31].

Johnson Ihyeh Agbinya (1996) discussed techniques and procedure of using wavelet transforms in speech compression. Higher the concentration of energy by the wavelet in the approximation part of the coefficients, the better will be its performance as a speech compressor. Concentration of energy into wavelet bands determines unvoiced, voiced and mixed speech frames. The energy gets concentrated mainly in 2 bands for voiced frames while its gets confined between 3 to *5* bands mixed frames and spreads across the entire bandwidth for unvoiced frames [32].

Yao, J. et al. (2002) have modified the wavelet transform that incorporates the active cochlear mechanism resulting in nonlinear adaptive time-frequency analysis, called as Bionic Wavelet Transform (BWT). Authors concluded several advantages of application of BWT in cochlear implants like improved recognition rates for vowels and consonants, reduction in the number of channels in cochlear implant, better noise tolerance, reduction in the average stimulation duration for words and higher speech intelligibility rates [33]. WPT (Wavelet packet transform) has been progressively used in speech enhancement techniques [34]. Authors discussed that wavelet analysis and subband decomposition tracks speech tone and transient component of speech signal respectively.

Lu and Wang, 2004 derived a robust weighting factor for each wavelet subband. This factor was used to keep the energy of residual noise lower than the noise masking threshold and the speech distortion smaller than the residual noise. Results showed that the proposed method improved the naturalness of enhanced speech [35].

A commercial Advanced Combinational Encoder (ACE) scheme for speech processing in cochlear implants was designed and incorporated using wavelet packet (WP) filter bank. Three different arrangements using different mother wavelets for the WP tree such as Haar, Db3 and mixed wavelet packet were applied. Decomposition of WP at each stage was done using different filter lengths. These three conFig.urations were implemented in ACE scheme. Seven cochlear implant recipients participated in speech intelligibility tests. Averaged tests results of these tests shows the performance of speech perception is better using mixed WP filter bank over Fast Fourier transform used in commercial ACE scheme [36].

Karmarkar A. et al. *(*2006) have proposed multi-resolution model of auditory excitation pattern and its evaluation of subjective wideband speech quality. In this model, time-frequency decomposition of input signal is done by wavelet packet transform. The critical band structure decides an optimality criterion to select wavelet packet tree so as to decrease the cost function. The models of altered auditory aspects such as spectral spreading and temporal smearing are conveyed for the multi-resolution framework [37].

Kolte M.T. et al. (2010) developed comb filters and modified wavelet packet algorithm using Symlet family. A MATLAB code was written for designing 512 coefficient comb filters while Simulink model was developed for modified wavelet packet (10 band). An offline experiment was carried out on seven hearing impaired subjects. For comb filter relative improvement in recognition scores varied from 1.78% to 4.44% and 3.33 to 22.23% for modified wavelet packet algorithm. The processing scheme based on comb filters were beneficial to subjects with low frequency hearing impairment and gradual sloping asymmetrical impairment while the processing scheme based on modified wavelet packets were beneficial to subjects with low and high frequency hearing impairment [38].

M. Mehrzad et al. (2012) has presented method for coding speech signals for the simulation of a cochlear implant, based on a wavelet packet decomposition strategy. Authors used wavelet packet Db4 for 7 levels, generated a series of channels with bandwidths exactly the same as nucleus device and applied an input stimulus to each channel. The results showed the power of this method in processing of the input signal for implant users with less complexity than other methods, while maintaining the contents of the input signal to a very good extent [39].

An algorithm to improve the speech recognition rate is been discussed by Zhimin et al. (2012).They proposed a single channel method based on multi-resolution wavelet so that an envelope of signal spectrum can be extracted and gain can be adjusted at each feature point. Satisfactory results were obtained over the traditional LPC based method [40].

Grisha et al. (2016) discussed a design technique for analysis and synthesis of multi rate filter banks. A structural design of Discrete Wavelet Packet Transform on FPGA was implemented using Simulink-Xilinx System Generator tool. The audiogram of hearing aid person was matched with the input of 24 kHz audio signal by amplification process [41].

2.2 Review of Real Time Implementation

Earlier literature survey shows that the digital signal processors (DSP) are being used to implement the FIR (Finite Impulse Response) filter banks [4], [21]. Few researchers have attempted to implement filter algorithms on FPGA (Field Programmable Gate Array). Kambalimath et al. (2014) have reported that FPGAs are more suitable than DSPs for applications involving highly parallel computations [42], [43].

Tiwari et al. (2012) implemented multiband frequency compression of complex spectral samples using fixed frame processing with least square error based estimation on 16bit fixed point DSP processor TMS320C5515. A reduction in processing delay and perceptual similarity in speech output when compared with pitch-synchronous processing was observed [40].

Stationary wavelet packet algorithm has been implemented by Nivin Ghamry (2013) for noise reduction in hearing aids. The performance evaluation has been carried out using MATLAB and the design has been implemented on Virtex-II platform. She concluded that the implementation of proposed system accomplishes the requisite of hearing aids such as intelligibility, portability, cost effective and good performance [46].

L. Bendaouia et al. (2014) have proposed new speech processing pattern to enhance the speech intelligibility for the impaired people. The scheme uses both Discrete Wavelet Transform (DWT) and Over-Lap and Add (OLA) technique executed on FPGA platform. The proposed method gives efficient real time results as compared to classical methods and shows improvement in speech intelligibility [47].

Valerie Hanson et al. (2014) proposed the binary mask based algorithm and subsequently implemented the spectral analysis 28 channel bank stage with 8th order band pass filters. The algorithm was implemented on the Spartan-3A FPGA, and it showed the improvement in intelligibility of speech by 85% in presence of noise. The average latency was found to be 8.5ms [48].

In [42], [43] 513-coefficient filters with sampling frequency of 10 kHz comb filters based on auditory critical bandwidths were implemented on Altera Cyclone II FPGA. It included 16bit codec and 15-bit integer filter coefficients and utilized 47% of combinational functions, 34% of logic registers and 53% of logic elements available on the board. The magnitude responses of this and offline floating-point implementation was similar. Further an efficient implementation of scheme was achieved using sequential multiply-accumulate operations.

2.3 Listening test material

Many researchers have conducted listening tests for the investigation of perception by normal hearing people with simulated hearing loss and hearing impaired listeners. For these tests they have used various types of speech materials. Speech material presents a range of acoustic, linguistic phonetic and lexical variables. They vary in complexity from nonsense syllables, single words, vowels, spondee to sentences. Fused speech stimuli provide independent control over spectral, temporal and intensity characteristics making it more attractive for experimentations [49]-[55].

Word recognition test mostly has fifty monosyllables representing familiar words that are equally difficult to understand. These word lists are also phonetically balanced to represent the frequency of occurrence of sounds in everyday life [56].

Many researchers in their listening tests have used the speech material containing nonsense syllables in different contexts like vowel-consonant (VC), consonant-vowel (CV) vowel-consonant-vowel (VCV) and consonant-vowel-consonant (CVC) [57]. The responses from these intelligibility tests can have open or closed set evaluation. The administration of the speech test may be in a fixed situation (for example same SNR) or varied situations.

The subject responses have been analyzed using different approaches. Miller et al. (1955) proved that information transmission analysis provides a measure of covariance between stimuli and responses. Information transmission analysis can be carried out for different feature grouping of stimulus and it has been used for evaluation in many studies [6], [57], [58].

Humes et al. (1987) use binary evaluation criterion for recognition scores to analyze subject response i.e. correct or incorrect [59]. Stimulus-response confusion matrix for response analysis proposed by ter Keurs et al. (1992), Tyler et al. (1992) takes care of chance scoring and errors in the pattern of phoneme confusions. Earlier researchers used different syllables like VC, CV, CVC and VCV in their experimentation. In case of VCV syllables, greater masking takes place in intervocalic consonants as vowels are present on both sides. Therefore, VCV syllables were considered to be the most appropriate test material in our research assisting to achieve our prime objective [50]-[51], [60]-[61].

Findings of Literature Survey

Researchers have reported various compression techniques to overcome the problem of acoustic amplification. Different compression techniques stated were automatic volume control, multiband compression systems, compression limiting and syllabic compression [1], [2], [8]. The hearing impairment is frequency dependent so multiband compression technique was found superior than others [2], [9]. In multi band compression method, 2-3 bands are sufficient to achieve adequate compensation for recruitment, since the compression reduces the spectral contrasts in complex signal [11]. With the objective to avoid the spectral flattening associated with Multi band compression and to compensate for narrow dynamic range, Asano et al. (1991) reported a digital hearing aid [10].

Researchers investigated frequency transposition techniques and frequency compression techniques, as most of the hearing loss is associated with high frequency region [1]. Frequency compression can help to improve the recognition of consonants, monosyllabic words and sentences in noisy environment, but it doesn’t improve the discrimination of consonants and shows adverse effect on vowel recognition, can be improved using frequency transposition technique [18].

Many hearing impaired subjects have sensorineural hearing loss in both ears and is overcome using dichotic presentation of speech signal and binaural hearing aid [19]. The split of speech signal in the odd–even frequency condition is more accurate than in the low–high condition for dichotic presentation [13].

An implementation a comb filter based on different critical bands, corresponding to the auditory pattern described by Zwicker [71] used dichotic presentation of speech signal [4], [25], [28]. P. N. Kulkarni et al. showed 1/3rd octave band and critical band based comb filters are superior to constant bandwidth filters [27]. Researchers implemented different coefficient FIR filters (comb filters) and evaluation showed improvement in response time, recognition score and information transmission analysis. These processing schemes were useful for subjects having mild to moderate deafness [4], [28].

The time-scale analysis of wavelets and properties of cochlea coincide with each other [30]. Researchers have used wavelet packet filters for coding of speech signal for simulation of cochlear implants [33], [36], [39].

Researchers used filters with time frequency decomposition of input speech signal is done by WPT [37], [38]. Kolte M. T. et al. developed 10 band modified wavelet packet algorithm using Simulink and evaluated on seven sensorineural hearing subjects. Results showed subjects with low and high frequency hearing impairment were benefitted [37].

Earlier survey of literature shows researchers used digital signal processors (DSP) to implement the FIR (Finite Impulse Response) filter banks [4], [21], while few researchers have reported that FPGAs are more suitable than DSPs as highly parallel computations are required [42], [46]-[48].

The evaluations of speech processing algorithms are carried out using listening tests. These tests use different types of speech material as it presents a range of acoustic, linguistic phonetic and lexical variables [49]. Earlier researchers have used different syllables like VC, CV, CVC and VCV in their experimentation. Greater masking takes place in intervocalic consonants as vowels are present on both sides, therefore, VCV syllables were considered to be the most appropriate test material in our research [60] –[61].

The combined technique of development and real time hardware implementation on FPGA platform of optimized wavelet packet using binaural presentation may reduce the probability of cochlear implant.

Due to difficulties in available methodologies, there is great need for further research in this area and to find novel solution for the said problem. It is revealed from the detailed literature review that the reduction in the spectral masking is better solution to improve the perception of hearing impaired. The core researchers from the relevant domain also asserted that the problems due to the spectral masking are not completely solved by using band pass filter banks and comb filters, except they are useful only for moderate deafness. To overcome the above limitations, there is an urge to propose an algorithm which improves the perception of hearing impaired people and can also be helpful for moderate to severe deafness. Further, it is also essential to realize the prototype with efficacy.

Chapter 3

PHYSIOLOGY AND MATHEMATICAL MODEL OF HUMAN AUDITORY SYSTEM

Hearing impaired persons, especially those with moderate or acute hearing loss, suffer from diminished opportunities for conversation, embarrassment of misunderstanding and reduced perception of voice tones, which convey feelings, emotions and difficulty of easy participation in discussion, groups and meetings.

This chapter describes detailed function of the auditory system (physiology of hearing). It also explains types of hearing impairment and perceptual consequences of sensorineural hearing impairment on speech perception. The mathematical modeling of auditory system along with outline of proposed schemes is explicated.

3.1 Function of the Auditory System

The entire auditory system embraces central and peripheral systems. The central auditory system comprises brainstem nuclei, thalamic nuclei and auditory cortex, which carries out processing and interpretation of auditory information. The peripheral auditory system is categorized as outer ear, middle ear, inner ear, and cochlear nerve [3].

Figure 3.1 shows the anatomy of peripheral auditory system. The tympanic membrane (also known as eardrum) separates outer ear and middle ear. The outer ear collects and channelizes acoustic signal to set vibrations in the eardrum. The middle ear transmits the vibration of the eardrum to the cochlea (filled with fluid) situated in the inner ear and provides an impedance matching between the two. Cochlea acts as Fourier analyser for incoming sound and produces neural impulses, which are carried by fibres in the cochlear nerve to the brain.

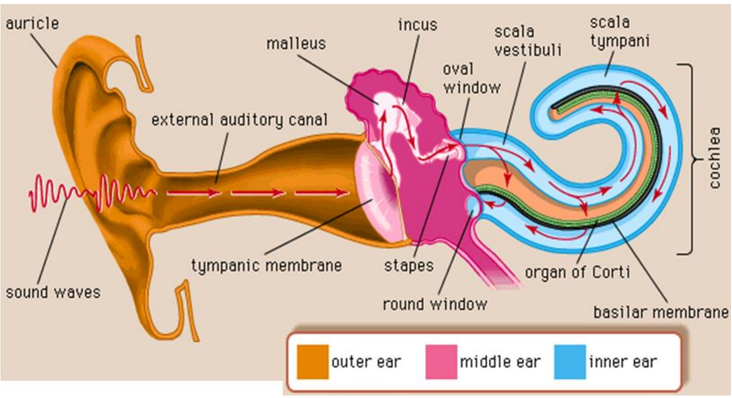


Fig. 3.1 Anatomy of peripheral auditory system [Source: 1997, Encyclopaedia Britannica]

3.1.1 Outer Ear

The outer ear is made up of auricle and external canal (meatus). Auricle lies at a 30-degree angle to the side of head. Pinna modifies the incoming signal at higher frequencies and thus helps in sound localization [2]. The auditory canal protects the ear from distant bodies and acts as filter transfer function. It contributes an increase in eardrum with a sound pressure of about 17 dB at frequency of 2700 Hz [3].

3.1.2 Middle Ear

An air filled cavity (middle ear), consists of tympanic membrane, with delicate chain of three tiny bones (ossicles) and middle ear muscles. The three little bones are the malleus, the incus and the stapes. Malleus is attached to the tympanic membrane at one end and to the incus at the other end. Incus makes contact with the stapes, the footplates of which are attached to the membrane of oval window (starting point of the inner ear). The eardrum vibrates in response to the change in pressure due to the impinging sound waves. Its vibratory pattern allows efficient transfer of sound waves from outer ear to the middle ear [63]. The vibrations are conducted through the air filled canal of the outer ear via the ossicles to the oval window (through the stapes footplate) and into the fluid filled cochlea of inner ear. There is an impedance mismatch between the air filled canal of outer ear and fluid filled inner ear. The tympanic membrane and the chain of tiny bones (ossicular) provide impedance matching for proper transmission of sound energy from air to fluid and it also reduces the amount of reflecting sound [2]. The impedance matching occurs through two mechanisms. First, the ratio of the area of eardrum to the area of footplates (19:1) amounts to an increase of approximately 25 dB in sound pressure within the middle ear. Second, the lever action of malleus and the incus (1.3:1) is equivalent to an increase in sound pressure of about 2 dB [3].

It has been projected that the high frequency limit of the human hearing system is set by the transmission characteristics of the middle ear. It operates as a resonant system tuned between 700 Hz to 1200 Hz with a falloff in transmission on either side of this frequency range. The muscles attaching with the ossicles contract for intense sounds and thus reduces the transmission of sound known as acoustic reflex. It plays an important role in protecting the inner ear from excessive intense sounds. However, due to slow activation of acoustic reflex, impulsive signals are likely to reach the inner ear and cause damage. The other function of acoustic reflex is the attenuation of self-generated sounds emitted during chewing and before vocalization [2].

3.1.3 Inner Ear

The inner ear is made up of three parts namely semi-circular canals, vestibule, and the cochlea. Cochlea is a snail shaped cavity with a spiral shaped structure, filled with incompressible fluids and also has bony rigid walls [3]. Figure 3.2 shows, the structure of the inner ear and a transverse section of the cochlea, with its three chambers: scala vestibuli, scala media, and scala tympani. It is divided along its length by two membranes, Reissner's membrane and basilar membrane (BM). Scala vestibuli and scala media are separated by a bony membrane called Reissner's membrane, and basilar membrane separates scala media from scala tympani. The base refers to start of cochlea at the location of oval window, while the other end is termed as the apex. Received signal results in the movement of oval window, the application of pressure differences at the tympanic membrane causing movement of cochlear fluid which generates up and down movement of the basilar membrane. The mechanical property of the basilar membrane along with fluctuating frequency of input signal influences the occurrence of vibration. The basilar membrane is less stiff and wider at the apex while it becomes comparatively stiff and narrow at the base. The low frequency signals generate maximum deflection at the apex whereas the high frequency signals generate maximum deflection at the base [2]. Figure 3.3 shows the vibrations of basilar membrane for low, medium and high frequencies.

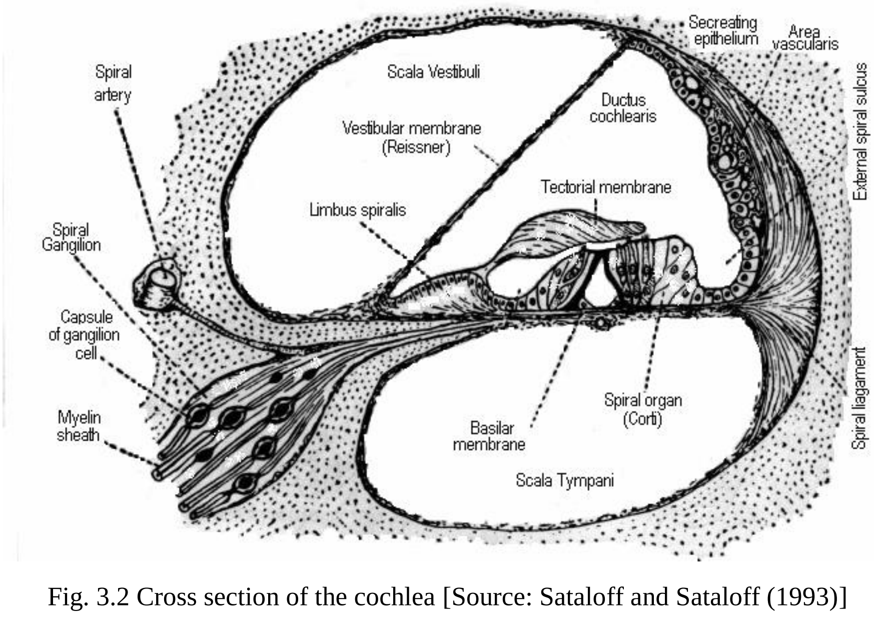


Fig. 3.2 Cross section of the cochlea [Source: Sataloff and Sataloff (1993)]

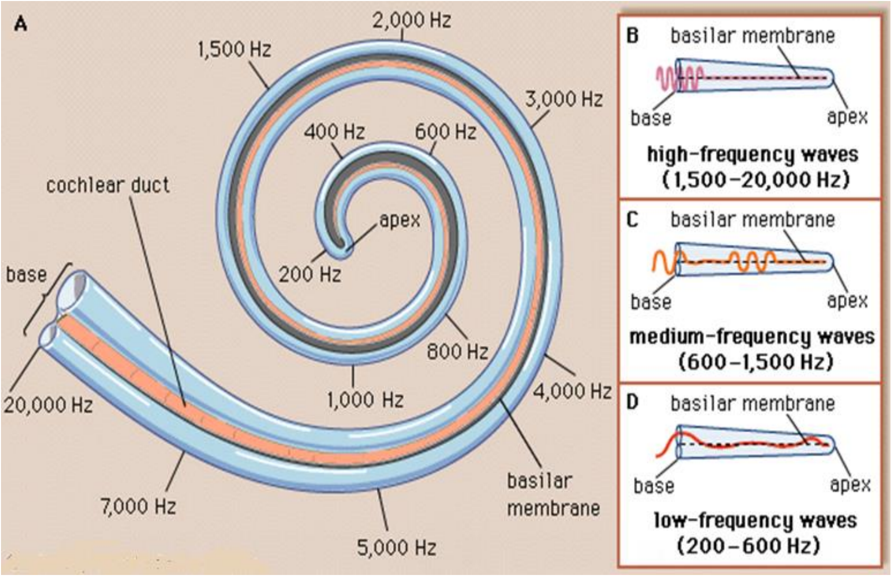


Fig. 3.3 Location of vibration at basilar membrane for different frequencies

[Source: 1997, Encyclopaedia Britannica]

Figure 3.4 shows Organ of Corti. It is covered by tectorial membrane and sits on basilar membrane containing sensory hair cells [65]. These hair cells are arranged as two rows of "inner" hair cells and three-to-five rows of "outer" hair cells.

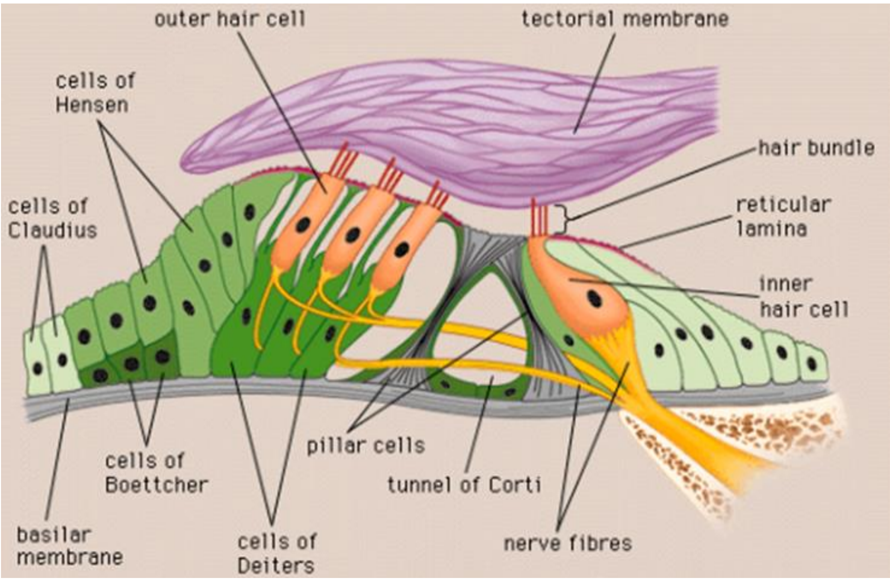


Fig. 3.4 Organ of Corti [Source: 1997, Encyclopaedia Britannica]

Outer hair cells are about 20,000 to 25,000, each with about 140 hairs and inner hair cells of about 3,500, each with about 40 hairs [2], [3]. Outer hair cells sharpen the tuning and improve the basilar membrane responses, thus producing high sensitivity and selectivity [2]. The information about sounds is carried via inner hair cells when the inner cells stimulate the afferent neurons.

3.2 Hearing Loss

Depending on the site of damage hearing losses are classified into four types, called as conductive loss, sensorineural loss, central loss, and functional loss. Conductive loss arises due to the damage of the auditory canal, eardrum, or middle ear. This loss causes reduction in the sound energy reaching to the cochlea. Conductive loss occurs due to wax, fluid, pus in the external auditory canal, or infections in the middle ear, or due to otosclerosis disease (abnormal rigidity of attachment of stapes to oval window). In conductive loss both hearing threshold levels and discomfortable loudness levels increases, but frequency selectivity does not reduce [66]. In most cases this loss can be treated medically or surgically [2], [3]. Sensorineural hearing loss occurs due to damage in cochlea (sensory) or auditory nerve (neural) of the inner ear. This loss is classified as cochlear (sensory) impairment and retrocochlear hearing impairment due to defect in the cochlea and in the auditory nerve, respectively.

Different causes for sensorineural hearing loss are congenital or hereditary factors, disease, tumors, acoustic trauma, old age, long term exposure to industrial noise, exposure to loud sound or toxic agents. The hearing thresholds are relatively elevated for the frequencies in the 1-4 kHz in the congenital sensorineural damage. Various diseases like typhoid, meningitis, etc indicate a peculiar pattern in audiogram (moderate to severe bilateral sensorineural loss is seen in typhoid) [3], [67].

A high threshold frequency in the range 3-6 kHz generally implies an acoustics trauma (hair cells loss, due to exposure to loud sound).With ageing basilar membrane hardening (hair cells damage nearer to the cochlea base) of the cochlea leads to degeneration of neurons in the complete auditory system and loss of hair cells in the organ of Corti. Impairment being bilaterally symmetrical has more hearing impairment at high frequencies and also increases difficulties in speech perception. With an increase in age the hearing loss at high frequencies continuous to increase with loss of sensitivity and becomes worse in time. The damage of auditory nerve fibers causes retrocochlear impairment resulting in tumor and hemorrhage (bleeding). Primary degeneration of auditory neurons may occur due to viral infection. Secondary degeneration is associated with peripheral processes of auditory neurons resulting in hair cells loss, mostly in the first half of basal turn. The probability of survival of neurons at apex is more [1].

The damage in the central nervous system results in central impairment. These patients have lower ability to interpret, compile or understand speech [3]. Central impairment occurs due to damage to auditory cortex, inflammation of the membranes covering the brain and spinal cord (meningitis), skull trauma, or congenital defects and may result in reduced speech comprehension ability [68]. Functional deafness is more psychological or emotional than physiological.

In acoustic measurements sound level is measured in Decibel (dB). Here sound pressure of 20 μPa is considered as reference level and sound level measured with respect to this is termed as sound pressure level (SPL).Sensational level is the sound level measured in dB when the threshold level of listener is taken as reference. The average hearing threshold of young adults is known as hearing level (HL). This hearing level is used as sound level of pure tones in clinical practices. Threshold measurement of hearing threshold level (HTL) is used to clinically identify hearing impairment. Audiogram is a plot of threshold as a function of tone frequency. Audiograms with other results are useful in computing the level of hearing impairment and also its cause [2]. The audiograms of the sensorineural hearing impairment have typical shapes and it depends on the pathology such as raised up thresholds at low and high frequencies. There are various impairments of moderate to profound bilateral sensorineural hearing impairment and unilateral very severe or profound deafness. In addition to the elevated thresholds, sensorineural hearing impairment is characterized by loudness recruitment, reduced frequency selectivity (increase in spectral masking) and reduced temporal resolution (increase in temporal masking) [4], [7]. These characteristics of sensorineural impairment are detailed in following sections.

3.2.1 Loudness Recruitment

Loudness recruitment is drastic increase of loudness associated with the rise in sound intensity. Hair cell damage and particularly damage to outer hair cells is possible cause for loudness recruitment. The sensitivity for low input sound levels is raised by outer hair cells leaving the response to high level sounds unchanged. The level at which the tone becomes unreasonably loud is termed as loudness discomfort level (LDL). For normal hearing persons it ranges from 100 to 110 dB SPL and may be same for people with sensorineural hearing loss [73]. The distinction between the loudness discomfort level and hearing threshold level is termed as dynamic range. In conductive loss, as both levels increase so there is no noticeable change in dynamic range. The hearing threshold level rises in case of sensorineural hearing impairment with no significant change in the loudness discomfort level; hence dynamic range may reduce drastically with loss of speech intelligibility [2].

3.2.2 Frequency Selectivity

Frequency selectivity refers to the ability to detect a tone of complex signals in the presence of other frequency bands. It is generally measured using different methods like critical bands, psychophysical tuning curves, masking patterns, and just-noticeable differences (JNDs) in frequency. Frequency selectivity is explained and quantified using masking effect. Masking is a phenomenon in which the perception of one sound is obscured by the presence of another. More specifically, when any person tries to listen two sounds simultaneously or concurrently, the threshold of hearing for other increases by the presence of one. Simultaneous sounds cause frequency masking, where a lower-frequency sound generally masks a higher-frequency sound. There is a possibility that one signal may get masked by another signal whose frequency components are close to or same as that of the previous one. It is crucial to distinguish between frequency selectivity and frequency discrimination. Ability to distinguish two tones when successively presented that differ in frequency content over time is termed as frequency discrimination [2]. Shift in the neuron firing patterns, may possibly result in the frequency discrimination [62]. It is found that frequency selectivity is reduced in person with sensorineural hearing loss [69].

Peripheral auditory system can be formulated as a bank of band pass filters, termed as auditory filters with overlapping pass bands. Individual location on the basilar membrane acts like a filter with different central frequency. Fletcher (1953) measured the threshold of a sinusoidal signal as a function of the bandwidth of a band pass noise masker. The noise was always centered at the signal frequency and the noise power density was held constant. Thus the total noise power increased with increase in bandwidth [70] as shown in Figure.3.5. With increase in the masking noise bandwidth, the signal threshold rises at first, but then flattens off. Further increase in masking noise bandwidth does not change the threshold.

Critical bandwidth (CB) concept is useful to explain the frequency selectivity of auditory system. It is the bandwidth at which the signal threshold stops to increase. For the center frequency below 500 Hz the critical bandwidth remains constant at 100Hz and above 1000 Hz it is 15–20% of center frequency [71], [73].

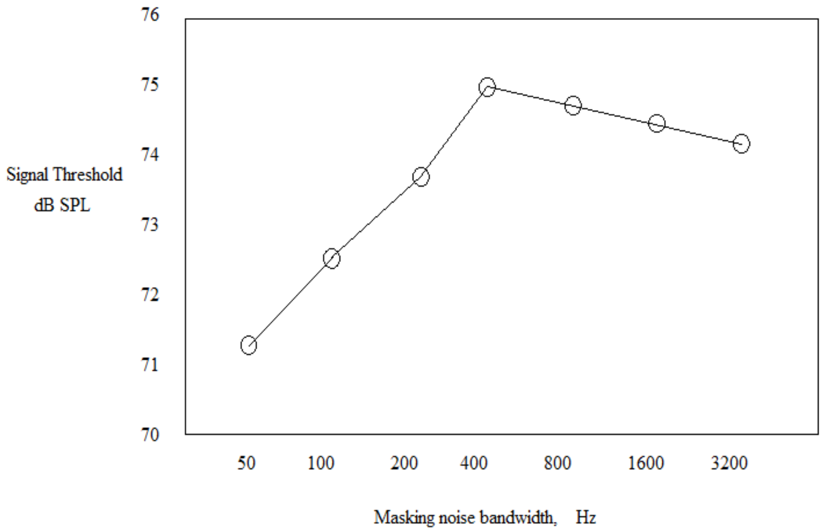


Fig. 3.5 The threshold of a 2000 Hz sinusoidal signal plotted as a function of the bandwidth

Patterson (1976) made use of “notch noise method” to determine threshold for sinusoid, centred in spectral notch of a noise as a function of notch width [74]. Auditory filter can be detailed in terms of equivalent rectangular bandwidth (ERB), on the basis of the results obtained from the above method. It is observed that ERBs are 11-20 % of centre frequencies. It is also observed that ERB and the CB described by Zwicker (1961) are almost similar, when frequencies are above 1000 Hz [71].

Figure 3.6 depicts the CB given by Zwicker (1961) and the estimates of ERB for auditory filter.

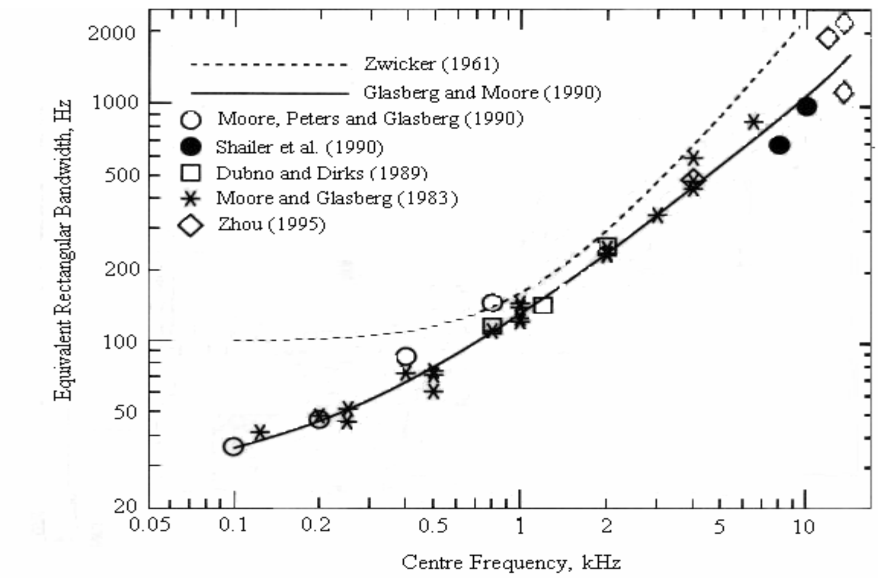


Fig. 3.6 The critical bandwidth described by Zwicker (1961) and the equivalent rectangular bandwidth (ERB) of the auditory filter [Source Moore (1997)]

The auditory filter in people with sensorineural hearing impairment is wider as compared to normal human beings. An increase in upward spread of masking (amount of masking raises nonlinearly on high frequency components sides) results in the widening of auditory filters in impaired persons. Speech perception is reduced in listeners with sensorineural hearing impairment as their frequency selectivity level is low as compared to normal people [72].

3.2.3 Temporal Resolution

The minimum detectable gap between two successive signals is defined as the temporal resolution. The gap detection method is normally used to measure it. Here the listener has to detect the gap between the two test signals, in the presence of continuous background broadband notched noise. A trial on hearing impaired and normal people showed that the temporal resolution is 2 to 3 ms for normal listener and gets degraded up to 8 ms for hearing impaired [75]. Forward and backward masking of weaker segments by intense one results in poor time resolution. In forward masking, signal follows the masker while, in backward masking the signal segment precedes the masker. In case of sensorineural impaired the above phenomenon is very intense. Forward and backward masking frequently occurs within 100 ms at beginning or the end of masking [73]. The effect of forward masking is most effective within 10 ms of the masker and becomes negligible after 100-200 ms. The backward masking, at most, extends over 20 ms before the masker. Both the masking effect exists due to temporal overlap of cochlear responses [2].

3.2.4 Effect of Damage to Sensory Receptors and Auditory nerve

Mechanical vibrations are converted into electrical signals by hair cell (also known as the sensory receptors of the inner ear) and the information is delivered to the brain in the form of neural impulses. The signals are detected by inner hair cells and result in excitation of nerve fibers. The basilar membrane responses to low level sounds are increased by the outer hair cells in auditory transduction. They also play an important role in fine tuning of auditory responses to a particular frequency [76]. Parameters like genetic disorder, aging, loud noise exposure, infection, ototoxic drugs may cause damage to the hair cells. Since these cannot be replaced by cell division their disappearance is associated with reduction of hearing sense. Hearing loss usually occurs at high frequencies, as sensory cells at base are more prone to damage than those at apex with low frequencies. The sensitivity to basilar membrane vibration reduces when inner hair cells are damaged; resulting in an increase of hearing threshold, but the frequency selectivity remains unaffected. The pronounced effect of damage to outer hair cells is loss in sensitivity along with reduced dynamic range and loudness recruitment. Damage to the auditory nerve fibers generally weakens the frequency coding resulting in loudness recruitment, reduced frequency selectivity and reduction in dynamic hearing range [4], [7].

3.3 Acoustic cues for Perception of Speech

Cues are defined as an essential stimulus pattern in a perceived occurrence. The study of acoustic cues includes all the aspects of speech signal necessary for speech perception by listeners. The physical quantity is represented by the acoustic cues, e.g. vowel quality is cued by relations between formant frequencies. For the identification of various segments normal hearing persons have to rely on cues like sequential intensities, duration and spectral composition [77]. Auditory processing deficit results in the loss of many acoustic cues to hearing impaired. The prime factors responsible for reduced recognition of acoustics cues are abnormal loudness growth, reduced discrimination due to poor frequency resolution and auditability due to increased threshold in frequency region of energy concentration [78].

3.3.1 Vowel Distinction

Formant pattern, i.e. relation between formant frequencies forms the principle cue for vowel quality. The relation between F1 formant frequency and F2 formant frequency is essentially important. The cue for front vowel includes large separation between F1 and F2 and a small gap between F2 and F3. On the other hand, for back vowels, the difference between F1 and F2 is small, while the difference between F2 and F3 is large. For central vowels the formant patterns are equally spaced i.e. uniform formant pattern. To differentiate the vowels, intensity difference serves as secondary cue. The vowels are articulated between consonants at rapid rates, while they are rarely in steady state in normal speech [2].

3.3.2 Differentiation in Voiced and Unvoiced consonant

Depending on the context, acoustic cues vary widely for voicing feature. Voice onset time (VOT) i.e. time interval between release burst and start of the vocal fold vibration, forms the main cue for word initial position consonant. For all contexts, relative intensity forms another cue for voiced and unvoiced differentiation. Voiced consonant has smaller intensity in comparison with its unvoiced counterpart [61].

3.3.3 Place of Articulation

Formant transitions/loci and noise spectrum are two important cues for place of articulation. Important information carried by all formant transition can be used as a cue for place of articulation. It was observed that for consonant differentiation, F2 transition locus is more important. Locus is defined as hypothetical intersection point derived by superimposing trajectories of formant transition from consonants in different vowel contexts. For all type of consonants, F1 rises from near- zero value to F1 frequency of subsequent vowel [61]. A spectral characteristic of noise is based on the location of noise generator and serves as cue for this location [79].

3.3.4 Manner of Articulation Distinction

Duration of fricative noise followed by burst release noise are used to differentiate stop and affricates. The duration of fricative noise is more in affricates as compared to stops. The cue for stops and affricates is the silence period ranging from 40 to 120 ms. The presence of turbulent noise ranging from 70 to 140 ms (on average) are used to mark the fricatives.

Periodic source parameters and absenteeism of noise are used to distinguish nasals, liquids and approximants. Very low F1, weak formants and presence of antiformants are cued for nasals and liquids. Approximants have vowel-like formant frequencies with rapid transitions. Unlike nasals and liquids, they do not drastically differ in their intensity from neighboring vowels [2].

3.4 Acoustic cues for Hearing Impaired

The consonants play a vital role for speech intelligibility and are very easily confused. The consonant are characterized by articulatory features like manner, duration, voicing and place of articulation [57]. There are various characteristics of the adjacent vowels on which the acoustic characteristics of consonant depends [79].

Listeners with sensitivity losses more than approximately 30 dB have reduced frequency resolution with significantly broad auditory filters. Reduced spectral contrast between peaks and valleys is used to describe vowels processed through above filters. The response of lower spectral amplitude components reduces with the presence of high amplitude spectral component at adjacent frequencies. This may function to increase the differences in amplitude near spectral peaks. For observed peak to valley difference of 1 to 2 dB, normal listener gained greater than 75% accuracy. A 6 to 7 dB difference for identification is required for listener with moderate hearing loss [80].

To discriminate vowel like stimuli presented in noise, it was observed that higher spectral contrast is required for listener with broad auditory filters [77]. It was observed by Baskent et al (2006) that the capability of resolving spectral peaks of rippled noise was correlated with recognition of vowels and consonants and input filter patterns, a measure of frequency resolution, were correlated with recognition of vowels. In natural speech, the difference in levels of back vowels lies between 5 to 7 dB, while in case of front vowels it is as large as 25 to 30 dB [62].

Formants F2 and higher ones are smoothed out with spectral smearing and upward spread of masking, leaving widened F1. To identify vowels with smaller difference between formant frequencies, hearing impaired listener makes use of cues like duration, which is redundant for normal listener. For both hearing impaired and normal hearing people there is reduction of accuracy of identification of stop voicing. For identification of place feature, acoustic cue within a consonant's onset are used by hearing impaired listener with gradually sloping or flat audiometric conFig.uration [72]. In case when only the transition cues are audible, hearing impaired listener gets confused between fricatives and stops. Reduced discrimination of F2 transitions by hearing impaired listeners may be caused by spectral and/or temporal masking. In case when the cue is at falling second formant transition, hearing impaired listeners face more difficulty to identify place of articulation for stop consonant rather than when it is at rising formant transition [81]. As the formant transitions are much slower in comparison with those for voiced stops, hearing impaired is easily able to recognize the place of articulation.

To discriminate fricatives like /s/, /z/ and /sh/, hearing impaired subjects with high frequency loss have difficulty since the energy in the spectra of these alveolar fricatives lies at 4 kHz and above (Pickett, 1999). The averaged power spectrum obtained by averaging square spectra in certain time intervals was used to measure the spectral amplitudes. The high amplitude spectral peaks believed to be more informative, are emphasized more than smaller peaks and are the merit of using power spectrum before averaging. From the locations of formant frequencies one may be able to expose the vocal tract conFig.uration at certain point in time during consonant-to-vowel or vowel-to-consonant transition and also can trace the first three formant frequencies [82].

3.5 Effects of sensorineural loss on Speech Perception

In sensorineural hearing loss, the dynamic range for hearing gets reduced with an abnormal increase in perceived loudness with an increase in sound level. A range of minimum 30 dB is covered by normal conversation speech and is larger than dynamic range of hearing in many cases of sensorineural loss. The continuous exposure to sound above loudness discomfort level may hamper residual hearing area. Speech reception is drastically affected by degradation in temporal resolution and increase in temporal masking. Sufficient temporal resolutions of subphonic segments like formant frequency transition and noise bursts are necessary for proper reception of consonants. The consonant with much information have low intensities, while vowels are more intense in speech signals [83]. So vowel may mask consonantal segments, stop bursts may get masked and result in poor discrimination of stop consonants, as they are 30 dB weaker than the subsequent vowel [84].

A dynamically varying broadband spectrum may be used to describe speech signals. The significant traits of acoustics are amplitude, frequency, and bandwidth of formants (spectral resonances specific to the vocal tract conFig.uration, nature of excitation (voiced/unvoiced), pitch (fundamental frequency of vibration of vocal chords) in voiced segments, and duration and frequency band of noise bursts, closure duration, and transition of formants).

Psychoacoustic measurements of the smallest detectable change or amplitude difference limen (DL) are reported by Flanagan (1972). For first and second formants the amplitude difference limen are 1.5 and 3 dB respectively. The DL for overall intensity of vowels and fricatives are about 1.5 and 0.4 dB respectively. DL for synthetic vowel fundamental frequency is 0.3-0.5%, while it is 3-5% and 20-40% for formant frequency and bandwidths, respectively [85].

Broadband noise with spectral peaks and notches related to the poles and zeros of the vocal tract filters are salient features of fricatives. *Q*'s (ratio of center frequency to bandwidth) should be greater than 5 for detectable spectral peaks and greater than 8 for notches. Moderate hearing impaired listeners require at least 7 dB, while, normal listeners are able to identify vowels with spectral peaks of only 2 dB above the spectral valleys. This indicates the effect of spectral masking. In natural speech, the vowels are characterized by spectral peak-to-valley differences of at least 8 to 10 dB; hence the identification of vowels by hearing impaired is better [80]. Uncertainty in the recognition of the transition of formants and frequency band of the noise burst may exist due to spectral masking.

High frequency noise of /t/ is differentiated from the low frequency noise of /p/ and /k/, in case of unvoiced stops /p, t, k/ with vowel /*a*/. The discrimination between stop /k/ is characterized by slightly higher frequency noise in comparison with /p/. The second formant of vowel /*a*/ initially rises for /b/, and thereafter falls for /d/ or /g/, in case of voiced stop consonants/b, d, g/. In case of /b/, the formant frequency is relatively more audible resulting in less confusion with /d/ or /g/, but it is difficult to distinguish /d/ and /g/. As compared to stop consonants, the nasals /m, n/ are somewhat more intense and slightly longer, whereas the fricatives /s, z/ are intense and longest. The fricatives /s, z/ have energy concentration at high frequencies as compared to fricatives /f, v/. To discriminate nasal /m, n/ the second formant transition at the end of vowels play a vital role [57].

From the above discussion, one can say that the consonantal segments, in which noise bands and formant transitions are not very different, will have perceptual confusions due to spectral masking. Amplitude or temporal cues are related to nasality and duration features. The frication and place characteristics are related to the presence of aperiodic noise and spectral differences respectively [50].

The studies on spectral spreading reveals that the poor speech perception by subjects with inner ear damage might be due to reduced frequency selectivity [76]. Hearing impaired people may face difficulty in identifying the place feature, as it is cued by spectral differences.

3.6 Speech Processing for Sensorineural Impairment

Hearing aids are used to moderately compensate the hearing impairment. It amplifies acoustic signals with frequency or gain characteristics, which helps to overcome the hearing deficiency [8]. All conventional hearing aids, analogue or digital, amplify sounds to make them audible for people with hearing loss. To overcome the problem of increased threshold, conventional hearing aids are frequently used. The basic components used in hearing aids are microphone, electronic filter, electronic circuitry, ear phone, and battery as a power source. It also includes earmold for coupling the earphone output to the external ear canal.

Depending on the size and way of wearing, conventional hearing aids are classified as, body-worn hearing aid, eyeglass hearing aid, behind-the-ear (BTE) hearing aid, in-the-canal (ITC) hearing aid and in-the-ear (ITE) hearing aid [1].

Recent developments in microelectronics and digital signal processing have resulted in development of superior digital hearing aids. The evolution in field programmable gate arrays (FPGA) technology, has led to the development of sophisticated features for better sound reproduction, maintaining lower power consumption of the devices and keeping small size.

The people with severe to profound hearing loss normally prefer BTE hearing aids. The components are enclosed in a small elliptical case that fits behind the ear. The signals are transferred to the ear canal through a flexible tube terminating in an earmold. In ITE hearing aid all components are enclosed in a small plastic case. It is fitted in concha and conserves few directional properties of pinna. Amongst all hearing aids, ITC hearing aid is smallest in size and completely fits in the ear canal. The available output power of this aid is low owing to its small size. So, this hearing aid is normally used by persons with moderate, flat, and gradual sloping impairment [1].

3.7 Mathematical model of Ear

Human auditory system is distributed into three sections, the outer ear, middle ear, and inner ear. The pinna (auricle) and the auditory canal (meatus) together form the outer ear. The middle ear, has an air filled cavity consisting of tympanic membrane, three ossicles and middle ear muscles. Ossicles are the malleus, incus and stapes. Inner ear or the cochlea resembles a cavity in the form of a spiral shaped structure and consists of a fluid. Outer ear acts as resonator. The pinna helps in sound localization by modifying the incoming signal at higher frequencies. The auditory canal protects the ear from foreign bodies. The amount of reflecting sound is reduced by impedance matching provided by the middle ear. The impedance matching provides sound pressure of about 25 dB in the middle ear. The frequency information for the brain is separated (Fourier analyser) by the inner ear. These parts are understood using the physics and mathematics in the following subsections.

3.7.1 Outer Ear as Resonator

The sound received by the external ear (pinna) passes directly into the ear canal. It resembles an amplifier with scrupulous dimensions, tuned to amplify the frequencies used in human speech. This amplification works through a process termed as resonance. All types of wave have resonance as its fundamental property. Whenever a travelling wave becomes trapped or confined in a small region, resonance occurs. The ear canal is fixed with ear drum at one end and is open at another end. An antinode at the open end is necessary for resonance to occur. One quarter of wave is the smallest portion of a wave that fits into the situation. This means that the wavelength of above wave should be four times more than the length of the ear canal [63]. This is shown in Figure 3.7

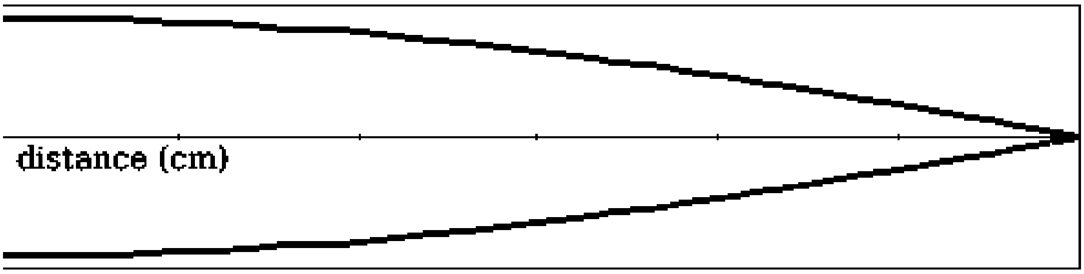


Fig. 3.7 Ear canal with open ended tube showing node and antinode

Equation 3.1 shows the relation between wavelength (λ) of the wave, its frequency (F) and speed of sound in air (C)

$$F = \frac{C}{\lambda} \qquad (3.1)$$

In case of resonant wave in the ear canal, λ=4L, hence substituting this value in above equation, we get

$$F = \frac{C}{4L} \qquad (3.2)$$

Where L = Length of ear canal

In above equation, after substituting the value of average length ear canal as 26 mm and speed of sound in air as 343.2 m/s, the value resonant frequency reaches around 3000 Hz. This value forms the peak of crucial frequencies produced in human speech. It indicates the resonation of the ear canal and the amplification of sound range [87].

$$F = \frac{343}{4 \times 26 \times 10^{-3}}$$

$$F = 3298.07\ Hz \qquad (3.3)$$

The ear canal, owing to its low resistance to sound, is able to prevent damping down frequencies which do not resonate inside it. It also simply behaves like a conduit to the ear drum for these frequencies.

3.7.2 Middle Ear for Impedance Matching

The middle ear transfers the sound waves from outer ear, filled with air medium, to the inner ear, filled with fluid medium. These two mediums are separated by a boundary. Part of the sound waves gets reflected on the boundary (P_r) while the remaining part passes through the boundary to the other medium (P_t). The transmission coefficient refers to that part of wave power that passes through the boundary. This is shown in Figure 3.8. In absence of middle ear, the sound wave gets attenuated by 30 dB (mathematically explained below). The middle ear uses two mechanisms to overcome this attenuation.

The transmission coefficient (T) is defined as ratio of total transmitted waves to the total incident waves. On similar lines, the reflection coefficient (R) is defined as ratio of total reflected waves to the total incident waves. The summation of reflection coefficient and transmission coefficient is unity. The mathematical expressions are given by the following equations.

$$T = \frac{P_t}{P_i} \qquad 3.4$$

$$R = \frac{P_r}{P_i} \qquad 3.5$$

$$R + T = 1 \qquad 3.6$$

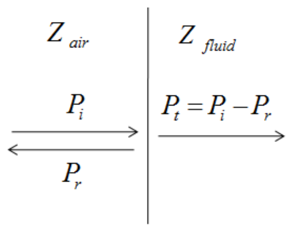


Fig. 3.8 Boundary separating Outer ear and Inner Ear

The acoustic impedance is defined as the product of density and wave speed. It is also the ratio of acoustic pressure to specific flow to the velocity

$$Z = \rho \times C = \frac{P}{u} \qquad 3.7$$

The relationship between incident acoustic flow velocity, reflected acoustic flow velocity and transmitted acoustic flow velocity is given by

$$u_i - u_r = u_t \qquad 3.8$$

The reflectivity coefficient and transmissivity coefficient in terms of fluid and air impedances are given by the equations 3.9 and 3.10 respectively

$$R = \frac{Z_{fluid} - Z_{air}}{Z_{fluid} + Z_{air}} \qquad 3.9$$

$$T = \frac{2 \times Z_{fluid}}{Z_{fluid} + Z_{air}} \qquad 3.10$$

The sound intensity (I) is the product of sound pressure (P) and velocity (u)

$$I = Px\ u \qquad 3.11$$

The average value of sound intensity over time (t_{av}) can be determine from the following equation

$$I = \frac{1}{t_{av}} \int_0^{t_{av.}} P \times u \times dt \qquad 3.12$$

From equation 3.7, the average value of sound intensity resulted to

$$I = \frac{1}{t_{av}} \int_0^{t_{av.}} \frac{P^2}{Z} \times dt \qquad 3.13$$

The above equation for intensity is rewritten by substituting the value of RMS for pressure as

$$I = \frac{P_{rms}^2}{Z} \qquad 3.14$$

Power transmission coefficient (τ) is the ratio of intensity of transmitted waves (I^t) to the intensity of incident waves (I^i).

$$\tau = \frac{I^t}{I^i} \qquad 3.15$$

Using equation 3.14, equation 3.15 becomes

$$\tau = \frac{\left(P_{rms}^t\right)^2 / Z_{fluid}}{\left(P_{rms}^i\right)^2 / Z_{air}} \qquad 3.16$$

The medium for transmitted wave is fluid, hence Z = Z_{fluid}, while that for incident wave is air, hence Z= Z_{air}.

Equation 3.16 can be rewritten in terms of transmission coefficient (T) (as defined in equation 3.4)

$$\tau = T^2 \frac{Z_{air}}{Z_{fluid}} \qquad 3.17$$

Using equation 3.10, the value of power transmission coefficient becomes,

$$\tau = \frac{4 \times Z_{air} \times Z_{fluid}}{\left(Z_{fluid} + Z_{air}\right)^2} \qquad 3.18$$

Putting the values of acoustic impedances of air (Z_{air}) and fluid (Z_{fluid}), equation 3.18 becomes [87]

$$Z_{air} = 4.15 \times 10^2 \text{ kgs/m}^2$$

$$Z_{fluid} = 1.44 \times 10^6 \text{ kgs/m}^2$$

$$\therefore \tau = \frac{4 \times 4.15 \times 10^2 \times 1.44 \times 10^6}{\left(1.44 \times 10^6 + 4.15 \times 10^2\right)^2} = 1 \times 10^{-3} = -30dB \qquad 3.19$$

The negative sign indicates the reduction in power by 30 dB when the sound travels from outer ear to inner ear i.e. its power reduces by 1000 times.

The middle ear uses two mechanisms to overcome the effect of sound attenuation. First mechanism includes pressure transfer from bigger surface area of the ear drum to the smaller surface area of oval window in ratio of 1:19, as shown in Figure 3.9 and expressed in equation 3.20

$$A_{oval} = \frac{1}{19} \times A_{drum} \qquad 3.20$$

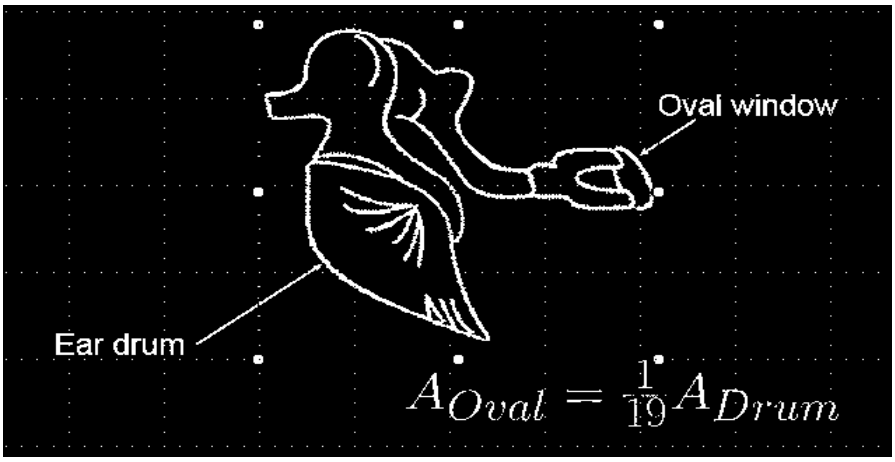


Fig. 3.9 Stapes Footprint

In second mechanism, the ear bones behave like compound lever. The product of applied force and the distance from the pivot point of lever is used here. Figure 3.10 shows the schematics of the arrangement of ossicles (bones).

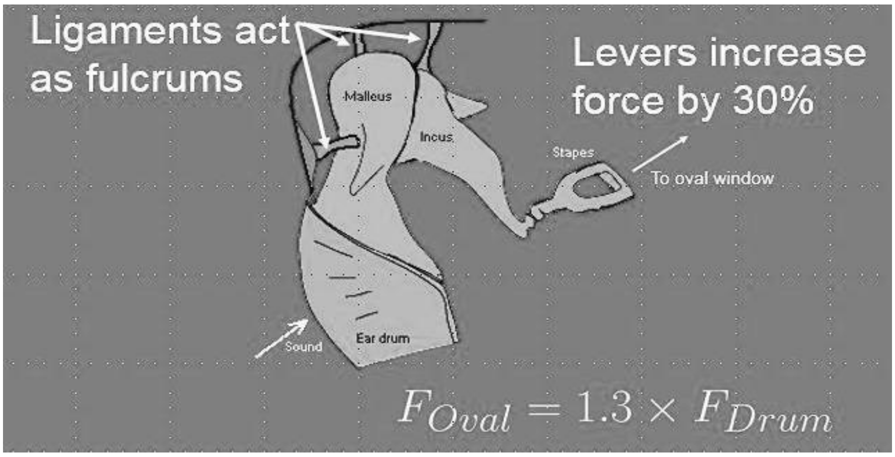


Fig. 3.10 Ossicles as Lever

The force on oval window get magnified by 1.3 times that on ear drum (The length of manubrium of the mallecus divided by the length of long process of incus (1.3:1)

$$F_{oval} = 1.3 \times F_{drum} \qquad 3.21$$

The pressure on oval window is given by the relation

$$P_{oval} = \frac{F_{oval}}{A_{oval}} \qquad 3.22$$

Substituting equation 3.20 and 3.21 in equation 3.22

$$P_{oval} = \frac{1.3 \times F_{drum}}{\frac{1}{19} x A_{drum}} \qquad 3.23$$

$$P_{oval} \cong 25 \times P_{drum} \qquad 3.24$$

We know, sound intensity is directly proportional to the square of pressure. For the applied pressure of 25 Pascal, the intensity becomes 625 or it corresponds to 28dB.

The two mechanisms of middle ear explained above assist to amplify the sound intensity from air medium (outer ear) to fluid medium (inner ear) by 625 times (28dB) partially overcoming the attenuation of 1000 times (30dB) [87].

3.7.3 Inner Ear as Fourier Analyser

Travelling wave theory forms the basis for explaining the inner ear (cochlea). The cochlea acts as a transducer that converts sound signals into electrochemical signals and acts as a Fourier Analyser separating the frequency information for the brain. The cochlea duct splits into two channel by the basilar membrane (BM) and are connected at the apex by the small aperture termed as helicotrema, as shown in Figure 3.11.

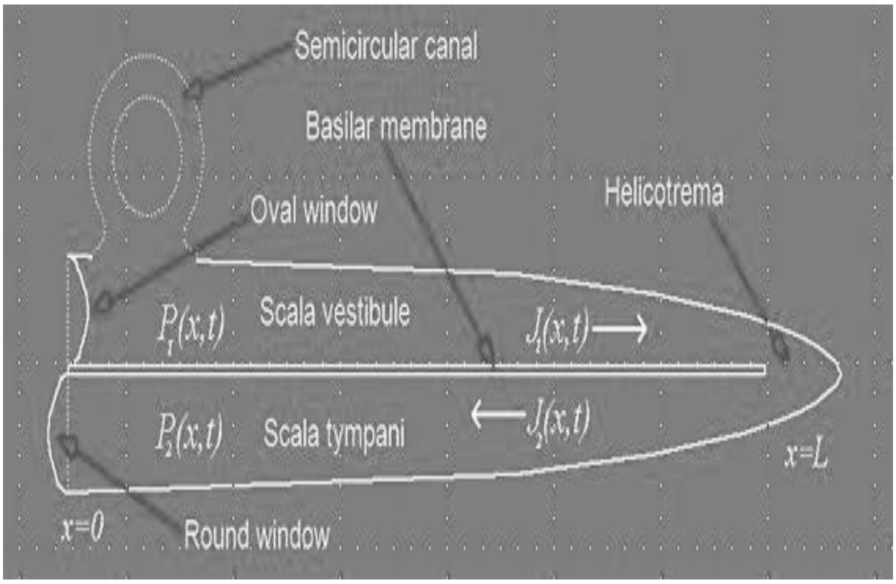


Fig. 3.11 Duct of Cochlea

A sound stimulus impinging on the oval window causes pressure $P_1(x, t)$ and $P_2(x, t)$ in both channels, t indicates the time and x shows the position along cochlea.

The pressure gradients include longitudinal currents $J_1(x, t)$ and $J_2(x, t)$ which flow in opposite directions shown in Figure 3.11.

The expression for relative current j and pressure difference p are shown below

$$j=J_1-J_2 \qquad 3.25$$

$$p= P_1-P_2 \qquad 3.26$$

A change in the relative current in time produces a change in the relative pressure in space as shown in equation 3.27.

$$\rho \frac{\partial}{\partial t} j = -bl \frac{\partial}{\partial x} p \qquad 3.27$$

Where,

b= fluid viscosity

l= scala height

ρ= fluid density

Negative sign indicates a pressure drop along the longitudinal direction.

Fluid incompressibility shows that change in relative current produces a corresponding change in the BM displacement h(x,t) as per the law of conservation of fluid given by equation 3.28

$$2b \frac{\partial}{\partial t} h + \frac{\partial}{\partial x} j = 0 \qquad 3.28$$

where h= Basilar membrane displacement

Taking time derivative of equation 3.28, we get

$$2b\frac{\partial^2}{\partial t^2}h=-\frac{\partial}{\partial x}\left(\frac{\partial}{\partial t}j\right) \qquad 3.29$$

From pressure current equation 3.27, equation 3.29 becomes

$$2\rho b\frac{\partial^2}{\partial t^2}h=\frac{\partial}{\partial x}\left(bl\frac{\partial}{\partial t}p\right) \qquad 3.30$$

$$\therefore\frac{\partial^2}{\partial t^2}h=\frac{l}{2\rho}\frac{\partial^2}{\partial x^2}p \qquad 3.31$$

The generalized form of the wave equation is given as below

$$\frac{\partial^2 A}{\partial t^2}=C\frac{\partial^2 B}{\partial x^2} \qquad 3.32$$

where, A=h(x, t),

B= p(x, t),

$C=\frac{l}{2\rho}$

Putting the value of the parameters,

A is function of BM displacement in terms of distance x and time t

B is function of pressure in terms of distance x and time t

C is constant as it is ratio of scala height to the fluid density, Equation 3.32 becomes

$$\frac{\partial^2}{\partial t^2}h(x,t)=\frac{l}{2\rho}\frac{\partial^2}{\partial x^2}p(x,t) \qquad 3.33$$

The above equation shows that the deformation of BM occurs due to the pressure difference. The response relation gets restricted to the basal end for passive response. Here the stiffness K(x) is high for BM. The relation between stiffness, pressure and displacement is shown below

$$P(x,t)=K(x)\,h(x,t) \qquad 3.34$$

We get linear wave equation for pressure by putting equation (3.34) in equation (3.33)

$$\therefore\frac{\partial^2}{\partial t^2}h(x,t)=\frac{l}{2\rho}\frac{\partial^2}{\partial x^2}K(x)\,h(x,t) \qquad 3.35$$

$$\text{but, } K(x)=\alpha\times e^{-\beta x} \qquad 3.36$$

The stiffness of BM for the distance ranging from x=0 to x=L is

$$K(x) = \alpha \qquad x=0 \qquad 3.37$$

$$K(x) = \alpha \times e^{-\beta L} \qquad x=L \qquad 3.38$$

$$\therefore \frac{\partial^2}{\partial t^2} h(x,t) = \frac{\alpha\, l}{2\rho} \frac{\partial^2}{\partial x^2} e^{-\beta x}\, h(x,t) \qquad 3.39$$

After simplifying the above equation (Numerically) the obtained results can be plotted graphically and are as shown in Figure 3.12. For three different frequencies, maximum deflection is produced by high frequency signals at the base (x=0) whereas low frequency signals generate maximum deflection at the apex (x=L)

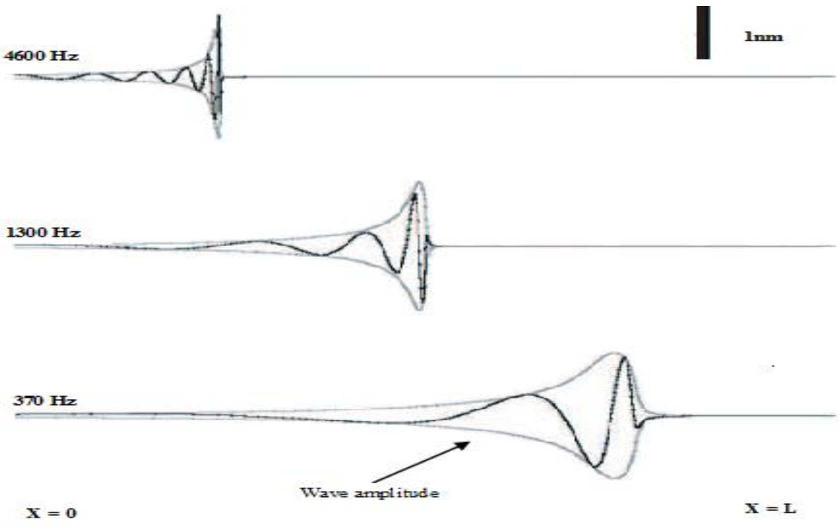


Fig. 3.12 Basilar membrane displacement patterns [Source: Duke et al. 2003]

3.8 Summary

The ear is remarkably wonderful piece of human auditory system that makes use of physical principles for the amplification, transmission and differentiation between complex sounds. These principles comprise of resonance, impedance matching and frequency separation. By keeping in mind the aforesaid objectives speech processing algorithms have been developed and implemented to overcome the difficulties in hearing by hearing impaired people.

3.9 Proposed Algorithm

Earlier reported schemes using binaural dichotic presentation to reduce the increased spectral masking effect are discussed in chapter 2. Spectral splitting scheme always stimulate sensory cells corresponding to alternate filter bands on the basilar membrane. It uses two complementary comb filters with pass bands corresponding to critical band auditory filter [4], [24].

We aim to develop and implement an algorithm which minimizes the masking effect and improves the perception of hearing impaired people. The research investigation was carried out in two phases. In first phase, the algorithm of splitting comb filters with 18 critical bands was designed having 512 coefficients. This algorithm was implemented in real time on FPGA platform and listening tests were carried out on seven hearing impaired people, referred as Experiment No. I.

Wavelet analysis is equivalent to a bank of band pass filters. In discrete wavelet transform the band pass filters divide the frequency axis into logarithmic bandwidth, while in wavelet packets the frequency gets divided into equal bandwidth. The theory of wavelet level is associated with particular frequency interval. The number of decimations carried out in the wavelet analysis may be thought of as an index label of filter banks. Efficient tools for speech analysis involve use of wavelet packets; they split the input signals into two-bands by filtering and down sampling at every decomposition level. Wavelet packet filter banks are designed by selecting the decomposition tree and then choosing the filters for every decomposition level of the tree. There is distinct time-frequency resolution for each decomposition level. The subsequent step involves selection of an appropriate wavelet filter for every decomposition level of the tree [95], [96]. Wavelet basis to be used for the decomposition wavelets like Daubechies, Coiflet, Biorthogonal and Symlet, of different order were assessed. The decomposition tree of 8 to 12 channel frequency bands was chosen and speech intelligibility was also checked.

In the investigation of second phase, decomposition tree of Daubechies, Symlet and Biorthogonal wavelets of different orders with 8 frequency bands were used. These wavelets were implemented using software based offline mode (MATLAB) and hardware based real-time mode (FPGA). In hardware based real time mode the testing was done on five normal people with simulated hearing loss, referred as Experiment No. II. The listening tests were also carried out on eight hearing impaired subjects using software based offline mode, referred as Experiment No. III and hardware based real time mode referred as Experiment No. IV.

Chapter 4

EVALUATION AND REAL TIME IMPLEMENTATION OF COMB FILTERS

Physiology of auditory system, mathematical model of ear, types of hearing impairment and characteristics of sensorineural hearing impairment on speech perception has been reviewed in previous chapter. D. S. Chaudhari et al. (1997), A. N. Cheeran et al. (2003) reported that binaural dichotic presentation of speech signal using comb filters based on critical bands, reduces the effect of increased spectral masking and improves the perception of sensorineural hearing impaired people. In chapter 2 binaural dichotic presentation schemes are highlighted along with a brief discussion on evaluation methods, listening test material and techniques. By carrying out the listening tests on people with sensorineural hearing impairment, a complete assessment of spectral splitting scheme with comb filters was done and reported in subsequent sections. In real-time processing on FPGA platform seven hearing impaired subjects participated for the evaluation of spectral splitting with unprocessed speech signal and with processed speech signal, referred as Experiment I. Latency and PSNR were analysed for the comb filter processing algorithm.

4.1 Comb Filter Design Method

For binaural splitting of speech signal, filters were designed using linear phase FIR filters. Frequency sampling technique was used to design these filters [90] - [92]. In the frequency sampling technique, the desired frequency response $H_d\ (e^{jw})$ is sampled at a set of uniformly spaced frequencies,

$$w = \frac{2\pi k}{N}$$

where $k = 0, 1, 2, \text{-----}\ \frac{N-1}{2}$ for N odd

$$= 0, 1, 2, \text{-----}\ \frac{N}{2} - 1 \qquad \text{for } N \text{ even} \qquad (4.1)$$

Inverse Discrete Fourier transform (IDFT) is used to determine the filter coefficients of this set of samples, i.e.

$$h(n) = \frac{1}{N}\sum_{k=0}^{N-1} H(k) e^{\frac{j2\pi nk}{N}} \qquad n = 0, 1, 2, \text{-----}\ N-1 \qquad (4.2)$$

The frequency response $H\ (e^{jw})$, calculated using N-point Finite Impulse Response (FIR) $h(n)$, will coincide with $H_d(e^{jw})$ at $w = 2\pi k/N$ and is given as

$$H(e^{jw}) = \sum_{n=0}^{N-1} h(n) e^{-jwn} \tag{4.3}$$

The L uniformly spaced samples of the frequency response $H(e^{jw})$ are given as

$$H(e^{j2\pi k/L}) = \sum_{n=0}^{L-1} h(n) e^{\frac{-j2\pi kn}{L}}, \qquad k = 0, 1, 2, \text{- - - -} L\text{-}1. \tag{4.4}$$

To generate N-filter coefficients for comb filter, a MATLAB program called “Critical_Band.m” has been used. It uses frequency sampling technique of linear phase FIR filter. The program calculates filter coefficients and is stored in “coeff.mat” file. The program provides the facility to observe the interpolated magnitude response in linear scale.

A MATLAB code compatible with Very high speed integrated circuit Hardware Description Language (VHDL) was developed which used “coeff.mat” file. A project file named “comb_filter.prj” was generated in HDL coder which included a test bench and function file. The test bench “combfilter_tb.m” used the unprocessed speech signals as input and “comb_filter.m” processed it to generate “.vhdl” files. These files were used to program FPGA to get processed speech signal as output.

4.2 Development of Algorithm using Comb Filters

The algorithm used in this research is based on the spectral splitting with comb filters based on auditory critical bandwidths. In initial stage, comb filters were designed with 512 coefficients., using frequency sampling technique of linear phase FIR filter [91].The filters are optimized for low perceived spectral distortion with pass band ripple less than 1dB, minimum stop band attenuation of 50 dB and crossover gain of -5dB to -7dB.The magnitude response of the pair of comb filters is shown in Figure 4.1. Each comb filter has nine pass bands up to 5 kHz, corresponding to auditory critical bandwidths [71]. The bandwidths of each of these nine bands for left and right ear are given in Table 4.1 along with the central frequencies.

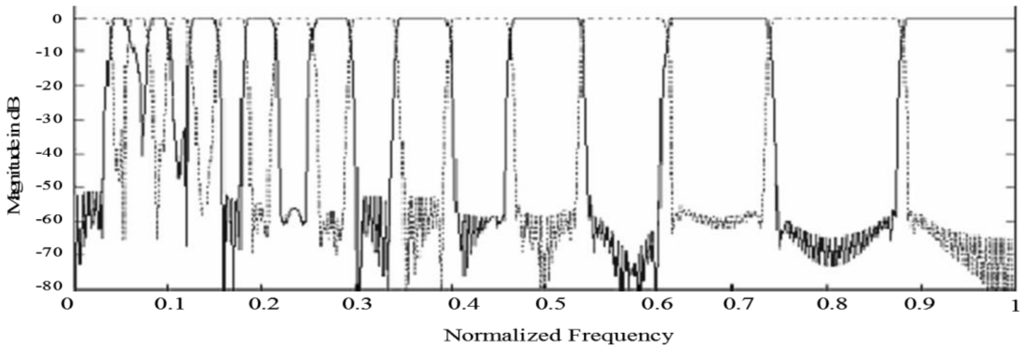


Fig. 4.1 Magnitude response of pair of Comb filters

Table 4.1 Eighteen critical bands of Comb Filter

Filter for left ear			Filter for right ear		
Band	Centre frequency kHz	Passband frequency kHz	Band	Centre frequency kHz	Passband frequency kHz
1	0.13	0.01-0.20	2	0.25	0.20-0.30
3	0.35	0.30-0.40	4	0.45	0.40-0.51
5	0.57	0.51-0.63	6	0.70	0.63-0.77
7	0.84	0.77-0.92	8	1.00	0.92-1.08
9	1.17	1.08-1.27	10	1.37	1.27-1.48
11	1.60	1.48-1.72	12	1.86	1.72-2.00
13	2.16	2.00-2.32	14	2.51	2.32-2.70
15	2.92	2.70-3.15	16	3.42	3.15-3.70
17	4.05	3.70-4.40	18	4.70	4.40-5.00

4.3 Implementation of Comb Filter on FPGA Platform

Filter designed in Section 4.2 can't be synthesized onto FPGA using MATLAB HDL coder [106]. Therefore, a novel approach has been proposed for designing the comb filter for synthesizable FPGA implementation. In this approach, the direct form–I filter with the same coefficients are used for odd and even filter banks. The filter structure used in this work is shown in Figure 4.2. The basic adder has only two inputs so an adder tree is used to add all multiplications results. This approach makes the MATLAB code synthesizable on FPGA.

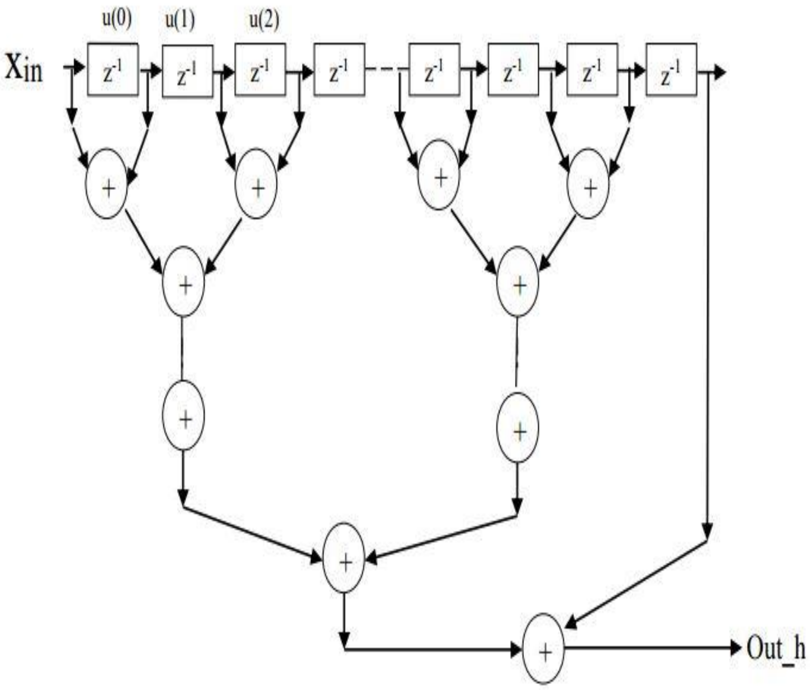


Fig. 4.2 Filter Structure

If X (z) is the filter input and $Y(z)$ is the filter output, then the transfer function H (z) is given in below,

$$H(z) = \frac{Y(z)}{X(z)} = \sum_{i=0}^{N} a_i z^{-i} \qquad (4.5)$$

where, a_i is a filter coefficient, N is order of filter. The length of filter is N+1.

The difference equation for odd filter bank is given by equation 4.6, as below

$$Out\_h = h(0)u(0) + h(1)u(n-1) + \ldots\ldots\ldots\ldots\ldots + h(512)u(n-512) \qquad (4.6)$$

where $\mathrm{h}(0), \mathrm{h}(1) \ldots\ldots\ldots.. \mathrm{h}(512)$ are odd filter bank coefficients.

Similarly, the difference equation for even filter bank is given by equation 4.7

$$Out\_g = g(0)u(0) + g(1)u(n-1) + \ldots\ldots\ldots\ldots\ldots + g(512)u(n-512) \qquad (4.7)$$

where, $\mathrm{g}(0), \mathrm{g}(1) \ldots\ldots\ldots.. \mathrm{g}(512)$ are even filter bank coefficients.

The stepwise workflow of the new approach of the comb filter for synthesizable FPGA implementation is presented in the following pseudo algorithm:

Pseudo-Algorithm:

1) Read audio input signal ***x(n)*** (with noise) of length *N*.
2) h1, g1 are the coefficients of comb filter pair, design using FIR filter (Frequency sampling technique)
3) For *i=1 to N* repeat following:

right_out(i)=conv(x(i), h1);

left_out(i)=conv(x(i), g1);

where *conv(x, f): step 1: multiplication*

for k=1 to M (length of comb filter f)

*m(k)=u(k)*f(k);*

step 2: addition(adder chain)

Repeat till last two terms m_1,m_2

for j=1 to M/2

add m(j),m(j+1);

out=m_1+m_2;

step 3: Shift input sample x(i) in persistent variable u

4.3.1 Real Time Implementation

The flow process of developed algorithm for real-time processing implemented by HDL coder toolbox available in MATLAB is shown in Figure 4.3. HDL coder toolbox was used to generate VHDL code from MATLAB using HDL Workflow advisor, while ModelSim was used to observe and verify the test bench. The generated VHDL code and test bench was used to program Spartan 6 (XC6SLX45CSG324C) FPGA. The user can use the generated ".vhdl" files through Xilinx environment to view the Register Transfer Level (RTL) and Technology schematic. The detailed implementation process has been discussed in Appendix B

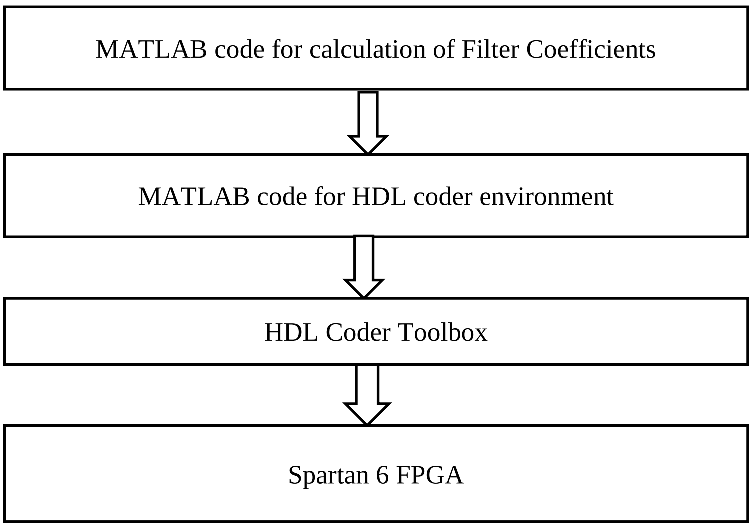


Fig. 4.3 Flow diagram for MATLAB-VHDL

In real-time processing, Atlys Circuit board of Digilent was used [107]. This board has numerous features of which Xilinx Spartan 6 LX45 FPGA, 10/100/1000 Ethernet PHY along with AC-97 Codec with line-in, line-out, headphone and mic, were used. It also includes Digilent's new Adept USB2 system which offers device programming, virtual I/O's, simplified user data transfer facilities etc. In our implementation AC 97 protocol based audio codec "LM4550" available on the circuit board uses 18-bit Sigma-Delta ADCs and DACs. The block schematic for real time implementation is shown in Figure 4 4.

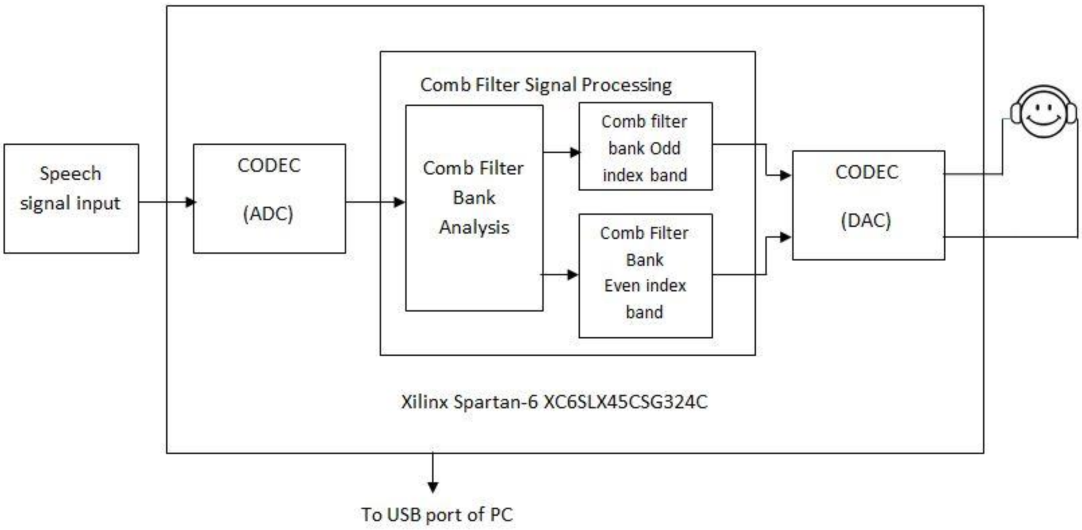


Fig. 4.4 Block schematic for Real time implementation

4.3.2 Real-Time Experimental Implementation

The interconnection of FPGA "Xilinx Spartan-6 XC6SLX45CSG324C" with audio codec "LM4550" on the board used for implementing the comb filter. The experimental setup accepts the unprocessed speech signal as an input processes the speech signal through comb filter implemented in FPGA via AC link and produces the processed speech output as shown in Figure 4.5

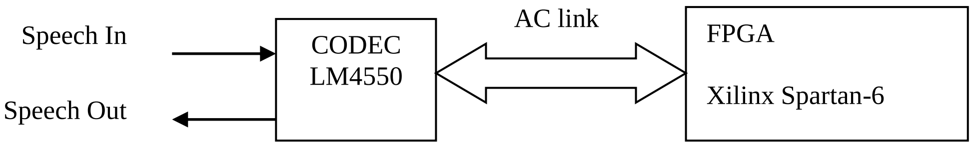


Fig. 4.5 Flow of unprocessed and processed signal using FPGA

The unprocessed speech signal was used as input and given through audio-in of Atlys Circuit board. The VHDL compatible MATLAB code was used to program the Xilinx 6 FPGA. The input unprocessed speech signal being continuous in nature had to be converted into digital form while the processed digital signal so obtained had to be converted back to continuous signal. So, we used an on-board AC'97 Codec (LM4550) for the same. The codec has input and output range of 1 V_{RMS}. It is programmed using AC Link Serial Interface Protocol for 18-bit analogue-to-digital and digital-to-analogue conversion. AC-link protocol consists of serial interface which is bidirectional and is 5-wire bus. The command and status information as well as all digital audio information are communicated over this serial bidirectional bus. It consists of serial data in (SDATA_IN), frame synchronization (SYNC), serial data out (SDATA_OUT), reset (RESET), and clock (BIT_CLK) as shown in Figure 4.6.The input frames are carried to input of AC’97 Digital Controller (Spartan 6 FPGA) from output of LM4550 codec on serial data in signal. Output of AC ’97 Digital Controller is connected to input of LM4550 Codec on which output frames are transferred using serial data out signal. The input frames make use of one tag slot followed by 12 data slots. Each data slot contains 12 bits while tag slot consists of 16 bits. Hence the overall frame consists of 256 bits. Slot 1 consists of 7-bit address of control register while Slot 2 consists of 16-bit control data. Slot 3 consists of PCM data for left channel and Slot 4 consists of PCM data for right channel. Remaining slots are unused and stuffed with zeros. A low to high transition of SYNC indicates arrival of new input/ output frame. It remains high for a total duration of 16 BIT_CLKs at the beginning of each audio frame. Serial data is transitioned on rising edge of BIT_CLK. The bidirectional link of AC’97 receiver samples each serial bit on the falling edge of BIT_CLK. The BIT_CLK is fixed at 12.288 MHz to provide same SYNC transition fixed at 48MHz. This provides the alignment of input and output frames. Codec operation such as output volume control, input volume control, mono volume control, etc can be controlled by writing value in Slot 2. For our Spartan 6 XCLS45, the Atlys circuit board consists of 18bit codec. The unprocessed continuous speech signal given through mic in of the circuit board is converted into 18 bit digital code. This code is processed through our algorithm. The output is converted back to continuous processed speech signal using 18bit DAC.

In our research, for Spartan LS45 the incoming signal from SDATA_IN gets split in 9 bits. Each 9-bit data is stored in two registers for each channel. The compatible 9 bit vhdl code (comb filter code) is implemented on the FPGA so as to process each register. Processed data is then sent on SDATA_OUT line. This data is then converted to continuous analog signal using 18bit DAC. The codec used here processes data in serial fashion while the FPGA processes data in parallel manner. Hence we have also written the codes for serial in parallel out (SIPO) and parallel in serial out (PISO).

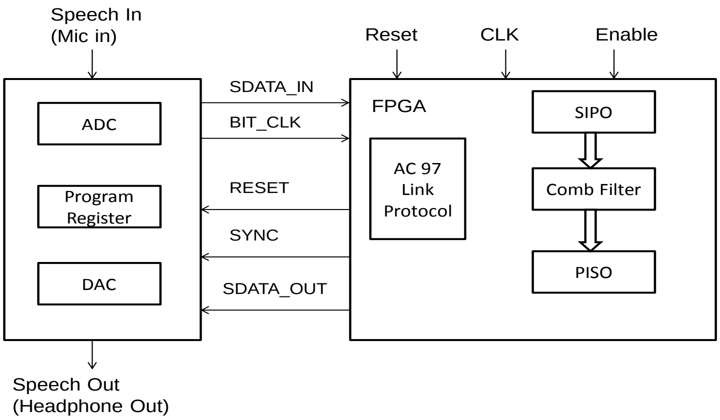


Fig. 4.6 FPGA implementation schematic

The real time experimental setup used in research is shown in Figure 4.7.

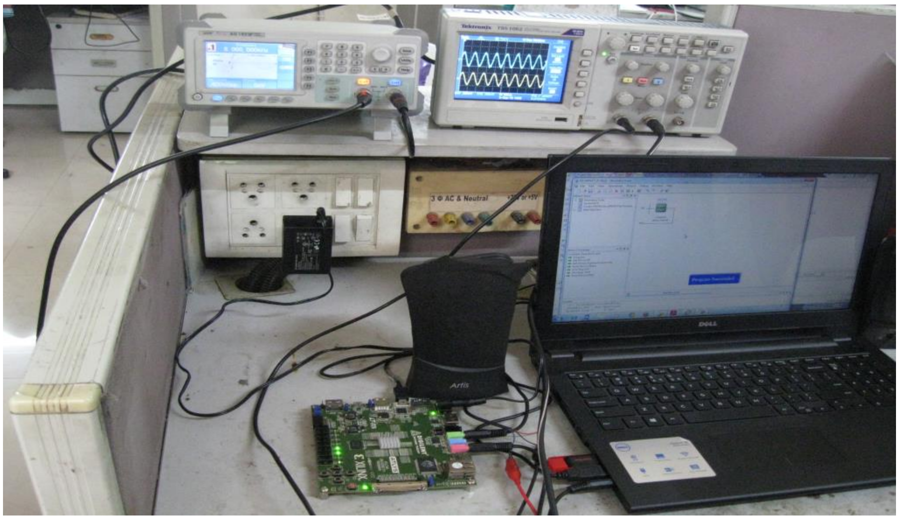


Fig. 4.7 Real time experimental setup

4.4 Experimental Results

The experimental results obtained after deploying the developed algorithm on FPGA platform are discussed in following subsection.

4.4.1 Simulation Results

The input speech signal is fed to the system in VCV context. The filter bank splits the input signal into even and odd bands. The magnitude response for critical band filter using MATLAB code for the test input signal “ada.wav” is shown in Figure 4.8.

Once the MATLAB simulated results are verified, the HDL coder is used to convert MATLAB code into VHDL code. The generated VHDL code is synthesized and the RTL schematic is generated using XILINX 14.2 for verification. The generated RTL schematic for comb filter is shown in Figure 4.9. The generated RTL schematic shows that the implemented filter structure in FPGA has an input of 18 bits called xin. The output signal obtained for even and odd bands are g_out and h_out, respectively, both of which are 18 bit. The comb filter is been implemented on Xilinx Spartan-6 FPGA. Figure 4.10 shows the ModelSim simulated test bench waveform of the same.

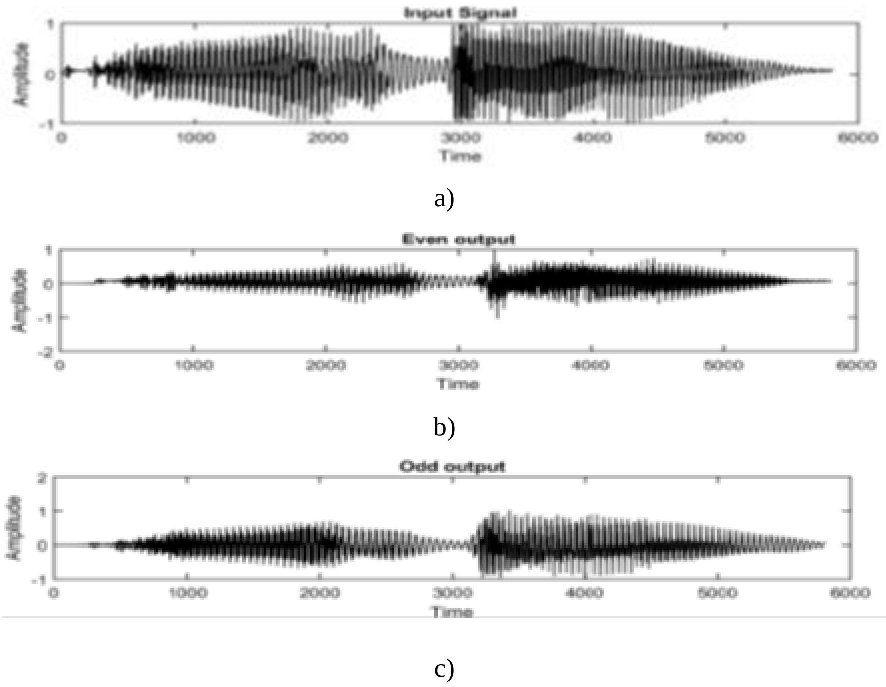


Fig. 4.8 Magnitude response of input speech signal ‘ada.wav’ a) unprocessed signal b) reconstructed signal from even bands for right ear c) reconstructed signal from odd bands for left ear

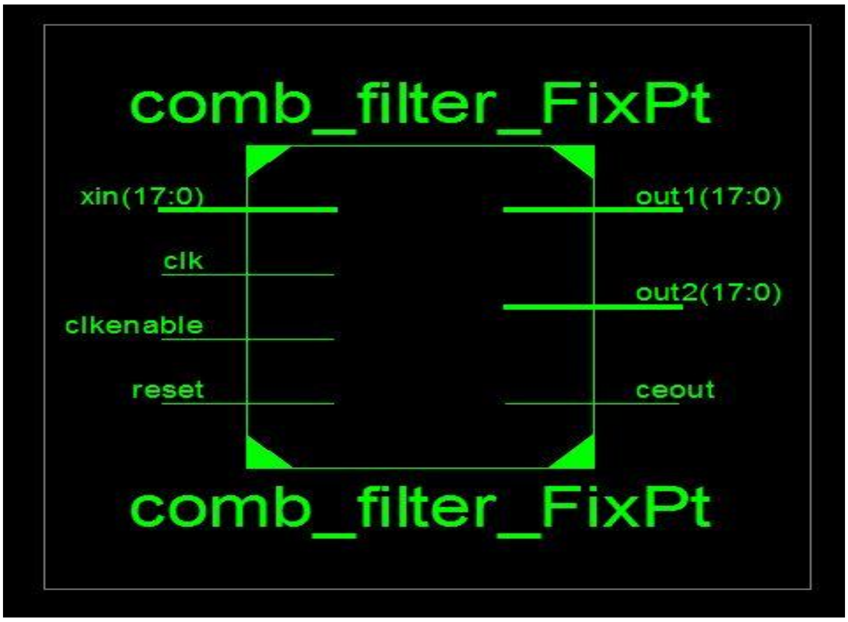


Fig. 4.9 RTL schematic for critical band filter

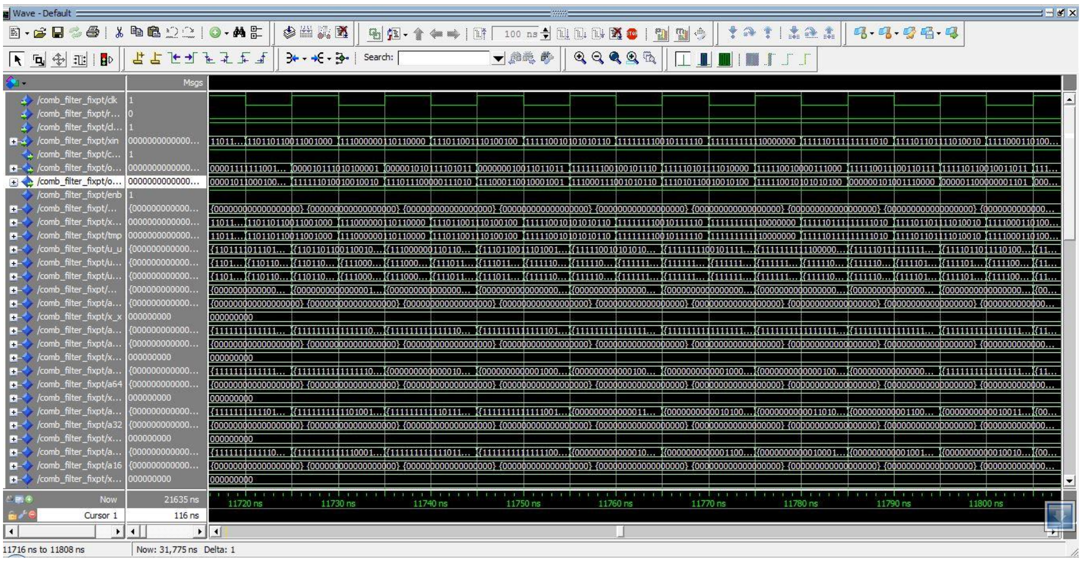


Fig. 4.10 Simulated test bench waveforms of Comb filter

4.4.2 Utilization Report of used hardware

In this work, for realization of comb filter into FPGA the input and output data lines split into 9 bits each. Due to this reason, the FPGA implementation of 512-coefficient comb filter utilize fraction of the resources available on FPGA spartan6 XC6SLX45 CSG324 i.e. (10% of total slice registers, 31% of slice LUTs, 25% of fully used LUT-FF pairs and 6% of bonded IoBs') with scope for implementation of other processing blocks of the hearing aid.

4.4.3 Latency

The latency of our algorithm was measured using speech signals (VCV context). Here, latency is defined as the time delay between the unprocessed signals deployed on FPGA and processed signal fed to ear. This delay was measured by giving the unprocessed signal and processed signal to Digital Storage Oscilloscope Textronix TBS1062 so that it could simultaneously record both the signals with high sampling frequency; the recorded signals were then analysed in MATLAB. We tested various VCV context speech signals having different frequency compositions. The latency obtained varied from 1.6 ms (minimum) to 14.08 ms (maximum). The average value of latency of our algorithm was found to be 8ms, noticeable to subjects [94].

4.4.4 Peak Signal to Noise Ratio

PSNR is defined as ratio of power of speech signal to power of noise signal. High value of PSNR indicates less amount of noise added to the signal while low value of PSNR indicates high amount of noise added to signal. Noise performance of our algorithm is assessed by considering the PSNR value. It was found to be 96.31 dB.

4.5 Evaluation with Real Time implementation

The real time implementation of the proposed algorithm used spectral splitting with comb filters based on auditory critical bandwidths. Comb filter was initially designed with 512 coefficients using MATLAB code. The designed MATLAB code was then converted to compatible VHDL code using HDL coder toolbox. The generated VHDL bit file was then programmed into Spartan6 FPGA. It used both diotic and dichotic presentation of unprocessed and processed signal respectively. This is referred as phase one, Experiment I.

4.5.1 Subjects and Test material

The assessment was carried out by conducting listening tests on seven subjects with bilateral 'mild' to severe sensorineural hearing loss. (APL: F 16, KUR: M 57, KSH: F 12, BHL: F 10, PDH M 74, PNK: M 45, BHV: M 12) Subjects KSH, BHL and BHV have severe and symmetrical hearing impairment. Subject KUR has mild to moderate and asymmetrical high frequency impairment (less loss in one ear and more loss in other ear). Subject PNK has mild and symmetrical sloping high frequency impairment. Subject APL has moderately severe and symmetrical hearing impairment while subject PDH has moderately severe and asymmetrical low frequency hearing impairment. Hearing thresholds of impaired subjects are mentioned in Appendix A.

The speech material used for the listening test consisted of a set of fifteen nonsense syllables in VCV context with consonants / p, b, t, d, k, g, m, n, s, z, f, v, r, l, y / and vowel /a/ as in 'farmer'. Responses were tabulated in the form of confusion matrix and response time was also recorded. Confusion matrices were used for calculating recognition scores and relative transmitted information. Further, the consonants were clustered according to the articulatory features [57] and the contributions of different features were analyzed. The features selected for this study were voicing (voiced: / b d g m n z v r l y / and unvoiced: / p t k s f /), place (front: / p b m f v /, middle: / t d n s z r l /, and back: / k g y /), manner (oral stop: / p b t d k g l y /, fricative: / s z f v r /, and nasals: / m n /), nasality (oral: / p b t d k g s z f v r l y /, nasal: /m n /), frication (stop: / p b t d k g m n l y /, fricative: / s z f v r /), and duration (short: / p b t d k g m n f v l / and long: /s z r y /).

4.5.2 Experimental procedure for Listening test

The experimental procedure to conduct listening tests is detailed in Appendix C. The listening tests were carried out in an isolated chamber and presentations were completed for the impaired subjects at the comfortable listening level. The tests were carried out on these subjects without addition of any noise to the speech stimuli. The subjects were informed about procedure of listening test. To familiarize with the speech material, the subjects could listen to the test material frequently at their wish. In a test fifteen stimulus items were displayed for five times in random manner, thus giving a total number of seventy-five presentations. For the binaural presentation of test stimuli, laptop based experimental setup was used. The stimuli were displayed every time on the laptop screen when the stimulus item was made available on headphone. A push button related to each stimulus was displayed on laptop screen. The response was obtained after pressing the correct push button by the subject. For every presentation the time utilized by the subject to respond was also noted. A stimulus response confusion matrix consisting of stimuli along rows and responses along the columns was obtained at the completion of each test. Every entry in the cell represents the occurrence of stimulus response pair. Correct responses are provided by diagonal elements whereas off diagonal elements provided incorrect responses. The addition of diagonal elements provided total number of correct responses. In addition to the confusion matrix, the percentage correct recognition score and response time statistics were also obtained.

Based on the accessibility and willingness of the subjects, a period of four to five months was necessary for the conduction of test session. Each subject took about one to one and half-hour's time for the compilation of the test under different test conditions. Subject's qualitative evaluation of the test stimuli is combined for various listening conditions to understand the speech qualitative analysis. For binaural dichotic presentation of processed speech and also binaural diotic presentation of unprocessed speech the listening tests were conducted.

The processed and unprocessed signal was graded by the people on a scale ranging from 1 to 5. The people gave their choice after receiving pair of signals in sequential manner. The comparison of the load on perception was made by response times. The information transmission analysis was carried out from the obtained confusion matrix (detailed in Appendix A).

4.6 Listening test Results

The following subsection includes results of listening tests for qualitative assessment, response time, recognition score and information transmission analysis. To find out the contribution of the features in improvement of speech perception, these tests were conducted for consonantal features.

4.6.1 Qualitative assessment

To compute qualitative assessments, five times pre-recorded test material in VCV context was heard by subjects. On the basis of quality of sound, they were asked to give the rating as ‘Outstanding’, ‘Good’, ‘Fair’, ‘Average’ and ‘Below Average'. These ratings were indexed from 5 to 1; 5 being ‘Outstanding’ and ‘1’ being ‘Below Average’. Averages of ratings were computed to find mean opinion score of each subject. Results of qualitative assessment of seven subjects are given in Table 4.2. From the Table 4.2 it can be seen that, six subjects APL, KSH, BHL, PDH and PNK ranked the quality of processed signal as higher than the unprocessed signal. Subject KUR has ranked the unprocessed signal same as processed signal. Graphical representation of qualitative assessment (mean opinion score) of each subject is shown in Figure 4.11.

Table 4.2.Experiment I - Qualitative assessment

Subjects	Unprocessed Signal (US)	Processed Signal (PS-Comb Filter)
APL	3.75	4.80
KUR	5.00	5.00
KSH	2.75	3.00
BHL	3.00	3.75
PDH	3.25	4.00
PNK	4.50	5.00
BHV	2.50	2.25

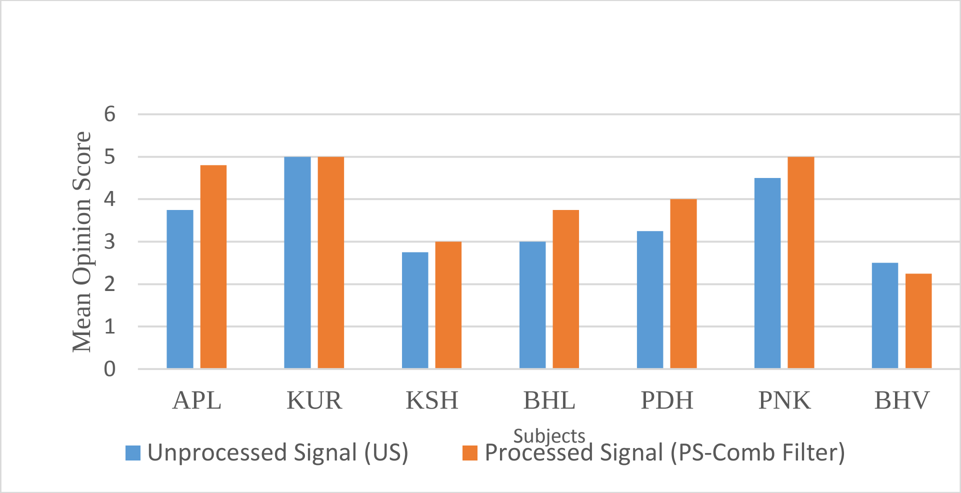


Fig. 4.11 Experiment I - Qualitative assessment (MOS)

4.6.2 Response time

Response time is the time interval between speech materials presented dichotically to subjects and the response given by them. It reduces significantly showing reduction in load on perception process. Response time for seven subjects is given in Table 4.3 and graphically represented in Figure 4.12 It shows variation from 4.08 seconds to 8.6 seconds for unprocessed signal and 4.68 seconds to 7.95 seconds for processed signal. There was reduction in response time for processed signals. This relative decrease is from -7.35 to 33 %. Relative decrease in response time is significant for the subjects KUR and KSH as shown in Table 4.4 and graphically presented in Figure 4.13

Table 4.3 Experiment I - Response time

Subjects	Unprocessed Signal (seconds)	Processed Signal (Comb Filter) (seconds)
APL	5.82	5.35
KUR	6.99	4.68
KSH	7.25	5.66
BHL	4.83	4.83
PDH	5.51	5.19
PNK	4.08	4.38
BHV	8.6	7.95

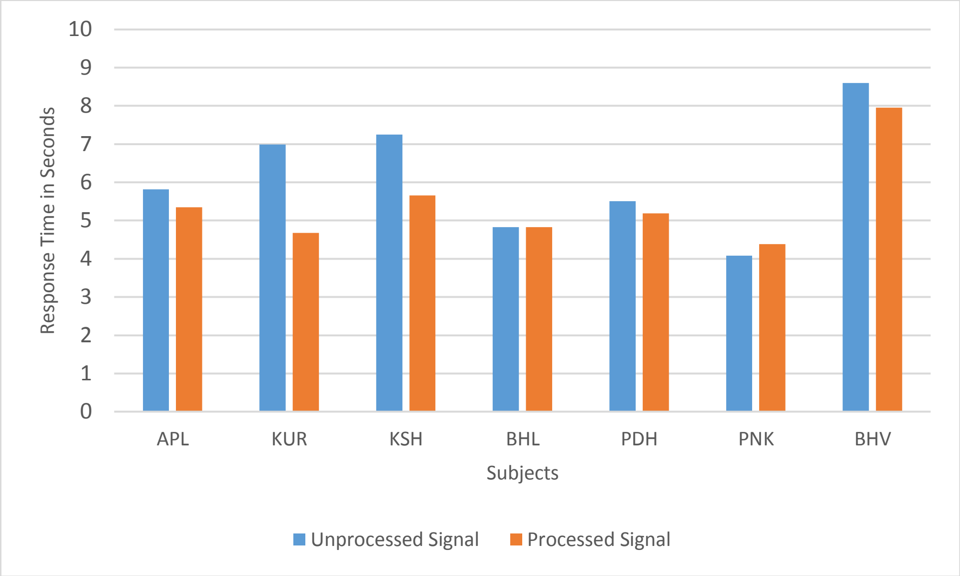


Fig. 4.12 Experiment I - Response time

Table 4.4 Experiment I - Relative decrease in Response time

Subjects	Relative Decrease (%)
APL	8.07
KUR	33.04
KSH	21.93
BHL	0.00
PDH	5.80
PNK	-7.35
BHV	7.55

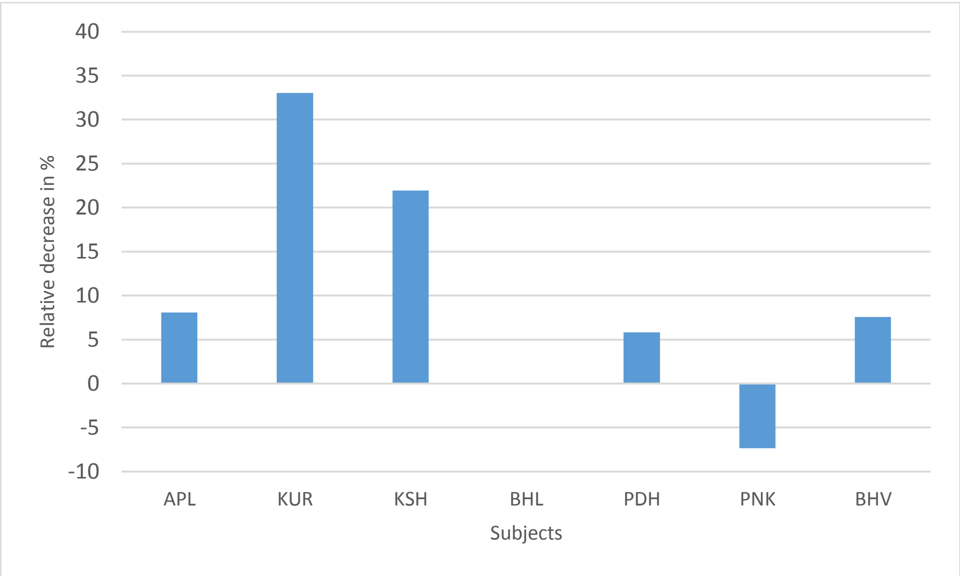


Fig. 4.13 Experiment I - Relative decrease in Response time

4.6.3 Recognition scores

The recognition scores and percentage relative improvement were obtained from the confusion matrix as shown in Table 4.5 and Table 4.6 respectively. Plot of recognition scores and percentage relative improvement is shown in Figure 4.14 and Figure 4.15 respectively. For the impaired subjects, recognition score is observed in the range of 48.33% to 90% for unprocessed signal while for processed signal the recognition score increases from 60% to 96.66%. Three subjects (BHL, APL, and KUR) having mild to severe frequency impairment have shown maximum relative improvement and one subject (PNK) with symmetrical high frequency hearing impairment have shown maximum decrease in recognition score.

Table 4.5 Experiment I - Recognition scores

Subjects	Unprocessed Signal	Processed Signal (Comb Filter)
APL	53.33	70.00
KUR	84.44	96.66
KSH	48.33	60.00
BHL	68.88	86.66
PDH	61.10	70.00
PNK	90.00	93.33
BHV	53.33	60.00

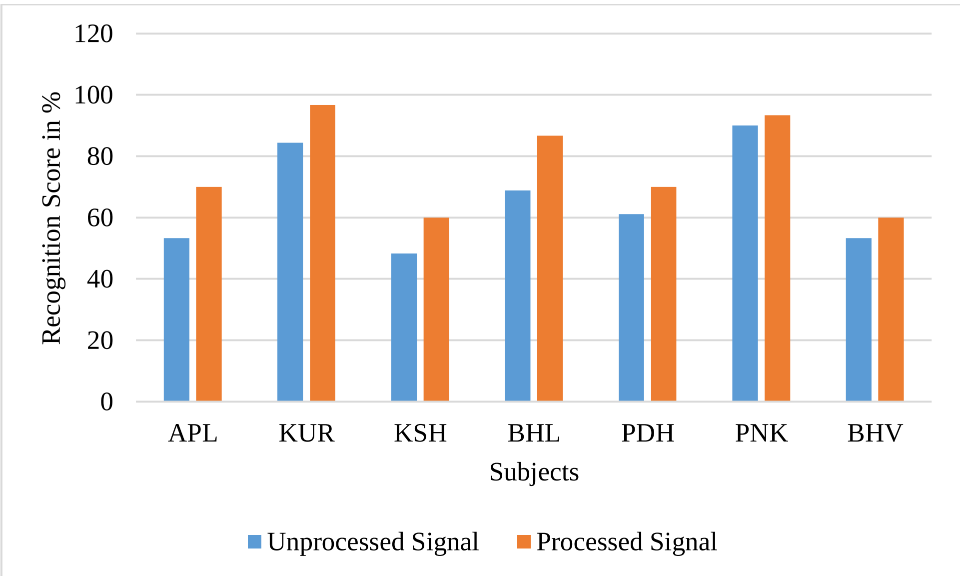


Fig. 4.14 Experiment I - Recognition scores in percentage

Table 4.6 Experiment I - Percentage Relative Improvement

Subjects	Relative Improvement in %
APL	16.67
KUR	12.22
KSH	11.67
BHL	17.78
PDH	8.90
PNK	3.33
BHV	6.67

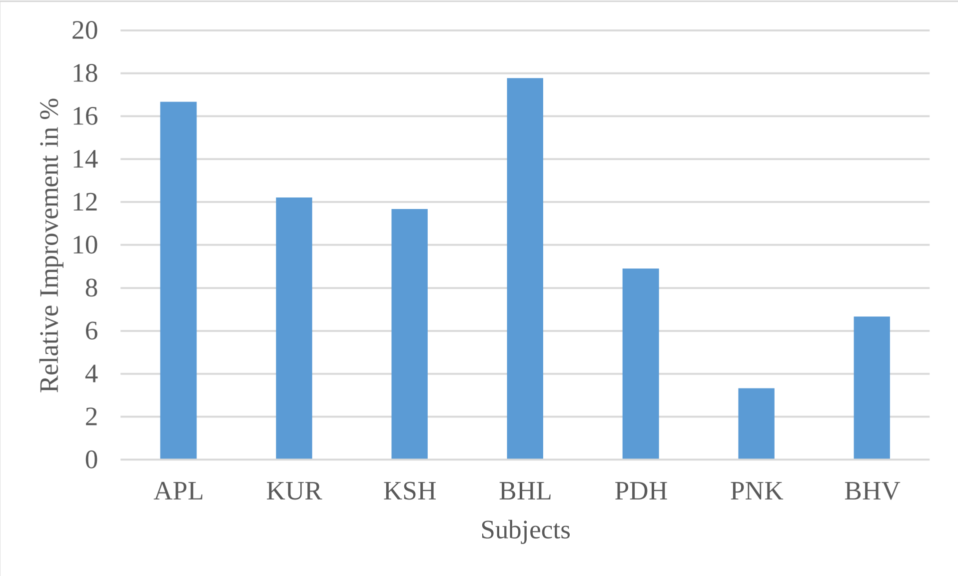


Fig. 4.15 Experiment I - Relative improvement in percentage

4.6.4 Information transmission analysis

Combined confusion matrices of each subject were used to evaluate information transmission analysis. Relative information transmitted for consonantal features is given in Table 4.7 and plotted in Figure 4.16 to Figure 4.23.

Table 4.7 (a to h) Experiment I - Information Transmission Analysis for VCV context. US: Unprocessed speech, PS-Comb: Processed Signal using Comb filters and RI: Relative improvement in percentage with respect to unprocessed signal for different features

a) Feature: Overall					b) Feature: Continuance			
Subject	US	PS-Comb	RI		Subject	US	PS-Comb	RI
APL	71.15	79.95	8.80		APL	33.15	44.43	11.28
KUR	89.77	98.16	8.39		KUR	41.12	86.49	45.37
KSH	62.00	80.28	18.28		KSH	26.02	44.43	18.41
BHL	84.30	92.95	8.65		BHL	39.84	76.35	36.52
PDH	74.69	84.34	9.65		PDH	42.53	35.61	-6.92

PNK	94.24	96.59	2.35		PNK	44.43	60.42	15.99
BHV	74.19	71.62	-2.57		BHV	11.42	37.59	26.16
Averages	78.62	86.27	7.65		Averages	34.07	55.04	20.97
c) Feature: Duration					d) Feature: Frication			
Subject	US	PS-Comb	RI		Subject	US	PS-Comb	RI
APL	41.12	62.95	21.83		APL	26.46	38.31	11.85
KUR	58.00	100.00	42.00		KUR	53.77	86.94	33.16
KSH	41.29	24.72	-16.57		KSH	21.17	53.35	32.18
BHL	41.16	100.00	58.84		BHL	38.72	78.94	40.22
PDH	76.74	100.00	23.26		PDH	57.73	64.90	7.16
PNK	76.35	100.00	23.65		PNK	48.96	64.90	15.93
BHV	44.43	58.26	13.83		BHV	10.52	52.43	41.91
Averages	54.16	77.99	23.83		Averages	36.76	62.82	26.06
e) Feature: Manner					f) Feature: Nasality			
Subject	US	PS-Comb	RI		Subject	US	PS-Comb	RI
APL	40.45	38.48	-1.97		APL	57.10	35.78	-21.32
KUR	70.73	91.43	20.70		KUR	100.00	100.00	0.00
KSH	40.37	57.88	17.51		KSH	71.40	63.68	-7.72
BHL	50.54	74.25	23.71		BHL	67.57	67.57	0.00
PDH	64.02	66.16	2.14		PDH	74.54	67.57	-6.96
PNK	68.05	78.42	10.37		PNK	100.00	100.00	0.00
BHV	15.55	57.50	41.95		BHV	20.52	67.31	46.78
Averages	49.96	66.30	16.34		Averages	70.16	71.70	1.54
g) Feature: Place					h) Feature: Voicing			
Subject	US	PS-Comb	RI		Subject	US	PS-Comb	RI
APL	27.64	38.70	11.07		APL	100.00	78.94	-21.06
KUR	61.11	100.00	38.89		KUR	71.86	100.00	28.14
KSH	18.52	24.48	5.96		KSH	63.98	82.45	18.47
BHL	52.58	88.27	35.69		BHL	86.94	100.00	13.06
PDH	79.67	100.00	20.33		PDH	37.01	71.68	34.67
PNK	89.30	100.00	10.70		PNK	100.00	100.00	0.00
BHV	25.06	30.95	5.89		BHV	21.80	26.75	4.95
Averages	50.55	68.91	18.36		Averages	68.80	79.97	11.18

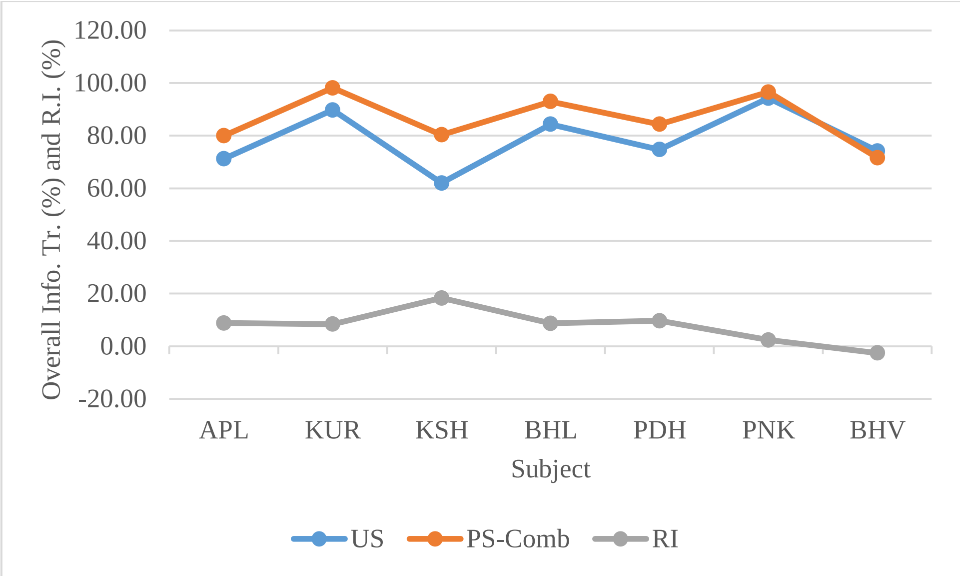


Fig. 4.16 Experiment I - Relative Information Transmitted for Overall

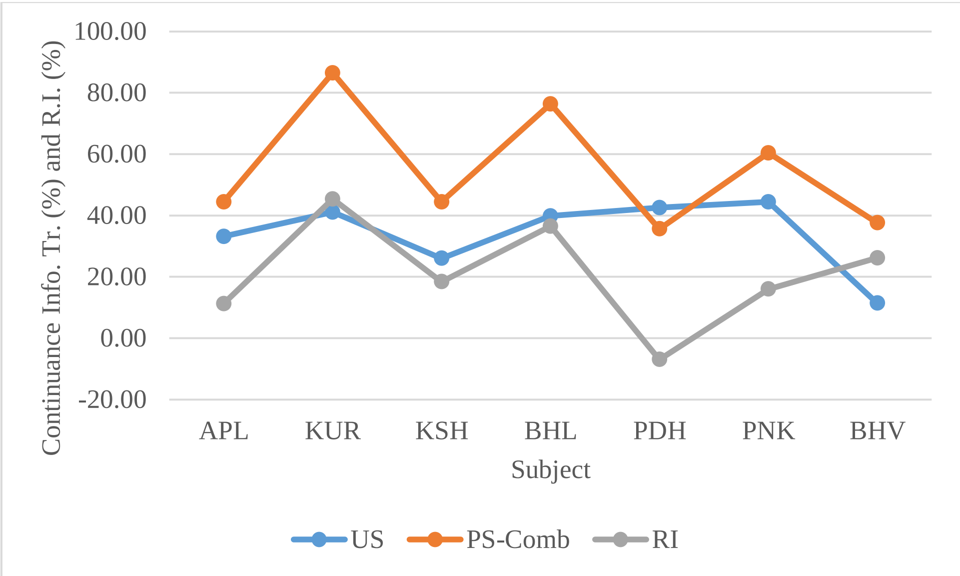


Fig. 4.17 Experiment I - Relative Information Transmitted for Continuance

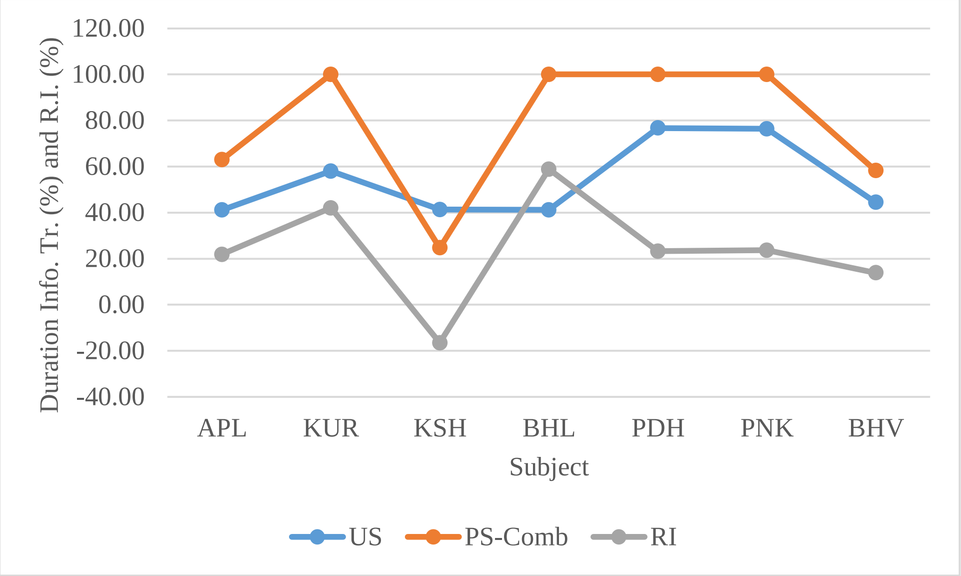


Fig. 4.18 Experiment I - Relative Information Transmitted for Duration

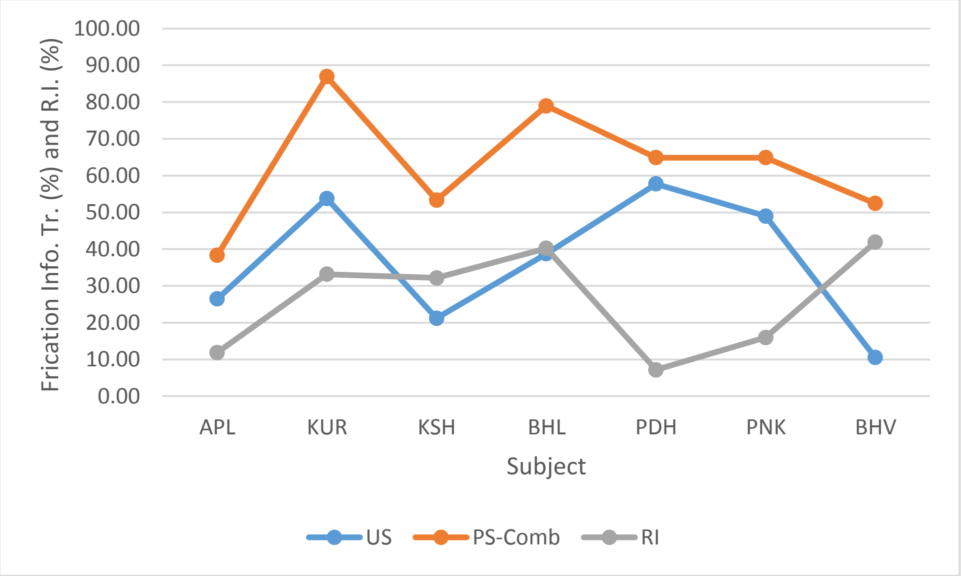


Fig. 4.19 Experiment I - Relative Information Transmitted for Frication

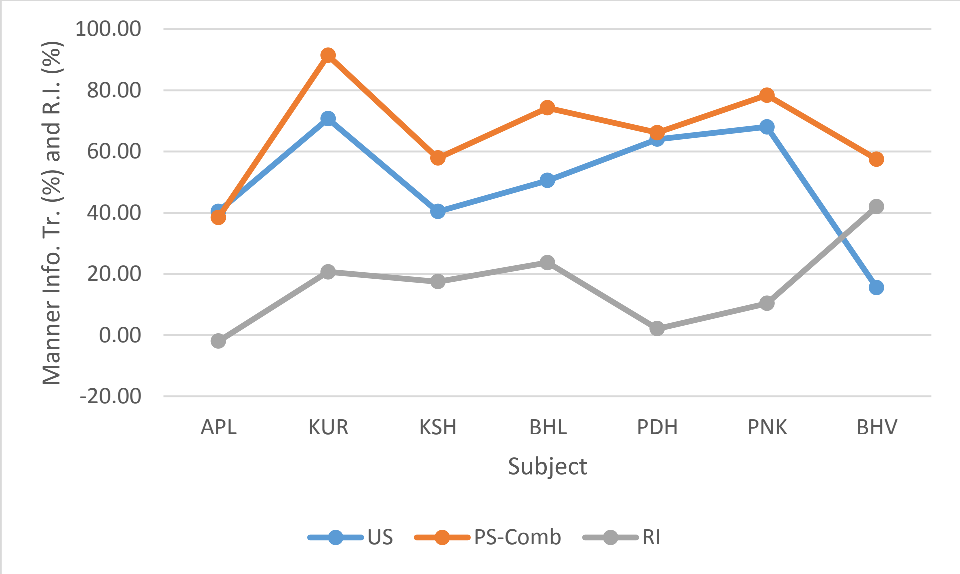


Fig. 4.20 Experiment I - Relative Information Transmitted for Manner

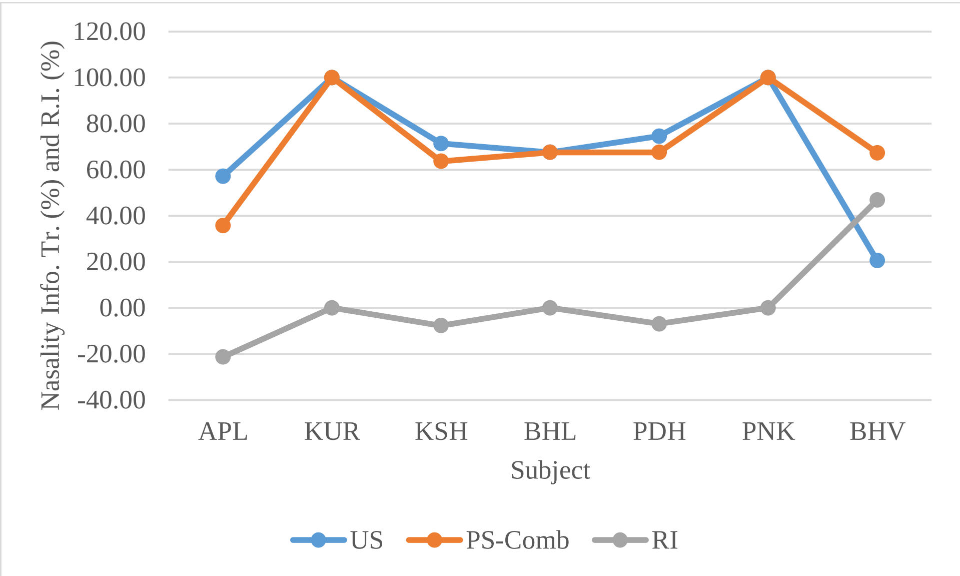


Fig. 4.21 Experiment I - Relative Information Transmitted for Nasality

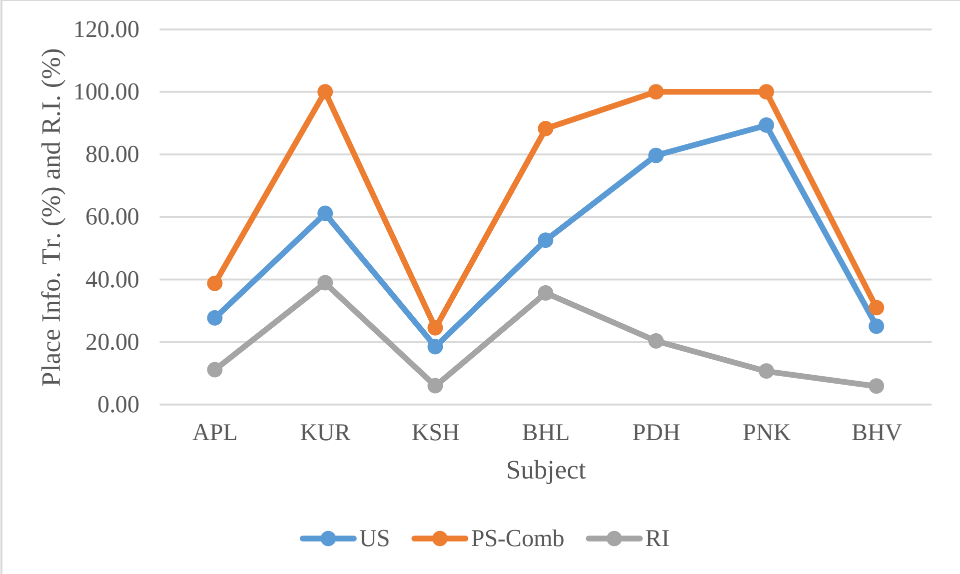


Fig. 4.22 Experiment I - Relative Information Transmitted for Place

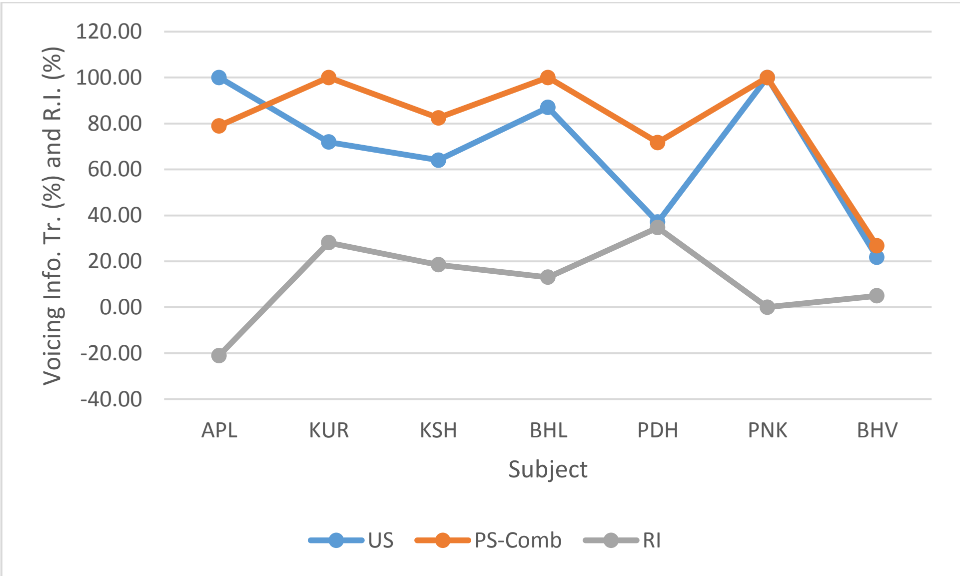


Fig. 4.23 Experiment I - Relative Information Transmitted for Voicing

Overall: From Figure 4.16, the overall relative information transmission varied from 62.00% to 94.24% for unprocessed speech signal while for processed speech signal it ranged from 71.62% to 98.16%. The relative improvement was varied from -2.57% to 18.28% for all the subjects.

Continuance: The relative information transmission for continuance varied from 11.42% to 44.43% for unprocessed speech signal while it ranged from 37.59%

to 86.49% for processed speech signal as shown in Figure 4.17. For the processed signal, the relative improvements were high for KUR and low for PDH.

Duration: Figure 4.18 shows the relative information transmitted for duration feature. It varied from 41.12% to 76.35% for unprocessed speech signal while for processed speech signal it ranged from 24.72% to 100%. The relative improvements for the subjects BHL and KUR are high for the processed signal and were 58.84% and 42% respectively.

Frication: The relative information transmitted for frication feature is shown in Figure 4.19. For unprocessed speech signal varied from 10.52% to 57.73% while for processed speech signal it ranged from 38.31% to 86.94%. Subject PDH showed relative improvement of 7.16% (minimum) while subject BHL showed relative improvement of 40.22% (maximum)

Manner: Figure 4.20 gives graphical representation of the relative information transmitted for manner feature. It varied from 40.37% to 70.73% for unprocessed speech signal while for processed speech signal it ranged from 57.50% to 91.43%. The relative improvement was varied from -1.97% to 41.95%.

Nasality: Figure 4.21 shows the relative information transmitted for nasality feature. For unprocessed signal, information transmitted varied from 20.52% to 100%. Its relative improvement varied from -21.32 to 46.78%, for processed signal.

Place: From Figure 4.22, the relative information transmission of place feature varied from 18.52% to 89.30% for unprocessed speech signal while for processed speech signal it ranged from 24.48% to 100%. For subjects KUR and BHL the relative improvements were high and for all subject it varied from 5.89% to 38.89%.

Voicing: The relative information transmission for voicing feature is shown in Figure 4.23. It varied from 21.80% to 100% for unprocessed speech signal while for processed speech signal it ranged from 26.75% to 100%. The highest relative improvement was observed for subject PDH and lowest was observed for subject APL.

Discussion

The listening tests were conducted on seven subjects with mild to severe bilateral sensorineural hearing impairment, to assess the spectral splitting algorithm using comb filters in real time mode.

Fifteen English consonants in VCV context were used for the listening tests. The parameters like qualitative assessment (speech quality), subject's response times, recognition scores and information transmitted for consonantal features were evaluated. Ranking of speech quality of the test material was evaluated by mean opinion score when presented to subjects. The processing algorithm resulted in reduction of response times indicating reduced load on perception process as compared to the response time with unprocessed signal. The usage of spectral splitting scheme was prominent as relative reduction in response time was observed. This improvement was highest for the subject with asymmetrical high frequency hearing impairment. The improvement in consonantal identification by the binaural dichotic presentation is shown by the recognition scores along with highest improvements. Three subjects having mild to moderately severe hearing impairment showed relative improvement in recognition scores.

For unprocessed signal the information transmission analysis indicates deterioration in relative information transmitted for manner and nasality features. The processed signals resulted in an improvement of transmission of all the consonantal features, particularly place, frication and duration. The relative improvements of subjects KUR and BHL were high for place, frication and duration features. Subject BHL has severe symmetrical hearing impairment and subject KUR has mild to moderate and asymmetrical high frequency hearing impairment. Dichotic presentation improves the reception of tough consonantal features like voicing, manner, and place for all subjects. Among these, place feature showed maximum improvement. The relative improvement for place feature varied from 5.89% to 38.89% with an average value of 18.36% across the seven subjects. Since the place information is linked to frequency resolving capacity of auditory process, the effect of spectral masking has been reduced

It is found from the analysis of recognition scores and information transmission that, the scheme which gives benefit by reduction of increased masking effects is based on the individual hearing loss conFig.uration. The spectral splitting scheme benefits subject with mild to moderate and asymmetrical high frequency hearing impairment.

The potential to improve the speech perception for people using binaural hearing aids lies in the processing algorithm for dichotic presentation. The evaluation of algorithm with hearing impaired subjects showed that they are able to compile dichotic speech signals and improved the speech perception. Dichotic presentation reduces the load on the perception process and is shown by the improvement in response time. The improvement in consonantal reception and reduction in response time follow distinct trends, in case of hearing impaired subjects. The grading of processed signal is higher than the unprocessed signal for the test involving qualitative ranking. To estimate the merits of spectral splitting extended tests with hearing impaired subjects are necessary.

Chapter 5

EVALUATION AND IMPLEMENTATION OF WAVELET BASED FILTERS

In the last chapter, an overall evaluation of the schemes of spectral splitting using comb filters was carried out by conducting listening tests on subjects with mild to severe sensorineural hearing impairment. It resulted in maximum improvements in the sensorineural hearing impairment having mild to moderate hearing loss. In this chapter, an overview of wavelet packets, multi resolution analysis and selection of wavelet functions are explained. The listening tests were conducted on five normal people with simulated hearing loss and eight sensorineural hearing impaired subjects. The processing scheme for wavelet packets with different wavelet basis functions were used for the evaluation of listening tests.

As described ahead, assessment has been conducted in two phases. In the first phase the listening test were carried out on normal people with simulated hearing loss. This is referred as Experiment II. The second phase of assessment was divided into two parts. In the first part, the listening test, were conducted as software based off line experimentation, called as experiment III and in second part the similar tests were carried out using real time hardware implementation on seven hearing impaired subjects, called as experiment IV.

5.1 Overview of Wavelet

Recently, wavelet transforms have been used in various fields of speech processing. The cochlea can be considered as a fine wavelet analysis organ. The properties of cochlea resembles with that of time-scale analysis of wavelets therefore they are used in the speech signal processing [30] [10]. The wavelet transform employ a variable width window (wide at low frequencies and narrow at high frequencies) that allows focusing on very short duration high frequency phenomena like spikes, transients in signals [95]. In this section, multi resolution analysis, Wavelet packets and wavelet properties that influence the type of wavelet basis functions that is appropriate for a particular application is examined.

Researchers in signal processing domain have introduced wavelet analysis in the decade of 1970’s. However similar ideas can be traced back to the work of Haar (1910) and Gabor (1946). The field finds its usage in many areas and part of its theory have been developed independently [97]. The main contribution of wavelets is its effectiveness to a wide variety of problems with common characteristics. The representation of wavelet basis is more efficient than the Fourier basis. The wavelet essentially acts as a band pass function providing a flexible way of analyzing a signal across various frequency regions and at various resolutions [96]. This flexibility is especially important in speech processing as the wavelet transform can provide an analysis of the audio signal in accordance to the critical band resolution of the inner ear and usually provides a method that can adapt to the time-varying nature of the signal [101].

Wavelet transforms (WT) are broadly classified into three categories namely: continuous wavelet transform (CWT), discrete wavelet transform (DWT) and multi-resolution-based wavelets.

5.1.1 Continuous Wavelet Transform

The continuous wavelet transform (CWT) is termed as the sum over all time of the signal multiplied by scaled, shifted versions of the wavelet function ψ:

$$C(\text{scale,position}) = \int_{-\infty}^{\infty} f(t)\Psi(\text{scale}, \text{position}) \quad (5.1)$$

The results of the CWT are many wavelet coefficients C, which are a function of scale and position. Multiplying each coefficient by the appropriately scaled and shifted wavelet yields the constituent wavelets of the original signal

The CWT of real signal s(t) with respect to the wavelet function Ψ(t) is defined as,

$$S(b,a)=\frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} \Psi' \frac{(t-b)}{a} s(t)dt \quad (5.2)$$

Where Ψ' denotes the complex conjugate of Ψ, parameter 'a' corresponds to scale of analyzing wavelet and the parameter 'b' corresponds to time shifts.

5.1.2 Discrete Wavelet Transform

In the DWT the scale and shift parameters are discretized as $a=a_0^m$ & $b=nb_0$ and the wavelet function becomes

$$\Psi_{m,n}(t)=a_0^{-\frac{m}{2}} \Psi\frac{(t-nb_0)}{a_0^m} \quad (5.3)$$

where m and n are integer values.

The discrete wavelet transform and its inverse transform are defined as follows:

$$S_{m,n}=\int_{-\infty}^{\infty} \Psi'_{m,n}(t)s(t)dt \quad (5.4)$$

$$S(t)=k_{\Psi}\sum_m \sum_n s_{m,n} \Psi_{m,n}(t) \quad (5.5)$$

where, K_Ψ is constant value for normalization.

The function $\Psi_{m,n}(t)$ provides sampling points on the scale time plane :linear sampling in the time (b-axis) direction but logarithmic in the scale (a-axis) direction. The most common situation is that a_0 is chosen as

$$a_0=2^{1/v} \tag{5.6}$$

Where V is an integer value and that V pieces of $\Psi_{m,n}(t)$ are as processed as one group which is called as Voice. The integer V is the no of voices per octave it defines a well-tempered scale in the sense of Music this is analogous to the use of a set of narrowband filters in conventional Fourier analysis.

5.1.3 Multi-resolution Analysis

In this section, the relationship between the scaling function $\phi(t)$ and the wavelet function $\psi(t)$ is discussed [97], [99], [100]. In the following discussion, L^2 refers to the space of square-integrable signals. Multi-resolution analysis involves the approximation of functions in sequences of nested linear vector spaces $\{V_k\}$ in L^2 that satisfy the following properties.

1. Ladder property: …. $V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2$ …
2. $\bigcap_{j=-\infty}^{\infty} V_j = \{0\}$.
3. Closure of $\bigcap_{j=-\infty}^{\infty} V_j$ is equal to L^2.
4. Scaling property: $x(t) \in V_j$ if and only if $x(2t) \in V_{j+1}$. Because this implies that "$x(t) \in V_0$ if and only if $x(2^{-j}t) \in V_j$", all the spaces V_j are scaled versions of the space V_0. For $j > 0$, V_j is a coarser space than V_0.
5. Translation invariance: If $x(t) \in V_0$, then $x(t-k) \in V_0$; i.e., the space V_0 is invariant to translation by integers. The scaling property implies that V_j is invariant to translation by $2^{-j}k$.
6. Special Orthonormal basis: A function $\phi(t) \in V_0$ exists such that the integer shifted version $\{\phi(t-k)\}$ forms an orthonormal basis for V_0. Using the scaling property means that $\left\{2^{-\frac{j}{2}}\phi(2^{-j}t-k)\right\}$ is an orthonormal basis of V_j. The function $\phi(t)$ is called the scaling function of multi-resolution analysis.

The scaling function $\phi_{j,k}(t) = 2^{-\frac{j}{2}}\phi\,(2^{-j}t-k)$ spans the space V_j. To better describe and parameterize signals in this space, a function that spans the difference between the spaces spanned by various scales of the scaling function is needed. Wavelets are these functions.

The space W_j spanned by the wavelet function has the following properties [100].

1. $\{\psi(t-k)\}$ is an orthonormal basis of W_0, given the orthogonal complement of V_0 in V_1, i.e., $V_1 = V_0 \oplus W_0$, where V_0 is the initial space spanned by $\phi(t)$.
2. If $\psi(t) \in W_0$ exists, then $\psi_{j,k}(t) = 2^{-\frac{j}{2}}\psi(2^{-j}t-k)$ is an orthonormal basis of the space W_j. W_j is the orthogonal complement of V_j in V_{j+1},

   i.e., $V_{m+1} = V_m \oplus W_m = V_0 \oplus W_0 \oplus W_1 \oplus \ldots \oplus W_m$
3. $L^2 = V_0 \oplus W_0 \oplus W_1 \oplus \ldots$

Using the scaling function and the wavelet function, a set of functions that span all of L^2 can be constructed. A function $x(t) \in L^2$ can be written as a series expansion in terms of these two functions as

$$x(t) = \sum_{k=-\infty}^{\infty} c(j,k)\phi_{j,k}(t) + \sum_{j=0}^{\infty}\sum_{k=-\infty}^{\infty} d(j,k)\psi_{j,k}(t) \qquad (5.7)$$

Here j is the coarsest scale. In the above expression, the first summation gives an approximation to the function $x(t)$ and the second summation adds the details. The coefficients $c(j, k)$ and $d(j, k)$ are the discrete scaling coefficients and the discrete wavelet coefficients of $x(t)$ respectively [100].

5.2 Wavelet analysis in practice

The low-frequency content is the crucial part for many signals. The signal gets its identity from this, whereas flavor is given by high-frequency content. The voice sounds different in case the high- frequency components are taken out. But on the removal of sufficient part of low- frequency components, the sound heard became vague. Approximations and details form an integral part of wavelet analysis. The approximations are the high-scale, low-frequency components of the signal while details are the low-scale, high-frequency components. Figure 5.1 shows the filtering process, at the initial level.

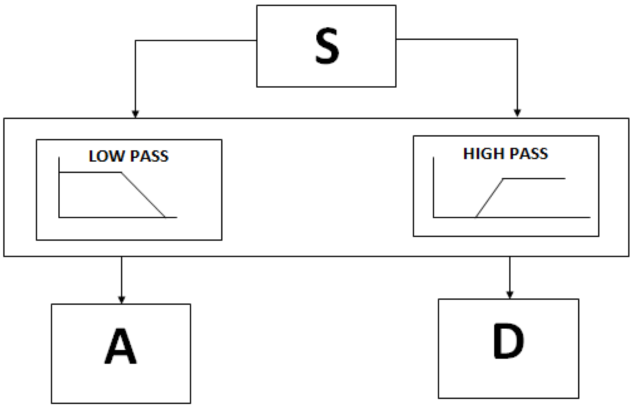


Fig. 5.1 Wavelet Transform decomposition at initial level

The input signal S will emerge as a combination of two signals as it passes through two complementary filters. If above operation is carried out on real digital signal, we are left with twice as much data as we started with. As an example, if original signal S consists of 1000 samples of data, then the final signals will have 1000 samples each, giving 2000 samples. This decomposition with wavelets can be carried out in more subtle way. The detailed information can be obtained by keeping only one point from two in each of the 2000 samples. This is idea of down sampling. Two sequences are generated called cA and cD as shown in Figure 5.2.

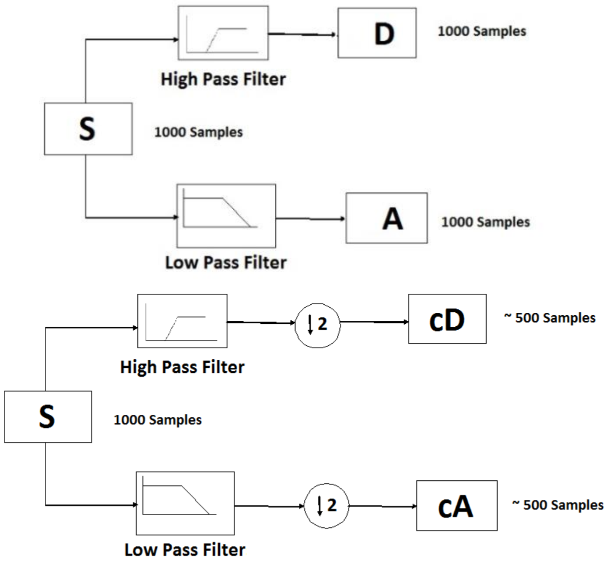


Fig. 5.2 Wavelet transform decomposition with down sampling

5.2.1 Analysis of Discrete Wavelet Transform

Logarithmic frequency resolution is generated by the discrete wavelet transform (DWT). Narrow bandwidth is characterized by the low frequencies and wide bandwidth is characterized by high frequencies. The scaling coefficients of the present level are separated by down sampling and filtering, so as to obtain the successive level coefficients in DWT decomposition. The spectrum is split by the first stage into two equal bands i.e. a high pass and low pass, in decomposition scheme. The low pass spectrum is split by a pair of filters into lower low pass and band pass spectra in second stage. Figure 5.3 shows the logarithmic set of bandwidth generated as a result of splitting. The multilevel decomposition (analysis) by DWT is shown in Figure 5.4.

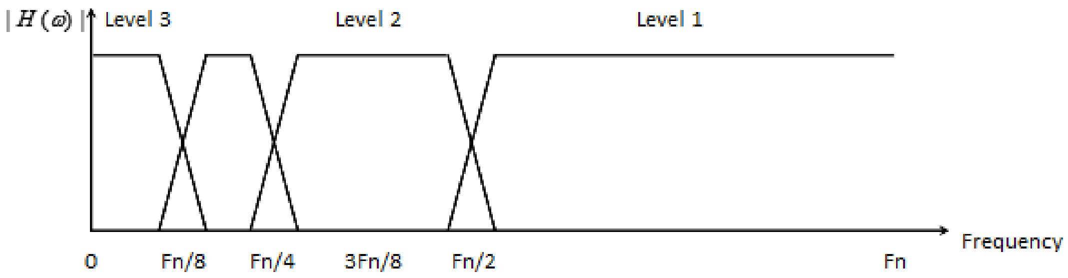


Fig. 5.3 Frequency response for level 3 DWT decomposition

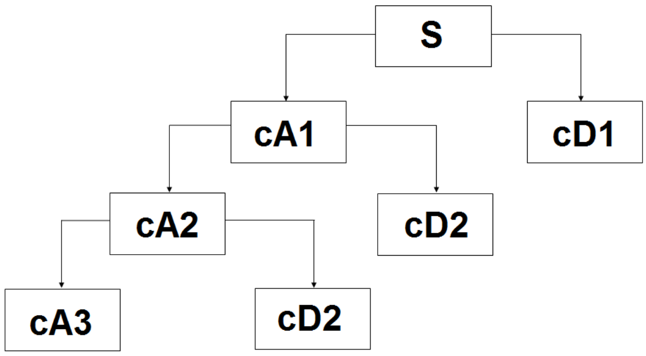


Fig. 5.4 Decomposition by DWT up to Level 3

The process to assemble the components into the original signal without losing any information is termed as reconstruction or synthesis. The inverse discrete wavelet transform (IDWT) is mathematical tool to manipulate the synthesis effects. Figure 5.5 shows a method of wavelet reconstruction.

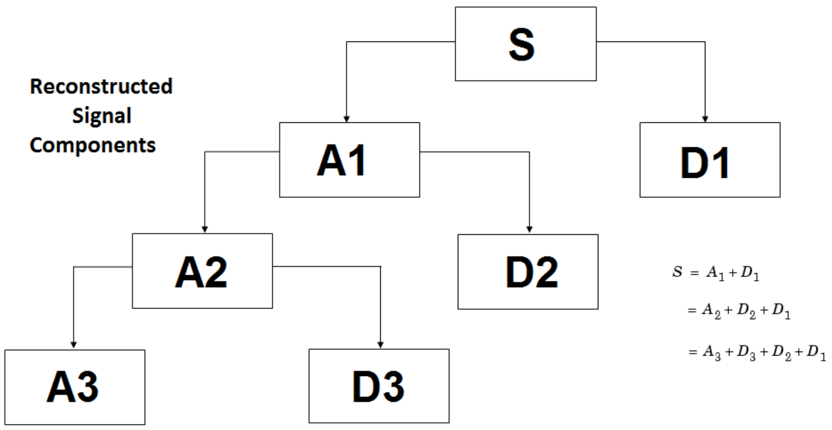


Fig. 5.5 Reconstruction by DWT

5.2.2 Analysis of Wavelet Packet

The analysis (decomposition) of wavelet packet allows the segmentation of low and high frequencies into smaller bands. The wavelet coefficients are split by filtering and down sampling of input signal. Figure 5.6 shows frequency spectra of full binary tree generated due to splitting of low (approximations) and high (details) frequency. Figure 5.7 shows multiple decomposition of wavelet packet (level 3). For a n-level decomposition, there are n+1 possible ways to decompose or encode the signal. The wavelet packets offer a more complex and flexible analysis, because in wavelet packet analysis, the details as well as the approximations are split. This yields more than 2e2e(n-1) different ways to encode the signal. For instance, wavelet packet analysis allows the signal S to be represented as A1 + AAD3 + DAD3 + DD2. This is an example of a representation that is not possible with ordinary wavelet analysis. Choosing one out of all these possible encodings presents an interesting problem. The wavelet packet reconstruction scheme is achieved by up sampling, filtering with appropriate filters and adding coefficients.

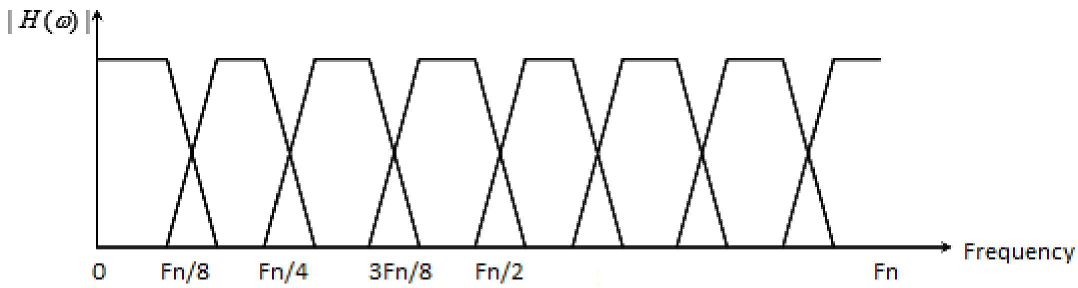


Fig. 5.6 Frequency response for level 3 Wavelet Packet decomposition

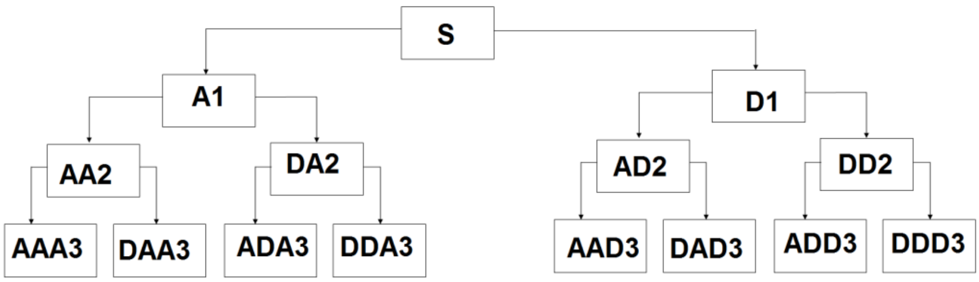


Fig. 5.7 Decomposition by Wavelet Packets up to Level 3

In our work, we did not use down sampling operation during the decomposition. Also we use a *frequency range* criterion to select the most suitable decomposition of a given signal. This means we look at each node of the decomposition tree and quantify the frequency range of the information to be gained by performing each split.

5.3 Selection of a Wavelet Basis

Proper selection of wavelet basis will produce a sparse representation with many wavelet coefficients being near to zero [102]. The number of vanishing moments, support size of $\psi(t)$ and regularities of the analyzed signal $x(t)$ are the important parameters on which lies the ability of wavelet analysis to produce a large number of non-significant coefficients. Mallat formed a relation between the number of vanishing moments and support size to the wavelet coefficient amplitude [102]. Wavelet function $\psi(t)$ has p vanishing moment if

$$\int_{-\infty}^{\infty} t^k \psi(t)dt = 0 \qquad \text{for } 0 \le k < p \qquad (5.8)$$

If $\psi(t)$ has enough vanishing moments and $x(t)$ is regular, then the wavelets coefficients $d(j, k)$ = $\langle x(t), \psi_{j,k} \rangle$ are small at fine scale. In size of support, if $x(t)$ has an isolated singularity (a point at which the derivative does not exist although it exists everywhere else) at t_0 and if t_0 is inside the support of $\psi_{j,k}(t)$, then $d(j, k)$ = $\langle x(t), \psi_{j,k} \rangle$ may have large amplitude. If $\psi(t)$ has a compact support of size K, there are K wavelets $\psi_{j,k}(t)$ at each scale 2^j whose support includes t_0. The number of large amplitude coefficients may be minimized by reducing the support size of $\psi(t)$. If $\psi(t)$ has p vanishing moments, then its support size is at least $2p - 1$ [102]. A reduction in the support size of $\psi(t)$ unfortunately means a reduction in the number of vanishing moments of $\psi(t)$. There is a trade off in the choice of $\psi(t)$. A high number of vanishing moments is preferred if the analyzed signal $x(t)$ has few singularities. If the number of singularities of $x(t)$ is large, a $\psi(t)$ with a short support size is a better choice.

5.3.1 Properties of Filter Banks

The type of errors introduced into the signal along with the properties provided by the filters is a salient feature of the filter banks. The reconstruction error in synthesizing is made up of three components viz. amplitude distortion, phase distortion and aliasing distortion. Filters used for audio coding should satisfy the properties of orthogonality, linear phase, and finite support as they are beneficial in two ways. First, these properties can be possibly of some value in reconstruction and secondly these properties can be transferred to the generated wavelets. For e.g., orthogonality ensures quantization noise to remain independent in two different channels, linear phase provides constant group delay and finite support leads to stable-simple implementation. So, if filter bank with a certain set of properties is designed, then the wavelet basis generated will also have the same properties. [103]

For the perfect reconstruction filter bank, it is observed that any two of the three properties can be simultaneously satisfied. Figure 5.8 shows this limitation with a different solution to the perfect reconstruction filter bank. Form the Figure it is observed the three properties cannot be simultaneously satisfied except at the center, where the three properties merge (Haar Solution). In addition to above said properties, the design of filter banks should include important properties like, sharp cut-off rate, stop-band ripples and low pass-band, stop-band attenuation and short delay [104].

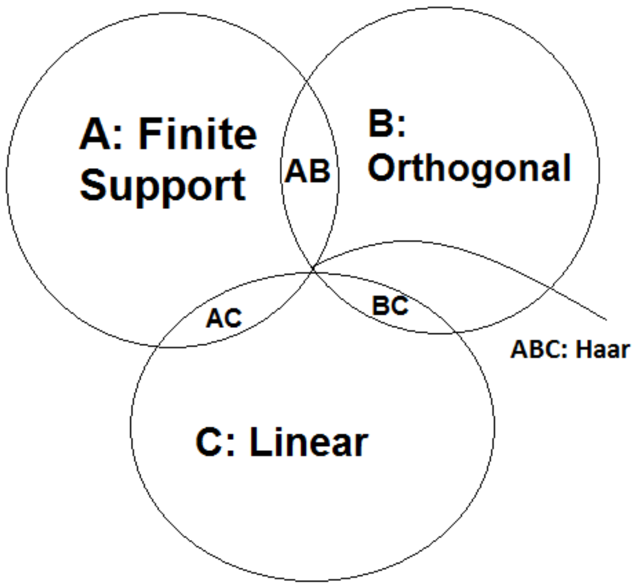


Fig. 5.8 Perfect reconstruction filter bank. Venn diagram for 1) finite support, 2) orthogonality, and 3) linear phase

5.3.2 Examples of Wavelets Basis

The orthogonal wavelets like Daubechies and Symlets possess highest number of vanishing moments for certain support width. Compactly supported wavelets such as Symlet have least asymmetry and Daubechies wavelets are far from symmetry. As per the nomenclature Daub-*i* or Sym-*i*, here *i* stand for the order e.g. Daub-4 indicates 4-th order Daubechies wavelet while Sym-9 is the 9-th order Symlet wavelet. The wavelet with a certain order has a filter length 2i with support width of 2i-1 and vanishing moments of ‘i’ [96].

The biorthogonal wavelets are generated from octave band filter bank design using two channel biorthogonal perfect reconstruction quadrature mirror filter (QMF) bank. FIR filter provides symmetry and exact reconstruction to compactly supported biorthogonal spline wavelets. This is not possible in orthogonal case except for Haar and shows the characteristic of linear phase [97].

5.4 Optimized Wavelet Packet

The logarithmic frequency resolution is obtained from discrete wavelet transform. Low frequencies have narrow bandwidth while high frequencies have wider bandwidth. The segmentation of higher frequencies into narrower bands is allowed by wavelets packets [100]. For speech analysis, wavelet packets prove to be an efficient tool. The selection of decomposition tree followed by selection of filters for every decomposition level of the tree is involved in the designing of wavelet packets. After the selection of decomposition tree, the immediate step involves selection of appropriate wavelet filters for every decomposition level of the tree. Different time-frequency resolution exists at each level. In our work, we have not used down sampling operation during the decomposition. We have used a *frequency range* criterion to select the most suitable decomposition of a given speech signal. This means we look at each node of the decomposition tree and quantify the frequency range of the information to be gained by performing each split.

In our research we use discrete wavelet transform and wavelet packet at various levels of decomposition to develop optimized wavelet packets as per the frequency criterion as shown in Figure 5.9 and Figure 5.10. MATLAB software was used to develop codes for optimized wavelet packet algorithms for different wavelets. In our research three different MATLAB codes were developed based on optimal wavelet packet using Daubechies, Symlet and Biorthogonal wavelet families. Daubechies wavelet is orthogonal wavelet that has the highest number of vanishing moments for a given support width. Symlet are compactly supported wavelets with least asymmetry. Figure 5.10 shows optimized wavelet packet tree structure. Biorthogonal wavelets chosen such that symmetry and exact reconstruction are possible using FIR filters. The inverse wavelet packet transform was used to synthesize speech components from the wavelet packet representation. To synthesize the speech component, wavelet coefficients were used.

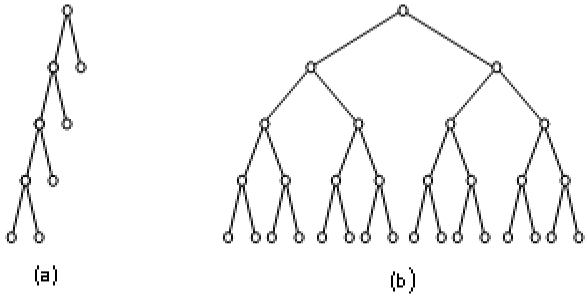


Fig. 5.9 Decomposition tree structure upto 4 level for a) DWT b) WP

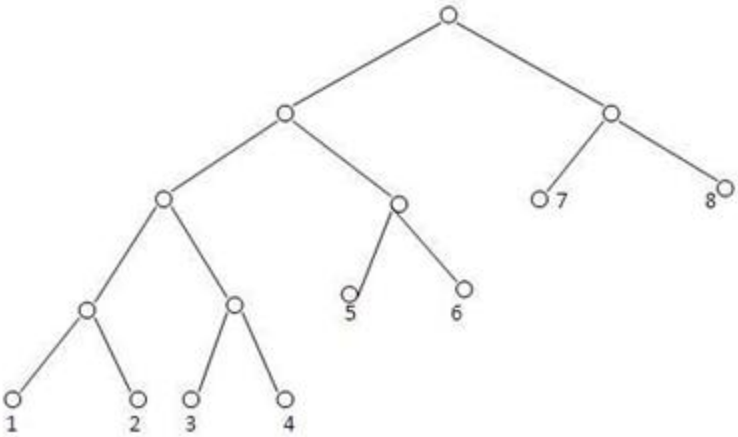


Fig. 5.10 Decomposition tree structure for WP (optimized)

The processing algorithms were developed as spectral splitting with optimized wavelet packets based on eight frequency bands as the performance by hearing-impaired subjects saturated around eight channels, while performance by normal-hearing subjects sustained to 12–16 channels in higher background noise [62]. The number of channels desired to obtain high levels of speech understanding is still the subject of discussion [105].

Table 5.1 shows the entire eight bands in alternate fashion for even-odd index with center and pass band frequency for each band in KHz.

Table 5.1 Pass band frequencies

Filter for the left ear			Filter for the right ear		
Band	Centre frequency (KHz)	Passband frequency (KHz)	Band	Centre frequency (KHz)	Passband frequency (KHz)
1	0.15625	0-0.3125	2	0.46875	0.3125-0.625
3	0.78125	0.625-0.9375	4	1.0937	0.9375-1.25
5	1.5625	1.25-1.875	6	2.1875	1.875-2.5
7	3.125	2.5-3.75	8	4.375	3.75-5

Optimized wavelet packets were developed using the pseudo algorithm as follows:

- Read audio input signal $\boldsymbol{x(n)}$ of length N.
- Perform wavelet packet decomposition of $x(n)$ up to level 4 as shown in Figure 5.9(b)
- Construct the optimized wavelet packet tree T_{opti} by rejoining following nodes of the original tree T: [11, 12, 13, 14] and [9, 10, 5, 6]. The optimized tree will have only 8 nodes as shown in Figure 5.10
- Selectively reconstruct the optimized wavelet tree to get two output signals - one for left ear and other for right ear, as follows.
  - In optimized tree, make all 4 approximate coefficients nodes [15,17,9,5] zero while keeping detail coefficients nodes as it is and reconstructed that tree.
  - In optimized tree, make all 4 detail coefficients nodes [16,18,10,6] zero keeping approximate coefficients nodes as it is and reconstructed that tree.

5.5 Implementation of Wavelet filters on FPGA Platform

The wavelet filter algorithms developed in section 5.4 cannot be synthesized on to FPGA using MATLAB HDL coder. Therefore, the novel approach has been proposed for designing the optimized wavelet packet filter for synthesizable FPGA implementation. In this approach, the direct form –I filter with the wavelet coefficients have been used for odd and even filter banks. Convolution has been implemented by its basic steps of multiplication, shift and addition. This approach makes the MATLAB code synthesizable into FPGA. The transfer function H(z) used for implementation is given by equation 5.9

$$H(z) = \frac{Y(z)}{X(z)} = \sum_{i=0}^{N} a_i Z^{-i} \qquad (5.9)$$

where, a_i is a filter coefficient, N is order of filter and hence $N+1$ is length of filter.

The difference equation for filter bank is given by equation 5.10.

$$Out\_h = h(0)u(0)+h(1)u(n-1)+\ldots\ldots+h(7)u(n-7) \qquad (5.10)$$

where $h(0), h(1)\ldots.h(7)$ are filter coefficients.

Filter structure used in designing the algorithm is shown in Figure 5.11. The basic adder has only two inputs, so we require adder tree to add result of all multiplications. The flow chart of the algorithm is shown in Figure 5.12

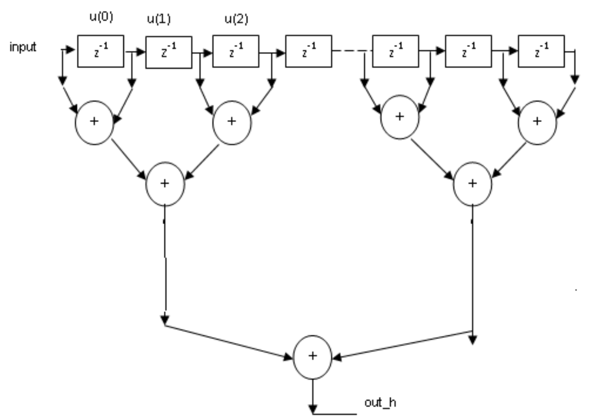


Fig. 5.11 Filter Structure

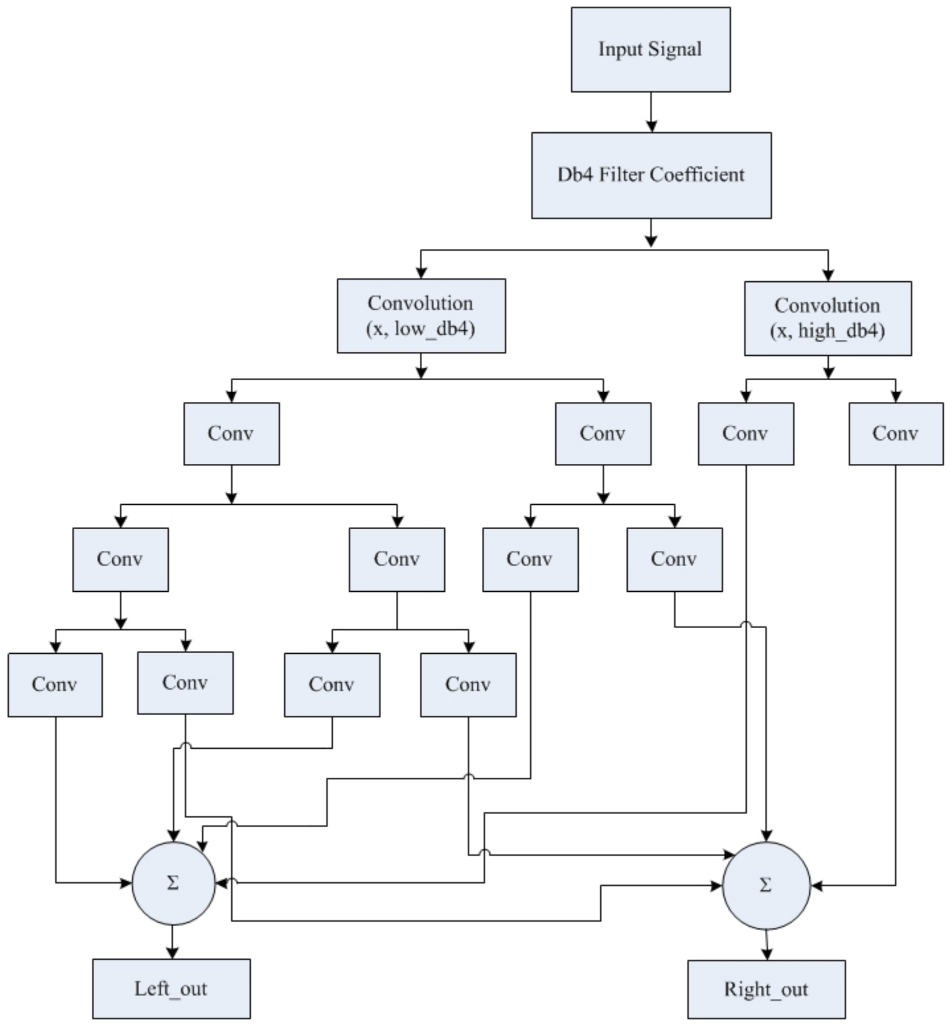


Fig. 5.12 Flowchart of proposed algorithm

5.5.1 Real Time Implementation of Wavelet based filters

The flow diagram of developed algorithms for real-time processing implemented by HDL coder toolbox is shown in Figure 5.13. The HDL coder toolbox was used to generate VHDL code from MATLAB using HDL Workflow advisor, while ModelSim was used to observe and verify the test bench. The generated VHDL code and test bench was used to program Spartan 6 (XC6SLX45CSG324C) FPGA. The user can use the generated ".vhdl" files through Xilinx environment to view the Register Transfer Level (RTL) and Technology schematic. The detailed implementation process has been discussed in Appendix B

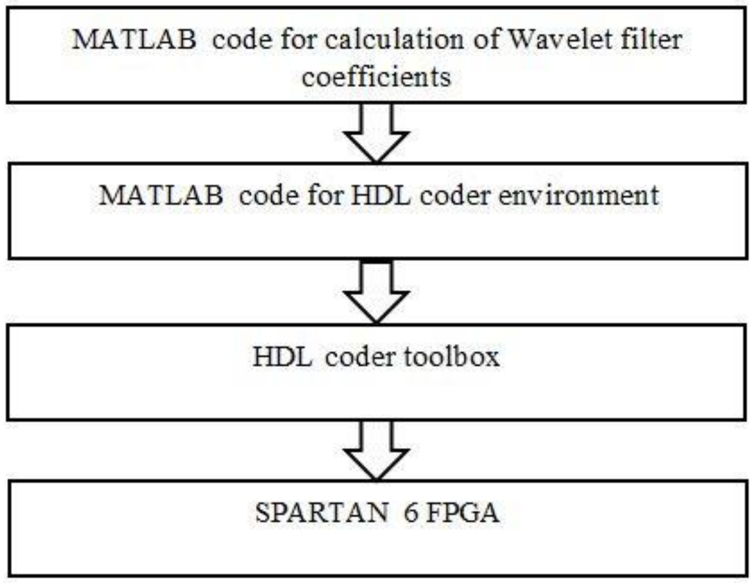


Fig. 5.13 Flow diagram for MATLAB-VHDL in Wavelet based filters

In real-time processing, Atlys Circuit board of Digilent was used [107]. This board has numerous features of which Xilinx Spartan 6 LX45 FPGA, 10/100/1000 Ethernet PHY along with AC-97 Codec with line-in, line-out, headphone and mic, were used. It also includes Digilent's new Adept USB2 system which offers device programming, virtual I/O's, simplified user data transfer facilities etc. In our implementation AC-97 protocol based audio codec "LM4550" available on the circuit board uses 18-bit Sigma-Delta ADCs and DACs. The block schematic for real time implementation is shown in Figure 5.14.

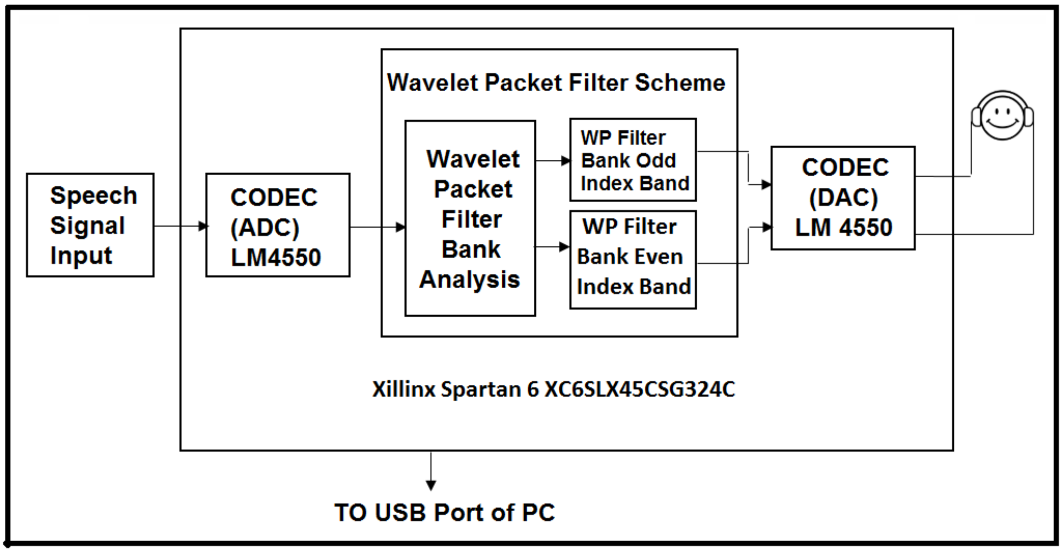


Fig. 5.14 Block schematic for Real time implementation of Wavelet based filters

5.5.2 Real Time Experimental Implementation of Wavelet based filters

The interconnection of FPGA “Xilinx Spartan-6 XC6SLX45CSG324C” with audio codec “LM4550” on the board used for implementing the wavelet filters is shown in Figure 5.15. The experimental setup accepts the input unprocessed speech signal, processes it through wavelet filters implemented on FPGA via AC link and produces the processed speech output. The input unprocessed speech signal being continuous in nature had to be converted into digital form while the processed digital signal so obtained had to be converted back to continuous form. So, we used an on-board AC-97 Codec (LM4550) for the same. The detailed explanation of interfacing AC-97 codec (LM4550) with FPGA is discussed in Chapter 4, section 4.3.2. For Spartan 6 XCLS45, the Atlys circuit board consists of 18bit codec. The unprocessed continuous speech signal given through mic-in of the circuit board is converted into 18 bit digital code. This code is processed through our wavelet based algorithms. The output is converted back to continuous processed speech signal using 18bit DAC.

In our research, for Spartan 6 XCLS45 the incoming signal from SDATA_IN gets split in 9 bits. Each 9-bit data is stored in two registers for each channel. The compatible 9 bit VHDL code (wavelet based filters code) is implemented on the FPGA so as to process each register. Processed data is then sent on SDATA_OUT line. This data is then converted to continuous analog signal using 18bit DAC. The codec used here processes data in serial fashion while the FPGA processes data in parallel manner. Hence we have also written the codes for serial in parallel out (SIPO) and parallel in serial out (PISO) in our algorithms.

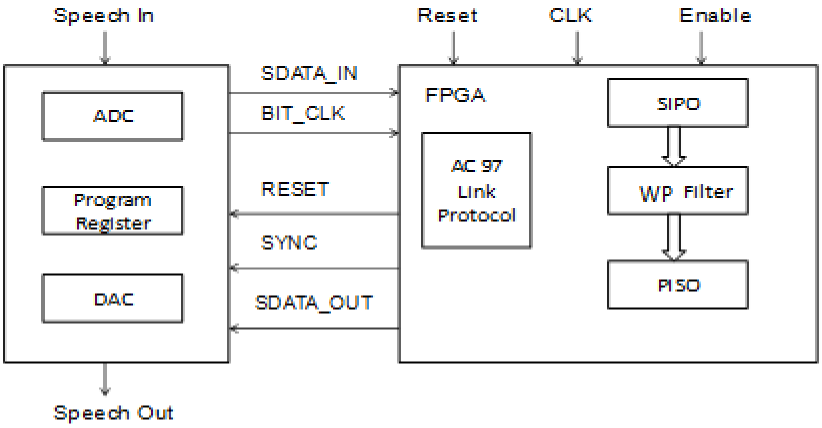


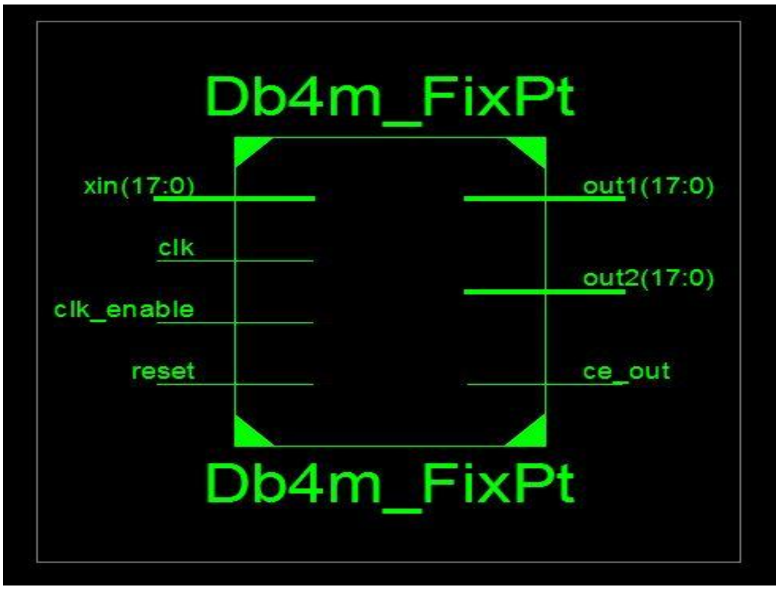
Fig. 5.15 FPGA implementation schematic of Wavelet based filters

5.6 Experimental Results

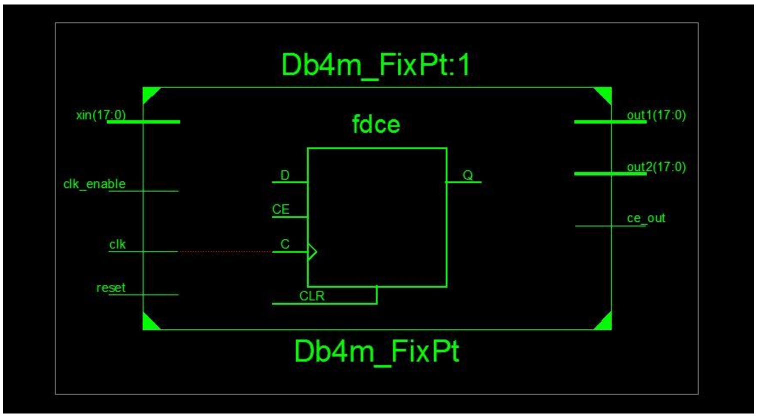
Experimental results were obtained after implementing the developed algorithms (db, sym, and bior) effectively on FPGA. The results obtained are discussed in subsequent sections for Daubechies wavelet based filter.

5.6.1 Simulation Results

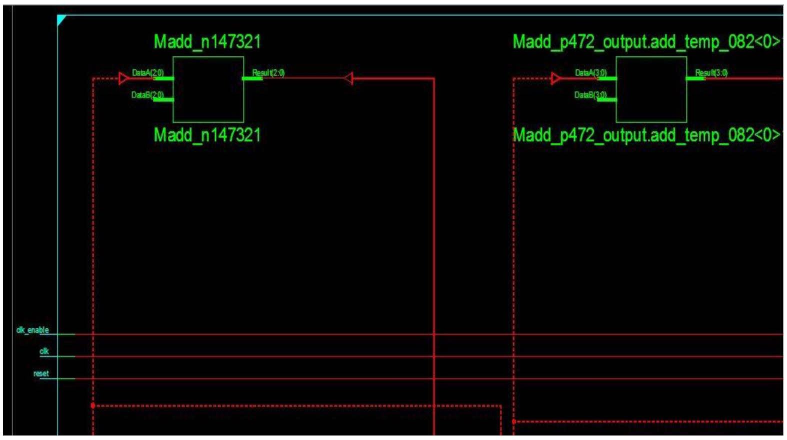
Once the MATLAB simulated results are verified the HDL coder is used to convert the MATLAB code into VHDL code. The generated VHDL code is synthesized and the RTL schematic is generated using XILINX 14.2 for verification. The generated RTL schematic and upto 2 level down for db4 wavelet based filter is shown in Figure 5.16. The generated RTL schematic shows that the implemented filter structure in FPGA has input ‘xin’ of 18 bit wide. The output signal is obtained for even and odd bands g_out and h_out respectively, both of which are 18 bit wide. Figure 5.17 shows the simulated test bench waveform using ModelSim for the implemented db4 wavelet based filter.



a)



b)



c)

Fig. 5.16 a) RTL Schematic for db4 Wavelet based filter b) One down level schematic c) Second down level schematic

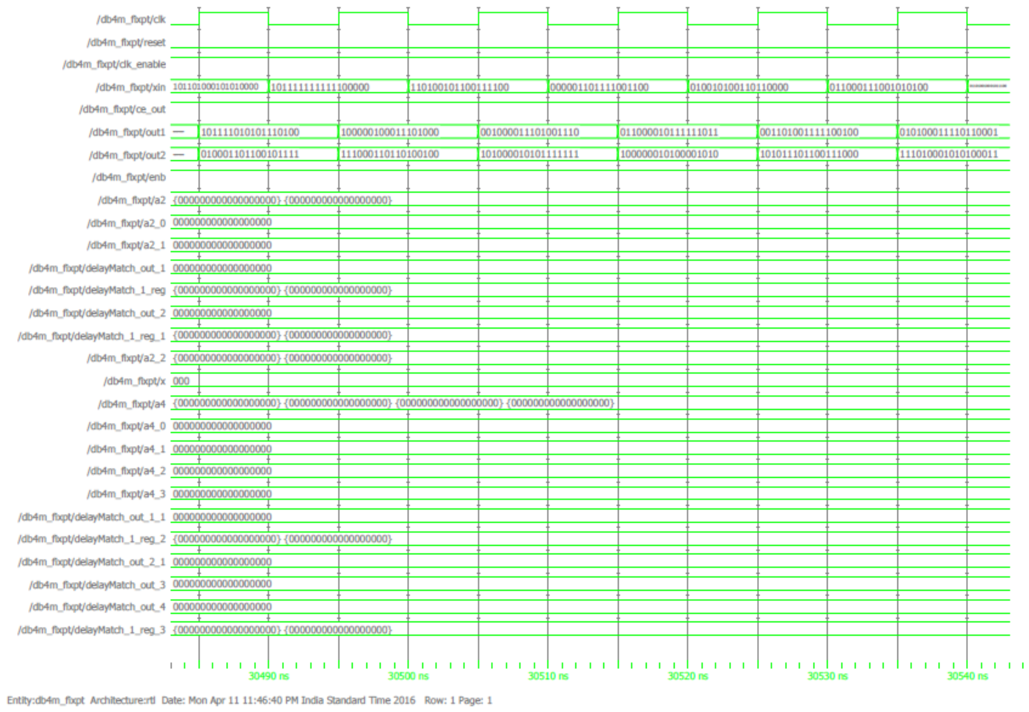


Fig. 5.17 Simulated test bench waveform for db4 Wavelet based filter

5.6.2 Hardware Implementation

In this work, for realization of db4 wavelet based filter into FPGA the input and output data lines splits into 9 bits each. Due to this reason, the FPGA implementation utilize fraction of the resources available on FPGA Spartan6 as shown in Table 5.2 with scope for implementation of other processing blocks of the hearing aid.

Table 5.2 Summary of Device Utilization

Logic	Used	Available	Utilization (%)
No. of Slice Register	2,125	54,576	3
No. of Slice LUT	10,679	27,288	39
No. of LUT/FF pairs	1,397	11,296	12
No. of IOBs	14	218	6

5.6.3 Latency

The latency of db wavelet based algorithm was measured using sinusoidal waveform of frequencies ranging from 20Hz to 5000Hz. Latency, is defined as measurement in the time delay between the unprocessed signal which enters in FPGA kit and the processed signal is made available. This delay was measured by giving the unprocessed signal and processed signal to Digital Storage Oscilloscope Textronix TBS1062 so that an oscilloscope could be used to record both signals with high sampling frequency as shown in Figure 5.18. Further, these recorded signals were analyzed in MATLAB. Test results varied with change in sinusoidal frequencies as shown in Table 5.3 and found average latency of 1.47ms. This amount of delay is evident to listeners [94], and also it does not take any delay contributed by other part of the hardware.

Table 5.3 Latency for different frequencies

Frequency in Hz	Delay(sec)
20	7.34E-03
200	1.00E-05
500	7.20E-06
1000	4.00E-06
5000	1.40E-06

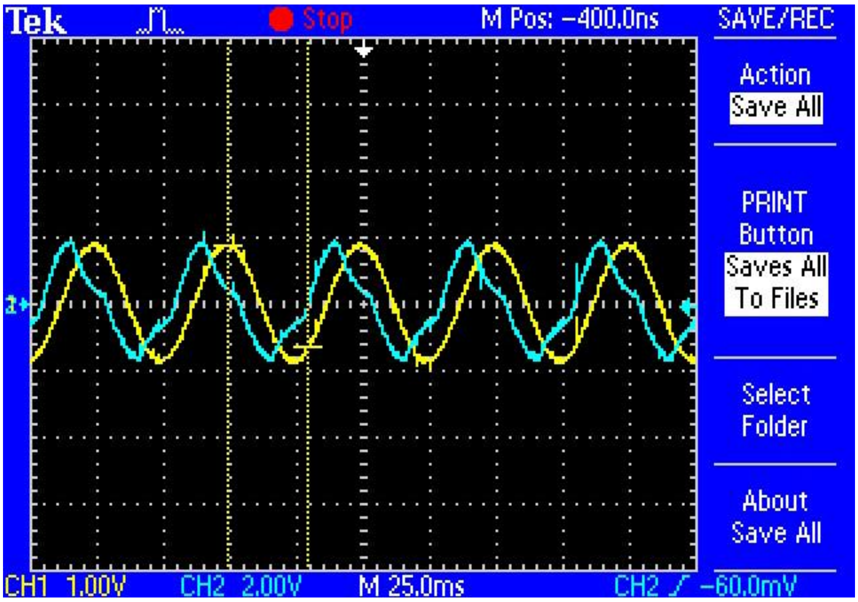


Fig. 5.18 Measurement of Latency for db4 Wavelet based filter

5.6.4 Peak Signal to Noise Ratio

PSNR is defined as ratio of power of speech signal to power of noise signal. High value of PSNR indicates less amount of noise added to the signal while low value of PSNR indicates high amount of noise added to signal. Noise performance of VCV context speech signal (ada.wav) in db, sym and bior wavelet based filter algorithms is assessed by considering the PSNR value and is shown in Table 5.4.

Table 5.4 PSNR values of different Wavelet based filters

Parameters	db	sym	bior
PSNR	99.6722	99.6754	99.2949

The PSNR value was found to be better when compared with comb filters.

5.7 Evaluation of Wavelet based Filter Algorithms

The evaluation and implementation of wavelet based filter algorithms are regarded as second phase of this research. In this phase, three experiments viz. Experiment II, Experiment III and Experiment IV were performed. Experiment II was implemented in real time and evaluation was carried out by conducting the listening tests on five normal people with simulated hearing loss. Experiment III was a software based offline implementation while Experiment IV was hardware based real time implementation. The listening tests of these experiments were carried out on eight hearing impaired subjects to evaluate the performance of algorithms.

5.7.1 Test subjects for Wavelet based Filters

In Experiment II, five normal hearing people (VG: M 36, SA: M 22, MF: M 19, MJ: M 36, BB: M 71) participated in the listening tests. Tests were conducted with simulated hearing loss by adding a broad band noise to the speech signal with signal to noise ratio (SNR) values of 3dB, 0dB, -3dB, -6dB and -9dB. All the subjects had pure tone thresholds less than 20 dB in the frequency range of 125 Hz to 6 KHz.

In Experiment III and Experiment IV, the assessment was carried out by conducting listening tests on eight subjects with bilateral 'mild' to severe sensorineural hearing loss. (APL: F 16, KUR: M 57, KSH: F 12, BHL: F 10, PDH M 74, PNK: M 45, BHV: M 12, SWK: F 50) Subjects KSH, BHL and BHV have severe and symmetrically hearing impairment. Subject KUR has mild to moderate and asymmetrical high frequency impairment (less loss in one ear and more loss in other ear). Subject PNK has mild and symmetrically sloping high frequency impairment. Subjects APL and SWK have moderately severe and symmetrical hearing impairment while subject PDH has moderately severe and asymmetrical low frequency hearing impairment. Hearing thresholds of impaired subjects are mentioned in Appendix A.

5.7.2 Test material for Wavelet based Filters

The speech material used for the phase 2 consisted of a set of fifteen nonsense syllables in VCV context with consonants / p, b, t, d, k, g, m, n, s, z, f, v, r, l, y / and vowel /a/ as in 'farmer'. To reduce the contribution of linguistic factors, nonsense syllables were used. Responses were tabulated in the form of confusion matrix and response time was also recorded. Confusion matrices were used for calculating recognition scores and relative transmitted information. Further, the consonants were clustered according to the articulatory features [57] and contribution of different features were analysed. The features selected for this study were voicing (voiced: / b d g m n z v r l y / and unvoiced: / p t k s f /), place (front: / p b m f v /, middle: / t d n s z r l /, and back: / k g y /), manner (oral stop: / p b t d k g l y /, fricative: / s z f v r /, and nasals: / m n /), nasality (oral: / p b t d k g s z f v r l y /, nasal: /m n /), frication (stop: / p b t d k g m n l y /, fricative: / s z f v r /), and duration (short: / p b t d k g m n f v l / and long: /s z r y /).

5.7.3 Listening test parameters

The outcome of the listening tests of phase II has been analyzed with respect to four parameters viz. qualitative assessment (mean opinion score), recognition score, response time and information transmission analysis. The qualitative assessment has been used to obtain the subjects opinion on the quality of the unprocessed and processed speech signal. Response time is the time interval between speech material presented dichotically to subjects and the response given by subjects. Recognition score is the accuracy of recognition of the speech material by the hearing impaired subjects. If this parameter is found nearly equal then response time statistics can be used to compare the effectiveness of the algorithms. The information transmission analysis was obtained as it was not affected by subject's response bias. The stimuli were combined in groups and the resulting matrices were analyzed for reception of the consonantal features.

The procedure and listening test results of these experiments are discussed in section 5.8, section 5.9 and section 5.10 respectively.

5.8 Experiment II: Listening tests on Normal hearing subject with simulated hearing loss

The objective of the experimental analysis was to assess the usefulness of the developed algorithms. Traditionally this experimentation is accomplished on hearing impaired subjects but they may be time consuming and may cause fatigue in elderly people. So, preliminary evaluations of the developed algorithms were conducted on normal people with simulated hearing loss and are considered as Experiment II in this work.

5.8.1 Experiment II: Procedure

Five normal hearing people (VG: M 36, SA: M 22, MF: M 19, MJ: M 36, BB: M 71) participated in the listening tests. These listening tests were conducted in an acoustically isolated room The loss was simulated by adding broadband noise to the speech stimuli with five different SNRs. SNR conditions used were 3dB, 0dB, -3dB, -6dB and -9dB. The wavelet based filter algorithms (db4, sym9 and bior2.4) were implemented in real time on FPGA platform. Listening tests were carried out for finding the confusion among the set of fifteen English consonants. To make these tests user friendly MATLAB based GUI was developed with a provision to manually enter the SNR conditions and interfaced with Spartan6 FPGA Atlys circuit board. The final results were collected in confusion matrix to evaluate the response times, recognition scores and information transmission analysis.

The time taken by every person for conducting the listening test of one algorithm with all SNR conditions was about ninety minutes. So, every person required approximately four to five hours to complete the listening tests for the three algorithms. Based on accessibility and readiness of the five normal people, the test sessions were spread over a period of approximately 1-2 months

5.8.2 Experiment II: Listening test Results

The following subsection includes results of listening tests for response time, recognition score, relative improvement in recognition score and information transmission analysis results for unprocessed and processed speech with all SNR conditions.

5.8.2.1 Response time

The response time for the processed and unprocessed speech signal for various SNR conditions is presented in the Table 5.5 and its graphical presentation is provided in Figure 5.19. Table 5.6 shows the average response time of three algorithms for the processed and unprocessed signals. From Table 5.5 it is observed that as SNR reduces, the response time increases for unprocessed and processed signals. For SNR conditions at -9 dB, -6dB, -3dB, 0dB, 3dB the corresponding decrease in average response time were 4.29sec, 3.73sec, 3.72 sec, 3.22 sec and 3.05sec for processed signals and similarly for unprocessed signal the values were 4.40 sec, 3.83 sec, 3.78 sec, 3.60 sec and 3.28 sec. The average response time increased with increase in the level of masking noise for all subjects. It was observed that response time for processed speech was significantly lower than that for unprocessed one.

Table 5.5 Experiment II - Response time for Normal People

Wavelets	US					PS				
	SNR: 3dB	SNR: 0dB	SNR: -3dB	SNR: -6dB	SNR: -9dB	SNR: 3dB	SNR: 0dB	SNR: -3dB	SNR: -6dB	SNR: -9dB
db	3.18	3.80	3.24	3.18	4.20	2.95	2.90	2.95	3.58	4.00
bior	3.18	3.80	3.24	3.18	4.20	3.20	3.13	2.68	2.50	3.66
sym	3.18	3.80	3.24	3.18	4.20	3.22	2.35	3.62	3.93	4.30
db	2.70	3.23	3.48	4.02	4.59	2.35	3.12	4.61	3.70	4.57

bior	2.70	3.23	3.48	4.02	4.59	2.70	3.24	3.36	3.65	4.32
sym	2.70	3.23	3.48	4.02	4.59	2.12	3.20	3.38	3.65	4.56
db	2.88	3.20	3.73	3.17	3.63	2.20	2.27	2.61	2.66	3.46
bior	2.88	3.20	3.73	3.17	3.63	2.13	2.36	4.43	2.51	4.20
sym	2.88	3.20	3.73	3.17	3.63	2.21	2.66	3.27	3.51	3.45
db	3.38	3.45	3.71	3.80	4.13	3.30	3.52	3.62	3.68	3.84
bior	3.38	3.45	3.71	3.80	4.13	3.11	3.25	3.72	3.78	3.92
sym	3.38	3.45	3.71	3.80	4.13	3.71	3.82	3.65	4.20	3.90
db	4.24	4.33	4.73	4.97	5.44	4.09	4.12	4.35	4.80	5.10
bior	4.24	4.33	4.73	4.97	5.44	4.17	4.26	4.72	4.92	5.40
sym	4.24	4.33	4.73	4.97	5.44	4.26	4.13	4.85	4.92	5.60

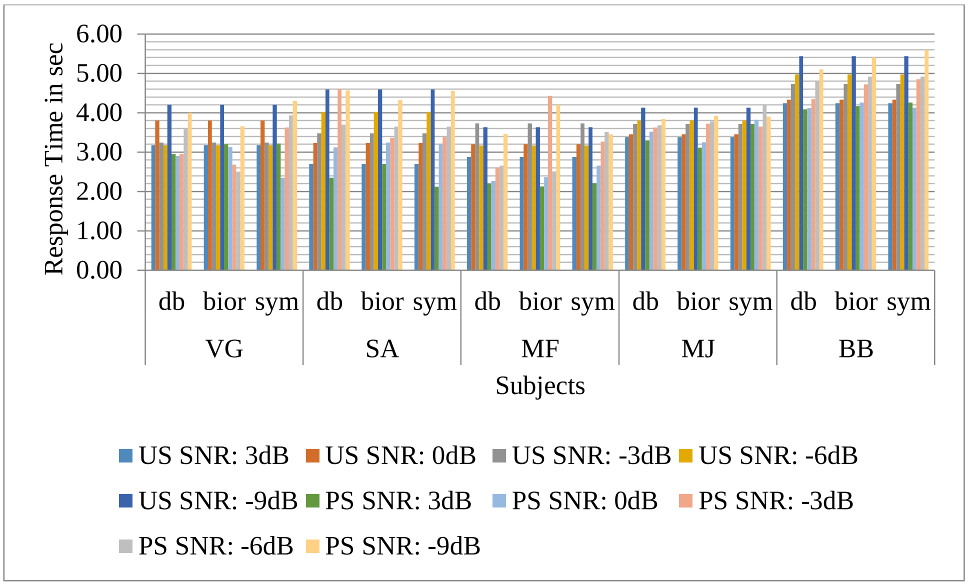


Fig. 5.19 Experiment II - Response time

Table 5.6 Experiment II - Average Response time

	SNR: 3dB	SNR: 0dB	SNR: -3dB	SNR: -6dB	SNR: -9dB
US	3.28	3.60	3.78	3.83	4.40
PS	3.05	3.22	3.72	3.73	4.29

5.8.2.2 Recognition scores

Table 5.7 shows the recognition scores of individual subjects, average of recognition scores and average relative improvement across five subjects at different SNR conditions. The graphical presentation of recognition score is depicted in Figure 5.20 and Figure 5.21 shows the average relative improvement of recognition score in percentage. The improvements in processing with all the three algorithms were highly significant when compared to unprocessed ones. The average relative increase for db algorithm was 1.38, 1.83, 2.14, 3.51, and 4.23% with SNR conditions 3, 0, 3, -6 and -9dB respectively. Similarly for bior the average relative increase were 2.60, 1.27, -0.37, 0.43 and 3.28% with SNR conditions 3, 0, -3, -6 and -9dB respectively and for sym, these values were 1.93, 1.60, -0.26, 1.11and 3.28% with SNR conditions 3, 0, -3, -6,and -9dB respectively.

For all the subjects, as the SNR decreases (masking noise level increases), the recognition score generally decreases. It is observed that the improvements due to algorithms are more for higher levels of masking noise and was found for db algorithm at -9dB SNR condition.

Table 5.7 Experiment II - Recognition scores for Normal people

Subject		VG	SA	MF	MJ	BB	Avg	Avg RI
SNR: 3dB	US	100.00	97.33	100.00	96.00	82.00	95.07	
	Pr-db	100.00	100.00	100.00	100.00	82.22	96.44	1.38
	Pr-bior	100.00	100.00	100.00	100.00	88.33	97.67	2.60
	Pr-sym	100.00	100.00	100.00	100.00	85.00	97.00	1.93
SNR: 0dB	US	100.00	97.33	100.00	98.66	76.00	94.40	
	Pr-db	100.00	97.14	100.00	100.00	84.00	96.23	1.83
	Pr-bior	100.00	100.00	100.00	100.00	78.33	95.67	1.27
	Pr-sym	100.00	100.00	100.00	100.00	80.00	96.00	1.60
SNR: -3dB	US	100.00	98.66	100.00	92.00	77.33	93.60	
	Pr-db	100.00	98.67	100.00	100.00	80.00	95.73	2.14
	Pr-bior	100.00	96.67	100.00	100.00	69.49	93.23	-0.37
	Pr-sym	100.00	96.67	100.00	100.00	70.00	93.33	-0.26
SNR: -6dB	US	98.66	97.33	97.00	93.33	66.66	90.60	
	Pr-db	98.67	98.67	98.67	98.53	76.00	94.11	3.51
	Pr-bior	100.00	96.61	96.67	98.53	63.33	91.03	0.43
	Pr-sym	100.00	95.00	96.67	98.53	68.33	91.71	1.11
SNR: -9dB	US	90.66	93.33	96.00	94.66	58.67	86.66	
	Pr-db	93.24	94.67	97.33	93.24	76.00	90.90	4.23
	Pr-bior	96.67	100.00	96.67	93.06	63.33	89.94	3.28
	Pr-sym	98.33	95.00	98.33	93.06	65.00	89.94	3.28

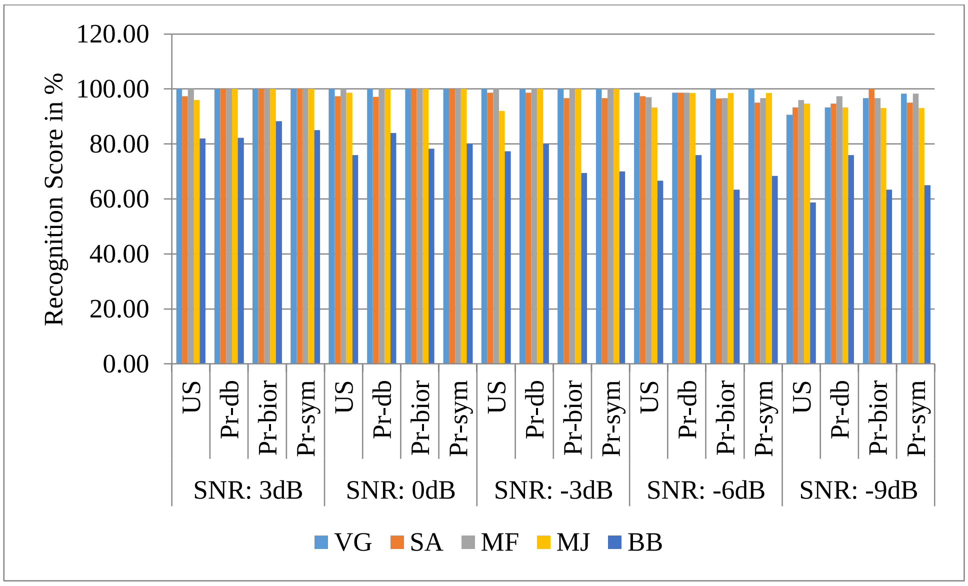


Fig. 5.20 Experiment II - Recognition scores for Normal People

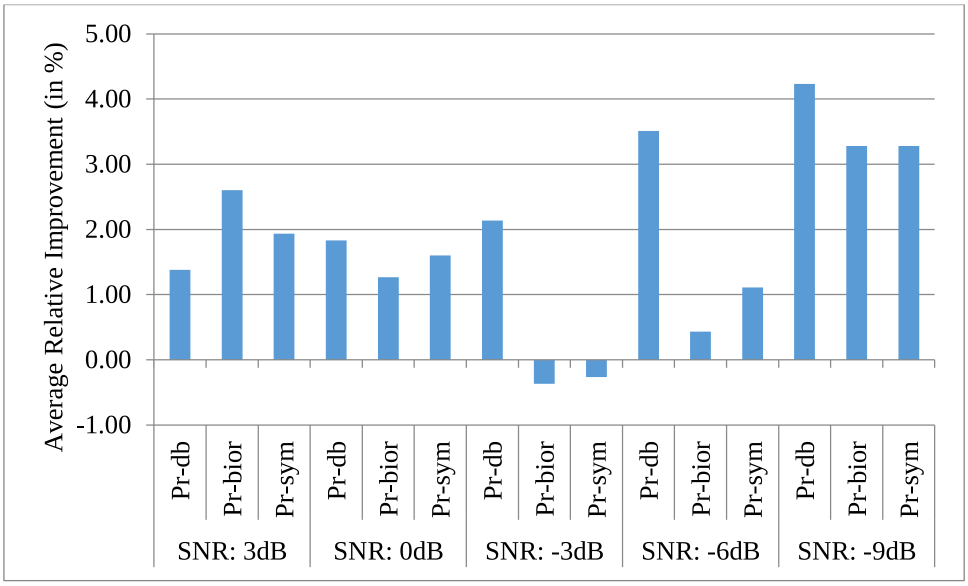


Fig. 5.21 Experiment II - Average Relative improvement (in %)

5.8.2.3 Information Transmission Analysis

Combined confusion matrix with consistency in score was subjected to information transmission analysis for every SNR condition. Table 5.8 (a to h) gives the relative information transmitted for the different consonantal features along with averaged over five subjects and average relative improvement across five subject with different SNR conditions. The graphical representation of these features is shown in Figure 5.22 to Figure 5.29.

Table 5.8 Experiment II - Relative Information transmitted (in %) for (a). overall, (b). continuance, (c). duration, (d). frication, (e). manner, (f). nasality, (g). place and (h).voicing

	Subject		VG	SA	MF	MJ	BB	Avg	Avg. R.I.
a) Overall	SNR: 3dB	US	100.00	100.00	97.65	96.21	88.55	96.48	
		Pr-db	100.00	100.00	100.00	100.00	86.52	97.30	0.82
		Pr-bior	100.00	100.00	100.00	100.00	91.63	98.33	1.85
		Pr-sym	100.00	100.00	100.00	100.00	88.24	97.65	1.17
	SNR: 0dB	US	100.00	97.34	100.00	98.67	82.26	95.65	
		Pr-db	100.00	97.07	100.00	100.00	90.54	97.52	1.87
		Pr-bior	100.00	100.00	100.00	100.00	87.82	97.56	1.91
		Pr-sym	100.00	100.00	100.00	100.00	87.34	97.47	1.82
	SNR: -3dB	US	100.00	98.67	100.00	92.63	81.24	94.51	
		Pr-db	97.26	93.04	100.00	100.00	91.89	96.44	1.93
		Pr-bior	100.00	96.92	100.00	100.00	86.76	96.74	2.23
		Pr-sym	100.00	96.92	100.00	100.00	84.86	96.36	1.85
	SNR: -6dB	US	100.00	98.46	100.00	96.79	79.18	94.89	
		Pr-db	98.54	92.26	100.00	98.60	82.72	94.42	-0.46
		Pr-bior	100.00	97.03	96.92	98.60	83.58	95.23	0.34
		Pr-sym	100.00	95.54	96.92	98.60	85.45	95.30	0.42

	SNR: -9dB	US	91.99	94.69	97.08	95.10	75.96	90.96	
		Pr-db	93.16	88.52	98.46	93.78	81.44	91.07	0.11
		Pr-bior	97.65	100.00	96.92	93.78	76.92	93.05	2.09
		Pr-sym	98.46	95.38	98.46	93.78	77.51	92.72	1.76

b) Continuance	Subject		VG	SA	MF	MJ	BB	Avg	Avg. R.I.
	SNR: 3dB	US	100.00	100.00	100.00	73.36	35.61	81.79	
		Pr-db	100.00	100.00	100.00	100.00	69.63	93.93	12.13
		Pr-bior	100.00	100.00	100.00	100.00	100.00	100.00	18.21
		Pr-sym	100.00	100.00	100.00	100.00	76.35	95.27	13.48
	SNR: 0dB	US	100.00	100.00	100.00	100.00	25.94	85.19	
		Pr-db	100.00	100.00	100.00	100.00	60.42	92.08	6.90
		Pr-bior	100.00	100.00	100.00	100.00	76.35	95.27	10.08
		Pr-sym	100.00	100.00	100.00	100.00	76.35	95.27	10.08
	SNR: -3dB	US	100.00	90.75	100.00	58.39	27.25	75.28	
		Pr-db	86.76	100.00	100.00	100.00	49.48	87.25	11.97
		Pr-bior	100.00	89.07	100.00	100.00	100.00	97.81	22.53
		Pr-sym	100.00	89.07	100.00	100.00	100.00	97.81	22.53
	SNR: -6dB	US	100.00	100.00	100.00	73.36	60.21	86.71	
		Pr-db	100.00	100.00	100.00	100.00	35.61	87.12	0.41
		Pr-bior	100.00	88.97	100.00	100.00	100.00	97.79	11.08
		Pr-sym	100.00	89.07	100.00	100.00	100.00	97.81	11.10
	SNR: -9dB	US	73.36	81.95	89.07	79.37	60.42	76.84	
		Pr-db	100.00	65.12	100.00	87.65	66.36	83.83	6.99
		Pr-bior	76.35	100.00	100.00	87.65	60.42	84.89	8.05
		Pr-sym	86.21	89.07	100.00	87.65	60.42	84.67	7.84

c) Duration	Subject		VG	SA	MF	MJ	BB	Avg	Avg. R.I.
	SNR: 3dB	US	100.00	100.00	100.00	79.37	100.00	95.87	
		Pr-db	100.00	100.00	100.00	100.00	65.52	93.10	-2.77
		Pr-bior	100.00	100.00	100.00	100.00	76.35	95.27	-0.60
		Pr-sym	100.00	100.00	100.00	100.00	67.93	93.59	-2.29
	SNR: 0dB	US	100.00	100.00	100.00	100.00	65.29	93.06	
		Pr-db	100.00	100.00	100.00	100.00	70.62	94.12	1.07
		Pr-bior	100.00	100.00	100.00	100.00	67.93	93.59	0.53
		Pr-sym	100.00	100.00	100.00	100.00	67.93	93.59	0.53
	SNR: -3dB	US	100.00	100.00	100.00	100.00	39.48	87.90	
		Pr-db	100.00	88.24	100.00	100.00	81.95	94.04	6.14
		Pr-bior	100.00	89.07	100.00	100.00	43.81	86.58	-1.32
		Pr-sym	100.00	89.07	100.00	100.00	38.17	85.45	-2.45
	SNR: -6dB	US	100.00	100.00	100.00	100.00	10.93	82.19	
		Pr-db	100.00	71.64	100.00	88.15	71.23	86.20	4.02
		Pr-bior	100.00	100.00	100.00	88.15	26.65	82.96	0.77

		Pr-sym	100.00	100.00	100.00	88.15	38.17	85.26	3.08
	SNR: -9dB	US	79.52	91.95	100.00	100.00	7.12	75.72	
		Pr-db	78.69	78.56	100.00	84.24	28.56	74.01	-1.71
		Pr-bior	100.00	100.00	100.00	84.24	13.89	79.63	3.91
		Pr-sym	100.00	75.53	100.00	84.24	13.89	74.73	-0.99

	Subject		VG	SA	MF	MJ	BB	Avg	Avg. R.I.
d) Frication	SNR: 3dB	US	100.00	100.00	100.00	75.00	64.90	87.98	
		Pr-db	100.00	100.00	100.00	100.00	84.57	96.91	8.94
		Pr-bior	100.00	100.00	100.00	100.00	87.69	97.54	9.56
		Pr-sym	100.00	100.00	100.00	100.00	71.50	94.30	6.32
	SNR: 0dB	US	100.00	100.00	100.00	100.00	51.26	90.25	
		Pr-db	100.00	100.00	100.00	100.00	49.88	89.98	-0.28
		Pr-bior	100.00	100.00	100.00	100.00	71.50	94.30	4.05
		Pr-sym	100.00	100.00	100.00	100.00	71.50	94.30	4.05
	SNR: -3dB	US	100.00	91.12	100.00	61.67	32.97	77.15	
		Pr-db	88.58	88.24	100.00	100.00	78.94	91.15	14.00
		Pr-bior	100.00	82.45	100.00	100.00	54.01	87.29	10.14
		Pr-sym	100.00	82.45	100.00	100.00	55.19	87.53	10.37
	SNR: -6dB	US	100.00	100.00	100.00	75.00	28.78	80.76	
		Pr-db	100.00	78.49	100.00	88.15	64.90	86.31	5.55
		Pr-bior	100.00	89.36	100.00	88.15	55.19	86.54	5.78
		Pr-sym	100.00	89.47	100.00	88.15	55.19	86.56	5.80
	SNR: -9dB	US	75.00	82.45	89.47	80.93	31.24	71.82	
		Pr-db	100.00	83.97	100.00	89.13	51.97	85.01	13.20
		Pr-bior	78.94	100.00	100.00	89.13	29.30	79.47	7.66
		Pr-sym	87.69	82.45	100.00	89.13	29.30	77.71	5.90

	Subject		VG	SA	MF	MJ	BB	Avg	Avg. R.I.
e) Manner	SNR: 3dB	US	100.00	100.00	100.00	83.91	78.42	92.47	
		Pr-db	100.00	100.00	100.00	100.00	90.38	98.08	5.61
		Pr-bior	100.00	100.00	100.00	100.00	92.30	98.46	5.99
		Pr-sym	100.00	100.00	100.00	100.00	82.41	96.48	4.01
	SNR: 0dB	US	100.00	93.54	100.00	100.00	65.11	91.73	
		Pr-db	100.00	100.00	100.00	100.00	69.46	93.89	2.16
		Pr-bior	100.00	100.00	100.00	100.00	82.41	96.48	4.75
		Pr-sym	100.00	100.00	100.00	100.00	82.41	96.48	4.75
	SNR: -3dB	US	100.00	94.17	100.00	75.78	58.15	85.62	
		Pr-db	93.10	92.97	100.00	100.00	86.93	94.60	8.98
		Pr-bior	100.00	86.10	100.00	100.00	71.66	91.55	5.93
		Pr-sym	100.00	86.10	100.00	100.00	72.09	91.64	6.02
	SNR: -6dB	US	100.00	100.00	100.00	83.91	56.40	88.06	
		Pr-db	100.00	86.77	100.00	92.93	78.42	91.62	3.56
		Pr-bior	100.00	93.02	100.00	92.93	72.09	91.61	3.54

		Pr-sym	100.00	87.70	100.00	92.93	72.09	90.54	2.48
	SNR: -9dB	US	83.91	88.48	93.09	87.80	57.77	82.21	
		Pr-db	94.95	81.78	100.00	93.31	70.25	88.06	5.85
		Pr-bior	86.93	100.00	100.00	93.31	56.53	87.35	5.14
		Pr-sym	92.30	88.48	100.00	93.31	56.53	86.12	3.91

	Subject		VG	SA	MF	MJ	BB	Avg	Avg. R.I.
f) Nasality	SNR: 3dB	US	100.00	100.00	100.00	100.00	100.00	100.00	
		Pr-db	100.00	100.00	100.00	100.00	100.00	100.00	0.00
		Pr-bior	100.00	100.00	100.00	100.00	100.00	100.00	0.00
		Pr-sym	100.00	100.00	100.00	100.00	100.00	100.00	0.00
	SNR: 0dB	US	100.00	82.40	100.00	100.00	88.62	94.20	
		Pr-db	100.00	100.00	100.00	100.00	100.00	100.00	5.80
		Pr-bior	100.00	100.00	100.00	100.00	100.00	100.00	5.80
		Pr-sym	100.00	100.00	100.00	100.00	100.00	100.00	5.80
	SNR: -3dB	US	100.00	100.00	100.00	100.00	100.00	100.00	
		Pr-db	100.00	100.00	100.00	100.00	100.00	100.00	0.00
		Pr-bior	100.00	78.94	100.00	100.00	100.00	95.79	-4.21
		Pr-sym	100.00	78.94	100.00	100.00	100.00	95.79	-4.21
	SNR: -6dB	US	100.00	100.00	100.00	100.00	100.00	100.00	
		Pr-db	100.00	100.00	100.00	100.00	100.00	100.00	0.00
		Pr-bior	100.00	100.00	100.00	100.00	100.00	100.00	0.00
		Pr-sym	100.00	86.67	100.00	100.00	100.00	97.33	-2.67
	SNR: -9dB	US	100.00	100.00	100.00	100.00	100.00	100.00	
		Pr-db	88.26	81.06	100.00	100.00	100.00	93.86	-6.14
		Pr-bior	100.00	100.00	100.00	100.00	100.00	100.00	0.00
		Pr-sym	100.00	100.00	100.00	100.00	100.00	100.00	0.00

	Subject		VG	SA	MF	MJ	BB	Avg	Avg. R.I.
g) Place	SNR: 3dB	US	100.00	100.00	100.00	86.39	74.25	92.13	
		Pr-db	100.00	100.00	100.00	100.00	61.79	92.36	0.23
		Pr-bior	100.00	100.00	100.00	100.00	63.79	92.76	0.63
		Pr-sym	100.00	100.00	100.00	100.00	54.24	90.85	-1.28
	SNR: 0dB	US	100.00	90.89	100.00	100.00	71.27	92.43	
		Pr-db	100.00	88.25	100.00	100.00	83.94	94.44	2.00
		Pr-bior	100.00	100.00	100.00	100.00	42.73	88.55	-3.89
		Pr-sym	100.00	100.00	100.00	100.00	44.90	88.98	-3.45
	SNR: -3dB	US	100.00	100.00	100.00	78.38	59.65	87.61	
		Pr-db	89.55	83.14	100.00	100.00	80.75	90.69	3.08
		Pr-bior	100.00	93.58	100.00	100.00	33.70	85.46	-2.15
		Pr-sym	100.00	93.58	100.00	100.00	34.50	85.62	-1.99
	SNR: -6dB	US	100.00	93.05	100.00	90.89	21.58	81.11	
		Pr-db	93.69	76.71	100.00	94.29	48.49	82.64	1.53

		Pr-bior	100.00	100.00	87.48	94.29	26.05	81.57	0.46
		Pr-sym	100.00	93.05	87.48	94.29	31.77	81.32	0.21
	SNR: -9dB	US	77.40	83.51	88.16	84.20	19.93	70.64	
		Pr-db	72.90	81.54	93.58	76.18	42.41	73.32	2.68
		Pr-bior	90.83	100.00	93.58	76.18	22.49	76.62	5.98
		Pr-sym	94.37	86.71	93.58	76.18	25.00	75.17	4.53

	Subject		VG	SA	MF	MJ	BB	Avg	Avg. R.I.
h) Voicing	SNR: 3dB	US	100.00	100.00	78.94	91.12	100.00	94.01	
		Pr-db	100.00	100.00	100.00	100.00	100.00	100.00	5.99
		Pr-bior	100.00	100.00	100.00	100.00	71.50	94.30	0.29
		Pr-sym	100.00	100.00	100.00	100.00	61.58	92.32	-1.70
	SNR: 0dB	US	100.00	91.12	100.00	89.69	70.13	90.19	
		Pr-db	100.00	100.00	100.00	100.00	86.94	97.39	7.20
		Pr-bior	100.00	100.00	100.00	100.00	61.58	92.32	2.13
		Pr-sym	100.00	100.00	100.00	100.00	61.58	92.32	2.13
	SNR: -3dB	US	100.00	100.00	100.00	80.93	86.94	93.57	
		Pr-db	100.00	63.38	100.00	100.00	100.00	92.68	-0.90
		Pr-bior	100.00	100.00	100.00	100.00	63.84	92.77	-0.81
		Pr-sym	100.00	100.00	100.00	100.00	64.90	92.98	-0.59
	SNR: -6dB	US	100.00	100.00	100.00	100.00	35.22	87.04	
		Pr-db	100.00	75.85	100.00	90.03	100.00	93.18	6.13
		Pr-bior	100.00	87.62	89.47	90.03	64.90	86.40	-0.64
		Pr-sym	100.00	87.69	89.47	90.03	64.90	86.42	-0.63
	SNR: -9dB	US	100.00	100.00	100.00	100.00	27.91	85.58	
		Pr-db	100.00	78.28	100.00	100.00	42.46	84.15	-1.44
		Pr-bior	78.94	100.00	87.69	100.00	33.64	80.05	-5.53
		Pr-sym	87.69	78.94	100.00	100.00	33.64	80.05	-5.53

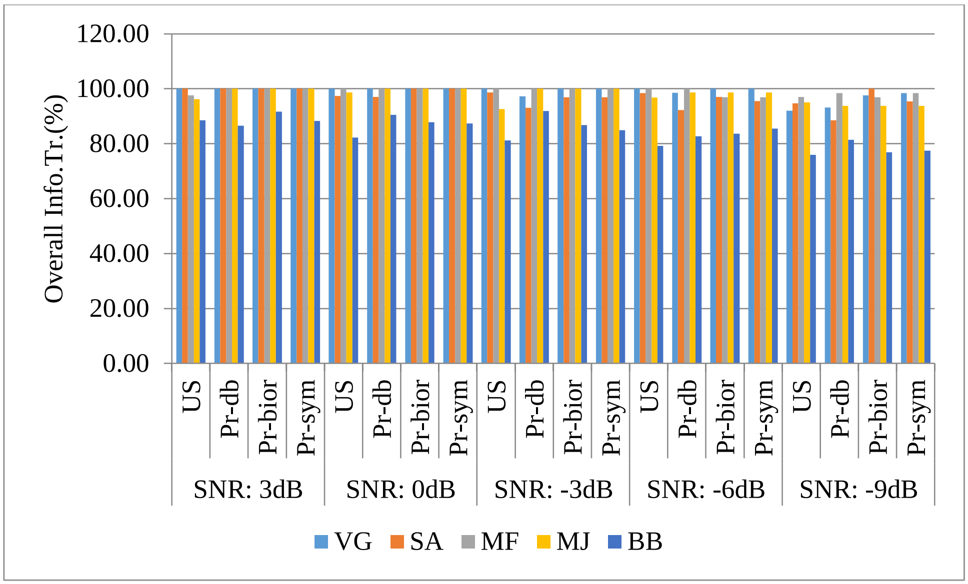


Fig. 5.22 Experiment II - Relative Information Transmitted for Overall

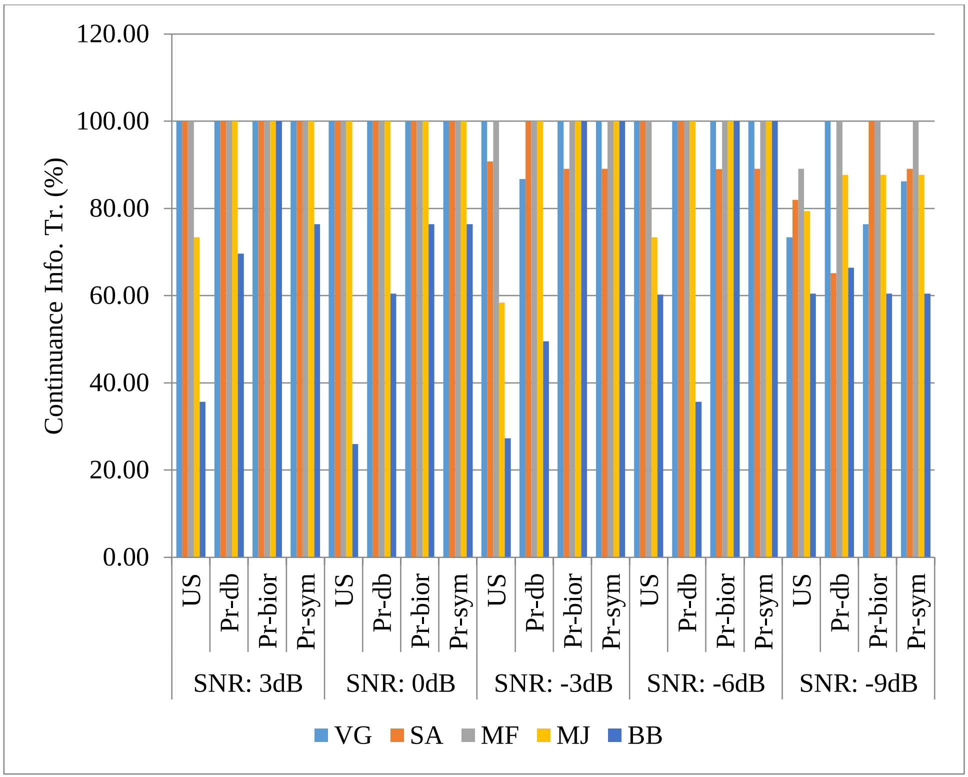


Fig. 5.23 Experiment II - Relative Information Transmitted for Continuance

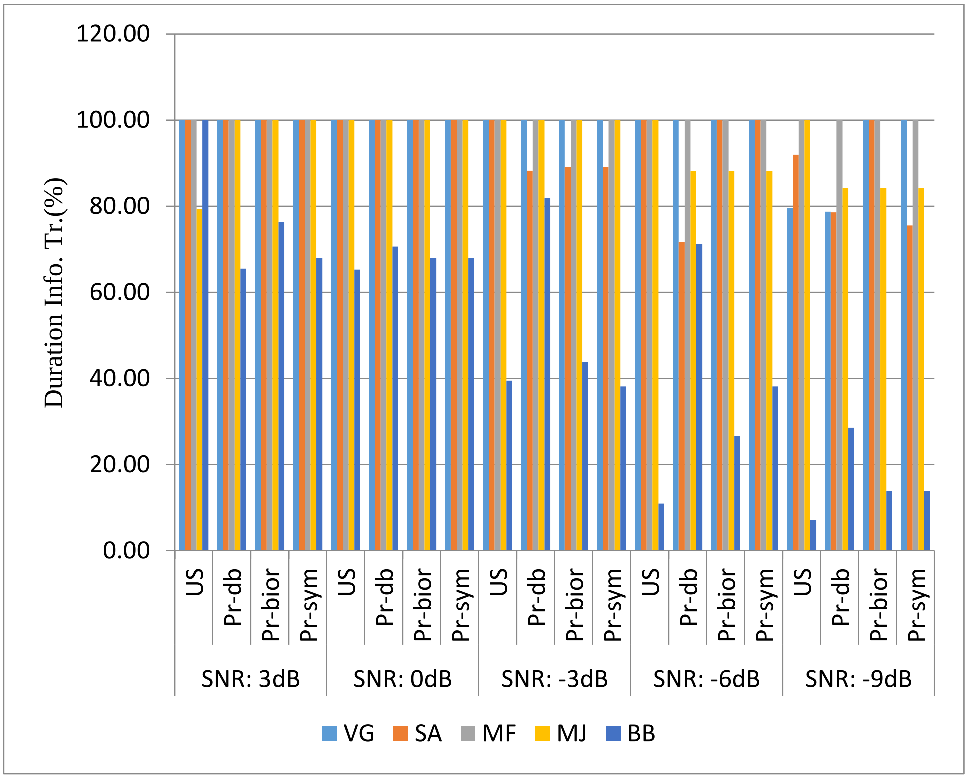


Fig. 5.24 Experiment II - Relative Information Transmitted for Duration

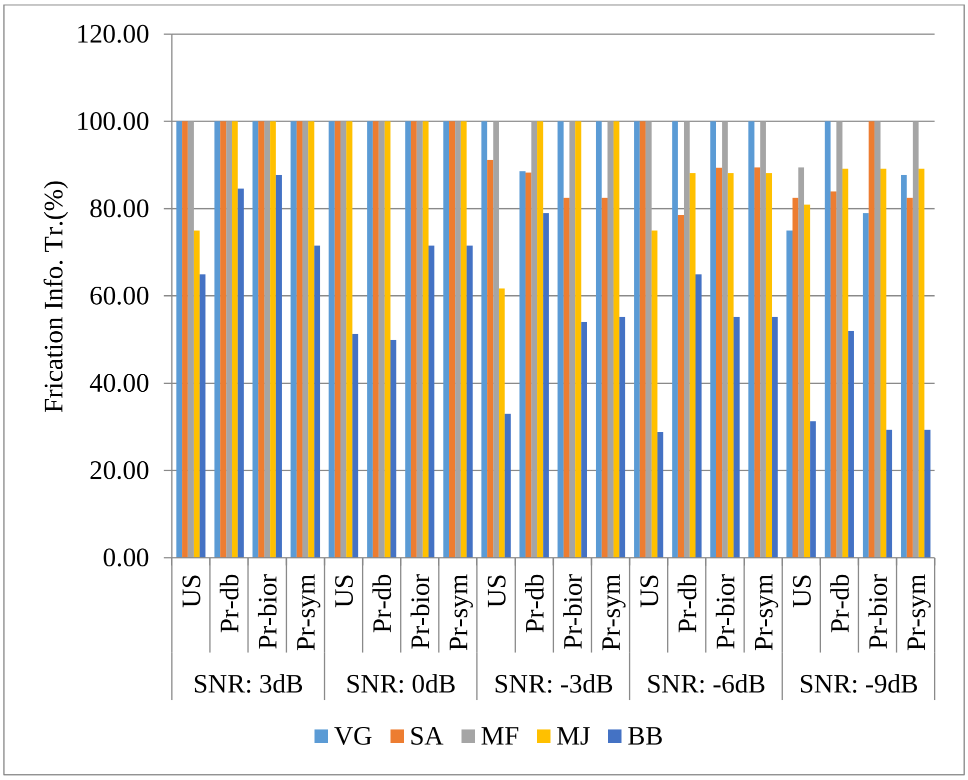


Fig. 5.25 Experiment II - Relative Information Transmitted for Frication

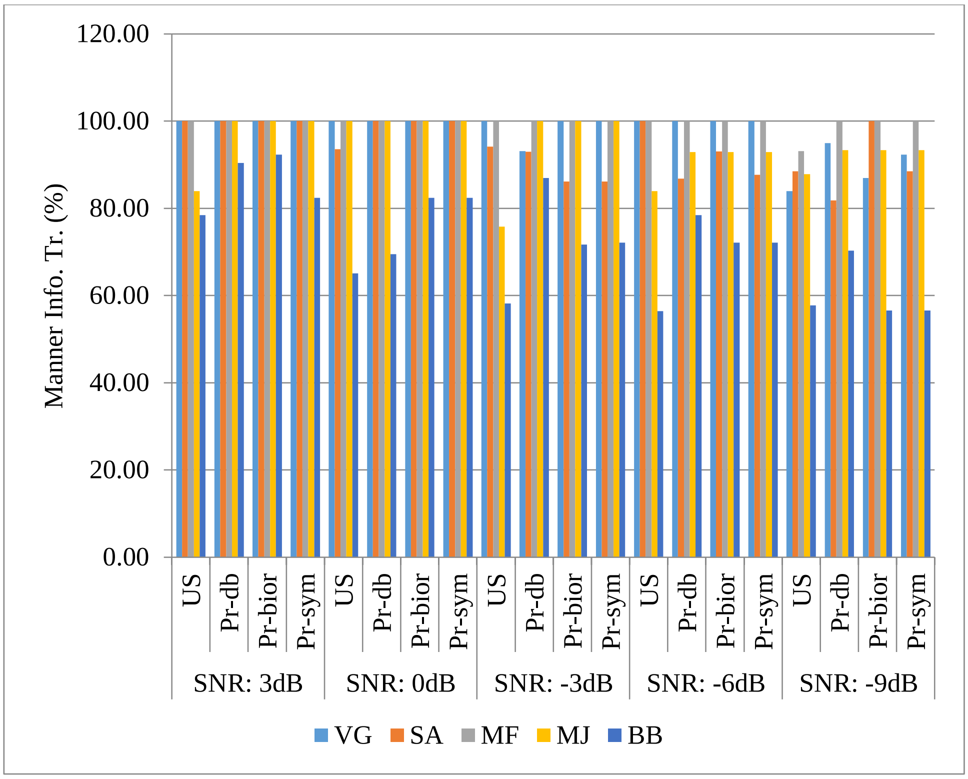


Fig. 5.26 Experiment II - Relative Information Transmitted for Manner

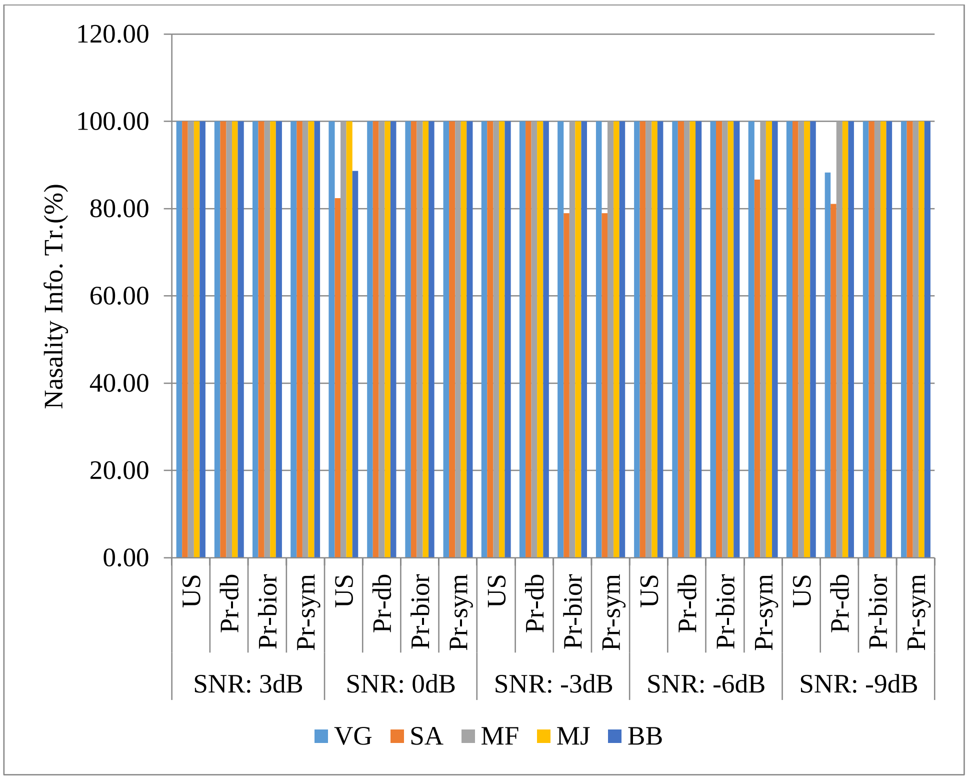


Fig. 5.27 Experiment II - Relative Information Transmitted for Nasality

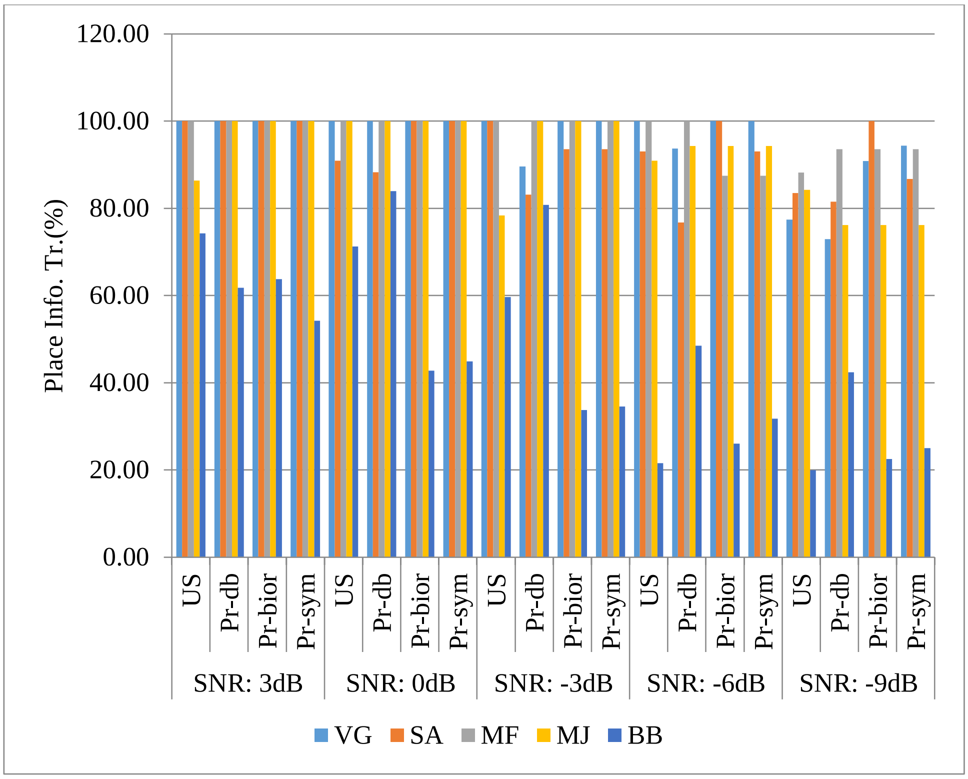


Fig. 5.28 Experiment II - Relative Information Transmitted for Place

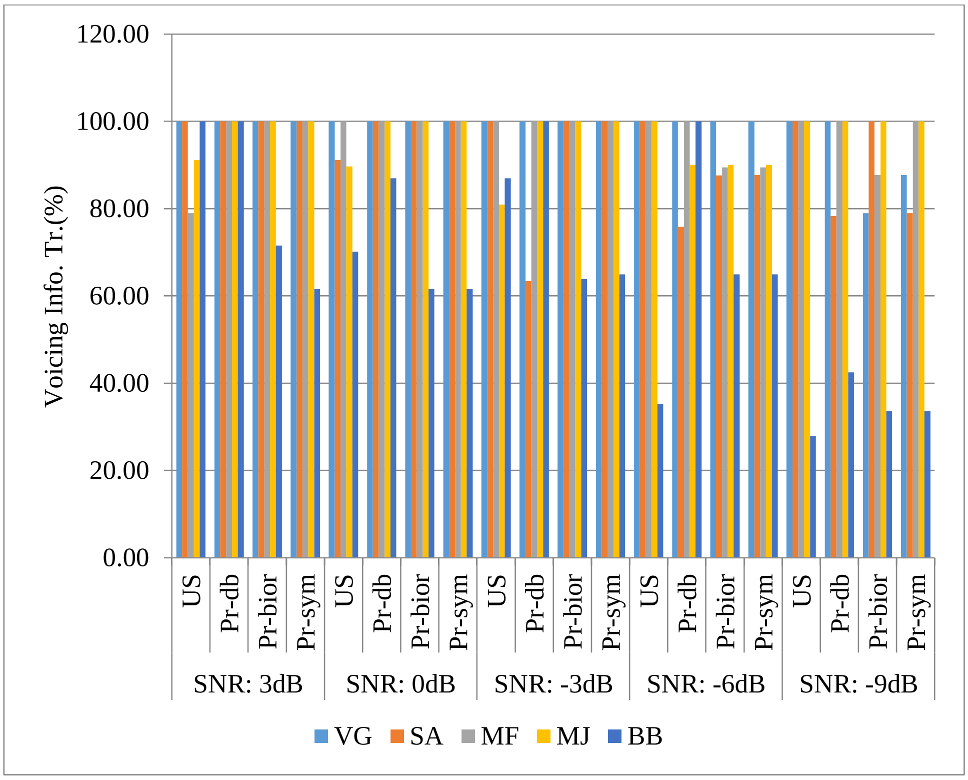


Fig. 5.29 Experiment II - Relative Information Transmitted for Voicing

Overall: With decrease in SNR condition related to unprocessed speech there was degradation in the overall information transmitted. For SNR conditions of 3, 0, -

3, -6 and -9dB the values of overall information transmitted were for db algorithms 97.30, 97.52, 96.44, 94.42, 91.07% similarly, the values were 98.33, 97.56, 96.74, 95.23 and 93.05% for bior and 97.65, 97.47, 96.36, 95.30 and 92.72% for sym respectively.

Continuance: The continuance was least affected feature among other features. The average relative improvement improved as the SNR reduced from 3dB to -9dB. For db algorithm, the values observed were 83.83, 87.12, 87.25, 92.08, 93.93% as masking of noise reduced i.e. for SNR -9, -6,-3, 0 and 3 respectively. Similarly an improvement in continuance feature was observed for bior and sym algorithms. Of these three algorithms, bior algorithm shows more average relative improvement over other algorithms.

Duration: As the SNR value reduced from 3dB to -9dB, the recognition score of unprocessed speech changed from 95.87% to 75.72%. For SNR conditions of 3, 0, -3, -6 and -9dB the values of information transmitted were for db algorithms 93.10, 94.12, 94.04,86.20 and 74.01% similarly, the values were 95.27, 93.59,86.58,82.96 and 79.63 % for bior and 93.59,93.59,85.45,85.26 and 74.73 % for sym respectively.

Frication: For this feature, the relative information transmitted for unprocessed speech ranged from 75% to 100%, as masking noise reduced (-9dB to 3 dB). The average relative increase in information transmitted in frication at 3dB of db, bior and sym over unprocessed signals was 5.61, 5.99 and 4.01 respectively while for -9dB it was found to be 5.85, 5.25 and 3.91%. The relative information transmitted for unprocessed speech almost remained same across all SNR conditions among all subjects.

Manner: The relative information transmitted for manner features with SNR conditions 3dB was 92.47% for unprocessed and SNR condition -9dB was 82.21%. For SNR conditions of 3, 0, -3, -6 and -9dB the values of overall information transmitted were for db algorithms 98.08, 93.89, 94.60, 91.62, 88.06% similarly, the values were 98.46,96.48, 94.60, 91.61 and 87.35% for bior and 96.48, 96.48, 91.64, 90.54 and 86.12% for sym respectively.

Nasality: The average relative improvement in this feature is either reduced or has no change for three algorithms except at 0dB SNR and was found to be 5.80% for all three algorithms. At 3dB SNR, the relative information transmitted for this feature was almost perfect.

Place: For unprocessed speech the relative information transmitted was 92.13% for 3dB and it reduced to 70.64% as masking level increased (SNR -9dB.). For db algorithm the average relative improvement in place feature 0.25, 2, 3.08, 1.53 and 2.68 for SNR condition 3, 0,-3, -6, and -9dB.respectively.The relative improvement was found maximum for the bior and sym as 5.98 and 4.53 respectively at -9dB SNR. Most of the subjects have shown improvement in place feature as SNR reduces.

Voicing: For voicing features with SNR condition 3dB, the information transmitted was 100, 94.30, 92.32% for processed (db, bior and sym) signal and it decreased to 84.15, 80.05 and 80.05% for SNR -9dB.The maximum relative improvement was found in db algorithm.

For the experimental analysis, it was observed that the algorithms improved the perception of most of the consonantal features with different SNR conditions. The relative information transmitted is near perfect with unprocessed speech and improves with processed speech for higher value of SNR. With lower values of SNR, the relative information transmitted with unprocessed speech decreases and improvements are observed with the processed speech. However, most of the subjects indicated the maximum improvement for the duration, frication and manner features. For lower values of SNR, relatively better improvement was observed for place feature. The reception of the place feature is related to frequency resolving capacity of the auditory processing so it can be implied that the algorithms have reduced the effect of spectral masking.

5.9 Experiment III: Listening tests on hearing impaired subjects (software based offline)

The objective of this experiment is to assess the usefulness of the developed algorithms to improve the perception of hearing impaired subjects. This software based offline experiment is referred as Experiment III. The assessment was carried out by conducting the listening tests on eight hearing impaired subjects with different degree of hearing loss. The experimental procedure for software based offline process and the result of listening tests is discussed in following subsections.

5.9.1 Experiment III: Procedure

Appendix B contains the description of software based offline experimental procedure. For the subjects presentations were completed at their comfortable listening level. The explanation regarding the procedure of listening test and test instruction given to hearing impaired subjects is explained in detail in Appendix C.

To familiarize the subjects with the stimuli, they could listen to the test material frequently as per their wish. Listening test included a presentation of fifteen stimulus items for five times in random manner, thus giving seventy five presentations for every subject. For binaural presentation of the test stimuli, a laptop based experimental setup (MATLAB based GUI) was used as described in Appendix B. The stimuli were displayed on subjects' laptop screen along with appropriate push button when it was presented over the headphones. The subjects' responded to the stimuli by pressing the appropriate push button. For every presentation response time of the subject was noted. A combination of stimuli along the rows with responses along the columns was used to obtain a stimulus response confusion matrix. Each entry in the cell corresponds to the stimulus-response pair showing error by the off-diagonal elements, while giving correct responses along the diagonal elements. The percentage correct recognition score was given by summation of the diagonal elements form the generated confusion matrix. The response time statics was also made available in the confusion matrix format.

Based on the accessibility and willingness of the subjects, a period of four to five months was necessary for the conduction of test session. Each subject took about one to two hour for the compilation of the test under different test conditions. Subjects' qualitative evaluation of the test stimuli is combined for various listening conditions to understand the speech qualitative analysis.

To compare the load on perception, response times were used. The confusion matrix obtained was used to evaluate the recognition score. The information transmission analysis was also carried out as it provides a measure independent of subject's biasing for response (as described in Appendix A). Features like voicing, nasality, frication, place, manner, duration, and continuance were used to group confusion matrices.

5.9.2 Experiment III: Listening test Results

The following subsection discusses the results of listening tests evaluated by software based offline experiment (Experiment III) on hearing impaired subjects. It includes mean opinion score, response times, recognition scores (in %), relative improvement in response times and recognition scores, information transmission analysis for unprocessed and processed speech signals for each hearing impaired subjects.

5.9.2.1 Experiment III: Qualitative assessment

This parameter gives quality of the unprocessed and processed speech material when presented to the hearing impaired subjects. The pre-recorded test material in VCV context was heard four times by subjects. On the basis of quality of sound, they were asked to give the rating as ‘Outstanding’, ‘Good’, ‘Fair’, ‘Average’ and ‘Below Average'. These ratings were indexed from 1 to 5; ‘1’ being ‘Below Average’ and 5 being ‘Outstanding’. Averages of these ratings were computed to find mean opinion score of each subject. Results of qualitative assessment of eight subjects are given in Table 5.9. It can be observed that, five subjects APL, KSH, BHL, PDH and PNK ranked the quality of processed signal as higher than the unprocessed signal for all three algorithms. Subject KUR ranked the processed signal same as unprocessed signal. Graphical representation of mean opinion score of each subject is shown in Figure 5.30.

Table 5.9 Experiment III - Qualitative assessment

Subjects	US	PS-db	PS-bior	PS-sym
APL	3.75	4.00	4.25	4.00
KUR	5.00	5.00	5.00	5.00
KSH	2.75	3.75	4.00	4.00
BHL	3.00	4.00	4.00	3.75
PDH	3.25	4.25	4.25	4.00
PNK	4.50	5.00	5.00	4.50
BHV	2.50	3.50	3.50	2.50
SWK	3.50	4.00	3.75	3.50

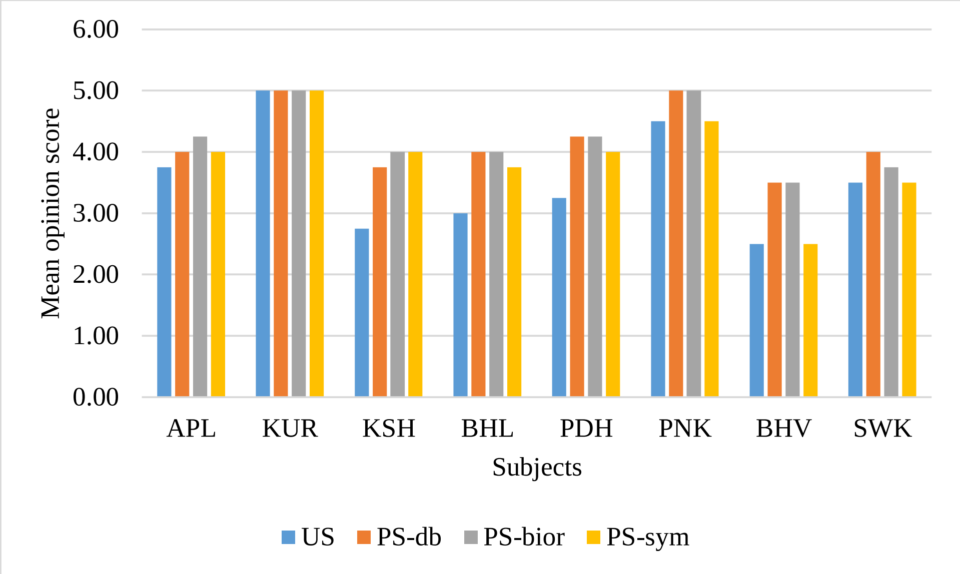


Fig. 5.30 Experiment III - Qualitative assessment

5.9.2.2 Response time

Response time is defined as time interval between speech materials presented dichotically to subjects and the response given by them. It reduces significantly showing reduction in load on perception process. Response time for eight subjects is given in Table 5.10 and graphically represented in Figure 5.31. It shows variation from 4.08 seconds to 8.6 seconds for unprocessed signal and minimum 3.71 seconds to maximum 8.58 seconds across all algorithms for processed signal. There was reduction in response time for processed signals for almost all hearing impaired subjects across all the algorithms. The relative decrease was from -22.50 to 27.59 %, -1.18 to 44.21% and-7.35 t o46.92% for db, bior and sym algorithms respectively. Relative decrease in response time is significant for the subjects KUR and KSH as shown in Table 5.11 and is graphically presented in Figure 5.32. One of the subjects, PNK showed relative increase in response time.

Table 5.10 Experiment III - Response time

Subjects	US	PS-db	PS-bior	PS-sym
APL	5.82	5.03	5.13	5.08
KUR	6.99	4.37	3.90	3.71
KSH	7.25	5.25	4.75	5.19
BHL	4.83	4.56	4.03	4.91
PDH	5.51	5.26	5.19	5.67
PNK	4.08	5.00	4.13	4.38
BHV	8.60	7.95	8.20	8.58
SWK	5.30	4.58	4.80	4.90

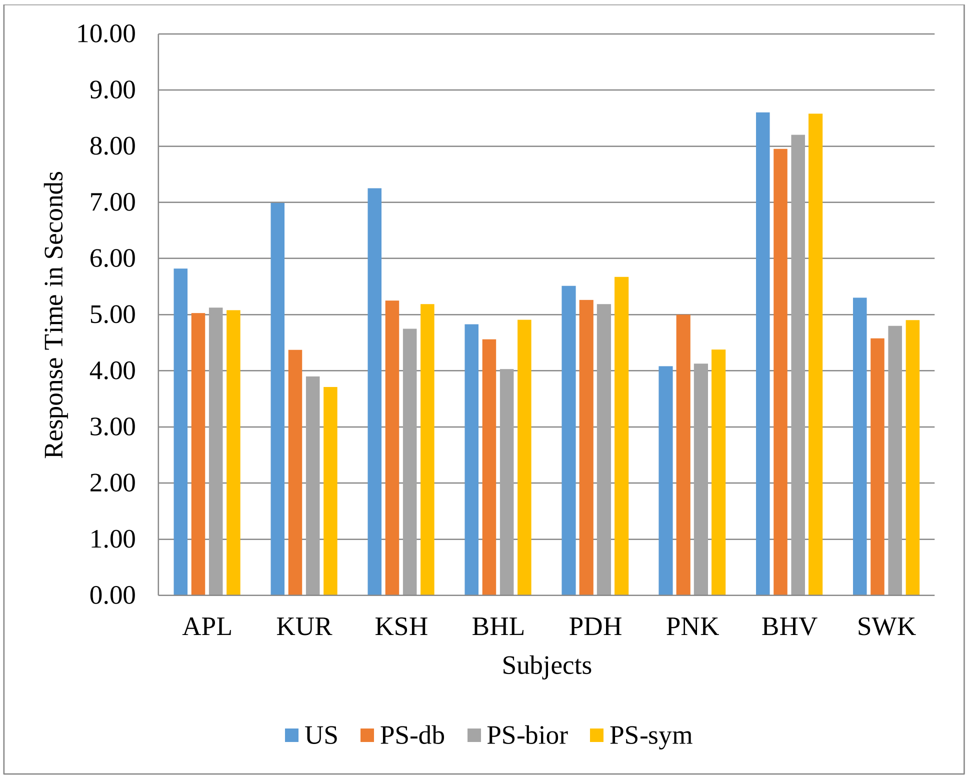


Fig. 5.31 Experiment III - Response time

Table 5.11 Experiment III - Relative Decrease in Response time

Subjects	PS-db	PS-bior	PS-sym
APL	13.57	11.94	12.71
KUR	37.48	44.21	46.92
KSH	27.59	34.48	28.41
BHL	5.59	16.56	-1.66
PDH	4.54	5.81	-2.90
PNK	-22.55	-1.18	-7.35
BHV	7.56	4.65	0.23
SWK	13.58	9.43	7.55

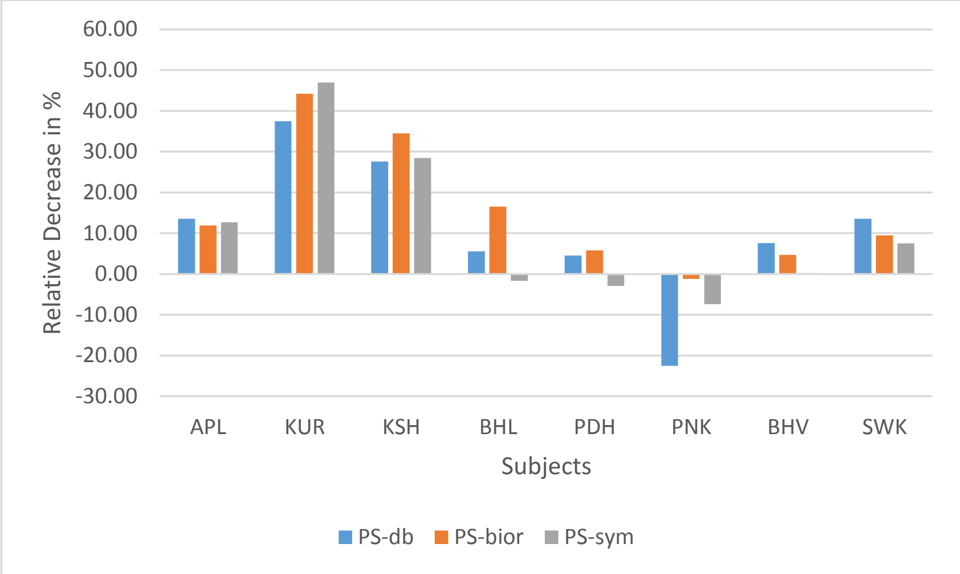


Fig. 5.32 Experiment III - Relative Decrease in Response time

5.9.2.3 Recognition scores

The recognition scores, percentage relative improvement, and averages of recognition score were obtained from the confusion matrix as shown in Table 5.12. Figure 5.33 shows percentage recognition scores for unprocessed and processed signals of all three algorithms and percentage relative improvement in recognition scores. For the impaired subjects, recognition score was low and varied from 53.33% to 90.00% for unprocessed signal while for processed signal the recognition score increased from 68.88% to 93.33% for db, 53.33% to 93.33% for bior and 60.00% to 97.77% for sym. The relative improvement in the recognition scores for db, bior and sym algorithms were ranged from 0 to 25%, 0 to 13.33% and 0 to 17.78% respectively. The maximum relative improvement among all the algorithms was found in db. In db algorithm, almost all the subjects showed better improvement in recognition score except PNK subject which showed no improvement at all.

Three subjects (APL, KSH and BHL) having mild to severe frequency impairment have shown maximum relative improvement in recognition score and one subject (PNK) with symmetrical high frequency hearing impairment have shown no relative improvement in db and sym while a slight relative improvement in recognition score is been seen in bior algorithm.

Table 5.12 Experiment III - Recognition scores and its Relative improvement (in %)

Subjects	US	PS-db	RI-db	PS-bior	RI-bior	PS-sym	RI-sym
APL	53.33	68.88	15.55	66.66	13.33	66.66	13.33
KUR	84.44	93.33	8.89	93.33	8.89	97.77	13.33

KSH	48.33	73.33	25.00	56.66	8.33	60.00	11.67
BHL	68.88	86.66	17.78	80.00	11.12	86.66	17.78
PDH	61.10	73.33	12.23	70.00	8.90	70.00	8.90
PNK	90.00	90.00	0.00	91.66	1.66	90.00	0.00
BHV	53.33	57.77	4.44	53.33	0.00	60.00	6.67
SWK	80.00	90.00	10.00	80.00	0.00	86.66	6.66
Averages	67.43	79.16	11.74	73.96	6.53	77.22	9.79

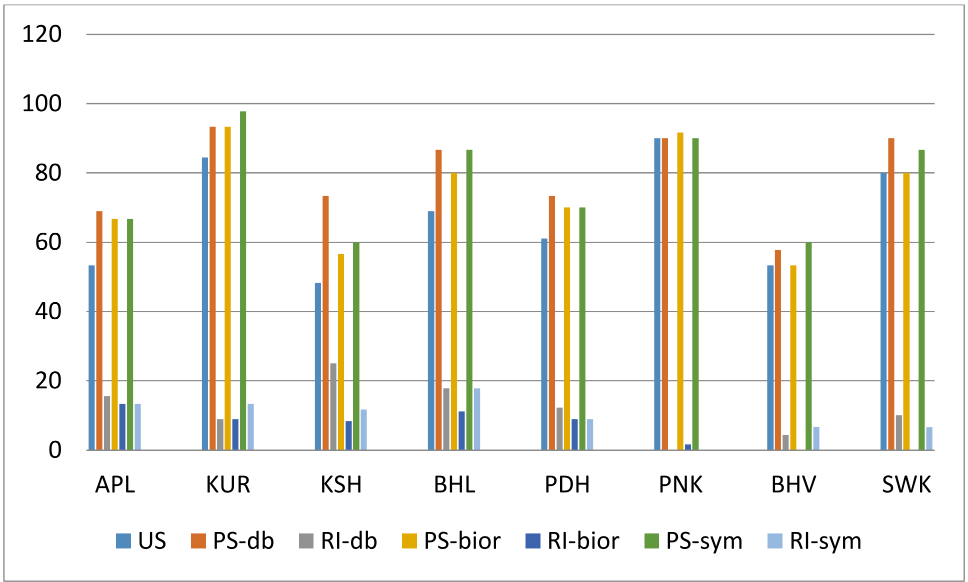


Fig. 5.33 Experiment III - Recognition scores and its Relative improvement (in %)

5.9.2.3 Information transmission analysis

Combined confusion matrices of each subject were used to evaluate information transmission analysis of all algorithms. Relative information transmitted for consonantal features of all algorithms, its relative improvement and average information transmitted for all algorithms is shown in Table 5.13 (a to h) and plotted in Figure 5.34 to Figure 5.41.

Table 5.13 (a to h) Experiment III - Information Transmission Analysis of consonantal features and Relative improvement (RI) in percentage

a) Feature: Overall							
Subject	US	PS-db	RI-db	PS-bior	RI-bior	PS-sym	RI-sym
APL	71.15	80.60	9.45	79.95	8.80	84.76	13.60
KUR	89.77	94.74	4.98	94.47	4.70	98.16	8.39
KSH	62.00	81.57	19.57	78.91	16.91	84.42	22.42
BHL	84.30	93.18	8.88	90.83	6.53	91.25	6.94
PDH	74.69	84.45	9.76	83.28	8.59	81.82	7.13
PNK	94.24	94.24	0.00	95.05	0.81	92.95	-1.29

BHV	74.19	70.61	-3.58	70.05	-4.14	74.19	0.00
SWK	89.77	94.24	4.47	87.32	-2.45	91.89	2.12
Averages	80.01	86.71	6.69	84.98	4.97	87.43	7.42

b) Feature: Continuance							
Subject	US	PS-db	RI	PS-bior	RI	PS-sym	RI
APL	33.15	33.15	0.00	35.61	2.46	23.93	-9.22
KUR	41.12	82.71	41.59	70.62	29.50	100.00	58.88
KSH	26.02	76.35	50.33	35.83	9.81	76.35	50.33
BHL	39.84	100.00	60.16	81.95	42.12	76.35	36.52
PDH	42.53	44.41	1.88	35.61	-6.92	38.36	-4.17
PNK	44.43	60.42	15.99	53.58	9.15	100.00	55.57
BHV	11.42	31.09	19.66	29.10	17.67	35.61	24.18
SWK	60.42	76.35	15.93	47.26	-13.16	60.42	0.00
Averages	37.37	63.06	25.69	48.69	11.33	63.88	26.51

c) Feature: Duration							
Subject	US	PS-db	RI	PS-bior	RI	PS-sym	RI
APL	41.12	39.48	-1.64	59.28	18.16	59.28	18.16
KUR	58.00	82.71	24.71	86.49	28.49	100.00	42.00
KSH	41.29	100.00	58.71	35.61	-5.68	44.43	3.14
BHL	41.16	71.23	30.07	71.23	30.07	81.95	40.79
PDH	76.74	100.00	23.26	100.00	23.26	89.07	12.32
PNK	76.35	100.00	23.65	89.07	12.72	100.00	23.65
BHV	44.43	60.42	15.99	60.42	15.99	69.13	24.70
SWK	100.00	100.00	0.00	72.02	-27.98	100.00	0.00
Averages	59.89	81.73	21.84	71.77	11.88	80.48	20.60

d) Feature: Frication							
Subject	US	PS-db	RI	PS-bior	RI	PS-sym	RI
APL	26.46	21.37	-5.09	48.96	22.50	32.75	6.29
KUR	53.77	84.57	30.80	84.57	30.80	100.00	46.23
KSH	21.17	53.35	32.18	26.43	5.26	53.35	32.18
BHL	38.72	100.00	61.28	82.45	43.72	62.18	23.45
PDH	57.73	73.88	16.14	64.90	7.16	61.58	3.85
PNK	48.96	64.90	15.93	64.90	15.93	100.00	51.04
BHV	10.52	45.53	35.01	23.72	13.20	45.53	35.01
SWK	64.90	78.94	14.05	53.35	-11.54	64.90	0.00
Averages	40.28	65.32	25.04	56.16	15.88	65.04	24.76

e) Feature: Manner							
Subject	US	PS-db	RI	PS-bior	RI	PS-sym	RI
APL	40.45	42.83	2.38	55.85	15.40	45.55	5.10
KUR	70.73	90.38	19.65	90.38	19.65	100.00	29.27
KSH	40.37	57.04	16.67	48.23	7.86	50.94	10.57
BHL	50.54	86.88	36.34	79.89	29.35	75.97	25.43
PDH	64.02	73.92	9.91	66.16	2.14	63.48	-0.53
PNK	68.05	78.42	10.37	78.42	10.37	100.00	31.95
BHV	15.55	53.65	38.10	29.11	13.56	56.14	40.59
SWK	78.42	86.93	8.51	71.52	-6.90	78.42	0.00
Averages	53.52	71.26	17.74	64.94	11.43	71.31	17.80

f) Feature: Nasality							
Subject	US	PS-db	RI	PS-bior	RI	PS-sym	RI
APL	57.10	74.54	17.44	67.57	10.47	67.57	10.47
KUR	100.00	100.00	0.00	100.00	0.00	100.00	0.00
KSH	71.40	59.41	-11.98	78.76	7.36	44.91	-26.49
BHL	67.57	67.57	0.00	78.76	11.18	100.00	32.43
PDH	74.54	74.54	0.00	67.57	-6.96	67.57	-6.96
PNK	100.00	100.00	0.00	100.00	0.00	100.00	0.00
BHV	20.52	67.57	47.05	24.48	3.96	74.54	54.01
SWK	100.00	100.00	0.00	100.00	0.00	100.00	0.00
Averages	73.89	80.45	6.56	77.14	3.25	81.82	7.93

g) Feature: Place							
Subject	US	PS-db	RI	PS-bior	RI	PS-sym	RI
APL	27.64	32.32	4.68	39.52	11.88	36.45	8.81
KUR	61.11	76.98	15.88	85.53	24.42	91.34	30.23
KSH	18.52	50.51	31.99	32.50	13.98	28.51	9.99
BHL	52.58	80.75	28.17	65.71	13.13	80.75	28.17
PDH	79.67	92.03	12.36	88.27	8.59	88.27	8.59
PNK	89.30	89.30	0.00	93.05	3.76	82.73	-6.57
BHV	25.06	32.63	7.57	41.97	16.91	28.90	3.84
SWK	78.11	78.11	0.00	63.02	-15.09	88.27	10.15
Averages	54.00	66.58	12.58	63.70	9.70	65.65	11.65

h) Feature: Voicing							
Subject	US	PS-db	RI	PS-bior	RI	PS-sym	RI
APL	100.00	78.50	-21.50	100.00	0.00	100.00	0.00

KUR	71.86	86.94	15.08	71.86	0.00	100.00	28.14
KSH	63.98	100.00	36.02	78.94	14.96	82.45	18.47
BHL	86.94	100.00	13.06	100.00	13.06	64.90	-22.04
PDH	37.01	65.86	28.85	71.68	34.67	60.13	23.12
PNK	100.00	100.00	0.00	100.00	0.00	82.45	-17.55
BHV	21.80	26.46	4.66	32.75	10.95	47.63	25.82
SWK	71.68	100.00	28.32	100.00	28.32	78.94	7.26
Averages	69.16	82.22	13.06	81.90	12.75	77.06	7.90

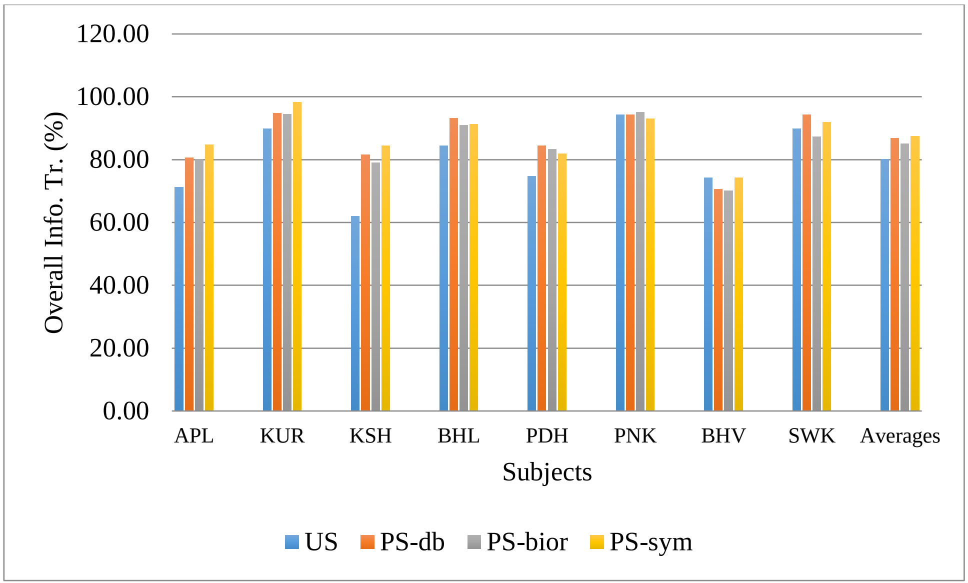


Fig. 5.34 Experiment III - Relative Information transmitted for Overall

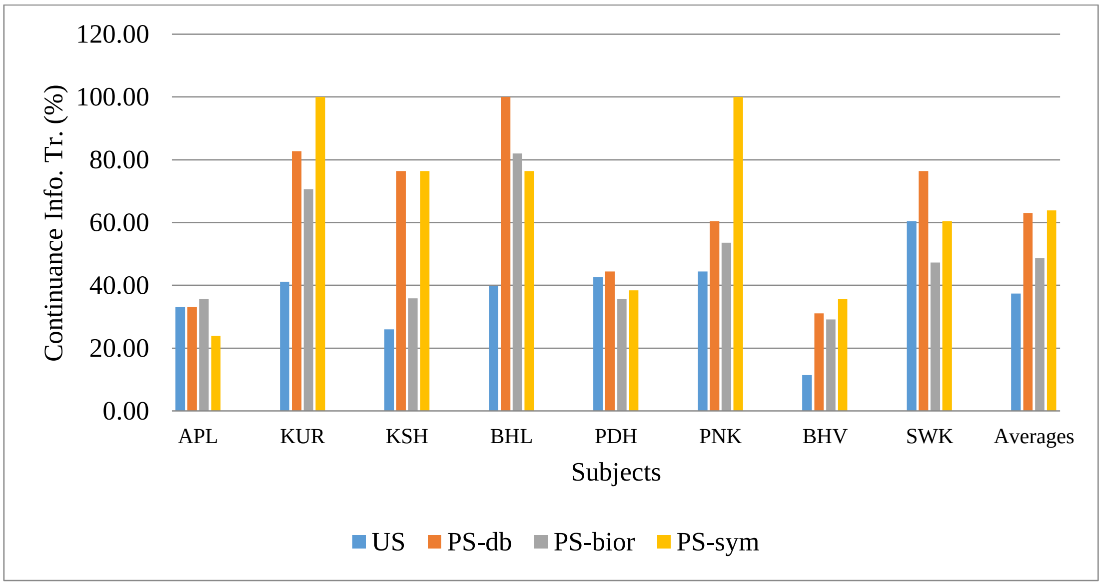


Fig. 5.35 Experiment III - Relative Information transmitted for Continuance

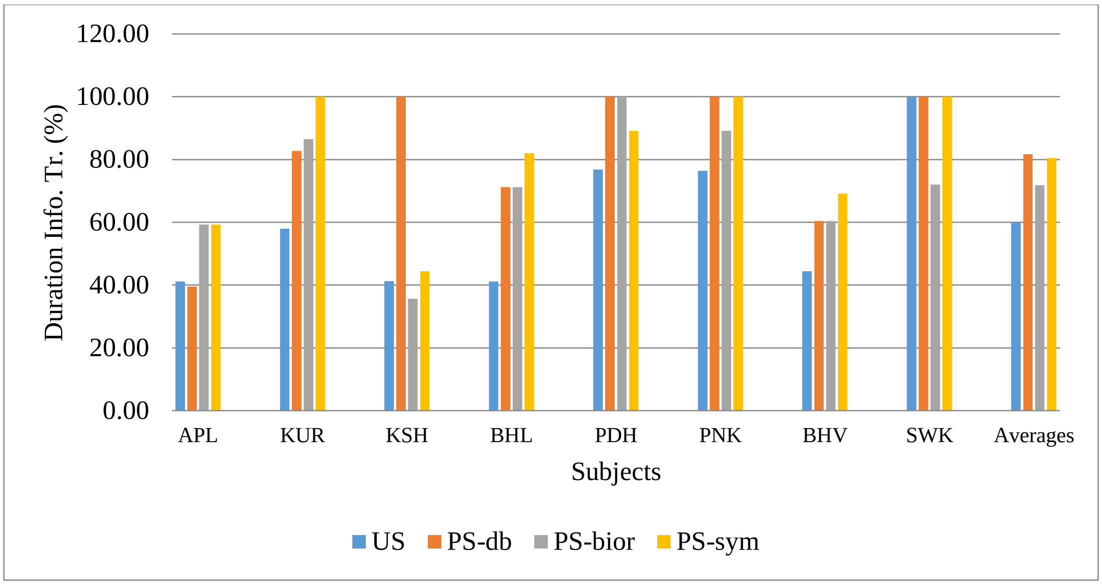


Fig. 5.36 Experiment III - Relative Information transmitted for Duration

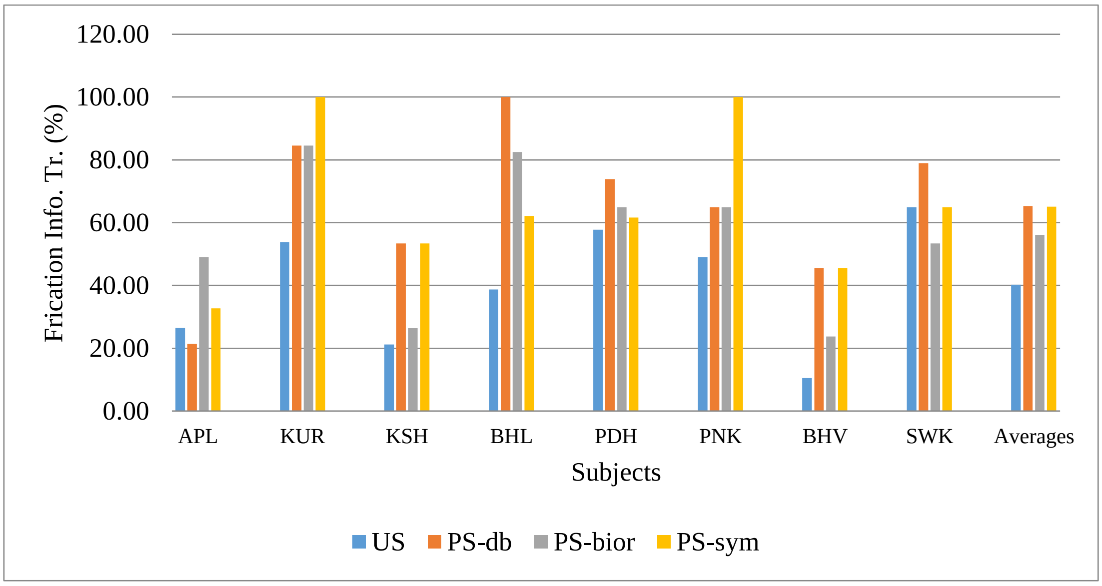


Fig. 5.37 Experiment III - Relative Information transmitted for Frication

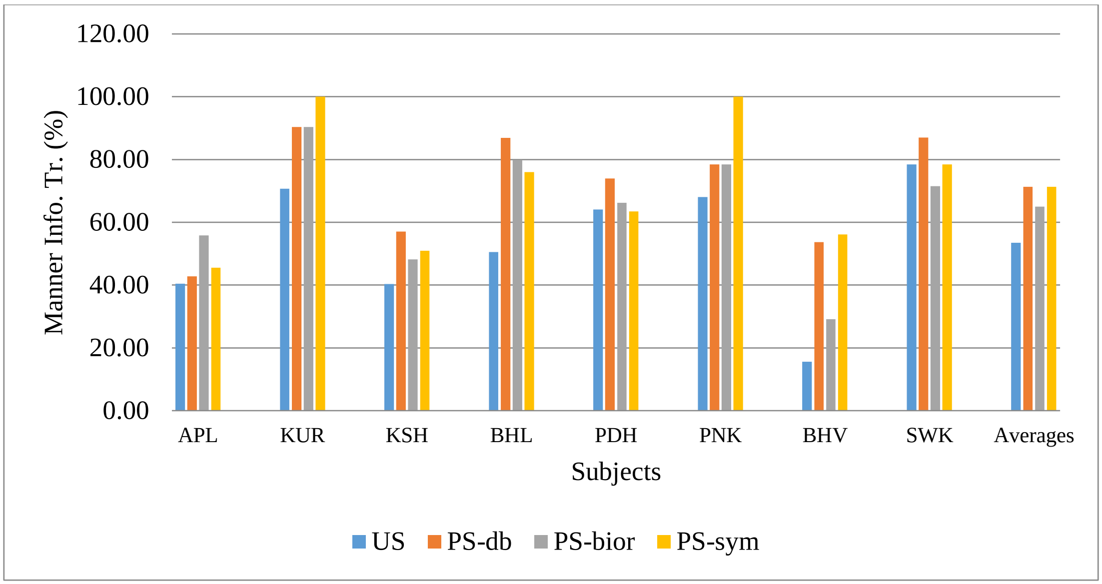


Fig. 5.38 Experiment III - Relative Information transmitted for Manner

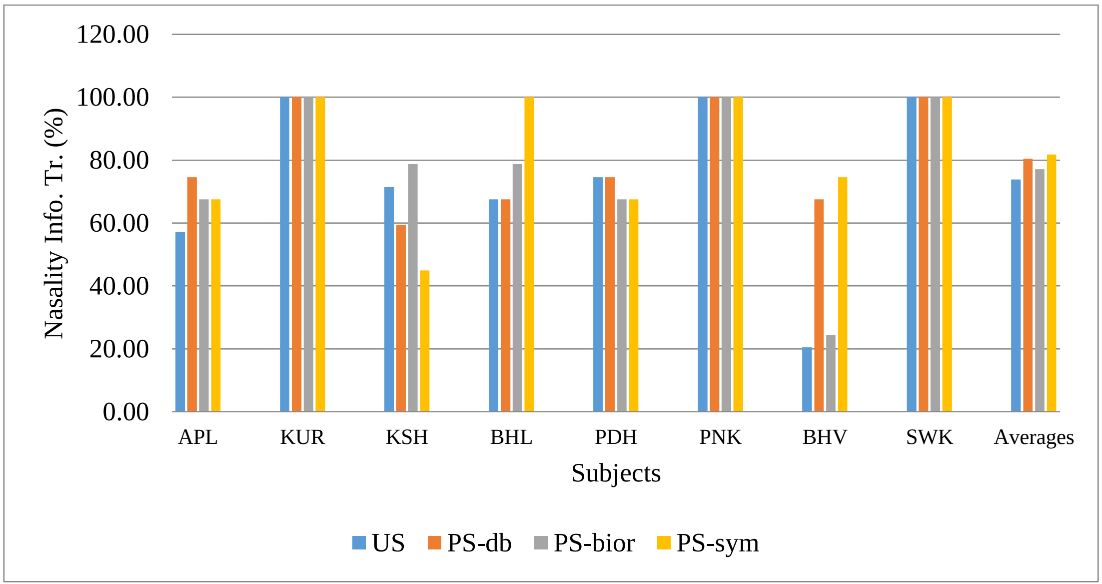


Fig. 5.39 Experiment III - Relative Information transmitted for Nasality

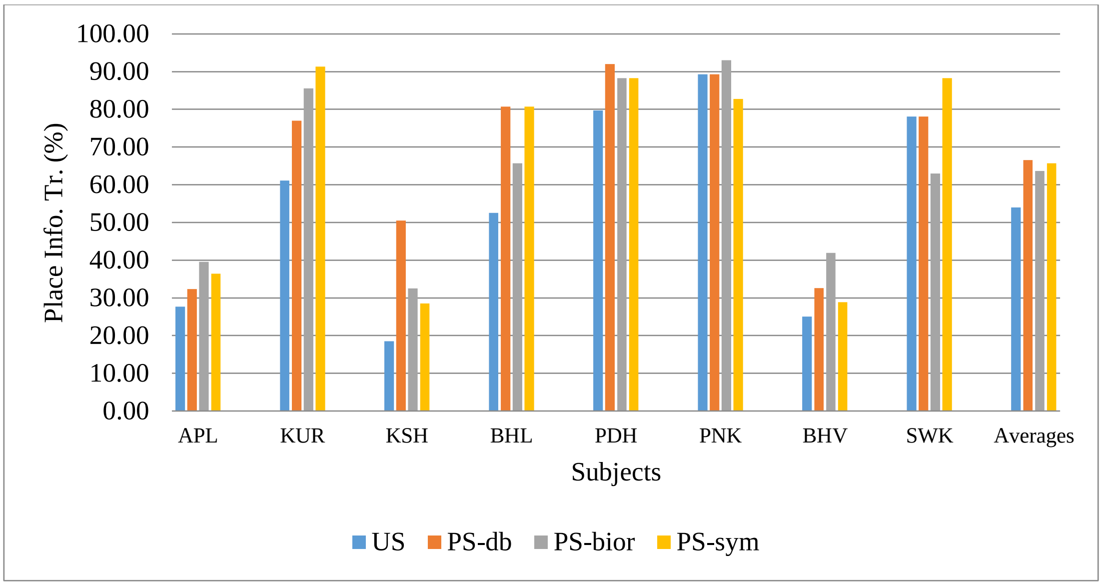


Fig. 5.40 Experiment III - Relative Information transmitted for Place

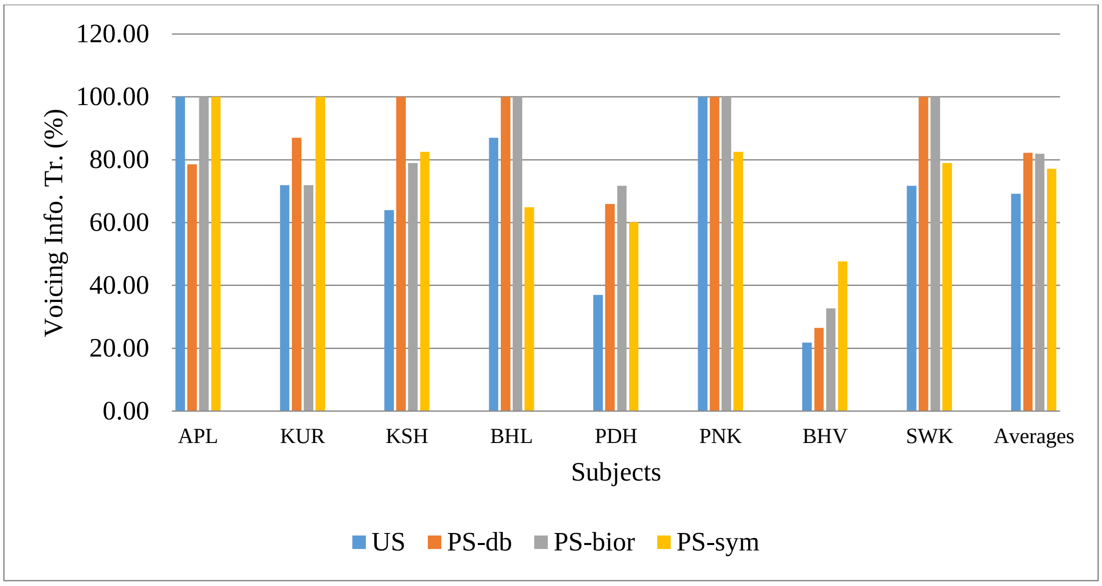


Fig. 5.41 Experiment III - Relative Information transmitted for Voicing

Overall: The relative information transmitted for overall feature for all the subjects is shown in Figure 5.34. The overall information transmitted for unprocessed signal varied from 62% to 94.24%. Its relative improvement varied from -3.58 to 19.57%, -4.14 to 16.95% and -1.29 to 22.42% for db, bior and sym algorithms respectively.

Continuance: The relative information transmitted for continuance feature is shown in Figure 5.35. The information transmitted for this feature for unprocessed signal varied from 11.42 to 60.42%. The relative improvements are high for subjects KUR, KSH and BHL for all algorithms.

Duration: The relative information transmitted for duration feature is shown in Figure 5.36. The relative information transmitted for this feature for unprocessed signal varied from 41.16 to 100%. Its relative improvement varied from -1.64 to 58.71%, -27.98 to 30.07% and 0 to 42% for db, bior and sym algorithms respectively. The algorithm db shows the improvement in relative information transmitted for this feature among all the subjects except APL and SWK.

Frication: Figure 5.37 shows the relative information transmitted for frication feature. For unprocessed signal, information transmitted for this feature varied from 10.52 to 64.90%. For the subject BHL, relative improvement was highest for db and bior algorithm and was found to be 61.28% and 43.72% respectively, while PNK subject shows highest relative improvement in algorithm sym and was found to be 51.04%.

Manner: Figures 5.38 show the relative information transmitted for manner feature. For unprocessed signal, information transmitted for this feature varied from 15.55% to 78.42%. Its relative improvement varied from 2.38 to 38.10%, -6.90 to 29.35% and -.0.53 to 40.00% for db, bior and sym algorithms respectively. Subjects BHL and BHV showed the maximum relative improvement for all the algorithms over other subjects.

Nasality: The relative information transmitted for this feature is shown in Figure 5.39. The relative information transmitted for this feature for unprocessed signal varied from 20.52 to 100%. Subject BHV showed highest relative improvement for db and sym algorithms and were 47.05and 54.01 % respectively. Subjects KUR, PNK and SWK showed no improvement for all the algorithms.

Place: The relative information transmitted for place feature is shown in Figure 5.40. For unprocessed signal, information transmitted for this feature varied from 18.52 to 89.30%. All subjects, except SWK and PNK show maximum relative improvement for all the algorithms. Subject KSH showed 31.99 % relative improvement in db algorithm while subject KUR showed 24.42 % and 30.23 % relative improvement in bior and sym algorithms respectively.

Voicing: Figure 5.41 shows the relative information transmitted for voicing feature. For unprocessed signal, information transmitted for this feature varied from 21.80 to 100%. The relative improvement varied from -21.50 to 36.02% for db, 0 to 34.67% for bior and -22.04 to 28.14% for sym. For this feature subjects KSH and PDH show maximum relative improvement for all the algorithms.

5.10 Experiment IV: Listening tests on hearing impaired subjects (hardware based real time)

The software based off-line implementation of the algorithms and their results from listening tests have been presented and discussed in section 5.9. The algorithms were found helpful in improving quality of speech, response time, recognition scores, and information transmission of consonantal features. On the basis of these results obtained from software based off-line, the algorithms were implemented in real-time on FPGA platform. This can be used as part of wearable binaural hearing aid. In the following subsections the procedure and listening test results are discussed. The procedure to carry out the listening tests to evaluate the algorithms is discussed in section 5.10.1 and section 5.10.2 deals with its results.

5.10.1 Experiment IV: Procedure

The procedure for experiment IV was same as it was implemented for software based off-line except FPGA based hardware. The detailed explanation is given in section 5.9.1. The real time hardware based experimental set up consists of Digilent’s Atlys Cirucit board and Laptop, was used for binaural presentation of the speech material, for displaying the response choices, and recording subject’s responses in MATLAB based GUI (Refer Appendix B). The response choices available in confusion matrix format were used to evaluate the recognition score. To compare the load on perception, response times were used. The information transmission analysis was also carried out as it provides a measure independent of subject’s biasing for response (as described in Appendix A). Features like voicing, nasality, frication, place, manner, duration, and continuance were used to group confusion matrices.

5.10.2 Experiment IV: Listening test Results

The following subsection discusses the results of listening tests evaluated by real time hardware experiment (Experiment IV) on eight hearing impaired subjects. It includes qualitative assessment, recognition scores (in %), relative improvement in recognition scores and information transmission analysis for unprocessed and processed speech signals for each hearing impaired subjects.

5.10.2.1 Qualitative assessment

This parameter gives quality of the unprocessed and processed speech material when presented to the hearing impaired subjects. The pre-recorded test material in

VCV context was heard four times by subjects. On the basis of quality of sound, they were asked to give the rating as ‘Outstanding’, ‘Good’, ‘Fair’, ‘Average’ and ‘Below Average'. These ratings were indexed from 1 to 5; ‘1’ being ‘Below Average’ and 5 being ‘Outstanding’. Averages of these ratings were computed to find mean opinion score of each subject. Results of qualitative assessment of eight subjects are given in Table 5.14. It can be observed that, subject KUR ranked the processed signal same as unprocessed signal while six subjects APL, KSH, BHL, PDH, PNK and SWK ranked the quality of processed signal higher than the unprocessed signal for all three algorithms. Graphical representation of mean opinion score of each subject is shown in Figure 5.42.

Table 5.14 Experiment IV - Qualitative assessment

Subjects	US	PS-db	PS-bior	PS-sym
APL	3.75	4.50	4.25	4.25
KUR	5.00	5.00	5.00	5.00
KSH	2.75	4.00	3.75	4.25
BHL	3.00	3.75	4.00	4.25
PDH	3.25	4.25	4.25	4.50
PNK	4.50	5.00	5.00	5.00
BHV	2.50	3.50	2.50	2.50
SWK	3.50	4.25	3.75	4.00

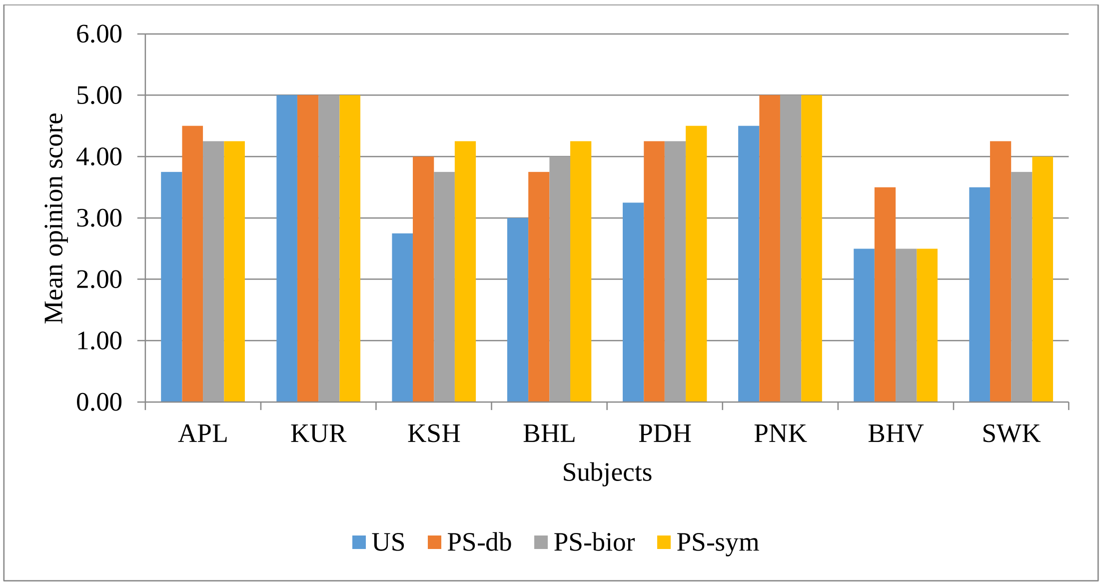


Fig. 5.42 Experiment IV - Qualitative assessment

5.10.2.2 Experiment IV: Recognition Score

Table 5.15 shows the recognition scores, percentage relative improvement, and averages of recognition score calculated from the confusion matrix. Figure 5.43 shows percentage recognition scores for unprocessed and processed signals of all three algorithms and percentage relative improvement in recognition scores. The recognition scores were improved with almost all the algorithms. The relative improvement in the recognition scores for db, bior and sym algorithms were ranged from 3.33 to 28.33%, 0 to 16.67% and 3.33 to 17.78% respectively. The maximum relative improvement among all the algorithms was found in db algorithm.

Subject KSH having severe frequency hearing impairment have shown 28.33% (maximum) relative improvement in recognition score for db algorithm while Subjects APL, BHL and PDH having mild to severe frequency hearing impairment have shown better improvements in recognition scores for all three algorithms. Subject PNK with symmetrical high frequency hearing impairment have shown no relative improvement in bior while a slight relative improvement in recognition score is been seen in db and sym algorithms.

Table 5.15 Experiment IV - Recognition scores

Subjects	US	PS-db	RI-db	PS-bior	RI-bior	PS-sym	RI-sym
APL	53.33	64.44	11.11	70.00	16.67	70.00	16.67
KUR	84.44	93.33	8.89	93.33	8.89	95.55	11.11
KSH	48.33	76.66	28.33	56.66	8.33	63.33	15.00
BHL	68.88	86.66	17.78	83.33	14.45	86.66	17.78
PDH	61.10	75.55	14.45	73.33	12.23	73.33	12.23
PNK	90.00	93.33	3.33	90.00	0	93.33	3.33
BHV	53.33	60.00	6.67	53.33	0	62.22	8.89
SWK	80.00	93.33	13.33	80.00	0	90.00	10.00
Averages	67.42	80.42	12.98	74.99	7.57	79.30	11.87

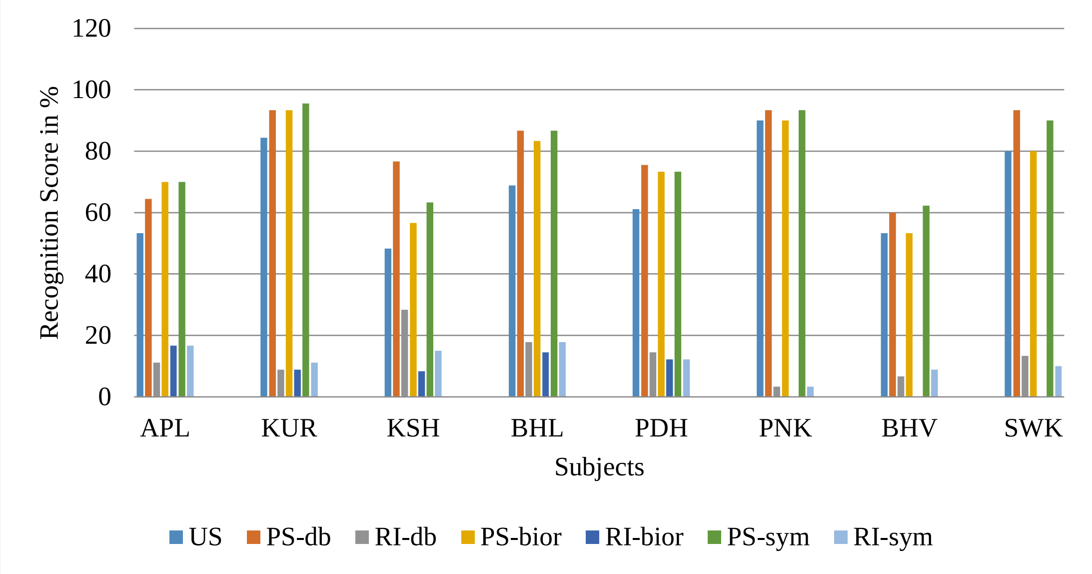


Figure 5.43 Experiment IV - Recognition scores and its Relative improvement (in %)

5.10.2.3 Information Transmission Analysis

Combined confusion matrices of each subject were used to evaluate information transmission analysis of all algorithms. Relative information transmitted for consonantal features of all algorithms and average information transmitted along with relative improvement over unprocessed signal is shown in Table 5.16 (a to h) and plotted in Figure 5.44 to Figure 5.51

Table 5.16 (a to h) Experiment IV - Information Transmission Analysis of consonantal features and Relative improvement (RI) in percentage

a) Feature: Overall							
Subject	US	PS-db	RI	PS-bior	RI	PS-sym	RI
APL	71.15	75.26	4.11	80.28	9.13	83.05	11.90
KUR	89.77	95.30	5.54	94.74	4.98	97.24	7.47
KSH	62.00	87.19	25.19	78.91	16.91	83.36	21.36
BHL	84.30	93.18	8.88	91.47	7.17	91.89	7.59
PDH	74.69	85.81	11.12	86.36	11.67	84.89	10.20
PNK	94.24	96.59	2.35	94.24	0.00	95.30	1.06
BHV	74.19	71.81	-2.38	70.05	-4.14	72.56	-1.63
SWK	89.77	96.59	6.82	87.32	-2.45	92.95	3.19
Averages	80.01	87.72	7.70	85.42	5.41	87.66	7.64

b) Feature: Continuance							
Subject	US	PS-db	RI	PS-bior	RI	PS-sym	RI
APL	33.15	35.61	2.46	35.61	2.46	29.10	-4.05
KUR	41.12	100.00	58.88	58.00	16.88	70.62	29.50

KSH	26.02	44.43	18.41	35.83	9.81	76.35	50.33
BHL	39.84	100.00	60.16	81.95	42.12	76.35	36.52
PDH	42.53	44.41	1.88	35.61	-6.92	38.36	-4.17
PNK	44.43	60.42	15.99	47.26	2.83	100.00	55.57
BHV	11.42	31.09	19.66	29.10	17.67	49.86	38.43
SWK	60.42	100.00	39.58	47.26	-13.16	76.35	15.93
Averages	37.37	64.49	27.13	46.33	8.96	64.62	27.26

c) Feature: Duration							
Subject	US	PS-db	RI	PS-bior	RI	PS-sym	RI
APL	41.12	50.36	9.24	59.28	18.16	59.28	18.16
KUR	58.00	100.00	42.00	100.00	42.00	100.00	42.00
KSH	41.29	100.00	58.71	35.61	-5.68	44.43	3.14
BHL	41.16	71.23	30.07	71.23	30.07	100.00	58.84
PDH	76.74	100.00	23.26	100.00	23.26	89.07	12.32
PNK	76.35	100.00	23.65	81.95	5.60	100.00	23.65
BHV	44.43	60.42	15.99	60.42	15.99	60.42	15.99
SWK	100.00	100.00	0.00	72.02	-27.98	76.35	-23.65
Averages	59.89	85.25	25.36	72.56	12.68	78.69	18.81

d) Feature: Frication							
Subject	US	PS-db	RI	PS-bior	RI	PS-sym	RI
APL	26.46	32.97	6.51	48.96	22.50	39.74	13.27
KUR	53.77	100.00	46.23	71.86	18.09	73.88	20.10
KSH	21.17	38.31	17.14	26.43	5.26	53.35	32.18
BHL	38.72	100.00	61.28	82.45	43.72	78.94	40.22
PDH	57.73	73.88	16.14	64.90	7.16	61.58	3.85
PNK	48.96	64.90	15.93	64.90	15.93	100.00	51.04
BHV	10.52	45.53	35.01	23.72	13.20	63.76	53.24
SWK	64.90	100.00	35.10	53.35	-11.54	78.94	14.05
Averages	40.28	69.45	29.17	54.57	14.29	68.77	28.49

e) Feature: Manner							
Subject	US	PS-db	RI	PS-bior	RI	PS-sym	RI
APL	40.45	46.31	5.87	55.85	15.40	49.95	9.50
KUR	70.73	100.00	29.27	82.06	11.33	83.85	13.11
KSH	40.37	41.42	1.05	48.23	7.86	50.94	10.57
BHL	50.54	86.88	36.34	79.89	29.35	78.55	28.01

PDH	64.02	73.92	9.91	66.16	2.14	63.48	-0.53
PNK	68.05	78.42	10.37	78.42	10.37	100.00	31.95
BHV	15.55	53.65	38.10	29.11	13.56	66.67	51.12
SWK	78.42	100.00	21.58	71.52	-6.90	86.93	8.51
Averages	53.52	72.58	19.06	63.90	10.39	72.55	19.03

f) Feature: Nasality							
Subject	US	PS-db	RI	PS-bior	RI	PS-sym	RI
APL	57.10	67.57	10.47	67.57	10.47	67.57	10.47
KUR	100.00	100.00	0.00	100.00	0.00	100.00	0.00
KSH	71.40	44.91	-26.49	78.76	7.36	44.91	-26.49
BHL	67.57	67.57	0.00	78.76	11.18	78.76	11.18
PDH	74.54	74.54	0.00	67.57	-6.96	67.57	-6.96
PNK	100.00	100.00	0.00	100.00	0.00	100.00	0.00
BHV	20.52	67.57	47.05	24.48	3.96	74.54	54.01
SWK	100.00	100.00	0.00	100.00	0.00	100.00	0.00
Averages	73.89	77.77	3.88	77.14	3.25	79.17	5.28

g) Feature: Place							
Subject	US	PS-db	RI	PS-bior	RI	PS-sym	RI
APL	27.64	29.81	2.17	44.60	16.97	41.59	13.96
KUR	61.11	79.10	17.99	91.34	30.23	85.53	24.42
KSH	18.52	49.03	30.51	32.50	13.98	36.65	18.13
BHL	52.58	80.75	28.17	74.70	22.12	78.11	25.53
PDH	79.67	100.00	20.33	100.00	20.33	100.00	20.33
PNK	89.30	100.00	10.70	88.27	-1.03	82.73	-6.57
BHV	25.06	35.10	10.04	41.97	16.91	31.77	6.71
SWK	78.11	78.11	0.00	63.02	-15.09	73.56	-4.55
Averages	54.00	68.99	14.99	67.05	13.05	66.24	12.25

h) Feature: Voicing							
Subject	US	PS-db	RI	PS-bior	RI	PS-sym	RI
APL	100.00	86.94	-13.06	100.00	0.00	100.00	0.00
KUR	71.86	100.00	28.14	100.00	28.14	100.00	28.14
KSH	63.98	78.94	14.96	78.94	14.96	82.45	18.47
BHL	86.94	100.00	13.06	100.00	13.06	100.00	13.06
PDH	37.01	65.86	28.85	71.68	34.67	60.13	23.12
PNK	100.00	65.86	-34.14	100.00	0.00	100.00	0.00

BHV	21.80	26.46	4.66	32.75	10.95	34.90	13.10
SWK	71.68	100.00	28.32	100.00	28.32	100.00	28.32
Averages	69.16	78.01	8.85	85.42	16.26	84.68	15.53

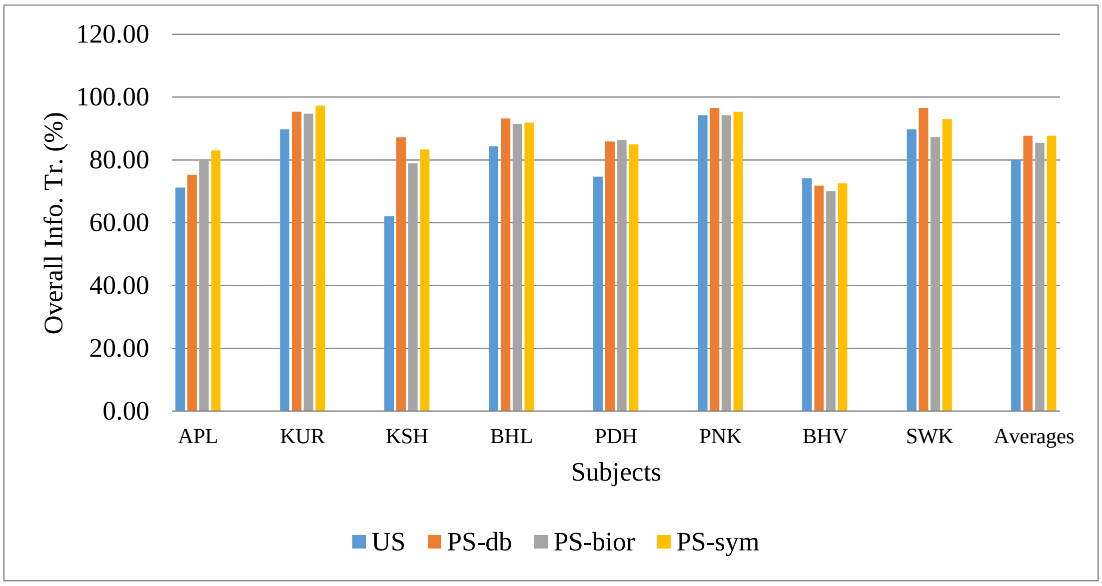


Fig. 5.44 Experiment IV - Relative Information transmitted for Overall

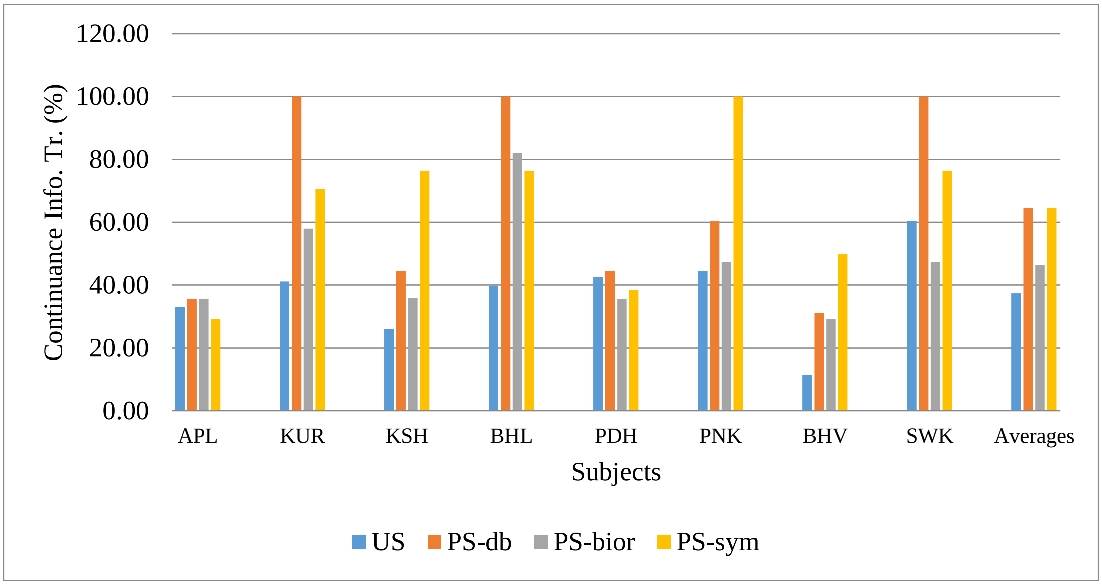


Fig. 5.45 Experiment IV - Relative Information transmitted for Continuance

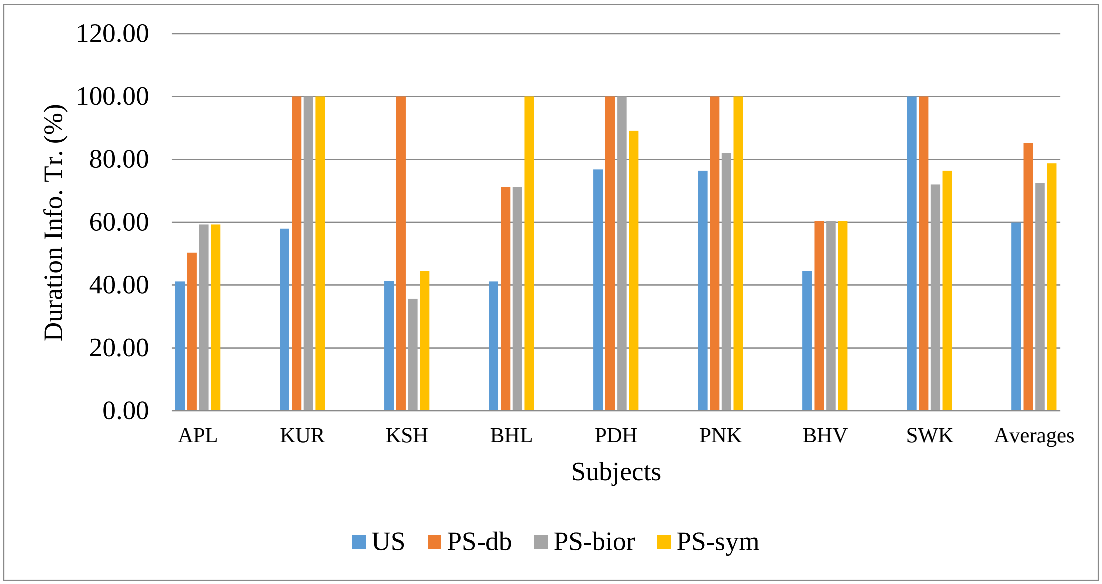


Fig. 5.46 Experiment IV - Relative Information transmitted for Duration

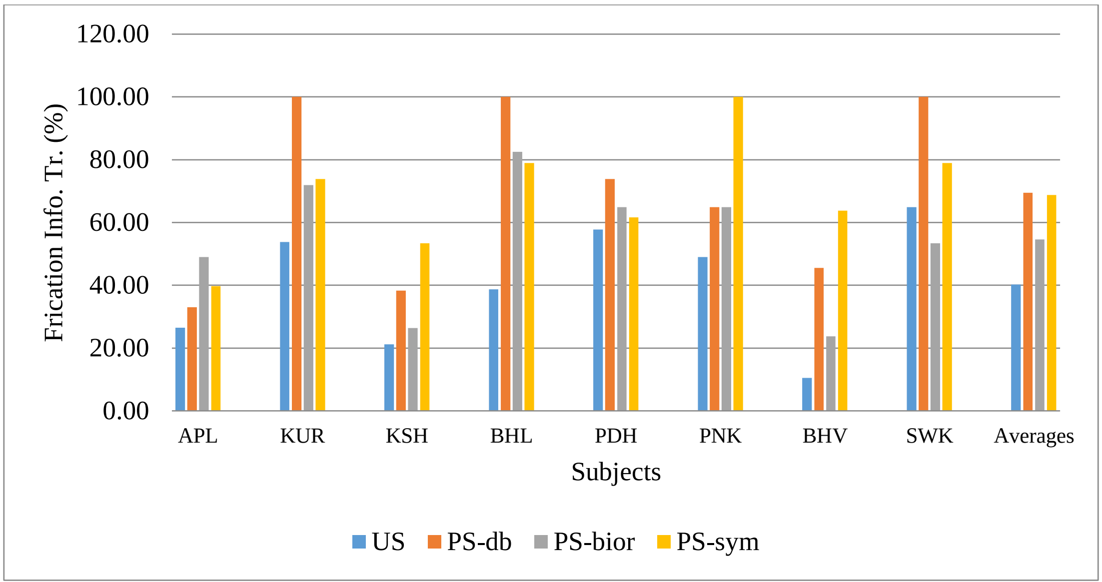


Fig. 5.47 Experiment IV - Relative Information transmitted for Frication

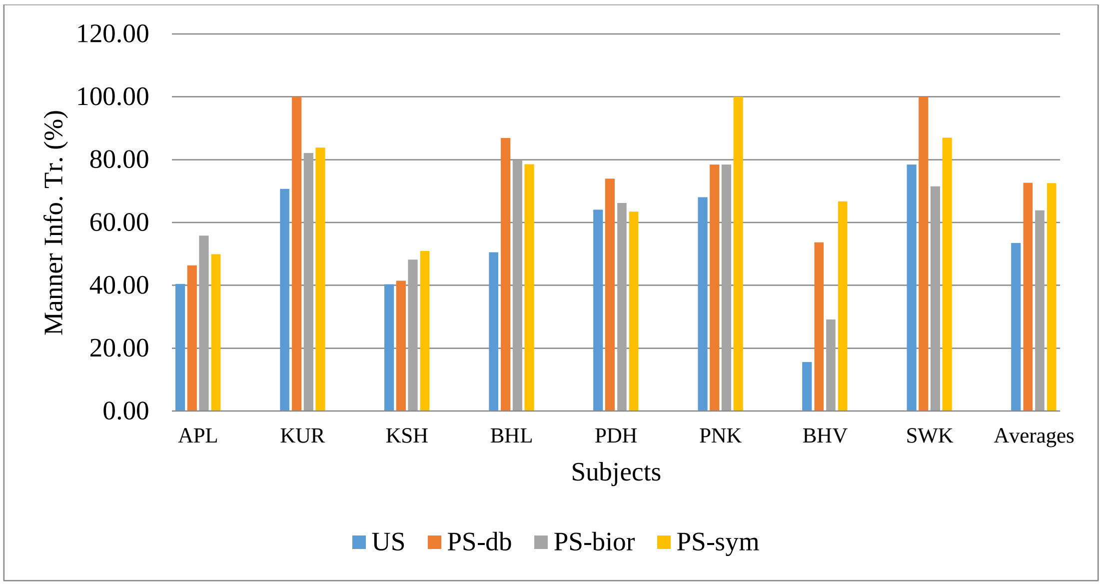


Fig. 5.48 Experiment IV - Relative Information transmitted for Manner

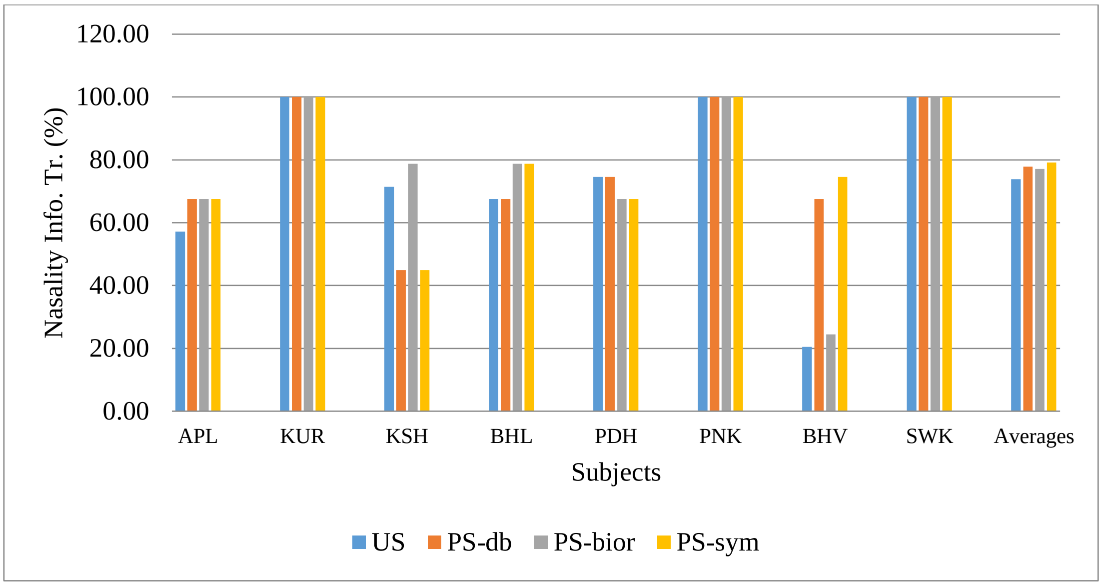


Fig. 5.49 Experiment IV - Relative Information transmitted for Nasality

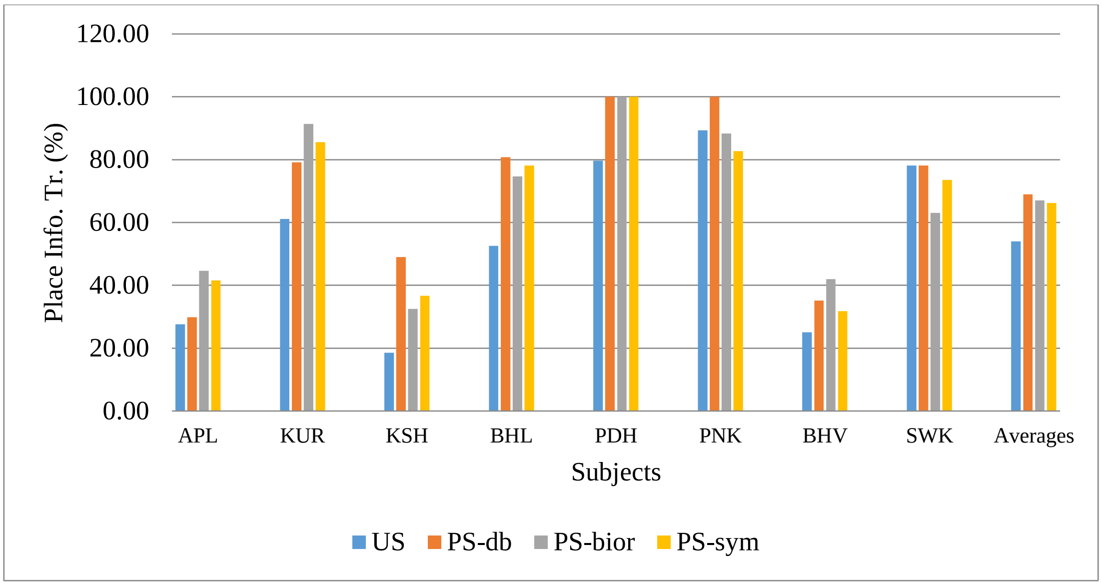


Fig. 5.50 Experiment IV - Relative Information transmitted for Place

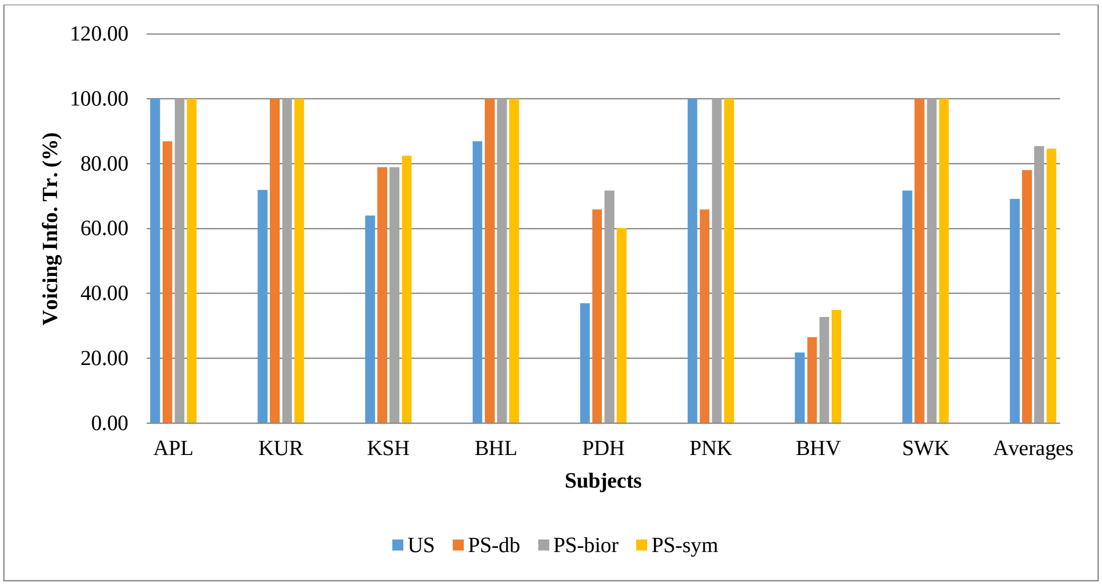


Fig. 5.51 Experiment IV - Relative Information transmitted for Voicing

Discussion

The evaluation of wavelet based algorithms (db, bior and sym) has been done by conducting listening tests on five normal people with 3dB, 0dB, -3dB, -6dB and -9dB SNR conditions (simulated hearing loss) and eight sensorineural hearing impaired subjects with mild to severe hearing loss. This was done with the objective, to assess the wavelet processing algorithms. For this purpose, the three processing algorithms were implemented using software based offline approach and real time based hardware approach.

Fifteen English consonants in VCV context were used to conduct the listening tests. Subject’s response times, recognition scores, and information transmitted for consonantal features were analyzed. Speech quality was ranked by the subjects in qualitative assessments. A summary of qualitative assessments about the test stimuli indicated that perceptual quality improved due to processing for dichotic presentation over the unprocessed diotic presentation.

The listening tests analysis carried out on normal people indicated a reduction in recognition scores with decrease in SNR for unprocessed speech. It was also seen that for a particular level of masking noise, the score for processed speech was significantly higher than that for the unprocessed one. For higher levels of masking noise (-9 dB SNR), i.e. higher levels of simulated sensorineural loss, it is seen that the relative improvements due to processing were more. The response time for processed speech was significantly lower than that for unprocessed one. The average response time increased with increase in the level of masking noise for all subjects. The relative information transmitted is near perfect with unprocessed speech and improves with processed speech for higher value of SNR. However, most of the subjects indicated the maximum improvement for the duration, frication and manner features.

Reduction in response time for all three algorithms is observed in analysis of listening tests conducted on sensorineural hearing impaired subjects when compared with unprocessed signals. This indicates reduction in burden on perception process. For the subject with mild to moderate and asymmetrical high frequency impairment, the relative decrease in response time is largest for all algorithms. The relative decrease in response time was statistically significant for db and bior algorithms. This shows the effectiveness of db and bior algorithms in load reduction. For the impaired subjects, recognition score was low and varied from 53.33% to 90.00% for unprocessed speech. An improvement was seen in recognition score for all the algorithms. The maximum relative improvement among all the algorithms was found in db. In db algorithm, almost all the subjects showed better improvement in recognition score except PNK with symmetrical high frequency hearing impairment subject which showed no improvement at all, while Subjects APL, BHL and PDH having mild to severe frequency hearing impairment have shown better improvements in recognition scores for all three algorithms.

For unprocessed signal, relative information transmitted for continuance, frication, manner and place features gets deteriorated most. The transmission of all the consonantal features gets improved by all algorithms, particularly the frication, continuance, manner and place features. Generally the db algorithm shows the highest improvement for frication, manner and place features, while the sym algorithm shows the highest improvement for continuance features. Subjects BHL and KUR shows maximum relative improvement in information transmitted for continuance, frication, manner and place for all algorithms. And the maximum improvement was found in db and sym algorithms. Subjects BHL and BHV showed the maximum relative improvement in every algorithm over other subjects for manner feature. For place feature all subjects, except SWK and PNK show maximum relative improvement for all the algorithms. Subject KSH showed 31.99 % relative improvement in db algorithm while subject KUR showed 24.42 % and 30.23 % relative improvement in bior and sym algorithms respectively. As the place information is related to frequency resolving capacity of the auditory processing, the implemented algorithms have reduced the effect of spectral masking and improved the perception of hearing.

From the analysis of recognition scores and information transmission, it is observed that the algorithm which gives maximum benefit by the reduction of the effect of increased masking is based on individual hearing impairment conFig.uration. Dichotic presentation results in the reception of relatively robust consonantal features (voicing, manner, and nasality). The potential to improve the speech perception for subjects using binaural hearing aids lies in the processing algorithm for dichotic presentation.

In listening tests with hearing impaired subjects, the subjects are able to perceptually integrate the dichotically presented speech signal, and the presentation results in improved speech reception. The improvements in response time, shows that the dichotic presentation reduces the load on perception process. It is observed that for hearing impaired subjects, the improvement in consonantal reception and reduction in response time do not follow the same trend. So, extensive tests on the hearing impaired subjects are necessary to estimate the merits of processing algorithms.

Chapter 6
CONCLUSION

Sensorineural hearing impairment occurs due to damage to the transduction mechanism of the inner ear. It is based on reduction in frequency resolving capacity of the auditory system due to spread of spectral masking. This causes a difficulty in perception of consonantal features. The hearing impaired subjects may face difficulties in the identification of the consonantal place feature as it is cued by the spectral differences. The effect of increased temporal masking is to cause an increase in forward and backward masking of weak acoustic segments by strong ones. Vital cues necessary for consonant identification are masked by initial and subsequent vowel segments. These masking occurs mainly at peripheral auditory system. Thus the overall effect of two types of masking is difficult in discriminating consonants, which results in relative reduction of speech perception by subjects with sensorineural hearing impairment.

In speech perception there is an integration of information gathered from both ears. The speech signal is split into two complementary parts along with the addition of the signals from alternate bands when presented to both ears. Hence, the two adjacent bands are likely to mask each other, helping in reduction of increased masking. Thus binaural dichotic presentation helps to improve speech reception by subjects with moderate bilateral sensorineural hearing impairment.

Previous research shows that, the speech processing scheme related to the spectral splitting of speech signal by use of comb filters with complementary pass band, help reduced the increased spectral masking effect. A probable solution to reduce increased masking effect is to split the speech signal and compress the frequency bands.

In this work, wavelet based algorithms have been developed and investigated to split the speech signal into two components for binaural dichotic presentation. The perception of various consonantal features in binaural dichotic presentation may be improved by the combination of the spectral splitting schemes with compression of frequency bands. Implementation and evaluation of these algorithms have been presented in the previous chapters. This chapter presents a summary of the investigation, conclusion based on the discussion of results and some suggestions for further work.

6.1 Summary of investigation

The effect of spectral splitting schemes with compression of frequency bands was studied. Different levels of optimized wavelet packets were investigated to synthesize the components of speech signal, which recognised transient information. In the developed algorithms, speech components are associated with different frequency ranges. These ranges are interpreted as ‘nodes’ in the wavelet packet analysis and as ‘levels’ in the wavelet transform analysis. The merit of wavelet packets is that they provide a fine and an evenly spaced division of the frequency spectrum. The energy of speech components synthesized using wavelet transform and wavelet packets gets concentrated in the frequency ranges and recognizes each kind of information. This information is usually said to have narrow bandwidths, as the wavelet packets split the spectrum into finer frequency. In sensorineural hearing impaired subjects, the frequency bands are reduced compared to normal subjects. Hence in our research the number of frequency bands is limited to eight.

The signal components were obtained using decomposition of speech signal by combination of discrete wavelet transform and wavelet packet. It involves different time-frequency resolution at each decomposition level. Depending on orthogonality and symmetry properties, appropriate wavelet basis were chosen. In this research, Daubechies, Symlet and Biorthogonal wavelet families are used. It helped in reducing the perception of spectral and temporal masking simultaneously, thereby improving the speech perception.

To reduce the spectral masking effect, algorithm of spectral splitting was implemented. The first step to design better filters included the designing of comb filters with 512 coefficients. For spectral splitting, the chosen comb filters were based on auditory critical bandwidths [71].

The initial investigation involved study of the comb filters, its hardware based real time implementation, experimental assessment and results as discussed in chapter 4. The successive investigation phase, reported in chapter 5, dealt with software based and real time implementation of wavelet based algorithms, its evaluation and listening tests results.

The listening tests were conducted to assess the effectiveness of the developed algorithms. In chapter 4, the listening tests were conducted on seven sensorineural hearing impaired subjects and in chapter 5 the tests were conducted on five normal people with simulated hearing loss and eight moderate to severe bilateral sensorineural hearing impaired subjects, for the assessment of the developed algorithms. An experimental setup with Laptop and Digilent Altys Circuit board was used for binaural presentation of the test stimuli for conducting and evaluating the listening tests. The subject responses were noted in the form of stimulus-response confusion matrices. The performance was noted for the subjects' qualitative assessment, recognition scores, response times, overall information transmitted and relative information transmission for consonantal features.

6.2 Conclusions

For the assessment of off-line and real time implementation of the wavelet based algorithms in dichotic presentation, three experimental sets of listening were conducted and evaluated. In Experiment II, listening tests were carried out on five normal hearing subjects with 3 dB, 0 dB, -3 dB, -6 dB and -9 dB SNR conditions. The listening tests in Experiment III (software based off-line) and Experiment IV (hardware based real time) were conducted on eight subjects with bilateral sensorineural hearing loss.

The overall conclusions drawn from experiment II are for unprocessed speech, the recognition scores decreased with decrease in SNR. It was observed that for a particular level of masking noise, the score for processed speech was significantly higher than that for the unprocessed one. It was also observed that the relative improvements due to processing were more for higher levels of masking noise (- 9 dB SNR), i.e. higher levels of simulated sensorineural loss. The average response time increased with increase in the level of masking noise for all subjects. It was observed that response time for processed speech was significantly lower than that for unprocessed one. The relative information transmitted is near perfect with unprocessed speech and improves with processed speech for higher value of SNR. However, many subjects indicated the maximum improvement for the duration, frication and manner features. Finally it can be concluded that in noisy environment the load on speech perception process is reduced and speech perception by normal people gets improved by wavelet based algorithms having binaural dichotic presentation.

From Experiment III and Experiment IV, the overall conclusions are: the summary of qualitative assessments about the speech material indicates improvement in perceptual quality due to processing for dichotic presentation over the unprocessed diotic presentation. The response times of processed signals were reduced when compared with unprocessed signals. This indicates reduction in burden on perception process. For the subject with mild to moderate and asymmetrical high frequency hearing impairment, the relative decrease in response time is largest for all algorithms. The relative decrease in response time was statistically significant for db and bior algorithms. This shows the effectiveness of db and bior algorithms in load reduction.

For the impaired subjects, recognition score was low and varied from 53.33% to 90.00% for unprocessed speech signal. An improvement was seen in recognition score for most of the subjects across all the algorithms. Algorithm db shows maximum relative improvement in recognition score among all algorithms.

For unprocessed signal, relative information transmitted for continuance, frication, manner and place feature observed as deteriorated most. The transmission of all the consonantal features, were improved using all algorithms, particularly frication, continuance, manner and place features. Generally the db algorithm showed the highest improvement for frication, manner and place features, while the sym algorithm showed the highest improvement for continuance features. For place feature almost all subjects showed maximum relative improvement for all the algorithms. As the place information is related to frequency resolving capacity of the auditory processing, the implemented algorithms have reduced the effect of spectral masking and improved the perception of hearing. It was observed that the improvements due to different algorithms were based on individual's degree of hearing loss. The algorithms db and sym were observed to be beneficial for the subjects having low and high frequency hearing impairment respectively.

The investigations showed the ability of hearing impaired subjects to combine the dichotically presented speech signal. The processing of the algorithms to split the speech signal in a complementary manner, so as to reduce the effects of increased masking effects, resulted in an improvement in speech reception. The improvements in response time resulted in reduction of the perception process by the dichotic presentation. The reduction in response time and the improvement in consonantal reception did not have the same pattern in case of hearing impaired subjects.

6.3 Suggestions for further work

There remains a massive scope to further enhance the performance of algorithms, developed and implemented in this research. Some of the further work may include:

- Need to extend the evaluation of algorithms with use of mixed hearing loss subjects,
- To create different processing schemes for different environments (diotic or/and dichotic),
- To work with psychoacoustics to optimize hearing of most important sounds,
- To build the platform for ASIC with flexibility of selecting the processing parameters to best suit the individual

Appendix A

Performance Assessment techniques:

The parameters like response time, percentage correct recognition scores, percentage relative improvement in response times and recognition scores are calculated so as to assess the performance of speech processing algorithms (Miller and Nicely, 1955).

A.1 Information transmission analysis

The effectiveness of speech processing scheme can be judge based on the recognition score. The pattern of errors can be detailed by stimulus response score represented in form of confusion matrix (Miller and Nicely, 1955). The responses are noted in the columns, while the stimuli are included along the rows and every entry in the cell represents either the frequencies or probabilities of stimulus response pair.

If s represents sets of n stimulus items $\{s_1, s_2, \ldots s_n\}$ and t represent sets of n responses $\{t_1, t_2, \ldots t_n\}$, and if $N(s_i)$, $N(t_j)$, and $N(s_i; t_j)$ represent the frequencies of stimulus s_i, response t_j and the stimulus-response pair $(s_i ; t_j)$ respectively in a test of N observations, then the probabilities can be obtained as

$$p(s_i ; t_j) = \frac{N(s_i;t_j)}{N} \tag{A.1}$$

$$p(s_i) = \frac{N(s_i)}{N} = \sum_{j=1}^{n} p(s_i;t_j) \tag{A.2}$$

$$p(t_j) = \frac{N(t_j)}{N} = \sum_{i=1}^{n} p(s_i;t_j) \tag{A.3}$$

In confusion matrix, the diagonal elements correspond to correct responses, while an off-diagonal element indicates incorrect responses. The recognition score R_s can be obtained from the sum of diagonal elements of a confusion matrix as,

$$R_s = \sum_{i=1}^{n} p(s_i;t_i) \tag{A.4}$$

The small confusion matrices are obtained by combination of the stimuli and responses into groups to study the pattern of confusions. The above combination is done in such manner that the confusion within the groups is more likely than confusion between the groups. The grouping of stimuli is related to certain common features. To study the reception of different features, recognition scores obtained for the smaller confusion matrices can be used.

The recognition scores may be influenced by persons biasing in response. The above problems are dealt with by information transmission analysis (Miller 1955), which provides a scale of covariance between stimuli and responses.

Mean logarithmic probability (MLP) measure of information is used as a measure of the information transmitted. The information measures of the stimulus *s* and response *t*, $I(s)$ and $I(t)$ respectively, are given by

$$I(s) = \mathrm{MLP}(s) = -\sum_i p(s_i) \log_2 p(s_i) \quad \text{bits} \tag{A.5}$$

$$I(t) = \mathrm{MLP}(t) = -\sum_j p(t_j) \log_2 p(t_j) \quad \text{bits} \tag{A.6}$$

MLP measure of covariance of stimulus-response is

$I(s;t) = \mathrm{MLP}(s) + \mathrm{MLP}(t) - \mathrm{MLP}(st)$

$$= -\sum_{i,j} p(s_i\,;\,t_j) \log_2 \frac{p(s_i)\,p(t_j)}{p(s_i;t_j)} \quad \text{bits} \tag{A.7}$$

The relative information transmission is given by the equation A.8

$$I_{rel}(s;t) = I(s;t)/I(s) \tag{A.8}$$

Since $I(s) \geq I(s;t) \geq 0$; $\quad 1 \geq I_{rel}(s;t) \geq 0$

Relative information transmission, like recognition score, can be derived for small confusion matrices, which can be obtained from the original matrix dependent on certain common features of stimulus and response items. The transmission performance in the context of specific features can be obtained from the relative information transmission.

The relation between relative information transmitted I_{rel} and recognition score *Rs* for special case involving stimulus items with equal probabilities, incorrect responses are equally distributed among the off diagonal cells, while the correct responses are equally distributed among the diagonal cell (Pandey 1987).

For ‘*n*’ number of stimuli, The entries in cells are shown below

$$p(s_i\,;\,t_j) = R_s/n\,, \qquad i = j$$

$$= \frac{1-R_s}{n^2-n} \qquad i \neq j \tag{A.9}$$

$$p(s_i) = p(t_j) = 1/n \tag{A.10}$$

It is observed that, zero scoring (R_s = 0) corresponds to a finite relative information transmission, $I_{rel}(s;t) = \log_2(n/(n-1))/\log_2(n)$, chance scoring, ($R_s = 1/n$) corresponds to $I_{rel}(s;t) = 0$ and perfect scoring (R_s=1) corresponds to $I_{rel}(s;t) = 1$.

A.2 Feature groupings for Information transmission analysis

Feature groupings for information transmission analysis used is shown in Table A.1 The features chosen for study were voicing (voiced: / b d g m n z v r l y / and unvoiced: / p t k s f /), place (front: / p b m f v /, middle: / t d n s z r l /, and back: / k g y /), manner (oral stop: / p b t d k g l y /, fricative: / s z f v r /, and nasals: / m n /), nasality (oral: / p b t d k g s z f v r l y /, nasal: /m n /), frication (stop: / p b t d k g m n l y /, fricative: / s z f v r /), and duration (short: / p b t d k g m n f v l / and long: /s z r y /).

Table A.1. Feature groupings for information transmission analysis

Feature	**Consonants**														
	p	b	T	D	k	g	m	N	s	z	f	v	r	l	Y
Duration (SH: short, LO: long)	0	0	0	0	0	0	0	0	1	1	0	0	1	0	1
Frication (ST: stop, FR: fricative)	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0
Nasality (OR: oral, NA: nasal)	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0
Manner (OS: oral stop, FR: fricative, NA: nasal)	0	0	0	0	0	0	2	2	1	1	1	1	1	0	0
Place (FN: front, MD: mid, BK: back)	0	0	1	1	2	2	0	1	1	1	0	0	1	1	0
Voicing (UV: unvoiced, VO: voiced)	0	1	0	1	0	1	1	1	0	1	0	1	1	1	1

A.3 Analysis programs

For the experiments carried out, a MATLAB program was used to combine the confusion matrices, so obtained. Analysis of these confusion matrixes was done using another MATLAB program for obtaining recognition score and information transmission and results are stored as '.xlsx' file. The response_rec.m program reads the necessary information, namely, total number of presentations, stimulus name, stimulus age, number of test stimuli, nature of test carried (unprocessed speech, software based offline or hardware based real time) and cell entries from the input file. It also contains information about the feature groupings. The program gives overall percentage scores for different speech features, overall information transmitted and information transmission analysis for the speech features.

The .xlsx file obtained after executing MATLAB based GUI is saved by subjects name_ age_gender and contains the following information-

- Total number of stimuli,
- Correct and incorrect entries of responses
- sum of the cell entries in the row and along the columns
- response times in seconds

Feature groupings file is of .xlsx format saved by result_sheet_number and contains the following information:

- number of feature classification,
- Feature classification information. Feature groups are represented by consecutive integers. The feature classification information consists of group numbers followed by feature and group labels.

A.4 Sample analysis results

A sample analysis confusion matrix is selected corresponding to the processing db algorithm for subject KUR is as follows,

i) Software Based Offline

1) Output Confusion Matrix

	ABA	ADA	AFA	AGA	AKA	ALA	AMA	ANA	APA	ARA	ASA	ATA	AVA	AYA	AZA	NO O/P	
ABA	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5
ADA	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5
AFA	0	0	3	0	0	0	0	0	0	0	0	2	0	0	0	0	5
AGA	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	5
AKA	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	5
ALA	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	5
AMA	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	5
ANA	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	5
APA	0	0	0	0	2	0	0	0	3	0	0	0	0	0	0	0	5
ARA	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	5
ASA	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	5
ATA	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	5
AVA	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	5
AYA	0	0	0	0	0	0	0	0	2	0	0	0	0	3	0	0	5
AZA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	5
NO O/P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	5	5	3	5	7	5	5	5	5	5	5	7	5	3	5	0	75

2) Confusion Matrix - Stimulus Response

	ABA	ADA	AFA	AGA	AKA	ALA	AMA	ANA	APA	ARA	ASA	ATA	AVA	AYA	AZA	NO O/P
ABA	3.91927	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ADA	0	3.61349	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AFA	0	0	5.33573	0	0	0	0	0	0	0	0	6.51777	0	0	0	0
AGA	0	0	0	2.91857	0	0	0	0	0	0	0	0	0	0	0	0
AKA	0	0	0	0	6.87308	0	0	0	0	0	0	0	0	0	0	0
ALA	0	0	0	0	0	3.16929	0	0	0	0	0	0	0	0	0	0
AMA	0	0	0	0	0	0	5.19027	0	0	0	0	0	0	0	0	0
ANA	0	0	0	0	0	0	0	2.62933	0	0	0	0	0	0	0	0
APA	0	0	0	0	4.81755	0	0	0	8.27129	0	0	0	0	0	0	0
ARA	0	0	0	0	0	0	0	0	0	3.09141	0	0	0	0	0	0
ASA	0	0	0	0	0	0	0	0	0	0	4.0352	0	0	0	0	0
ATA	0	0	0	0	0	0	0	0	0	0	0	3.03182	0	0	0	0
AVA	0	0	0	0	0	0	0	0	0	0	0	0	6.33061	0	0	0
AYA	0	0	0	0	0	0	0	0	9.09858	0	0	0	0	2.8715	0	0
AZA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3.44588	0
NO O/P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

3) Percentage Relative Information

	corr	err	ir	is	it	reltr	rerr	terr
overall	92	0.08	3.87307	3.90473	3.67912	94.2221	100	0.08
continuance	97.3333	0.02667	0.79475	0.83635	0.66971	80.0745	100	0.02667
duration	97.3333	0.02667	0.79475	0.83635	0.66971	80.0745	100	0.02667
frication	97.3333	0.02667	0.889	0.91801	0.75496	82.2386	100	0.02667
manner	97.3333	0.02667	1.37854	1.39915	1.2445	88.9467	100	0.02667
nasality	100	0	0.56622	0.56622	0.56622	100	100	0
place	92	0.08	1.48979	1.50539	1.11179	73.8537	100	0.08
voicing	97.3333	0.02667	0.94239	0.91801	0.78088	85.0628	100	0.02667

Table A2. Percentage Relative Information transmitted using db algorithm of Subject KUR using MATLAB

Continuance

55	0
2	18
Score	80.0745

Duration

55	0
2	18
Score	80.0745

Frication

50	0
2	23
Score	82.2386

Manner

40	0	0
2	23	0
0	0	10
Score	88.9467	

Nasality

65	0
0	10
Score	100

Place

21	2	2
0	35	0
2	0	13
Score	73.8537	

Voicing

25	0
2	48
Score	85.0628

ii) Hardware Based (Real Time)

1) Output Confusion Matrix

	ABA	ADA	AFA	AGA	AKA	ALA	AMA	ANA	APA	ARA	ASA	ATA	AVA	AYA	AZA	NO O/P	
ABA	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5
ADA	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5
AFA	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	5
AGA	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	5
AKA	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	5
ALA	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	5
AMA	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	5
ANA	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	5
APA	0	0	0	0	3	0	0	0	0	0	0	2	0	0	0	0	5
ARA	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	5
ASA	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	5
ATA	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	5
AVA	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	5
AYA	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	5
AZA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	5
NO O/P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	5	5	5	5	8	5	5	5	0	5	5	7	5	5	5	0	75

2) Confusion Matrix - Stimulus Response

	ABA	ADA	AFA	AGA	AKA	ALA	AMA	ANA	APA	ARA	ASA	ATA	AVA	AYA	AZA	NO O/P
ABA	3.41265	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ADA	0	3.49063	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AFA	0	0	4.64528	0	0	0	0	0	0	0	0	0	0	0	0	0
AGA	0	0	0	3.67478	0	0	0	0	0	0	0	0	0	0	0	0
AKA	0	0	0	0	5.41396	0	0	0	0	0	0	0	0	0	0	0
ALA	0	0	0	0	0	3.56622	0	0	0	0	0	0	0	0	0	0
AMA	0	0	0	0	0	0	4.85467	0	0	0	0	0	0	0	0	0
ANA	0	0	0	0	0	0	0	2.77806	0	0	0	0	0	0	0	0
APA	0	0	0	0	8.29063	0	0	0	0	0	0	6.13209	0	0	0	0
ARA	0	0	0	0	0	0	0	0	0	2.86904	0	0	0	0	0	0
ASA	0	0	0	0	0	0	0	0	0	0	3.76751	0	0	0	0	0
ATA	0	0	0	0	0	0	0	0	0	0	0	3.65776	0	0	0	0
AVA	0	0	0	0	0	0	0	0	0	0	0	0	6.68817	0	0	0
AYA	0	0	0	0	0	0	0	0	0	0	0	0	0	3.24569	0	0
AZA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4.06344	0
NO O/P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3.88708

3) Percentage Relative Information

	corr	err	ir	is	it	reltr	rerr	terr
overall	93.3333	0.06667	3.78724	3.90473	3.72255	95.3345	100	0.06667
continuance	100	0	0.83635	0.83635	0.83635	100	100	0
duration	100	0	0.83635	0.83635	0.83635	100	100	0
frication	100	0	0.91801	0.91801	0.91801	100	100	0
manner	100	0	1.39915	1.39915	1.39915	100	100	0
nasality	100	0	0.56622	0.56622	0.56622	100	100	0
place	93.3333	0.06667	1.50509	1.50539	1.19978	79.6987	100	0.06667
voicing	100	0	0.91801	0.91801	0.91801	100	100	0

Table A3 Percentage Relative Information Transmitted using db algorithm of Subject KUR using FPGA

Continuance	
55	0
0	20
Score	100

Duration	
55	0
0	20
Score	100

Frication	
50	0
0	25
Score	100

Manner		
40	0	0
0	25	0
0	0	10
Score	100	

Nasality	
65	0
0	10
Score	100

Place		
20	2	3
0	35	0
0	0	15
Score	79.6987	

Voicing	
25	0
0	50
Score	100

The average of pure tone thresholds of all the eight sensorineural hearing impaired subjects are given in Table A4.

Table A.4: Information about Hearing impaired Subjects. PTA: Average of pure tone thresholds in dB HL, taken at test frequencies 0.25, 0.5, 1, 2, 4 KHz.

Subject Code (Sex, Age)	Ear L=Left R=Right	Hearing thresholds (dB HL) Frequency (kHz)					PTA (dB)
		0.25	0.5	1	2	4	
APL : F 16	L	60	65	70	70	65	66
	R	65	65	70	80	75	71
KUR : M 57	L	25	20	25	35	40	29
	R	65	70	85	85	75	70
KSH : F 12	L	70	85	90	80	70	79
	R	70	80	85	75	60	74
BHL : F 10	L	65	70	85	75	65	72
	R	65	70	85	85	75	76
PDH : M 74	L	55	55	45	70	50	55
	R	85	80	70	60	55	70
PNK : M 45	L	20	20	15	15	70	28
	R	10	15	25	25	75	30
BHV : M 12	L	80	85	95	85	70	83
	R	90	90	90	80	75	85
SWK: F50	L	60	60	65	60	100	67
	R	55	70	75	65	90	71

Appendix B
Hardware and Software description

B.1 Hardware and Software

The experimental setup consists of hardware and software. The hardware used consists of Dell Laptop having 8GB RAM, i5 Processor with 2.20MHz, integrated sound card and Atlys Circuit Development Board. The software required is MATLAB 8.1.0.64 (R2013a) with Signal Processing Toolbox, HDL coder Toolbox and Wavelet Toolbox.

The Signal Processing Toolbox offers variety of inbuilt functions from generation of signals to design and filter it using filters. It also supports to transform the signals from one domain to another domain as well as estimate the power spectra and bandwidth. It can also be used to extract features, develop and validate custom algorithms so as to gain a better prospective of data. All of these inbuilt functions are written in M-files which become useful during the implementation. Instead of finding and utilizing the inbuilt functions, some of them have been converted to apps, which become convenient while implementing any code.

HDL Coder generates convenient and easily synthesizable VHDL and Verilog code from MATLAB functions and Simulink models. The generated HDL code of industry standards can be used for FPGA programming or prototyping/designing and ASIC. A workflow advisor automates the process of converting the MATLAB code into VHDL code so as to program Xilinx and Altera FPGAs. The HDL architecture and implementation can be chosen as per the requirements and its hardware resource utilization can be estimated. HDL Coder provides traceability between the Simulink model and the generated VHDL code, enabling code verification for high-integrity applications.

HDL coder establishes unidirectional link from MATLAB to Xilinx and FPGA devices. It allows the MATLAB user to create a project file that links to FPGA devices and real-time data, so that the user can process real-time data through FPGA
The steps to convert a MATLAB code into an equivalent VHDL code so as to program the FPGA hardware are as follows:

- Build, debug and simulate proposed algorithm in MATLAB.

- Generate a MATLAB file having user input and other MATLAB file having the actual logic of the proposed algorithm.
- Use HDL workflow advisor of the HDL coder toolbox to generate the equivalent VHDL/Verilog code and the testbench for the respective FPGA prototype.
- Verify HDL code with the linked simulation tool by using the test bench created
- Verify FPGA prototype by programming it.

Wavelet Toolbox is one of the advanced toolbox used in variety of applications ranging from denoising of images to compression of images. It also finds its usefulness in feature extraction techniques, communication applications and many more. Suitable apps and variety of inbuilt MATLAB functions help to explore the different algorithms used for continuous wavelet transforms and discrete wavelet transforms. It also helps the user to build customized wavelet to suit the respective application. The toolbox also allows analyzing the frequency content of data that changes over time and displays time-varying patterns common in multiple signals. Multi-resolution analysis can also be performed to extract neatly defined -scale features, to identify the discontinuities in the data, and identify change points or events that are invisible in the raw data.

Using these toolboxes the proposed algorithm was developed and prototyped on FPGA hardware. The FPGA hardware used was Atlys Circuit board from Digilent Inc. It is a ready to use development board based on Xilinx Spartan-6 LS45 FPGA. This board consists of many advanced peripherals ranging from Gbit Ethernet to Audio ports. It can also be used to design and implement wide range of digital signal processing based systems alongwith embedded processor designs based on Xilinx's MicroBlaze.

The Spartan-6 LX45 FPGA is upgraded for high performance logic and offers some of the following features:

- 58 DSP slices,
- 6,822 slices, each containing four 6-input LUTs and eight flip-flops,
- 500MHz+ clock speeds

- six phase-locked loops,
- four clock tiles (eight DCMs & four PLLs),
- 2.1Mbits of fast block RAM,

To initiate the usefulness of the proposed algorithms on hearing impaired subjects, the experiments were built using a customized user interface. It is easily developed with help of MATLAB's Graphical User Interface Design Environment (GUIDE). A GUI is a simple to design custom built interface for specific user to perform a specific task. It provides simplicity, consistency, familiarity and a dynamic interface to the user without knowing the applications underlying commands and graphical object used.

The MATLAB based GUI module contains the M-files developed for analysis of speech signals using the different processing schemes. The module contains the pre-recorded speech signals used in our experimentation. Based on the user response, the unprocessed pre-recorded speech signal gets processed and finally the output is stored for further analysis.

B.2 Listening test set-up

A GUI for normal hearing with simulated hearing loss was developed for easy experimentation on normal people as shown in Fig. B.1.Each user clicked the respective push button as well as gave input for different SNR values (-9, -6, -3, 0, 3) and their final results were collected in confusion matrix format. The confusion matrix was used to evaluate the response times, recognition scores and information transmission analysis.

Another GUI was developed for the offline software based and real time hardware based listening tests as shown in Fig. B.2. It assumes the usage of said hardware and software installed while performing the experiment on hearing impaired subjects. The GUI was developed in local dialect for easy assistance.

Each subject clicked the respective push button and their final offline software based and real time hardware based results were collected in confusion matrix format. The confusion matrix was used to evaluate the response time, recognition score and information transmission analysis.

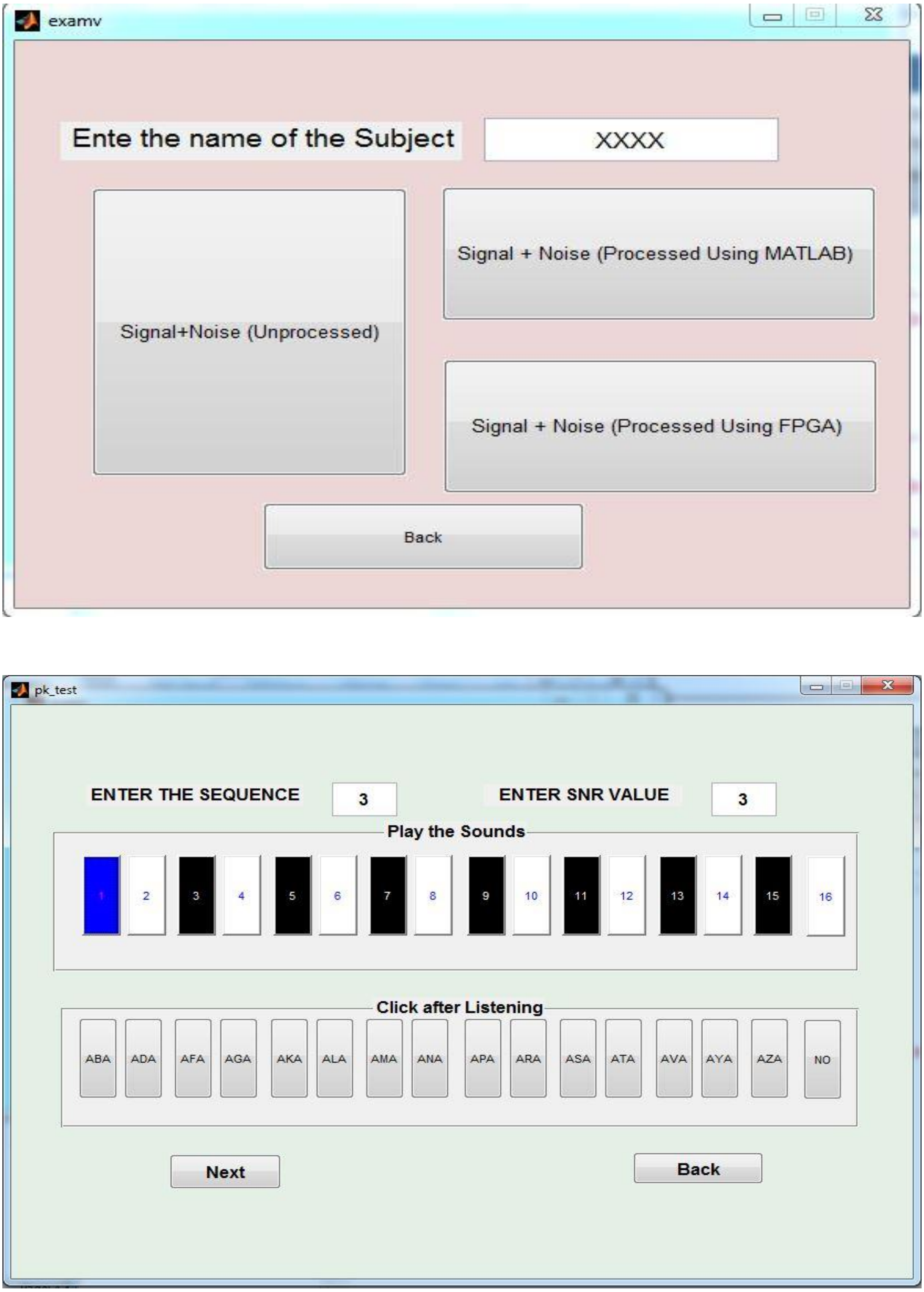
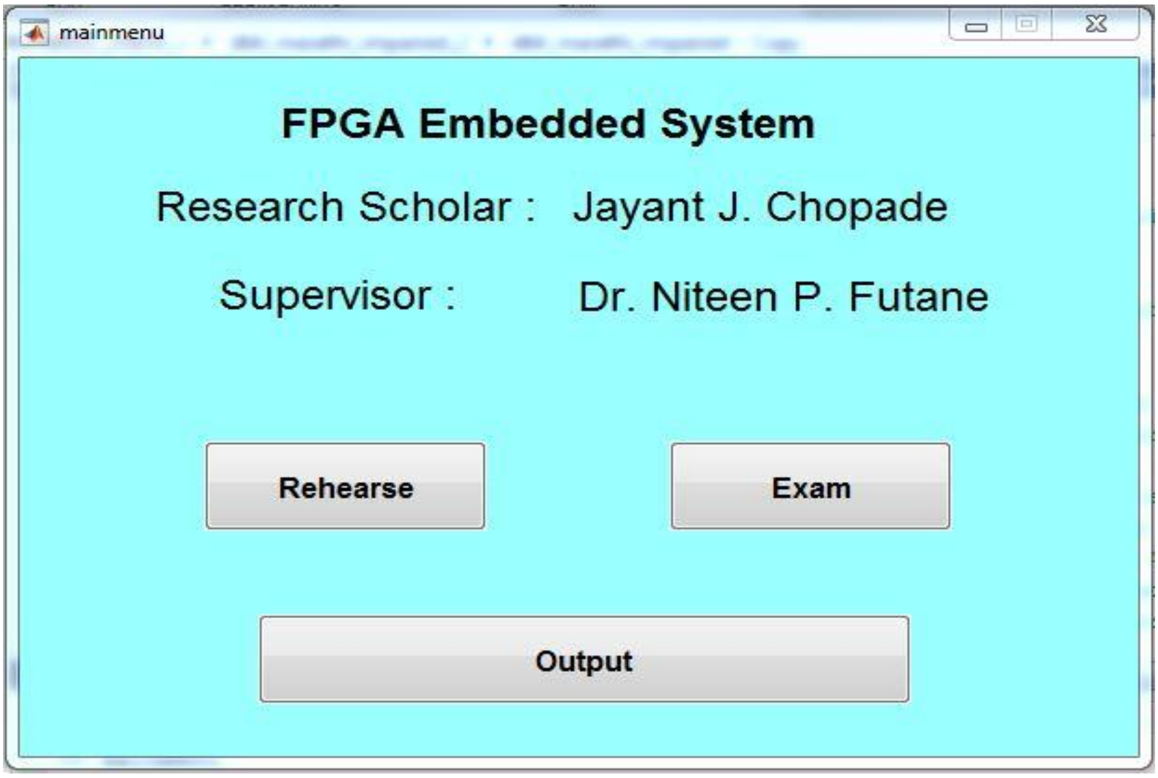
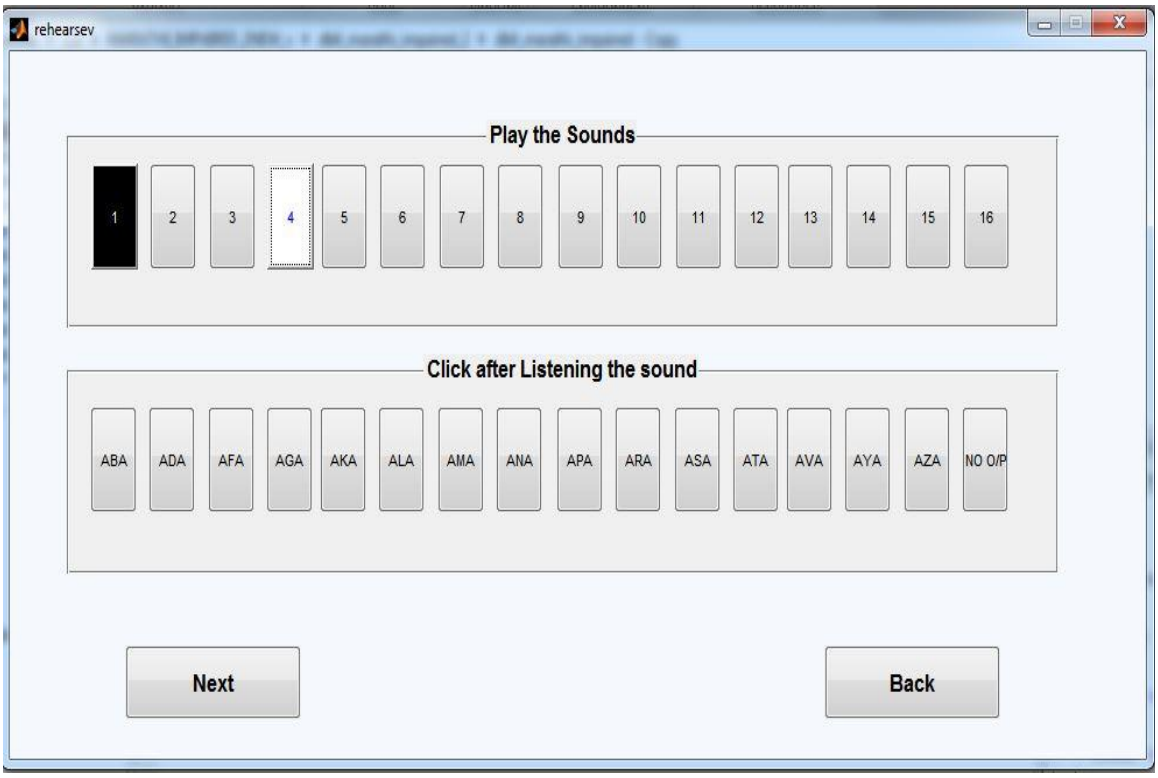
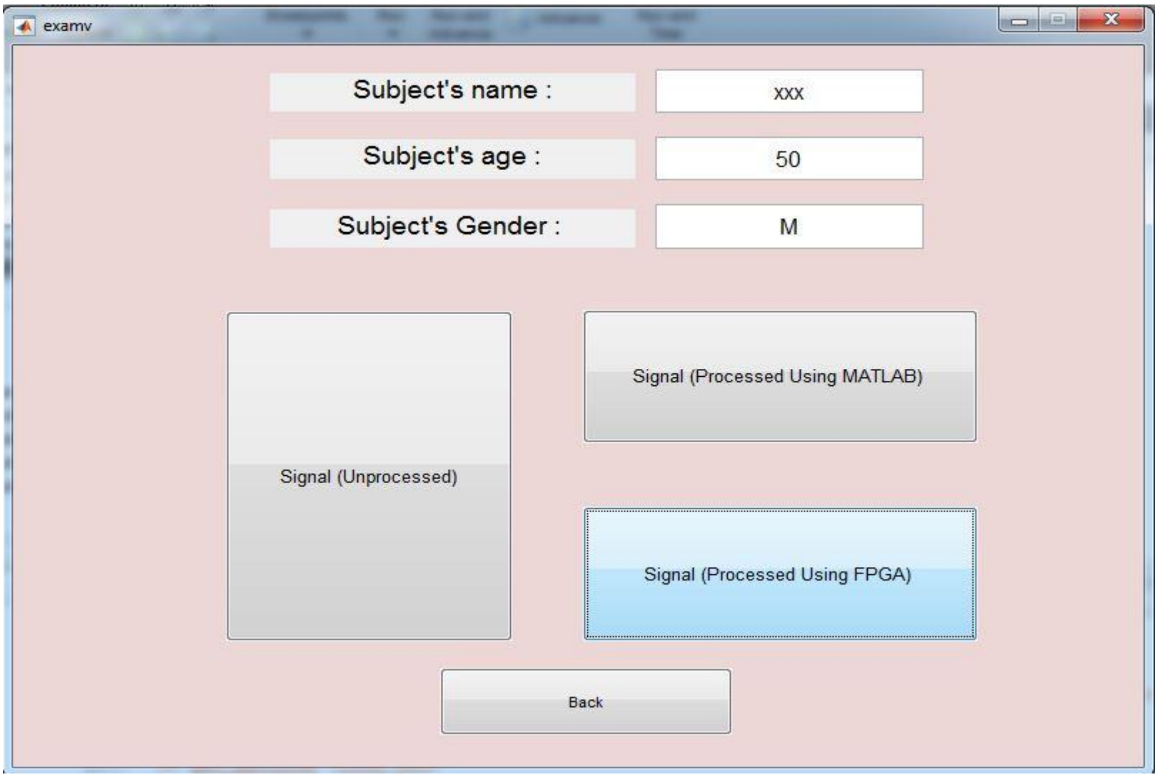
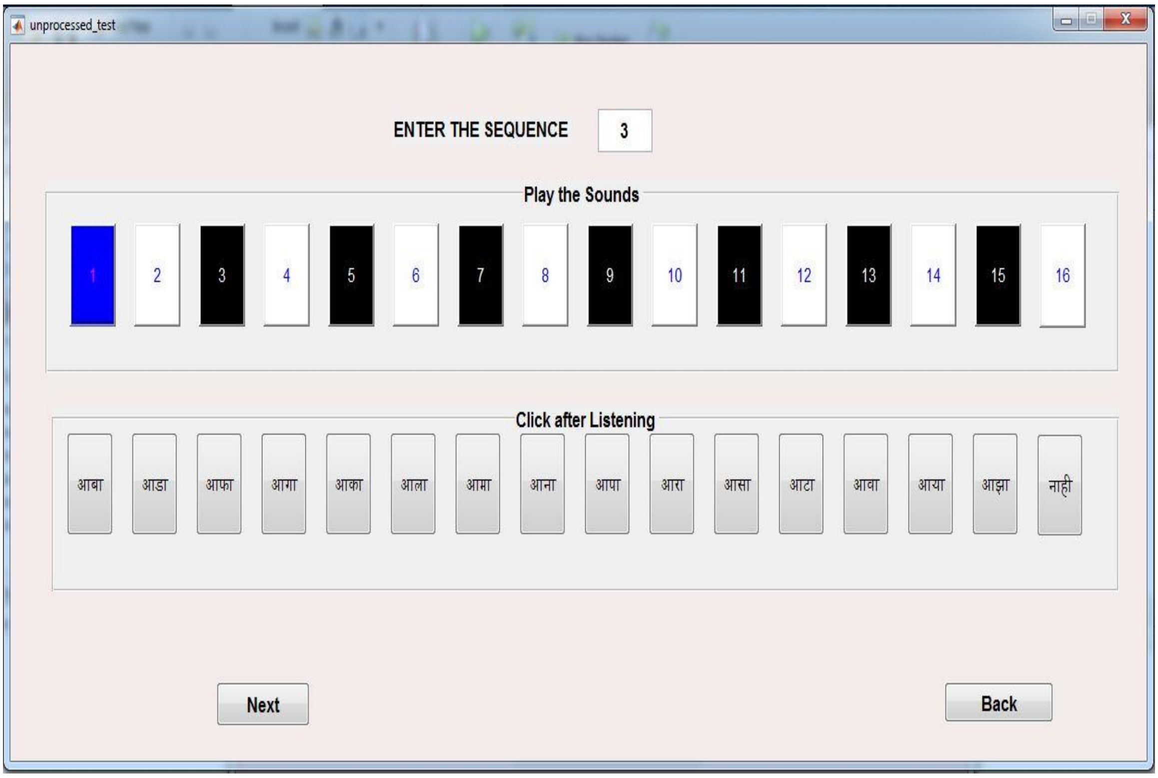


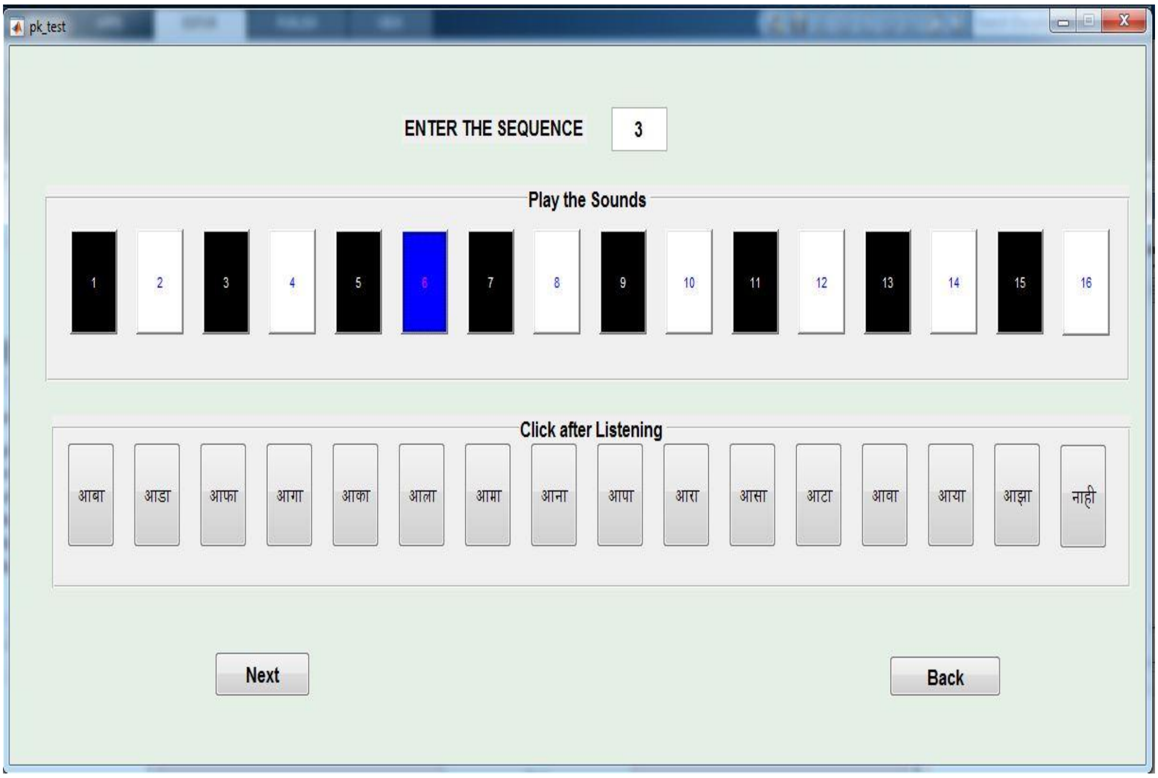
Fig. B.1: GUI for Experimentation on Normal People

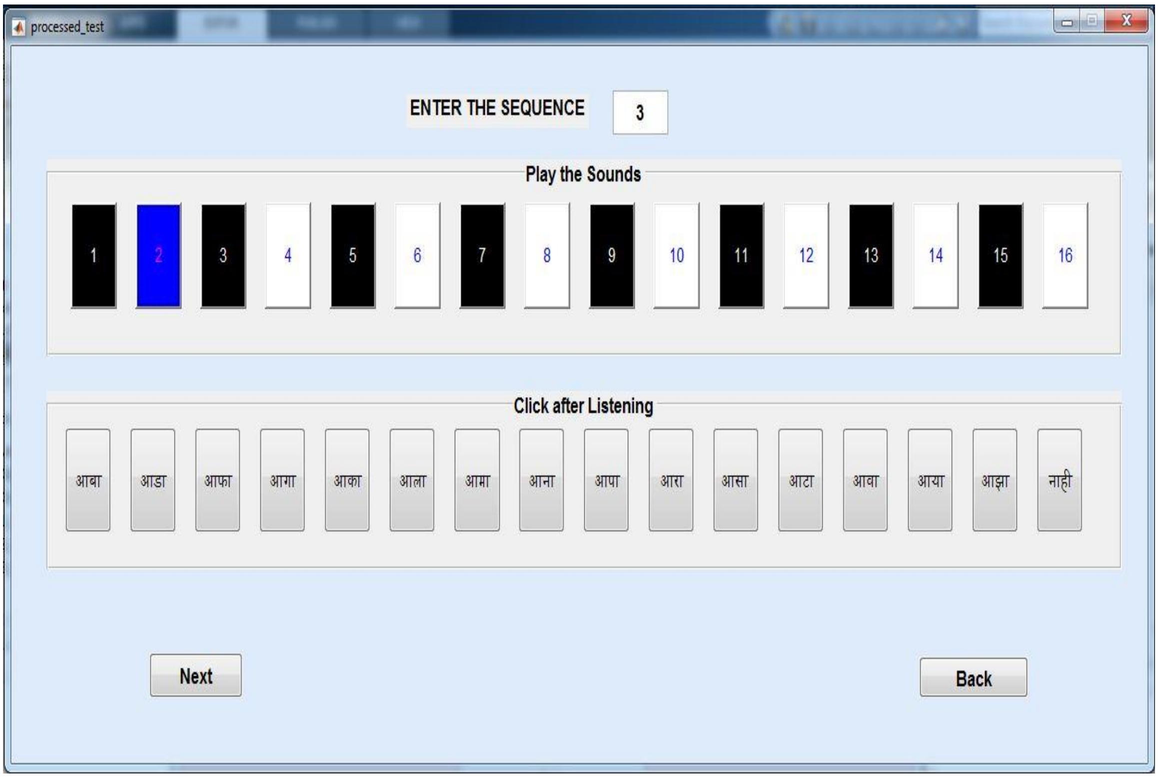

mainmenu
FPGA Embedded System
Research Scholar : Jayant J. Chopade
Supervisor : Dr. Niteen P. Futane
Rehearse
Exam
Output


rehearsev
Play the Sounds
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
Click after Listening the sound
ABA ADA AFA AGA AKA ALA AMA ANA APA ARA ASA ATA AVA AYA AZA NO O/P
Next
Back


examv
Subject's name :
xxx
Subject's age :
50
Subject's Gender :
M
Signal (Unprocessed)
Signal (Processed Using MATLAB)
Signal (Processed Using FPGA)
Back


unprocessed_test
ENTER THE SEQUENCE
3
Play the Sounds
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
Click after Listening
आबा आडा आफा आगा आका आला आमा आना आपा आरा आसा आटा आवा आया आझा नाही
Next
Back


pk_test
ENTER THE SEQUENCE
3
Play the Sounds
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
Click after Listening
आबा आडा आफा आगा आका आला आमा आना आपा आरा आसा आटा आवा आया आझा नाही
Next
Back


processed_test
ENTER THE SEQUENCE
3
Play the Sounds
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
Click after Listening
आबा आडा आफा आगा आका आला आमा आना आपा आरा आसा आटा आवा आया आझा नाही
Next
Back

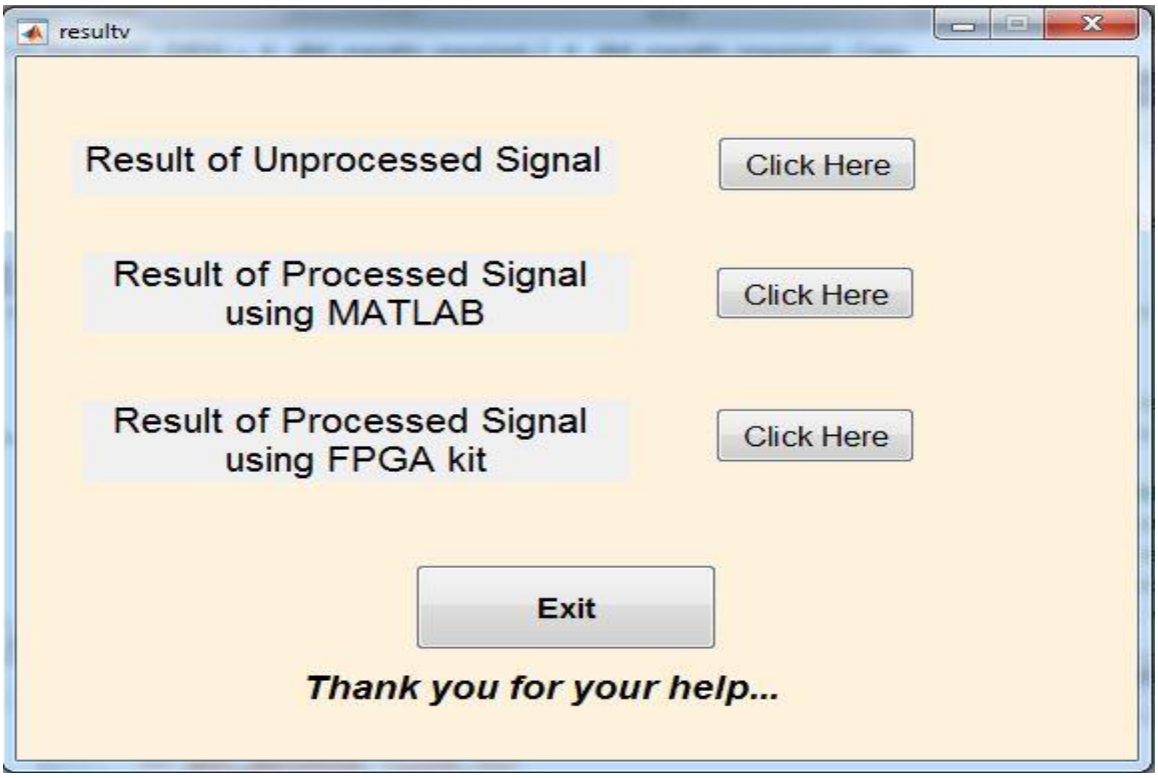


Fig. B.2 Listening test set-up for Hearing Impaired Subject

Appendix C
Test instructions to Normal and Hearing Impaired people.

The listening tests were carried out to determine the performance of comb based filter and wavelet based filters. The speech material used consisted of fifteen English nonsense syllables / aBa, aDa, aFa, aGa, aKa, aLa, aMa, aNa, aPa, aRa, aSa, aTa, aVa, aYa, aZa /, in VCV context with vowel /a/ as in father. A laptop based experimental setup with full automation was used. The instructions to the participant in the above listening test are mentioned in the paragraph given below.

The experimentation was carried out in an acoustically isolated room. A laptop is used to record the responses and headphone is used to hear the sounds.

Fifteen stimuli in scrambled manner are displayed in the interface (developed in MATLAB GUIDE) to the subjects for five times. So, the total number of presentation will be seventy five. To listen to the stimuli any number of times a provision was made, so as to get acquainted with it. Hereafter when they take up test, they have to press the corresponding key from the available choices displayed on the screen. The sequence and response choices with appropriate key to be pressed will be shown before every presentation. Pressing a key is necessary for the test to proceed further. If by mistake a presentation is skipped, you can press any key other than valid response key. Seventy five presentations will form one test run. For normal people the time taken to complete one test run is approximately 10 to 15 minutes. The same time may reach up to 20 to 25 minutes for hearing impaired subjects.

C.1 Prior Instructions to the Subjects for Listening tests

1. Welcome to the FPGA Embedded computerized listening session.
2. Instructions to follow
3. You can listen to the stimuli in any order you want by pressing the appropriate key as listed below.

Sounds	aBa	aDa	aFa	aGa	aKa	aLa	aMa	aNa	aPa	aRa	aSa	aTa	aVa	aYa	aZa	NO O/P
Key	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16

4. Hit ‘back’ key when you have finished the listening test. After listening the stimuli through headphones, press the corresponding button as quickly as possible.

5. A presentation will not be repeated. If you are not sure, you can guess.
6. The test will not proceed if you do not respond.
7. Please press button when you are ready for the test.

C.2 Form for Information about Normal and Hearing Impaired Subjects

Subject Information

Date:

Name: Subject Id:

Gender: M/F Age:

Contact Details:

Phone:

Occupation:

Mother Tongue:

Other Languages Known:

History of Noise Exposure:

History of hearing problems:

C.3 Form for subject's willingness to participate

<u>CONSENT FORM</u>

I have carefully read and understood the test instructions explained to me by Mr. Jayant J. Chopade for participation in listening experiments for evaluation of speech algorithms. I show my willingness to participate in the tests conducted by him.

Signature:

Name:

Contact Details:

Date:

PATENT FILED

Jayant J Chopade, Dr N P Futane, Indian Patent (Filed), 201621008170, 2016 " Hearing Aid for Hearing Impaired with minimum latency for better perception of the speech signal".

RESEARCH PUBLICATIONS

The following are the list of papers published and communicated

Papers under Review

1. Jayant J. Chopade, Dr N. P. Futane, “Effective FPGA Implementation of Comb Filters to Improve Perception of Sensorineural Hearing Impaired”, International Journal on Electrical Engineering and Informatics, ISSN 2085-6830 Submitted on 24/09/2015.

Papers Published in International Journals

1. Chopade, J. J., and Futane, N. P. (2016). Design of Optimized Wavelet Packet Algorithm to Improve Perception of Sensorineural Hearing Impaired. Journal of Signal and Information Processing, 7(01), p18-26
2. Chopade J., J., and Futane, N. P. (2015). Wavelet Based Scheme to Improve Performance of Hearing under Noisy Environment. International Journal of Computer Applications, 130 (6), 57-61.
3. J. J. Chopade, Dr. N. P. Futane, “Design of adaptive wavelet algorithm for Audibility enhancement”, International Journal of Applied Engineering Research, ISSN 0973-4562, Vol. 10, No. 7, 2015, pp-38343- 38348

Papers Published in International Conferences

J. J. Chopade, Dr. N. P. Futane, and M. N. Ingale,“Signal Processing Strategies for Speech Perception & Needed Research for Hearing Impaired”, Proceedings of 3rd International Conference on Recent Trends in Engineering & Technology, (ICRTET’2014), ISBN No:978-93-5107-22-8 date 28-30 March 2014 ELSEVIER Publication 2014.

REFERENCES

[1] CHABA, "Speech-perception aids for hearing-impaired people: Current status and needed research," J. Acoust. Soc. Am. vol. 90, 637–683, 1991.

[2] B. C. J. Moore, *An Introduction to Psychology of Hearing*, London, Academic, 4th ed., pp. 89-140, 1997

[3] Wall L. G., *Hearing for the Speech-Language Pathologist and Health Care Professional,* Butterworth-Heinemann, Boston, Massachusetts, 1995.

[4] D. S. Chaudhari and P. C. Pandey, "Dichotic presentation of speech signal using critical filter bank for bilateral sensorineural hearing impaired," Proceedings of 16th International Conf. on Acoustics, Seattle, Washington, Vol. 1, pp. 213-214, 1998.

[5] Chaudhari D. S., and Pandey P. C. "Dichotic presentation of speech signal with critical band filtering for improving speech perception," Proc. Of IEEE Int. Conf. Acoust., Speech and Signal Processing (ICASSP'98), Seattle, Washington, AE 3.1., 1998.

[6] Chaudhari D. S., and Pandey P. C. "Binaural audio in multimedia systems to improve auditory perception for the hearing impaired," Proc. 5th Int. Sym. Signal Processing and Its Applications (ISSPA'99), Brisbane, Australia., 1999.

[7] D. Chaudhari, "Binaural Dichotic Presentation of Speech Signal for Improving its Perception to Sensorineural Hearing - Impaired using Auditory Filters," IETE Technical Review, 2008.

[8] McAllister H. G., Black N. D., and Waterman N. "Hearing aids-a development with digital signal processing devices," Computing and Control Engineering Journal, vol. 6(6), pp. 283–291, 1995.

[9] Villchur E. "Signal processing to improve speech intelligibility in perceptive deafness," J. Acoust. Soc. Am. Vol. 53, pp. 1646–1657, 1973.

[10] Asano F., Suzuki Y., Sone T., Kakehata S., Satake M., Ohyama K., Kobayashi T., and Takasaka T., "A digital hearing aid system that compensates for lack of loudness," Proc. 15th Int. Sym. Speech and Hear. Science (ISSHS), pp. 122–133, 1991.

[11] Magotra N., Hamill T., and Swartz B., "Digital signal processing of speech for the hearing-impaired," Proc. of the Twenty-Ninth Asilomar Conference on Signals, Systems and Computers, vol. 2, pp. 1209–1213, 1995.

[12] Yasu K., Kobayashi K., Hishitani M., Arai T., and Murahara Y., "Critical band based frequency compression for digital hearing aids," J. Acoust. Sci. and Tech. vol. 25, pp. 61–63, 2004.

[13] Loizou P. C., Mani, Arun vijay, and Dorman M. F., "Dichotic speech recognition in noise using reduced spectral cues," J. Acoust. Soc. Am. vol. 114(1), pp. 475–483, 2003.

[14] Reed C. M., Hicks B. L., Braida L. D., and Durlach N. I., "Discrimination of speech processed by low-passed filtering and pitch-invariant frequency lowering," J. Acoust. Soc. Am. vol. 74, pp. 409–419, 1983.

[15] Munoz C. M. A., Nelson P. B., Rutledge J. C., and Gaga A., "Frequency Lowering Processing For Listeners With Significant Hearing Loss," Proc. of IEEE Int. Conf. on Electronics, Circuits and Systems (ASSP'99), vol. 2, pp. 741–744, 1999.

[16] Quatieri T. F., *Discrete-Time Speech Signal Processing Principles and Practice* (Pearson Education (Singapore) Pte. Ltd., Delhi, India), 2004.

[17] El Danasoury Iman MS, et al., "Efficacy of frequency transposition on speech perception and production in hearing impaired Arabic-speaking children," Egyptian Journal of Ear, Nose, Throat and Allied Sciences 14.2 pp. 129-135, 2013.

[18] Alnahwi M., Al Qudehy Z. A., "Comparison between frequency transposition and frequency compression hearing aids," The Egyptian Journal of Otolaryngology, 31(1), pp. 10–18, 2015.

[19] Nabelek A. K. and Robinson P. K., "Monoaural and binaural speech perception in reverberation for listeners of various ages," J. Acoust. Soc. Am. vol. 71(5), pp. 1242–1248, 1982.

[20] Lyregaard P. E., "Frequency selectivity and speech intelligibility in noise," Scand. Audiol. Suppl. 15, pp. 113-122, 1982.

[21] Lunner T., and Hellgren J. A, "Digital filter bank hearing aid- design, implementation, and evaluation," Proc. in IEEE ASSP, Vol. 5, pp. 3661–3664, 1991.

[22] Jangamashetti D. S. and Pandey, P. C., "Inter aural switching with different fading functions for dichotic presentation to reduce the effect of temporal masking in sensorineural hearing loss," 4th World multi conference on systemics, cybernetics and informatics (SCT2000), Orlando, F1.USA, pp. 346─353, 2000

[23] D. S. Jangamashetti, A. N. Cheeran, P. N. Kulkarni, and P. C. Pandey, "Simulation of increased masking in sensorineural hearing loss for a preliminary evaluation of speech processing schemes," Proc. 20th Int. Congress on Acoustics (ICA 2010), Sydney, Australia, paper. no. 980, 2010.

[24] Alice N. Cheeran, P. C. Pandey, and D. S. Jangamasheti,"Design of comb filters based on auditory filter bandwidths for binaural dichotic presentation for persons with sensorineural hearing impairment," in Proc. 14th Int. Conf. Digital Signal Processing (Santorini, Greece), pp. 971–974, 2002.

[25] A. N. Cheeran, and P. C. Pandey, "Dichotic presentation to overcome the effect of increased spectral masking and frequency dependent hearing threshold shifts in persons with bilateral sensorineural hearing impairment," *J. Acoust. Soc. Am*., vol. 112, pp. 23-59, 2003.

[26] P. N. Kulkarni and P.C. Pandey, "Optimizing the comb filters for spectral splitting of speech to reduce the effect of spectral masking," International conference on signal processing, Communication and Networking, Chennai, India, pp.69-73, Jan 2008.

[27] P. N. Kulkarni, P. C. Pandey, and D. S. Jangamashetti, "Multi-band frequency compression for reducing the effects of spectral masking," International Journal of Speech Technology, vol. 10(4), pp. 219-227, 2009.

[28] P. N. Kulkarni, P. C. Pandey, and D. S. Jangamashetti, "Study of perceptual balance for designing comb filters for binaural dichotic presentation," Proc. 20th Int. Congress on Acoustics (ICA 2010), Sydney, Australia, paper. no. 556, 2010.

[29] Kulkarni, Pandurangarao N., Prem C. Pandey, and Dakshayani S. Jangamashetti, "Binaural dichotic presentation to reduce the effects of spectral masking in moderate bilateral sensorineural hearing loss" International journal of audiology 51.4, pp. 334-344, 2012.

[30] Kaibao N., Ning L., and Shangkai G., "Wavelet-Based Feature Extraction Of Speech Signal For

Cochlear Implants,” Proc. of the First Joint BMES/EMBS Conference Serving Humanity, Advancing Technology, Atlanta, GA, USA, 1999.

[31] Drake, Laura A., Rutledge, Janet C., and Cohen, J., “Wavelet Analysis in Recruitment of Loudness Compensation,” IEEE Transactions on Signal Processing, vol. 41(12), pp. 3306–3309, 1993.

[32] Agbinya J. I., “Discrete Wavelet Transform Techniques in Speech Processing,” Proc. IEEE TENCON - Digital Signal Processing Applications, vol. 2, pp. 514–519, 1996.

[33] Yao J., and Zhang Y., “The Application of Bionic Wavelet Transform to Speech Signal Processing in Cochlear Implants using Neural Network Simulations,” IEEE Transactions on Biomedical Engineering, vol. 49(11), pp. 1299–1309, 2002.

[34] Singh L. and Sridharan S., “Speech enhancement for forensic application using dynamic time warping and wavelet packet analysis,” Proc. IEEE TENCON-SITCT, vol. 2, pp. 475–478, 1997.

[35] Lu C. T. and Wang H. C., **“**Speech enhancement using perceptually-constrained gain factors in critical band wavelet packet transform,” Electronics Letters, vol. 40 (6), pp. 394**–**396**,** 2004.

[36] Nogueira W., Giese A., Edler B. and Buchner A., “Wavelet Packet Filter bank for Speech Processing Strategies in Cochlear Implants,” Proc. IEEE Int. Conf. Acoustic Speech, Signal Processing (ICASSP ’06). vol. 5, pp. 14–19, 2006.

[37] Karmakar A., Kumar A. and Patney R. K., “A Multiresolution Model of Auditory Excitation Pattern and Its Application to Objective Evaluation of Perceived Speech Quality,” IEEE Transactions on Audio, Speech and Language Processing, vol. 14(5), pp. 1912–1923, 2006.

[38] Mahesh T. Kolte, D. S. Chaudhari, “Evaluation of speech processing schemes to improve perception of sensorinural hearing impaired,” Current Science, Vol.98, No 5, 2010.

[39] M. Mehrzad, M. D. Abolhassani, A. H. Jafari, J. Alirezaie, and M. Sangargir, “Cochlear Implant Speech Processing Using Wavelet Transform,” International Scholarly Research Network ISRN Signal Processing, Volume 2012, Article ID 628706, pp. 1-6.

[40] Zhimin Lin, Zhang Linghua and Yang Zi., "A Single-channel Loudness Compensation Method for Digital Hearing Aids Based on Multi-resolution Wavelet," Instrumentation, Measurement, Computer, Communication and Control (IMCCC), Second International Conference on. IEEE, 2012.

[41] Girisha G. K. and S. L. Pinjare, "Wavelet Transform Based Multi Rate Filter Bank for Human Auditory System," Emerging Research in Computing, Information, Communication and Applications. Springer India, pp. 549-556, 2016.

[42] Kambalimath, Shankarayya G., et al., "FPGA based implementation of comb filters for use in binaural hearing aids for reducing intraspeech spectral masking," 2013 Annual IEEE India Conference (INDICON), IEEE, 2013.

[43] S. G. Kambalimath, P. C. Pandey, P. N. Kulkarni, S. S. Mahant-Shetti, and S. G. Hiremath, “FPGA-Based Implementation of Comb Filters Using Sequential Multiply-Accumulate Operations for Use in Binaural Hearing Aids spectral masking,” Proc. Of IEEE INDICON, pp.1–5, 2014.

[44] Kambalimath, Shankarayya G., et al. "FPGA-based Design of a Hearing Aid with Frequency Response Selection through Audio Input." 29th International Conference on VLSI Design and 15th International Conference on Embedded Systems (VLSID). IEEE, 2016.

[45] N. Tiwari, P. C. Pandey, and P. N. Kulkarni, “Real-time implementation of multi-band frequency compression for listeners with moderate sensorineural impairment,” Proc. 13th Annual Conference of the International Speech Communication Association (Interspeech 2012), Portland, Oregon, September 9-12, paper no. 689, 2012.

[46] Nivin Ghamry, “FPGA Implementation of Hearing Aids using Stationary Wavelet-Packets for Denoising,” International Journal of Computer Applications (0975 – 8887) Volume 69, No.15, May 2013.

[47] L. Bendaouia, S.M. Karabernou, L. Kessal, H. Salhi, “Implementation on FPGA and Evaluation Of a Prosody Modification of Speech for Impaired Persons using DWT-OLA,” International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering. Vol.3, pp. 11389–11395, August 2014.

[48] Valerie Hanson and Kofi Odame, “Real-Time Embedded Implementation of the Binary Mask Algorithm for Hearing Prosthetics,” IEEE Transactions on Biomedical Circuits and Systems, Vol. 8, No.4, August 2014.

[49] J. R. Dubno and A. B. Schaefer, “Comparison of frequency selectivity and consonant recognition among hearing-impaired and masked normal-hearing listeners,” J. Acoust. Soc. Am., vol. 91(4), pp. 2110–2121, 1992.

[50] Tyler, R. S., and Moore, B. C. J., “Consonant recognition by some of the better cochlear-implant patients,” J. Acoust. Soc. Am. vol. 92, pp. 3068–3077, 1992.

[51] ter Keurs, M., Festen, J. M., and Plomp R., “Effect of spectral envelope smearing on speech reception. I” J. Acoust. Soc. Am. vol. 91, pp. 2872–2880. 1992.

[52] ter Keurs, M., Festen, J. M. and Plomp R. “Effect of spectral envelope smearing on speech reception. II,” J. Acoust. Soc. Am. vol. 93, pp. 1547–1552, 1993.

[53] Dubno J. R. and Dirks D. D., “Auditory filter characteristics and consonant recognition for hearing-impaired listeners,” J. Acoust. Soc. Am. vol. 85, pp. 1666–1675, 1989.

[54] Baer T. and Moore B. C. J., “Effects of spectral smearing on the intelligibility of sentences in noise,” J. Acoust. Soc. Am. vol. 94, pp. 1229–1241, 1993.

[55] Turner C. W. and Robb M. P., “Audibility and recognition of stop consonants in normal and hearing-impaired subjects,” J. Acoust. Soc. Am. vol. 81, pp. 1566–1573, 1987.

[56] Lippmann R., Braida L. and Durlach N., “Study of multichannel amplitude compression and linear amplification for persons with sensorineural hearing loss,” J. Acoust. Soc. Am. vol. 69, pp. 524–534, 1981.

[57] Miller G. A. and Nicely P. E., “An analysis of perceptual confusions among some English consonants,” J. Acoust. Soc. Am. vol. 27(2), pp. 338–352, 1955.

[58] Hou Z. and Pavlovic C. V., “Effects of temporal smearing on temporal resolution, frequency selectivity, and speech intelligibility,” J. Acoust. Soc. Am. vol. 96, pp. 1325–1340, 1994.

[59] Humes L. E., Dirks D. D., Bell T. S. and Kincaid G. E., “Recognition of nonsense syllables by hearing-impaired listeners and by noise-masked normal hearers,” J. Acoust. Soc. Am. vol. 81, pp. 765–773, 1987.

[60] Revoile Sally G., et al., "VCVs vs CVCs for stop/fricative distinctions by hearing-impaired and normal-hearing listeners," The Journal of the Acoustical Society of America 89.1, pp. 457-460, 1991.

[61] Pickett J. M., *The Acoustics of Speech Communication: Fundamentals, Speech Perception Theory, and Technology* (Allyn and Bacon, Boston, Massachusetts), 1999.

[62] Baskent, Deniz, “Speech recognition in normal hearing and sensorineural hearing loss as a function of the number of spectral channels,” J. Acoust. Soc. Am. vol. 120(5), pp. 2908–2925, 2006.

[63] Tipler P. A., *Physics for Scientists and Engineers, Vol. 1: Mechanics, Oscillations and Waves and Thermodynamics*, London, W. H. Freeman, 1998.

[64] Sataloff, R. T. and Sataloff J, *Hearing Loss* (Marcel Dekker, New York), 1993.

[65] Pickles J. O., *An Introduction to Psychology of Hearing* (Academic, London), 1982.

[66] Zwicker, Eberhard and Karin Schorn, "Psychoacoustical tuning curves in audiology," *Audiology* 17.2, pp. 120-140, 1978

[67] Wilson R.H., Arcos J.T. and Jones H.C., “Word recognition with segmented-alternated CVC words: A preliminary report on listeners with normal hearing,” J. Speech Hear. Res. 27, pp. 378-386, 1984.

[68] Levitt H., Pickett J. M. and Houde R. A., *Eds. Sensory Aids for the Hearing Impaired*, IEEE Press, New York, 1980.

[69] Baskent and Deniz, “Speech recognition in normal hearing and sensorineural hearing loss as a function of the number of spectral channels,” J. Acoust. Soc. Am., vol. 120(5), pp. 2908–2925, 2006.

[70] Fletcher H., *Speech and Hearing in Communication*, Van Nostrand, New York, 1953.

[71] E. W. Zwicker, “Subdivision of audible frequency range into critical bands,” (Freqenzgruppen), J. Acoust . Soc. Am., Vol. 33, pp. 248, 1961.

[72] Dubno J. R. and Schaefer A. B., “Comparison of frequency selectivity and consonant recognition among hearing-impaired and masked normal-hearing listeners,” J. Acoust. Soc. Am. vol. 91(4), pp. 2110–2121, 1992.

[73] Moore B. C. J. and Glass berg B. R., “Suggested formulae for calculating auditory filter bandwidths and excitation patterns,” J. Acoust. Soc. Am. vol. 74, pp. 750-753, 1983.

[74] Patterson, Roy D., "Auditory filter shapes derived with noise stimuli." The Journal of the Acoustical Society of America 59.3, pp. 640-654, 1976.

[75] Fitzgibbons P. J. and Wightman F. L., “Gap detection in normal and hearing-impaired listeners,” J. Acoust. Soc. Am. 72, pp. 761-765, 1982.

[76] Nejime Y. and Moore B. C. J., “Evaluation of the effects of speech-rate slowing on speech intelligibility in noise using a simulation of cochlear hearing loss,” J. Acoust. Soc. Am. vol. 103, pp. 572–576, 1998.

[77] Leek M. R. and Summers V., “Reduced frequency selectivity and the preservation of speech contrast in noise,” J. Acoust. Soc. Am. vol. 100, pp. 1796–1806, 1996.

[78] Zeng F. and Turner C. W., “Recognition of voiceless fricatives by normal and hearing-impaired-subjects,” J. Speech Hear. Res. 33, pp. 440-449, 1990.

[79] Dubno J. R. and Levit H., “Predicting consonant confusions from acoustic analysis,” J. Acoust. Soc. Am. vol. 69(1), pp. 249–261, 1981.

[80] Leek M. R., Dorman M. F. and Summerfield Q., “Minimum spectral contrast for vowel identification by normal-hearing and hearing-impaired listeners,” J. Acoust. Soc. Am. vol. 81, pp. 148–154, 1987.

[81] Hannley M. and Dorman M. F., “Susceptibility to intraspeech spread of masking in listeners with sensorineural hearing loss,” J. Acoust. Soc. Am. vol. 74, pp. 40–51, 1983.

[82] Loizou, P. C., “Mimicking the human ear,” IEEE Signal Processing Magazine, vol. 15(5), pp. 101–130, 1998.

[83] Crandall I. B., “The composition of speech,” Phys. Rev. 10(2), pp. 26-28, reprinted in Speech Intelligibility and Speaker recognition, edited by M.E. Hawley (Dowden Hutchinson Ross, Stroudsburg, Pennsylvania), 1917.

[84] Florentine M. and Buus S., “Temporal gap detection in sensorineural and simulated hearing impairments,” J. Speech Hear. Res. 27, pp. 449-455, 1984.

[85] Flanagan, J. L., *Speech Analysis Synthesis and Perception*, Springer-Verlag, New York, 1972.

[86] Tyler R. S. and Moore B. C. J., “Consonant recognition by some of the better cochlear-implant patients,” J. Acoust. Soc. Am. 92, pp. 3068-3077, 1992.

[87] Mullin W. J., George W. J., Mestre J. P. and Velleman S. L., *Fundamentals of sound with applications to speech and hearing*, Boston, Allyn & Bacon, 2003

[88] Duke, Thomas, Licher & Frank, “Active traveling wave in the cochlea,” Physical review letters, 90, 2003.

[89] Oppenheim A. V., Schafer R. W. and Buck J. R., *Discrete-time Signal Processing* (Prentice Hall, Englewood Cliffs, New Jersey), 1999.

[90] Proakis J. G. and Manolakis D. G., *Digital Signal Processing Principles, Algorithms, and Applications,* Prentice Hall, New Delhi, India, 1997.

[91] Rabiner L. R. and Schafer R. W., *Digital Processing of Speech Signals*, Prentice Hall, Englewood Cliffs, New Jersey, 1978.

[92] L. R. Rabiner and R. W. Schafer, *Digital Processing of the Speech Signals*, Englewood Cliffs, NJ: Prentice Hall, pp.35-53, 1978.

[93] *Altys Board Reference Manual* (REV.C of the board), Digilent Inc.

[94] Jeremy Agnew, Jeffrey M Thornton, “Just Noticeable and Objectionable Group Delays in Digital Hearing Aids,” J. Amer. Acad. Audio, vol.11, No.6, pp. 330-336, Jun.2000.

[95] Rao R. M. and Bopardikar A. S., *Wavelet Transform Introduction to Theory and Applications*, Addison Wesley Longmman Pte. Ltd., Delhi, 2001.

[96] Soman K. P. and Ramachandran K. I., *Insight into Wavelets from Theory to Practice,* Prentice-Hall of India, New Delhi, 2004.

[97] Daubechies I., *Ten Lectures on Wavelets,* Philadelphia: SIAM, CBMS- NSF Regional Conference in Applied Mathematics 61, 1992.

[98] Daubechies I. “Where do Wavelets Come From?-A Personal Point of View,” Proc. of IEEE, vol. 84(4), pp. 510–513, 1996.

[99] Vetterli M. and Kovacevic J., *Wavelets and Subband Coding*, Prentice Hall PTR, Englewood Cliffs, New Jersey, 1995.

[100] Burrus C. S., Gopinath R. A. and Guo H., *Introduction to Wavelets and the Wavelet Transforms A Primer* Prentice Hall, Upper Saddle River, NJ, 1998.

[101] Rioul O. and Vetterli M., "Wavelet and Signal Processing," IEEE SP Magazine, vol. 8(4), pp. 14–38, 1991.

[102] Mallat, S. G., *A Wavelet Tour of Signal Processing*, Academic Press, Chestnut Hill, MA, 1998.

[103] Herley C. *Wavelets and filter banks*, V. Madisetti and D. Williams Ed., The Digital Signal Processing Handbook CRC Press, 1999.

[104] Vaidyanathan P. P., *Multirate Systems and Filter Banks*, Prentice Hall Englewood Cliffs, NJ, 1993.

[105] Loizou P. C., "Signal processing techniques for cochlear implants," IEEE Magazine in Medicine and Biology Engg. Vol. 18(3), pp. 34–46, 1999.

[106] MATLAB HDL coder™ user guide, 2014b, Mathworks Inc., 2014

[107] Xilinx University Program XUPV5-LX110T Development System, available at: http://www.xilinx.com/support/university/boards-portfolio/xup-boards/AtlysBoard.html