SPEECH PROCESSING WITH DICHOTIC PRESENTATION FOR BINAURAL HEARING AIDS FOR MODERATE BILATERAL SENSORINEURAL LOSS

submitted in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

by

ALICE NOBLE CHEERAN
(Roll No. 99430303)

under the supervision of

Prof. P. C. Pandey

Biomedical Engineering Group
School of Biosciences and Bioengineering

Indian Institute of Technology, Bombay

2005

This thesis is dedicated to

the memory of my father

V.V. Samuel

Indian Institute of Technology, Bombay

Thesis Approval

Thesis entitled : Speech processing with dichotic presentation for binaural hearing aids for moderate bilateral sensorineural loss

by **Alice Noble Cheeran**

is approved for the award of the degree of **Doctor of Philosophy**.

Supervisor	: PCPandey	(Prof. P. C. Pandey)
Internal Examiner	: Preeti Rao	(Prof. P. Rao)
External Examiner	: S Sanyal	(Prof. S. Sanyal)
Chairman	: K Kurien Issac	(Prof. K. K. Issac)

Date: December 13, 2005

Indian Institute of Technology, Bombay, India

CERTIFICATE OF COURSE WORK

This is to certify that Ms. Alice Noble Cheeran was admitted to the candidacy of Ph.D. Degree in January 2000, after successfully completing all the courses required for the Ph.D. Degree Programme. The details of the course work done are given below.

Sr. No.	Course No.	Course Name	Credits
1.	BM 603	Physiology for Engineers	6.00
2.	BM 607	Physiological Measurements and Quantitative Physiology	10.00
3.	BM 695	Credit Seminar	4.00
4.	EE 603	Digital Signal Processing and its Applications	6.00
5.	EE 635	Applied Linear Algebra in Electrical Engg.	6.00
6.	BMS802	Seminar	6.00

I.I.T. Bombay

Dy. Registrar (Academic)

Dated:

A. N. Cheeran. “Speech processing with dichotic presentation for binaural hearing aids for moderate bilateral sensorineural loss”, *Ph.D. Thesis*, Biomedical Engineering Group, School of Biosciences and Bioengineering, IIT Bombay, 2005, Supervisor: P. C. Pandey.

Abstract

Sensorineural hearing impairment is characterized by frequency dependent shifts in hearing threshold, loudness recruitment, and increased spectral and temporal masking. Increased masking causes smearing of spectral peaks and suppression of cues like voice-onset-time, formant transitions, and burst durations, resulting in degraded consonantal identification. Masking takes place primarily at the peripheral level, while integration of information takes place at higher levels in the auditory system. For persons with moderate bilateral sensorineural impairment, the effect of increased masking may be reduced by splitting speech into two complementary signals, so that the signal components likely to mask or get masked are presented to different ears. The objective of the research is to investigate the use of dichotic presentation, with spectral, temporal, and combined splitting schemes, for improving speech perception by hearing impaired persons using binaural aids and by normal hearing persons under adverse listening conditions, in order to find the optimal splitting scheme and associated processing parameters.

For spectral splitting, perceptually balanced comb filters based on auditory critical bandwidths were designed as 256-coefficients linear phase FIR filters and had 1 dB passband ripple, 30 dB stop-band attenuation, and inter-band crossovers within 4 – 6 dB. Listening tests involving closed set identification of VCV syllables were conducted on five normal hearing subjects with simulated loss and on five hearing impaired subjects. On the basis of response time, recognition scores, and relative information transmission, the perceptually balanced comb filters were found to be superior to the comb filters with sharp inter-band transitions. Further investigations involved implementation and evaluation of the three speech processing schemes: (i) spectral splitting with perceptually balanced comb filters, (ii) temporal splitting with trapezoidal fading, 70 % duty cycle, 3 ms transition duration, and inter-aural switching period T_c = 20 – 80 ms, and (iii) combined splitting with time-varying comb filters, realized with a set of m (4 – 16) perceptually balanced comb filters swept over a cycle time T_c = 20 – 180 ms. The objective was to find the optimal value of T_c for temporal splitting, optimal values of T_c and m for combined splitting and to compare the three schemes. Listening tests were carried out, with phonetically balanced monosyllables, on 7 normal hearing subjects with simulated loss and 13 subjects with bilateral sensorineural hearing loss.

Test results showed that all the presentation schemes reduced the load on perception process and improved speech perception. The optimal processing conditions were T_c = 20 – 40 ms for temporal splitting, and T_c = 40 – 80 ms with m = 8 and 16 for combined splitting. The highest improvements in response time as well as recognition score were generally provided by spectral splitting, closely followed by combined splitting. On normal hearing subjects the SNR advantage at 60 % recognition score with best processing parameters was 5 dB for spectral splitting, 1.5 dB for temporal splitting, and 4 – 4.5 dB for combined splitting. For hearing impaired subjects the effectiveness of the schemes varied across subjects, which can be related to the extent and nature of loss. The relative improvements, for the optimal scheme and parameters for individual subject, ranged 6 – 138 % with an average of 35 %. It may be concluded that dichotic processing schemes can improve speech perception for persons using binaural hearing aids. The implementation should permit selection of the dichotic splitting scheme and fine-tuning of processing parameters.

Contents

Appendices

List of figures

List of tables

List of symbols and abbreviations

Symbols

$A_c(f)$	Adjustable magnitude response filter gain in dB.
d	duty cycle
$e(n)$	broadband noise
f_s	sample frequency
F	variance ratio
F_{rs}	frequency resolution
$H(e^{j\omega})$	frequency response
$h(n)$	impulse response
$I(x)$	Information measure of stimulus x, in bits
$I(y)$	Information measure of response y, in bits
$I(x, y)$	Information measure of stimulus x to response y, in bits
$I_{rel}\,(x; y)$	Relative information transmitted
L	"on" period in samples
m	number of shiftings
M	transition duration in samples
N	inter-aural switching duration in samples
p	significance level, probability
$p(x_i)$	probability of stimulus x_i
$p(y_j)$	probability of response y_j
$p(x_i, y_j)$	probability of (x_j, y_j)
R_s	recognition score
$s(t)$	analog signal
$s(n)$	digital signal
$S(n, k)$	short-time Fourier transform of $s(n)$
T_c	inter-aural switching period in temporal splitting with trapezoidal fading functions and sweep cycle duration in combined splitting with time-varying comb filters.
$w(n)$	fading function
$\alpha(f)$	hearing loss in dB
α_{max}	maximum hearing loss over 125 Hz to 5 kHz frequency range
α_{min}	minimum hearing loss over 125 Hz to 5 kHz frequency range
β	scaling factor
σ_s	short-time energy of signal $s(n)$
σ_e	short-time energy of broad band noise $e(n)$
Δf	transition width from pass to stop band in the comb filter magnitude response

Abbreviations

ADC	analog-to-digital converter
ANOVA	analysis of variance
AGC	automatic gain control
Avg.	average
AYJNIHH	Ali Yaver Jung National Institute for Hearing Handicapped
BTE	behind the ear
CB	critical band
CIC	completely-in-the-canal
CF	characteristic frequency
CV	consonant-vowel
CVC	consonant-vowel-consonant
CVR	consonant-to-vowel intensity ratio
DAC	digital-to-analog converter
DSP	digital signal processing
dB	decibel
df	degree of freedom
DL	difference limen
ERB	equivalent rectangular bandwidth
F1	first formant
F2	second formant
FFT	fast Fourier transform
FIR	finite impulse response
HL	hearing level
HTL	hearing threshold level
ITC	in the canal
ITE	in the ear
JND	just-noticeable difference
MCL	most comfortable level
MLP	mean logarithmic probability
NST	nonsense syllables test
PB	phonetically balanced
PC	personal computer
PTA	pure tone threshold average
PTC	psychophysical tuning curves
R.I.	relative improvement
RS	recognition score
RT	response time
S	subject
s.d.	standard deviation
SL	sensational level
SNR	signal-to-noise ratio
SpA	spectral splitting with comb filters with sharp transitions between bands.

Sp_AG-CS-T_c/m	combined splitting with time-varying comb filters of sweep cycle duration of T_c ms and number of shiftings *m*. with adjustable magnitude response filter cascaded
Sp_AG-SS	spectral splitting with perceptually balanced comb filters, with adjustable magnitude response filter cascaded
Sp_AG-TS-T_c	temporal splitting with trapezoidal fading functions of intra-aural switching period of T_c ms, with adjustable magnitude response filter cascaded
SpB	spectral splitting with perceptually balanced comb filters
SpC	spectral splitting with perceptually balanced comb filters, with adjustable magnitude response filter cascaded
Sp_CS_T_c/m	combined splitting with time-varying comb filters of sweep cycle duration of T_c ms and number of shiftings *m*.
SPL	sound pressure level
Sp_SS	spectral splitting with perceptually balanced comb filters
Sp_TS_T_c	temporal splitting with trapezoidal fading functions of intra-aural switching period of T_c ms
SRT	speech recognition threshold
SST	synthetic sentences
STFT	short time Fourier transform
Su	unprocessed speech
TI	Texas Instruments
UCL	uncomfortable level
VC	vowel-consonant
VCV	vowel-consonant-vowel
VOT	voice-onset time

Chapter 1

INTRODUCTION

1.1 Overview

Sensorineural hearing impairment, which is associated with loss of hair cells in the cochlea or degeneration of auditory nerve fibers or both, cannot be treated medically. The characteristics of this loss are frequency dependent shifts in hearing threshold, loudness recruitment, reduced frequency and temporal resolution, and increased spectral and temporal masking. Frequency selective amplification and amplitude compression techniques are being used in hearing aids to overcome frequency dependent shifts in hearing threshold and loudness recruitment.

Masking is the phenomenon in which presence of one signal component elevates the threshold of neighboring signal components. Increased spectral masking acts as though the auditory filters are broader than normal. The clarity of speech reduces due to smearing of peaks and valleys of the speech spectrum. Increased temporal masking leads to increase in forward and backward masking of weak acoustic segments by adjacent strong ones. Cues like voice-onset time, formant transitions, and burst duration, which are important for the identification of consonants, get masked by the following or preceding vowel segment, resulting in degraded speech perception.

Masking takes place primarily at the peripheral level, while integration of information takes place at higher levels in the auditory system. Splitting of speech into two complementary signals and presenting them dichotically, in such a way that the signal components, likely to mask or get masked, are presented to different ears can be used for reducing the effect of increased masking in persons with moderate bilateral sensorineural hearing loss, i.e. persons having some residual hearing in both the ears.

Speech processing schemes based on spectral splitting, using a pair of comb filters with complementary magnitude responses based on auditory critical bandwidths, have helped in improving the speech perception by reducing the effect of increased spectral masking (Chaudhari and Pandey, 1998a, b). These comb filters designed with sharp transition between bands had perceptual distortion due to passband ripple of 4 dB, stop-band attenuation of 10 dB and inter-band crossovers lying between 0 to 10 dB. A possible solution to this problem is to design perceptually balanced comb filters with improved magnitude responses. Spectral splitting using such comb filters may further improve the speech perception of sensorineural hearing impaired. Temporal splitting, in which speech was switched between two ears using trapezoidal fading function with an inter-aural switching period of 20 ms, has helped in reducing the effect of increased temporal masking (Jangamashetti, 2003; Jangamashetti and Pandey, 2000b). Effect of inter-aural switching period on the speech perception needs to be investigated.

In the scheme of spectral splitting, the sensory cells corresponding to alternate bands of the basilar membrane are always stimulated, whereas sensory cells of the other bands do not receive stimulation. In the temporal splitting scheme, all the sensory cells of the ears get relaxed alternately for some time. A combined splitting scheme can be suitably devised to provide all the sensory cells of the basilar membrane periodic relaxation from stimulation, and thereby achieve a simultaneous reduction in the adverse effects of increased spectral and temporal masking. In an investigation in our lab, combined splitting was implemented using a pair of time varying comb filters with pre-calculated sets of coefficients (Jangamashetti, 2003; Jangamashetti *et al.*, 2001; Pandey *et al.*, 2001), which were selected in steps for a cyclic sweeping of magnitude responses such that the pass bands of each of these comb filter pairs are shifted in a complementary manner along the frequency axis. The scheme was implemented for 2, 4, 8, and 16 shiftings, for a constant sweep cycle of 20 ms. There is a possibility that implementation with other sweep cycle duration with these number of shiftings, may provide better speech intelligibility.

An overall evaluation of the schemes of spectral, temporal, and combined splitting may help in determining the best possible scheme and its processing conditions, for improving the speech perception by persons with bilateral sensorineural loss, as well as by normal hearing persons under adverse listening conditions.

1.2 Research objective

The research objective is to investigate the use of binaural presentation for improving the speech perception by persons with moderate bilateral sensorineural hearing impairment and by normal hearing persons under adverse listening conditions. This has been carried out in two phases. In the first phase of this research, perceptually balanced comb filters have been designed with improved magnitude responses to minimize the changes in intensity perception with frequency. The design of filters involved adjustment in magnitude response at transition crossovers, reduction in passband ripple, and increase in stop-band attenuation. Spectral splitting scheme has been implemented and evaluated using these comb filters. The second phase of investigations involved implementation and evaluation of the three speech processing schemes namely, spectral splitting, temporal splitting, and combined splitting for different processing conditions on persons with normal hearing with simulated loss and on persons with moderate bilateral sensorineural hearing impairment. Use of adjustable magnitude response filter, to partly compensate for the frequency dependent shifts in hearing thresholds, alone and cascaded with other schemes, also has been evaluated.

1.3 Thesis outline

Chapter 2 deals with a short review of the peripheral auditory system, sensorineural hearing loss and its effects on speech perception, and some of the schemes for improving the speech perception by persons with sensorineural loss. In Chapter 3, speech processing schemes with binaural dichotic presentation are reviewed. Proposed scheme and evaluation methods are also described in this chapter.

First phase of the research is presented in Chapter 4. It deals with the design of perceptually balanced comb filters based on auditory critical bandwidth. Implementation and evaluation of spectral splitting schemes with comb filters with sharp transitions and perceptually balanced comb filters are also presented. Second phase of the research is presented in the fifth and sixth chapters. Chapter 5 gives a description of the combined splitting scheme, its implementation, and listening tests for selecting appropriate values for the processing conditions, the sweep cycle time and number of shiftings. Listening tests for overall evaluation of the three dichotic presentation schemes and related processing conditions are presented in Chapter 6. Chapter 7, the last chapter, gives a summary of the investigations, conclusions drawn from the results, and some suggestions for further work.

Appendices provide supplementary information. Appendix A deals with the spectrographic analysis set-up. Appendix B describes the performance evaluation methods. Comb filter design procedure is given in Appendix C. The hardware set-up and software for speech processing and listening tests are described in Appendix D and Appendix E respectively. Appendix F deals with the instruction for conducting the listening tests. The pure tone audiogram and the response of adjustable magnitude response filter used for the right and left ear for the sensorineural hearing impaired subjects participated in the listening tests are given in Appendix G.

Chapter 2

SENSORINEURAL HEARING IMPAIRMENT

2.1 Introduction

This chapter contains an overview of the peripheral auditory system, hearing impairment, and the identification of acoustic cues by listeners with hearing impairment. It also reviews some of the speech processing schemes for improving speech perception by such listeners.

2.2 Peripheral auditory system

The auditory system is grossly divided into peripheral and central auditory system. The peripheral auditory system converts acoustic signal to neural impulses. The central auditory system, which includes the auditory nerve, the central auditory pathways and the auditory cortex, carries out processing and interpretation of auditory information.

The structure of the peripheral auditory system is shown in Fig. 2.1. The peripheral auditory system is divided into outer ear, middle ear, and inner ear. The eardrum (also known as tympanic membrane) separates the outer and middle ears. The longitudinal cross section of the ear in Fig. 2.2 shows the path of sound transmission. The outer ear collects and directs acoustic signal to set vibrations in the tympanic membrane. The middle ear transmits the vibration of the tympanic membrane to the fluid filled cochlea situated in the inner ear and provides an impedance matching between the two. Cochlea spectrally analyzes the incoming sound, and produces neural impulses, which are carried by fibers in the auditory nerve to the brain.

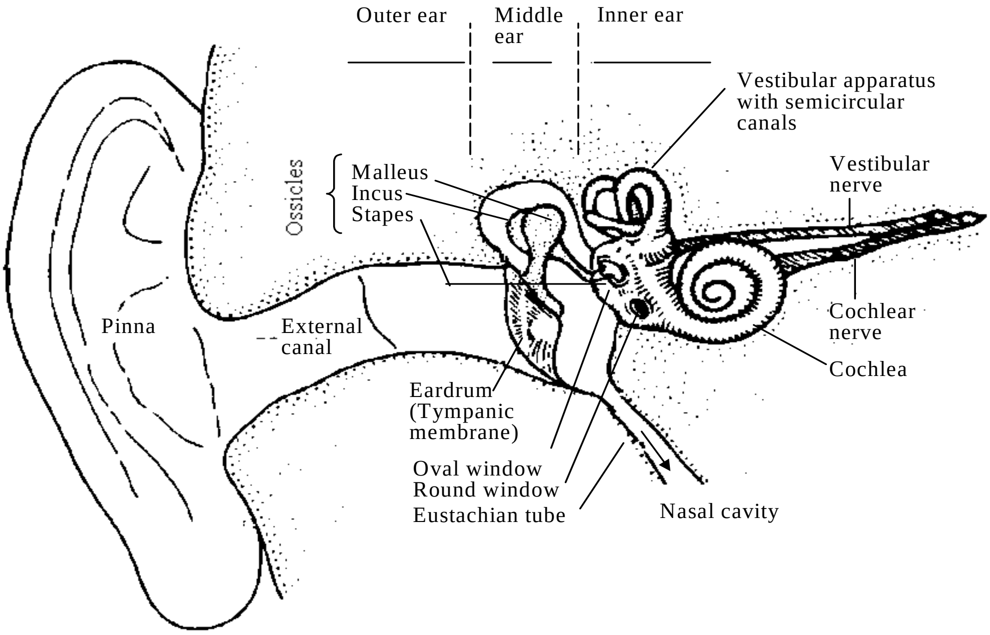


FIG. 2.1. Peripheral part of the auditory system. Source: Flanagan (1972), Fig. 4.1.

2.2.1 The outer ear and the middle ear

The outer ear consists of pinna and ear canal leading to the eardrum. The pinna with its shape collects and directs sound waves to set up vibrations in the tympanic membrane. Its asymmetrical response with respect to direction plays an important role in localization of the sources of sound (Williams *et al.*, 1989). The ear canal introduces an increase in level of about 10 to 20 dB in the frequency range from nearly 1.5 kHz to 7 kHz (Pickles, 1982; Wall, 1995; Yost, 1994).

The middle ear is an air filled cavity with three movable auditory ossicles and the Eustachian tube. The Eustachian tube provides a path to the exterior via nasopharynx and helps in equalizing the pressure on either side of the tympanic membrane (Ganong, 1993; Gelfand, 1998). The tympanic membrane vibrates in response to the pressure changes due to the impinging sound waves. Its vibratory pattern allows efficient transfer of sound waves from outer ear to the middle ear (Yost, 1994). The functioning of the three auditory ossicles, malleus, incus, and stapes can be considered to be a lever system that converts the vibrations of the tympanic membrane into movements of stapes and from stapes to oval window against the fluid filled cochlea of the inner ear. The tympanic membrane and the ossicular chain provide impedance matching for proper transmission of sound energy from air to fluid. The pressure is increased by approximately 33 dB by three structural characteristics namely, the lever action of the malleus and incus multiplies the force by 1.3 times, the ratio of effective area of the tympanic membrane to that of the stapes footplate multiples the pressure by 17 times, and the buckling action of the tympanic membrane due to its conical shape causes the malleus to move with twice the force (Gelfand, 1998; Pickles, 1982; Yost, 1994). This pressure transformation is significant across a wide range of frequencies around 2.5 kHz, which is important for the perception of speech.

The muscles in the middle ear are normally in a state of tension. At higher intensities, the contraction of the middle ear muscles, reduces the transmission of pressure through the ossicular chain. This contraction increases with increase in stimulus level. For loud sounds, the reduction amounts approximately 10 to 30 dB and has more effect at frequencies below 2 kHz.

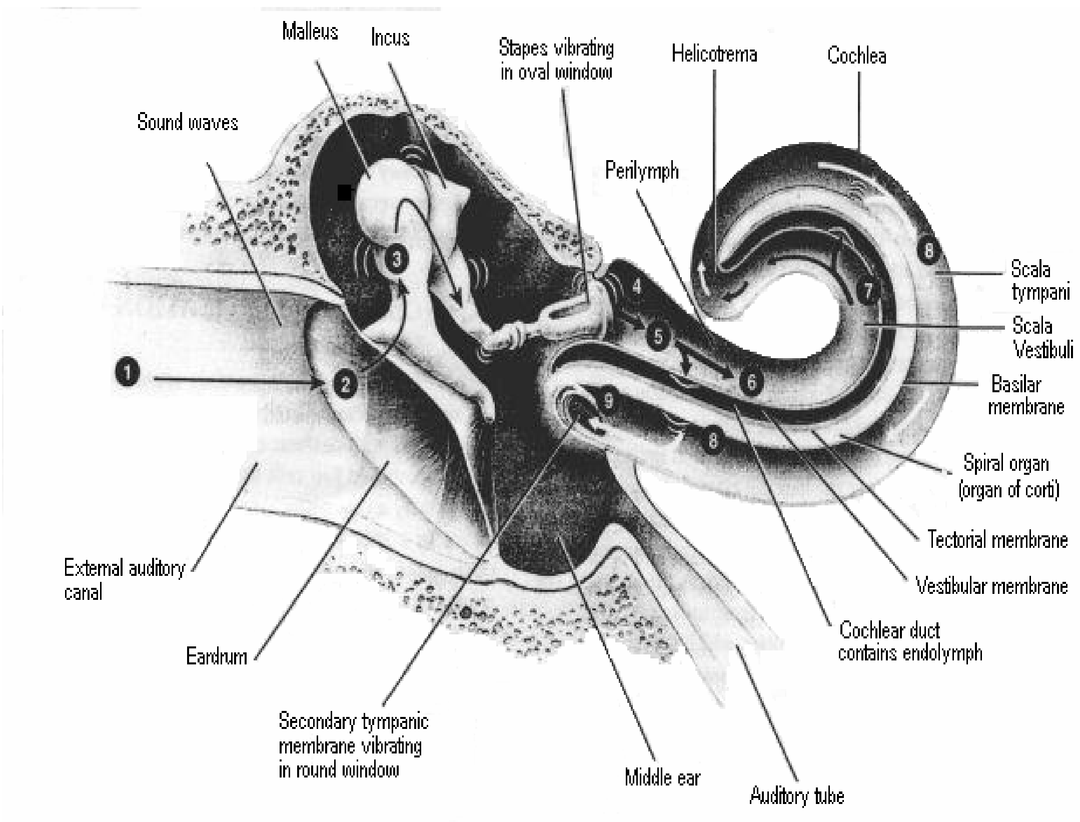


FIG. 2.2. The longitudinal cross section of the ear. Source: Williams *et al.* (1989), Fig. 16.20.

2.2.2 The inner ear

The inner ear has three parts: the semicircular canals, the vestibule, and the cochlea located in the temporal bone (Mackenna and Callender, 1997; Ganong, 1993; Gelfand, 1998; Yost, 1994). The cross-section of the cochlea, the essential organ of hearing, is shown in Fig. 2.3. Cochlea resembles a tube of decreasing diameter, which spirals 2¾ times round the pillar of a bone termed as modiolus. The modiolus contains the auditory nerve and the blood vessels that supply the cochlea. Along the length of the cochlea, the basilar membrane and Reissner’s membrane divide it into three chambers. The upper chamber, scala vestibuli, and the lower chamber, scala tympani, contain perilymph (similar to extracellular fluid) and are joined at the apex through a small opening known as helicotrema. The middle chamber, scala media, contains endolymph (similar to intracellular fluid) and forms a completely sealed sac called cochlear duct. Fluid in the scala vestibuli receives vibrations from the auditory ossicles through the oval window attached to the stapes. Resulting fluid displacements are relieved by the round window at the base of the scala tympani.

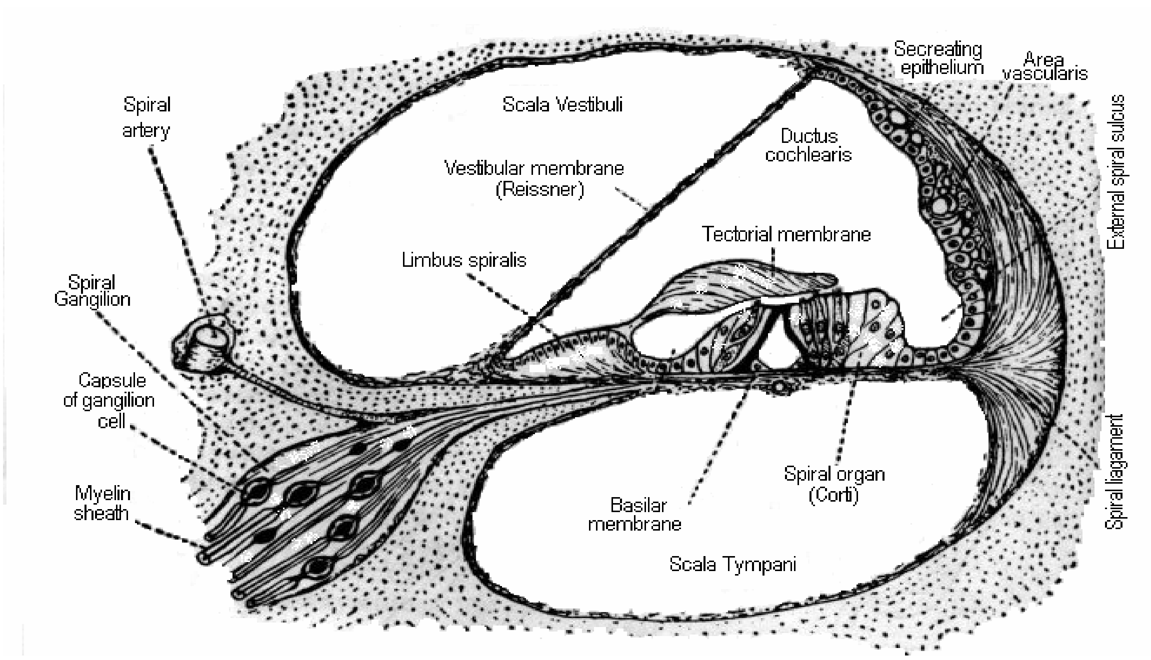


FIG. 2.3. Cross section of the cochlea. Source: Sataloff and Sataloff (1993), Fig. 3.2 (A).

Located on the basilar membrane extending from base to apex, is the organ of Corti containing receptors for hearing. Figure 2.4 shows the cross section of the organ of Corti. The dual auditory receptors, the inner and outer hair cells, have distinctly different cellular organizations and innervation patterns and similar general characteristics such as organization of stereocilia (hairs attached to the tip of hair cells) and gradation of stereociliary height along the length of cochlea (Kelly, 1991; Lim, 1986). The outer hair cells, about 20 thousand in number, are arranged in 3 rows lateral to the tunnel formed by the rods of Corti, while a single layer of inner hair cells, about 3.5 thousand in number, are arranged medial to the tunnel. The rows of hair cells are covered by the thin, viscous, and elastic tectorial membrane, in which the tips of the hairs of the outer hair cells are embedded.

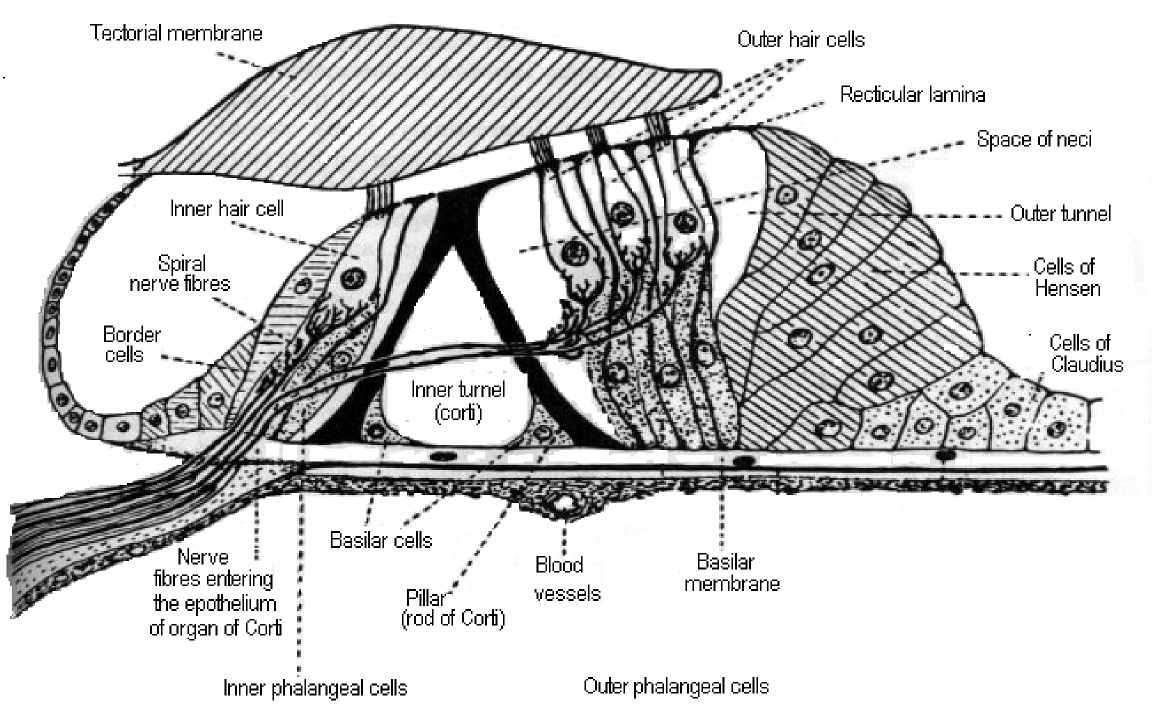


FIG. 2.4. Cross section of the organ of Corti. Source: Sataloff and Sataloff (1993), Fig. 3.2 (B).

The auditory stimuli reach the cochlea via the oval window in the form of vibrations. The vibration of the fluids causes movement of the basilar membrane in the form of a traveling wave. The basilar membrane gets progressively wider while the stiffness reduces from base to apex (Gelfand, 1998; Moore, 1997; Yost, 1994). The natural frequency of vibration of the basilar membrane decreases towards the helicotrema. Different frequencies will cause maximum vibration at different points along the basilar membrane, activating different population of tuned hair cells (Kelly, 1991; Moore, 1997). At a particular frequency an increase in the intensity is signaled by an increase in the amplitude of vibration, which leads to an increase in the number of activated cochlear nerve fibers and in their rate of discharge (Moore, 1997; Williams *et al.*, 1989). As basilar membrane vibrates, the tectorial membrane is displaced since it is hinged at one end only, causing a shear force on the stereocilia. The sensory hair cells of the cochlea are activated when the stereocilia are bent in a particular direction and inhibition occurs when they are bent in the opposite direction. Activation is associated with an increase in firing rate while inhibition contributes to decrease in firing rate of the auditory neurons connected to the hair cells (Gelfand, 1998; Kelly, 1991). There are approximately 30 thousand neurons in the human auditory system, which are primarily afferents. Efferent innervation is more for the outer hair cells compared to inner hair cells, reducing from base to apex.

A compressive non-linearity takes place for a range of input sound levels at the basilar membrane. While the mechanical behavior of the basilar membrane is mainly responsible for the broad discrimination between different frequencies, fine tuning is accomplished by the difference in the physiology of the two types of hair cells. Inner hair cells function as passive transducers, responsible for the detection of sound. The activities of the outer hair cells, as active transducers, play an important part in the sensitivity of the transduction mechanism (Kelly, 1991; Lim, 1986; Yost, 1994). Role of the two types of hair cells is discussed in Section 2.4.3.

Spontaneous firing of the neurons continues even in the absence of any stimulus (Gelfand, 1998; Moore, 1997; Yost, 1994). The rate of firing of the neurons increases with the increase in input signal level. The thresholds of a single neuron at any frequency, is the stimulus level required to increase the firing rate above the spontaneous firing rate, and the frequency at which the threshold is lowest is known as the characteristic frequency (CF) for that neuron, and it is determined by the site of innervation. The plot of threshold verses frequency for each neuron is known as its tuning curve, which has “V” shape. Tuning curves at high frequency side are steeper than those at low frequency side.

After leaving the cochlea, the neural impulses travel in the vestibulo-cochlear nerve (Cranial nerve VIII) to the cochlear nucleus where the nerve fibers form first synapse (Ganong, 1993; Mackenna and Callander, 1997; Yost, 1994). From the cochlear nuclei, the information from both the ears converge at both the ipsilateral and contralateral olivary complex, and at higher levels most of the neurons respond to inputs from both sides. Nerve fibres carrying auditory impulses pass via variety of pathways to inferior colliculi (the centers of auditory reflexes) and via the medical geniculate in the thalamus to the auditory cortex.

2.3 Hearing loss

Sound intensity expressed in dB with sound pressure of 20 μPa as reference is termed as sound pressure level (SPL). Sound level of pure tones expressed in dB with the average hearing threshold at the tone frequency, of young adults with normal hearing, as reference is known as hearing level (HL). When the reference level taken is the threshold level of the listener at the tone frequency, the measured sound level is referred as sensational level (SL) (Bess and Humes, 1995; Biswas, 1995, Moore, 1997; Newby, 1979; Rintelmann, 1991).

The audiogram is a graph showing hearing threshold levels (HTL) in dB HL as a function of frequency, measured using an audiometer (Bess and Humes, 1995; Biswas, 1995; Moore, 1997; Newby, 1979; Rintelmann, 1991; Sataloff and Sataloff, 1993). It shows the deviation from the normal hearing threshold. Hearing thresholds are normally measured for pure tones at 250, 500, 1000, 2000, 3000, 4000, 6000 and 8000 Hz. HTL averaged over frequencies of 500, 1000 and 2000 Hz is taken as pure tone average (PTA). Pure tone audiometry can also be used to determine the most comfortable level (MCL) and the uncomfortable level (UCL). A number of audiometric tests, e.g., tone decay test, alternate binaural loudness balance (ABLB) test, etc., are employed for diagnosing the nature of loss. Speech audiometric techniques, including speech recognition threshold (SRT) and speech discrimination tests are also used by clinicians, for the assessment of hearing impairment.

Hearing losses are mainly classified into four types depending on the site of damage, namely conductive loss, sensorineural loss, central loss, and functional loss (Levitt *et al*., 1980; Deutsch and Richards, 1979; Moore, 1997; Pickles, 1982; Sataloff and Sataloff, 1993).

Conductive loss causes reduction in the sound energy transmitted to the inner ear. Some of the reasons for this loss are plugging of the external auditory canal with wax or foreign bodies, middle ear infections, perforation of the auditory ossicles, and abnormal rigidity of the attachment of the stapes to the oval window known as otosclerosis (Ganong, 1993; Sataloff and Sataloff, 1993). This loss can normally be cured medically or surgically.

Sensorineural hearing loss is caused by the malfunctioning of cochlear sensory mechanism or the auditory nerve or both. This loss is subdivided into cochlear and retrocochlear depending on the location of the lesion causing the hearing loss (Deutsch and Richards, 1979; Pickles, 1982). Such defects are difficult to be treated medically. Congenital loss may be genetically transmitted. The shape of pure tone audiogram in such cases is normally trough-shaped. The loss can vary from mild to profound and may be degenerative. Prenatal hearing loss caused due to Rh incompatibility has decreased greatly in recent years, as remedies are available. Anoxia (lack of sufficient oxygen during birth) commonly causes damage to nervous system and sometimes sensorineural hearing loss (Biswas, 1995; Wall, 1995; Sataloff and Sataloff, 1993). Acquired sensorineural hearing loss can be due to many reasons. Viral and bacterial infections sometimes damage sensorineural mechanism to varying degrees. Meningitis (infection of the brain covering) may reach the inner ear and can cause wide spread damage to the cochlea and VIII nerve. Exposure to loud noise such as explosion may cause sensorineural hearing loss, which primarily affect the 3-6 kHz frequency range. Slowly progressing hearing loss is noticed in work related noise exposure. Presbycusis, the gradual reduction in hearing with advancing age, is due to sensorineural hearing loss (Deutsch and Richards, 1979; Sataloff and Sataloff, 1993; Wall, 1995). High frequencies are affected first.

Sensorineural hearing loss may be accompanied by chronic otitis (inflammation of the ear). Paget’s disease (a chronic disease in which bones become enlarged, weak, and deformed) can affect auditory nerve pathway. Bilateral sensorineural hearing loss of sudden onset is seen due to infections, drugs, emotionally induced illness, multiple sclerosis, etc. Meniere’s disease, believed to be due to excessive endolymph in the membranous labyrinth causes tinnitus and progressive sensorineural hearing loss, usually unilateral. Diseases like typhoid and mumps cause sensorineural hearing loss. Due to the intake of ototoxic drugs, hair cells and supporting cells get injured, causing high frequency hearing loss.

Central impairment is caused due to damage in the central nervous system. Such patients have reduced ability to interpret, integrate or understand speech (Deutsch and Richards, 1979; Wall, 1995). The causes of functional deafness are psychological or emotional rather than physiological.

2.4 Sensorineural hearing loss and its characteristics

Sensorineural hearing loss can affect the perception of speech signal in all its physical features like intensity, time and frequency. The main characteristics of sensorineural hearing loss are frequency dependent shifts in hearing threshold, loudness recruitment, poor frequency selectivity and increased spectral masking, and reduced time resolution and increased temporal masking (Baer and Moore, 1993; Dorman and Hannley, 1985; Hou and Pavlovic, 1994; Humes *et al.*, 1988; Minifie, 1994; Moore, 1998). The psychophysical parameters associated with cochlear impairment are inter-related and their effects on perception are difficult to separate from each other.

2.4.1 Hearing thresholds and dynamic range

Dynamic range is the difference between the threshold of hearing and the loudness discomfort level. In the transduction mechanism, outer hair cells increase the sensitivity of low-level signals, without affecting the response to high-level signals. This results in a significant compressive non-linearity and results in large dynamic range. Sensory loss, caused primarily by loss of outer hair cells, results in increase in threshold of hearing with no corresponding change in loudness discomfort level, reducing the dynamic range. The associated phenomenon of a rapid growth in loudness at higher sound levels is termed as loudness recruitment (Derleth *et al*., 2001; Moore, 1997; Oxenham and Plank, 1997; Pickles, 1982; Sandlin, 1988). The intensity variation of normal speech signal may lie in excess of 30 dB in the same utterance (Moore, 1997; Sandlin, 2000).

2.4.2 Spectral characteristics

Frequency selectivity is the ability to separate or resolve multiple spectral peaks in a complex sound. It is normally measured by psychophysical tuning curves, masking patterns, critical bands, and just-noticeable differences (JNDs) in frequency. Frequency selectivity can be considered to have two correlated psychophysical phenomena, frequency resolution and frequency discrimination. In quiet, the frequency selectivity of a normal ear seems to be much greater than what is required for clear perception, due to redundancy of features in speech. The high resolution of spectral contrasts is important in noisy environment (ter Keurs *et al.*, 1992).

The peripheral auditory system can be represented as a bank of band-pass filters with overlapping pass bands, termed as auditory filters (Baker and Rosen, 2002; Moore, 1986; Rosen *et al*., 1998). To determine the characteristic of the auditory filter, Fletcher (1953) measured the threshold of a tone as function of bandwidth of a band-pass noise masker centered at the tone frequency. As the bandwidth of the noise was increased keeping the center frequency constant, the threshold increased first, and later it remained constant. The noise bandwidth after which the threshold stopped increasing was termed as critical bandwidth. The critical bandwidths remain constant at 100 Hz for center frequencies below 500 Hz, and are between 15 –20 % of the center frequency above 1 kHz (Moore, 1986; Pickles, 1982; Zwicker, 1961). Patterson (1976) used "notch-noise method", which involves the determination of the detection threshold for a sinusoid, centered in a spectral notch of a noise, as a function of the width of the notch. On the basis of results obtained with this method, auditory filters can be described in terms of an equivalent rectangular bandwidth (ERB) as a function of center frequency. The ERBs are 11–20 % of the center frequency. The equivalent rectangular bandwidth and the critical bandwidth described by Zwicker (1961), are almost same for frequencies above 1 kHz. Figure 2.5 shows the critical bandwidth, described by Zwicker (1961), and the estimates (using notched-noise method) of the equivalent rectangular bandwidth (ERB) of the auditory filter. These auditory filters are highly nonlinear, having wider bandwidths with increasing level, which is responsible for high sensitivity and sharp frequency selectivity in normal auditory system (Hicks and Bacon, 1999; Jenison *et al*., 1991). Progressive increase in non-linearity was observed with increase in frequency.

Psychophysical tuning curves (PTC) indicate the masker level required to mask a tone of fixed frequency and level determined as a function of masker frequency (Carney and Nelson, 1983; Nelson, 1991; Nelson and Fortune, 1991; Moore, 1978). PTC has "V" shape similar to neural tuning curves. The psychophysical tuning curves obtained from sensorineural hearing impaired subjects may be (i) flattened on the low frequency side, showing upward spread of masking, (ii) flattened on the high frequency side, implying downward spread of masking, (iii) flattened at both sides or (iv) “w” shaped (Dorman and Hannley, 1985).

Masking is the phenomena by which, threshold of audibility of a signal component gets elevated due to the presence of a neighboring signal component. It takes place at the peripheral level of the auditory system. Simultaneous masking provides an insight into the frequency selectivity of the ear, since it reflects the excitation pattern along the basilar membrane (Gelfand, 1998; Moore, 1997).

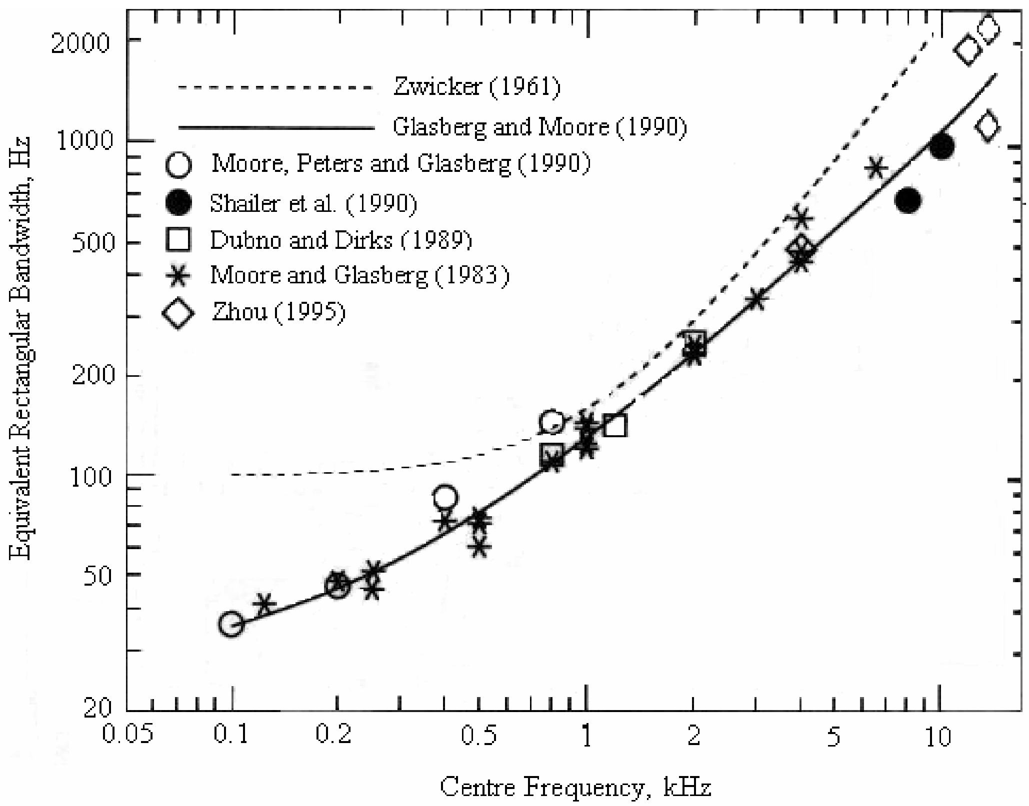


FIG. 2.5. The critical bandwidth described by Zwicker (1961) and the equivalent rectangular bandwidth (ERB) of the auditory filter. Moore (1997), Fig. 3.10.

Frequency resolution is the ability to resolve frequency components in a stimulus. Degraded frequency resolution acts as though the auditory filters are broader than normal (Dubno and Dirks, 1989; Leek and Summers, 1996; ter Keurs *et al.*, 1993). Further the impaired ears show much smaller change in auditory filter bandwidths with increasing stimulus level (Rosen *et al.*, 1998). Because of poor frequency selectivity, the peaks and the valleys of the speech spectrum get smeared. Also signal-to-noise ratio at the output of the auditory filter reduces (Leek and Summers, 1996). Broader excitation pattern would make the individual spectral features like formants harder to resolve (Moore *et al.*, 1992). Such listeners require more spectral contrasts for identification of vowel like sounds.

Frequency discrimination is the capacity to discriminate successively presented stimuli that differ in their frequency content (Tyler *et al.*, 1983). Frequency discrimination is measured as just-noticeable differences (JNDs) in frequency. The JND in frequency for two equally intense tones less than 1 kHz is 1 – 3 Hz and increases progressively as frequency increases (O'Shaughnessy, 2001). The JND is very often abnormally large in sensorineural hearing impaired (Tyler *et al*., 1983). Normal hearing and hearing impaired listeners demonstrated increases in JND with decrease in stimulus duration less than 200 ms (Hall and Wood, 1984).

Frequency discrimination depends primarily on the temporal coding (Tyler *et al*., 1983). Frequency discrimination helps in finding whether the fundamental pitch of one speech sound is higher or lower compared to the previous sound. Fundamental frequency analysis is important in the identification of voicing in speech, as well as the linguistic and paralinguistic identification of numerous aspects of speech and voice including: the phonetic voiced-voiceless distinction, intonation, tone, and speaker sex, age, identity, and emotional state. Frequency discrimination of transitions gets drastically reduced in listeners with hearing impairment.

2.4.3 Temporal characteristics

The temporal resolution is the minimum detectable gap between two successive signals. The temporal resolution for normal hearing persons is 2 – 3 ms and that for sensorineural impaired persons is about 8 ms (Fitzgibbons and Wightman, 1982). Due to poor time resolution, detection of acoustic events gets degraded. The temporal processing capability of the auditory system has been measured using different methods. Some of them are temporal integration, forward masking, backward masking, gap detection threshold, gap difference limen, etc (Dorman and Hannley, 1985; Tyler *et al*., 1982). Temporal integration is referred as the property in which ear operates as an energy integrator over a time frame (Gelfand, 1998; Oxenham *et al*., 1997, Tyler *et al*., 1982; Yost, 1994). A signal must have certain critical energy to be detected. For normal hearing, the integration duration is about 200 ms to attain threshold (Moore, 1997). Listeners with sensorineural hearing impairment have less temporal integration (Florentine *et al*., 1988; Oxenham *et al*., 1997).

The poor time resolution is associated with forward and backward masking of weaker segments by intense segments. In backward masking the signal segment precedes the masker segment, whereas in forward masking the signal follows the masker. Intense masker following a weak signal may reach higher neurological centers faster due to rapid neural transmission (Elliott, 1962). Backward masking occurs if the onset of the masker masks the offset of the previous signal segment. Neural activity due to the signal may persist after the signal offset (Oxenham and Plank, 1997). Backward masking is more severe than forward masking for very short masking intervals. In backward masking, the amount of masking decreases as the interval increases beyond 15 ms and occurs up to a temporal gap of 100 ms with masking effect increasing as the intensity level of masking tone.

The temporal resolution and forward masking are associated with auditory persistence (Smiarowski and Carhart, 1975). The sensitivity of a neuron is reduced after a stimulation, which leads to decrease in the detection ability of a following signal. The forward masking is most effective within 10 ms of the masker, reduces with time and is insignificant beyond 200 ms (Danaher *et al*., 1978; O'Shaughnessy, 1987). In normal auditory system, onset response is very fast, while the offset decays very slowly. The offset response of a prior burst can mask the onset response of a following burst if they are close enough (forward masking). The gap between two tone bursts will be obscured, if the temporal offset response of the two overlaps.

The gap detection threshold, an index of temporal acuity, is elevated in sensorineural hearing impaired (Fitzgibbons and Gordon-Salant, 1987; Fitzgibbons and Wightman, 1982; Glasberg and Moore, 1987; Irwin and McAuley, 1987; Tylor *et al*., 1982). The thresholds for gap detection in the presence of band-limited noise decrease with increase in center frequency of the band and also with increase in the bandwidth (Fitzgibbons, 1983; Fitzgibbons and Wightman, 1982; Shailer and Moore, 1983; 1985). Listeners with high frequency loss have larger minimum detectable gap. The minimum detectable gaps decrease very rapidly with increasing level up to about 80 dB SPL, above which they are nearly constant, in sensorineural hearing impaired listeners (Florentine and Buus, 1984). The difference in the minimum detectable gap reduces or gets eliminated when the presentation is made at the same SL in normal hearing and hearing impaired subjects (Glasberg *et al*., 1987).

Gap difference limens (DLs) for signals with 30 and 100 ms durations at 0.5 and 4 kHz, were found to be elevated by 20 to 25 ms at both frequencies and durations for a group of hearing impaired subjects (Tyler *et al,* 1982). The mean DL values for normal hearing subjects were 18.8 and 17.9 ms for tone frequencies of 500 Hz and 4 kHz respectively. The corresponding values for hearing impaired subjects were 48.0 ms and 41.8 ms.

2.4.4 Role of transduction mechanism in hearing loss

Hair cells, the sensory receptors of the inner ear, convert mechanical vibration into electrical signals, and the information is carried to brain in the form of neural impulses. Inner hair cells detect the signal and excite the nerve fibers. The role of outer hair cells in auditory transduction, is to enhance the basilar membrane responses to low-level sounds and in fine tuning of auditory responses to a particular frequency (Moore, 1998; Pickles, 1982; Wall, 1995). Hair cells may be damaged or lost due to genetic disorders, infections, ototoxic drugs, intense noise exposure, aging, etc. Since these cells are not replaced by cell division, their disappearance is associated with a decline of the sense of hearing.

Along the basilar membrane, sensory cells at the basal end corresponding to high-frequency are relatively more vulnerable to damage than those at apical end and corresponding to low frequency. So hearing loss usually begins at high frequencies. At any location on the basilar membrane, outer hair cells are more prone to damage. Damage to outer hair cells leads to loss in sensitivity, elevated threshold, and loss in cochlear compression resulting in reduced dynamic range and loudness recruitment. It also results in degradation in frequency tuning due to wider auditory filters. When inner hair cells are damaged, the sensitivity to basilar membrane vibrations reduces, causing increasing hearing thresholds, but frequency selectivity is not much affected. Since there is no loudness recruitment, dynamic range is not affected. The auditory nerve fibers reflect the frequency analysis of the basilar membrane. Damage to the auditory nerve fibers causes impairment to the frequency coding, resulting in reduced frequency selectivity, loudness recruitment, and reduction in dynamic range (Pickles, 1982).

2.5 Identification of speech cues by sensorineural hearing impaired

Speech perception requires an interface between an analysis of the incoming acoustic information and knowledge of linguistic information stored by the brain from various sources at different times (Minifie, 1994; O’Shaughnessy, 2001). Speech level varies significantly as a function of frequency, and the different regions of the spectrum contribute differently to intelligibility (Dubno and Schaefer, 1992). The speech cues are highly redundant in quiet surroundings and may be useful under adverse listening conditions for a normal hearing person (Borden *et al*., 1994; O'Shaughnessy, 2001; Revoile *et al*., 1982). Speech perception of hearing impaired listeners is affected by much lesser reverberation time and noise levels. Normal hearing persons depend on some of the many cues like sequential intensities, duration and spectral composition for the identification of various segments (Summers and Leek, 1992). For listeners with impaired hearing, many acoustic cues are lost or obscured due to auditory processing deficits. The basic factors causing reduced recognition of acoustic cues are poor audibility due to elevated thresholds in the frequency region of energy concentrations, abnormal loudness growth, and reduced discrimination due to poor frequency resolution and impaired temporal resolution (Zeng and Turner, 1990).

Vowels are associated with well-defined formant patterns and formant frequencies, with slowly changing spectrum, each having its own characteristic regions of prominence. Vowels tend to become the nuclei of syllables, having the greater amplitude (French and Steinberg, 1947; Pickett, 1999; Stevens, 1980). In a study by Leek *et al*. (1987), normal hearing listeners in quiet needed a 2 dB peak-to-valley level difference for 75 % correct score for vowel identification. For the same performance, normal hearing listeners needed a level difference of 4 dB in noise background sufficient to simulate moderate hearing loss. Hearing impaired listeners with moderate flat hearing loss required at least 6 dB level differences. Due to spectral smearing, the internal contrast is reduced for peaks lying closer to each other (Summers and Leek, 1994). In natural speech, the difference in levels of front vowels is as large as 25 – 30 dB, whereas it lies between 5 – 7 dB for back vowels. With spectral smearing and upward spread of masking, formants F2 and higher ones will be smoothed out, leaving a broadened F1. The vowels with only one formant are confused with back vowels (ter Keurs *et al*., 1992). Hearing impaired listeners make use of other cues like duration, which is redundant for normal listeners, for identification of vowels with less difference between formant frequencies.

The information content of consonants lies at higher frequencies with low intensities. The consonants are very important for speech intelligibility and are very easily confused (Crandall, 1917; Sandlin, 2000). The articulatory features that characterize consonants are manner, voicing, duration, and place of articulation (Ladefoged, 1982; Miller and Nicely, 1955; O'Shaughnessy, 2001). In the perception process, auditory system responds to these features in an integrated manner (Borden *et al.*, 1994; Dubno and Dirks, 1989). More than one acoustic cue contributes to the distinction of each of the features. Also an acoustical cue may characterize more than one feature. Two phonemes are more likely to be confused if they share more number of distinctive features. The ability to identify segmental and suprasegmental contrasts in connected speech context, is seen in listeners with sensorineural hearing loss as high as 105 – 114 dB (Boothroyd, 1984). All the cues are affected by the context of adjacent phonemes, the rate of speaking, and talker variability (Minifie, 1994; Pickett, 1999). Also, acoustic characteristics of consonants depend on several characteristics of the adjacent vowel (Dubno and Levitt, 1981). Recognition of consonants appearing with vowel /*a*/ is higher than those accompanied with /i/ and /u/. In syllables, words, and phrases, there is a continuous succession of rapid variations in intensity along time and frequency scales.

The manner feature categorizes consonants into semivowels, nasals, stops and fricatives. The cues for manner are amplitude, duration, formant structure, and the balance between low frequency voiced energy and high frequency frication. (O'Shaughnessy, 2001). Transition duration, the difference in time required for frequency change at onset and offset, may cue consonants with difference in manner but same place of articulation (Diehl and Walsh, 1989). Confusion in manner distinctions is relatively rare (Dubno and Levitt, 1981).

Semivowels are similar to vowels, but have much gradual change in formant structure (Kent and Read, 1992). They are characterized by weaker amplitude, wider bandwidth, and higher concentration of energy at low frequencies compared to vowels. Nasals are associated with formants and antiformants. Low frequency spectral pattern occurring during consonant constriction and in adjacent vowel formant transitions differentiate semivowels and nasal consonants. Nasal consonants have greater low frequency energy in relation to high frequency energy, than stop consonants (O'Shaughnessy, 2001; Stevens, 1980).

Fricatives are characterized by broadband noise-like interval with spectral peaks related to the place of articulation and these have fast transitions with respect to formant structure of the previous and following phonemes. Fricatives have lesser acoustic energy when voiced than voiceless (Ladefoged, 1982). The relatively long fricative noise is an important cue to differentiate fricatives from stops (Borden *et al*., 1994). Since the energy magnitude varies in this interval, audibility of fricatives may vary in hearing impaired, mistaking fricative noise as a burst of stop. Reduction in manner distinction for voiceless fricatives and stops is reported by Revoile *et al*, (1991a). This distinction is poorer in vowel-consonant-vowel (VCV) context compared to consonant-vowel-consonant (CVC) context for hearing impaired listeners. Fricative noises were more similar in mean duration to stop release in VCV contexts compared to CVC. Also greater masking takes place in intervocalic consonants due to presence of vowels on both sides. Also, consonants in the final position are harder to discriminate than those in initial position (Dubno and Levitt, 1981). For distinguishing alveolar consonants / d, n, l / spectral cues associated with murmur is more useful for hearing impaired listeners, since it is available for longer period of time relative to spectra of transitions (Revoile *et al.*, 1991b).

Voicing is cued primarily by harmonic structure, present at least up to 3 kHz (O'Shaughnessy, 2001). The perception of voicing feature is not affected in case of listeners with mild and moderate hearing loss (Dubno and Levitt, 1981; Johnson *et al*., 1984; Pickett, 1999). In English, timing plays a major role in voicing distinction. In voiced stops and fricatives, preceding vowel becomes longer with shorter consonants, whereas shorter vowel precedes longer voiceless consonants (Crystal and House, 1982; Ladefoged, 1982; Revoile *et al*., 1982). Voiced murmur during the occlusion of a voiced stop and its absence for voiceless stops is another important cue. The spectral cues associated with vowel transitions are not much useful for hearing impaired listeners, for consonant voicing distinction (Revoile *et al*., 1986). Generally the perception of final stop voicing is poorer for the hearing impaired listeners (Revoile *et al*., 1982).

The relative duration of a silent interval can cue several phonetic contrasts in English, e.g., the presence or absence of a stop consonant in a consonant cluster, the voicing of a stop consonant in word medial position, and the distinction between single and double stop consonants (Minifie, 1994). The duration of the stop closure and closure murmur, which normally varies between voiced and voiceless stops, was neutralized, the murmur was deleted and the silent interval made longer than normal for voiceless stops and shorter than normal for voiced stops (Revoile *et al*., 1982). The identification of stop voicing accuracy decreased slightly for both normal hearing and hearing impaired subjects.

For persons with hearing impairment, temporal cues may be more helpful than spectral cues (Revoile *et al*., 1982). The duration of articulatory closure lies between 50 – 100 ms and the burst is 5 – 10 ms long. Voice-onset time (VOT) is a major temporal cue for distinguishing voiced and voiceless stop consonant produced at the same place of articulation

(Minifie, 1994; Pickett, 1999; Revoile *et al.*, 1987; Tyler *et al.*, 1982). Voiced sounds typically have a VOT between 0 and 35 ms, in comparison with unvoiced which has a VOT between 30 and 150 ms depending on the place of articulation. Since the burst and formant transition cues differ for each place of articulation, the VOT boundary increases from labial to alveolar to velar (Massaro and Oden, 1980). Mild and moderately hearing impaired children did not alter the perception of VOT significantly (Johnson *et al.*, 1984). Low frequency periodicity is observed during consonantal closure interval in voiced stops except at initial position in a word (Crystal and House, 1982; Stevens, 1980). Since the F1 onset frequency is lower for alveolar than for labial stops for the same VOT, perception of alveolars are often confused (Massaro and Oden, 1980).

Stops are identified by the shape of the spectrum at burst release, the manner in which the spectrum changes over a short interval of time, and specific characteristics of formant transitions. It has been suggested that the sufficient cues for the identification of place feature is found in the gross shape of the spectrum in the first 25 – 35 ms of the waveform after the closure (Stevens and Blumstein, 1978; Van Tasell *et al.*, 1982). Differences in the F2 onset frequencies and transitions differentiate labial, alveolar, and velar stop consonants in CV syllables (Pickett, 1999; Stevens and Blumstein, 1978). Labial stop is characterized by low onset frequency and rising transition. High onset frequency and falling formant transition are the characteristics of velar stops. Alveolar stops have intermediate onset frequencies with slightly rising or straight formant transitions. The identification of velars is improved if the duration of the stimulus is longer than 10 – 20 ms (Stevens and Blumstein, 1978). In conversational speech, noise burst may not be released, which may reduce the perception of stop consonants by hearing impaired listeners (Pickett, 1999).

Consonants differing in place of articulation are easily confused by sensorineural hearing impaired even with hearing loss as low as 75 dB, since its spectral cues occur across relatively high frequency regions and in sub-phonemic segments that are brief and changing fast (Boothroyd, 1984; Turner *et al.*, 1997; Van Tasell *et al.*, 1982). The intensity level of individual stop consonants is typically 20 – 30 dB less than those of accompanying vowel (Turner and Robb, 1987). Turek *et al.*, (1980) reported errors in the identification of phonemes, which are primarily coded by formant transitions. Identification of consonants formed near the front of the mouth / p, f, b, v, m / are lower compared to those produced at middle and back (Dubno and Levitt, 1981). Hearing impaired listeners with either flat or gradually sloping audiometric configuration need the acoustic cues contained within the consonant's onset for identification of the place feature (Dubno *et al*., 1987). These listeners have little or no use of spectral cues in dynamically changing transition segments in discriminating stops and fricatives. The alveolar and palato-alveolar fricatives are characterized by well-defined and distinct spectral shapes while labio-dental and dental fricatives display relatively flat spectrum. For the identification of place of articulation of fricatives by normal hearing subjects and hearing impaired listeners, the steady state spectral cue in the frication portion may alone be sufficient (Zeng and Turner, 1990). Dynamic transition cue assists the identification at low signal levels. Hearing impaired listeners have difficulty in perceiving voiceless fricatives. Frication and transition portions are typically about 30 dB and 5 – 10 dB less than the vowel. Hearing impaired listeners confuse fricatives with stops in cases when only the transition cue is audible.

Simultaneous and/or temporal masking may cause reduced discrimination of F2 transitions by hearing impaired listeners. Hannley and Dorman (1983) have reported that hearing impaired listeners have more difficulty identifying place of articulation for stop consonants when the cue is a falling second formant transition than when the cue is a rising formant transition.

The place of articulation of semivowels is more easily identified by hearing impaired, since the formant transitions are much slower compared to those for voiced stops (Pickett, 1999). Anti-resonance is an important cue for distinguishing nasals according to place of articulation. The high frequency anti-resonance of /n/ may lead to its poorer perception compared to /m/.

The persistence of auditory activity limits the listener's ability to perceive a noise burst which follows another noise burst (Smiarowski and Carhart, 1975). The listener will be able to perceive a trailing burst as being temporally separate only if it is intense enough. Otherwise the listener will experience a fused sound. The persons with highly sloping high frequency audiogram configuration missed dental fricatives such as /θ/ and alveolar stops like /t/ most often (reviewed by Sandlin, 1988). Hearing impaired subjects with high frequency loss have difficulties in discriminating fricatives like /s/, /z/ and /sh/, since the energy in the spectra of these alveolar fricatives lies at 4 kHz and above. Hearing impaired subjects have poor gap detection thresholds. Gaps as wide as typical stop occlusion in continuous speech may be obscured and word pairs such as "speed" and "seed" may be confused. The distinction between /be/, /we/, /ue/ can be perceived by the rapidity of formant transitions, or by the rapidity of change of the amplitude envelope at release. Severe auditory impairment may cause smearing of faster changes in time.

Hearing impaired listeners prefer listening to adult male speakers (Levitt *et al*., 1980). The frequency spectrum of speech produced by adult male speaker has relatively more power to the low frequency side compared to that produced by female and child speakers.

From this review following conclusions can be made. Various acoustic features have energy concentrations in different regions. Degree of hearing loss (as indicated by audiogram) is the most commonly available measure of hearing status. But it is by no means, the only quantity that determines the perception capacity of the auditory system. Recognition of speech is affected in sensorineural hearing impaired persons, due to loss of audibility of parts of speech spectrum, abnormal relationship between intensity and loudness, reduced spectral resolution, and reduced temporal resolution. The identification of vowels are not normally affected much, if presented with good SNR and at higher SPL to take care of the increase in thresholds of hearing impaired listeners. In vowels itself, articulation height is better perceived than place of articulation (front, middle and back). In case of large amounts of spectral smearing, the perception of vowels may also get affected. Consonant in the final position is harder to discriminate than consonant in the initial position. Distinction of place feature in consonants is affected most. Listeners with hearing loss limited to the high frequencies generally can detect the voicing and manner of articulation cues. For such persons, lip reading helps to some extent. Hearing impaired listeners with flat hearing loss may have better discrimination of speech compared to those having sloping high frequency loss, since they are least affected by upward spread of masking.

2.6 Speech processing for sensorineural hearing loss

Hearing aids should improve the speech perception by the hearing impaired listeners, by making speech audible and comfortable. All conventional hearing aids, analog or digital, amplify sounds to make them audible for persons with hearing loss. The simplest hearing aid consists of a microphone, electronic circuitry, receiver (an earphone), a battery, and other acoustic components, like flexible tubing and an earmold. The electronic circuitry provides gain and overall shape of the frequency response, and limits the amplified signals to comfortable level. The performance of a hearing aid is characterized by acoustic gain, acoustic output, basic frequency response, frequency range, and distortion (Sataloff and Sataloff, 1993).

Conventional hearing aids are of limited benefit to persons with sensorineural deafness. Other sensory modalities like vision and touch also cannot provide much improvement in speech perception. Researchers have investigated different techniques for delivering electrical stimuli to auditory nerve fibers, to restore partial hearing to profoundly deaf people (Levitt *et al.*, 1980; Loizou, 1999; Pandey *et al.*, 1987; Spelman, 1999). A prosthetic device, called cochlear implant, can be implanted. These devices generally consist of a microphone that picks up the sound, a signal processor that converts the sound into the pattern for electrical stimulation of electrodes, a transmission system that transmits the electrical signals to the implanted electrodes, and an electrode or an electrode array (consisting of multiple electrodes) implanted for providing electrical stimulation to the auditory nerve fibers.

The miniaturization of hearing aid transducers and electronic circuits has led to a progressive development in hearing aids: body worn, eyeglass, behind-the-ear (BTE), in-the-ear (ITE), in-the-canal (ITC), and completely-in-the-canal (CIC). In addition to the cosmetic advantages, the size reduction has brought significant acoustic benefits. BTE hearing aids are very popular, normally used by persons with severe to profound hearing loss. The components are contained in a small elliptical case, which gets fitted behind the ear. A flexible tube, transfers the signal to the ear canal, terminates in the earmold. ITE hearing aids are placed in the concha and it preserves some directional properties of the pinna. Almost all directional cue of the external ear are preserved in ITC, which pick up sound directly from front of the blocked ear canal. This kind of hearing aid is normally used in cases of mild to moderate hearing loss. CIC placed in the ear canal, terminated beyond the second bend, is invisible externally. It provides greater overall sound level especially at high frequencies with about 5 to 10 dB less gain than ITC (CHABA, 1991; Sandlin, 2000; 1994; Studebaker and Hockberg, 1993; Wall, 1995).

The hearing aids usually provide linear amplification to increase the intensity level above the listener’s threshold and amplitude compression to accommodate the speech level into the reduced dynamic range of the hearing impaired. Some of the compression techniques used in hearing aids in the recent past, including single and multi-band compression, are discussed in Section 2.6.1. In another technique, high frequency information is transposed to low frequencies, in case of persons having some residual hearing in low frequencies (discussed in Section 2.6.2). Some schemes, that have tried to improve the speech perception by enhancing the spectral and temporal cues, are also discussed in Sections 2.6.3 and 2.6.4 respectively.

2.6.1 Compression techniques

Persons with sensorineural hearing loss do not have sufficient residual dynamic range for the normal intensity range of speech signals, due to increase in threshold of hearing with no corresponding increase in discomfort levels. Compression techniques are used to bring the intensity level of sounds in the audible range of normal hearing persons into the residual dynamic range of the hearing impaired, restricted to varying amounts in different frequency regions.

Compressors are classified as compression limiting, syllabic, and automatic gain control (AGC) according to their characteristics (Hickson *et al*., 1994; Lunner, 1997; Moore, 1998; Sandlin, 1988; Stone and Moore, 1992). Compression limiting with short time constants (attack time < 5 ms and release time between $20-100$ ms), high compression threshold, and a high compression ratio, operate as a linear amplifier for commonly occurring inputs and are activated by high level sounds to limit the level to discomfort level. Syllabic compressor preserves the overall shape of the speech envelope. In syllabic compression, the level differences between successive speech syllables are reduced. It is characterized by short time constants (typically $20-100$ ms), low compression threshold and low compression ratio, and are suitable for a listener with an extremely small dynamic range. AGC systems with long time constants (> 200 ms), low compression threshold, and a high compression ratio, adjust the gain of the amplifier as a function of the long term average level of the input. The gain of AGC remains low for sometime after a loud sound.

Single channel compression systems, cannot effectively deal with large threshold changes of subjects with strong frequency dependent hearing impairment. Multiband compression techniques in which the whole range of frequencies is separated into different bands, each frequency band having independent compression parameters, is thought to be superior, especially for severely hearing impaired (Moore *et al*., 1993; Villchur, 1973; Yund and Buckles, 1995b). Across the various studies, number of channels used lies between 2 and 16. Since there are many differences other than number of channels, it is difficult to come to a concrete conclusion regarding the optimal number of channels to be used. An 8-channel compression hearing aid with frequency dependent amplitude compression was found effective in some sensorineural hearing impaired listeners (Villchur, 1973). In case of many bands, input speech signals passing through many narrowband filters each with independent gain, undergo spectral distortion or flattening (Lippmann *et al*., 1981; Nabelek, 1983; Plomp, 1988). The discrimination of vowels and voiced stop consonants is reduced at high compression ratios and with more bands (Crain and Yund, 1995; Plomp, 1994). Yund and Buckles (1995a) found improvement in place feature recognition as subjects gained more experience with multi-band compression. Yund and Buckles (1995b) reported improvement in the discrimination of nonsense syllables in noise as the number of channels increased from 4 to 8 and remained same for 8 to 16. Their speculation was that with more than 8-channels the identification of certain speech cues e.g., manner of articulation, might improve, whereas certain others e.g., place of articulation, may be degraded. There is also a strong interaction between number of channels and the compression ratio. At high compression ratios, the negative effect of increasing number of channels increases (Crain and Yund, 1995).

A hearing aid referred to as CLAIDHA (Compensate for Loudness by Analyzing Input-signal Digital Hearing Aid) which employs frequency-dependent amplitude compression based on narrowband loudness compensation, has been reported to provide better speech intelligibility especially for subjects with flat and gradually sloping hearing loss (Asano *et al*., 1991; Hidaka *et al*., 1998).

2.6.2 Frequency lowering techniques

Frequency lowering is intended to take high frequency information and shift them into the low frequencies in case of persons having some residual hearing in low frequencies (CHABA, 1991; Sandlin, 2000). The basic methods are frequency transposition and frequency division (Levitt *et al*., 1980). In case of a transposer, a portion of the spectrum of the signal is separated out and resynthesized in a lower frequency band. In frequency division (compression), the frequencies of the signal are compressed by a fixed ratio. The perception of speech is not affected much by acoustic variation when the ratio relationship between energy peaks is maintained. Frequency compressed speech does not sound quite natural and identification requires intensive long-term training.

The discrimination of consonants in monosyllables processed with frequency lowering technique, was compared with low pass filtering of equivalent bandwidths by normal hearing listeners (Reed *et al.*, 1983). Although the overall performance was not different, the patterns of perception of articulatory features were different. Frequency lowering technique showed improved perception of fricatives but the perception of nasals and semivowels was degraded. Very often the frequency lowering schemes introduce different types of distortions, due to alteration of perceptually important characteristics of speech.

2.6.3 Processing to improve the perception of spectral cues

The perception of spectral cues get affected due to frequency dependent shifts in hearing threshold and due to increased spectral masking. This section reviews some of the schemes, which studied the effect of filtering the speech, to enhance certain spectral cues and improve speech perception. Gordon-Salant (1984) investigated a scheme for providing enhancement to the weak, high frequency spectral cues in the speech signal, by attenuating the more intense low frequency energy. Consonant identification performance was evaluated on subjects with flat hearing loss and high frequency hearing loss. Better recognition scores were obtained for persons having high frequency losses, with low frequency attenuation.

Horwitz *et al*. (1991) studied the effect of different frequency responses obtained by changing the relative amounts of low and high frequency amplification, for subjects with high frequency sensorineural hearing loss, and indicated no significant effect upon speech recognition scores once the audibility of the entire speech signal was ensured. Kamm *et al.* (1982) studied the effect of spectral shaping using three frequency responses namely, uniform frequency gain, high pass filtering, and a response shaped relative to each subject's loudness discomfort level curve. Speech recognition performance was conducted on hearing impaired listeners with flat and steep audiometric configuration, using tests involving nonsense syllables (NST) and tests involving synthetic sentences (SST), conducted at four levels (from 80 to 95 dB SPL) in the background of cafeteria noise. No significant difference was observed in NST performance, across different spectral shapes in the two groups of hearing impaired subjects. With SST, the performance with uniform frequency gain, improved for listeners with flat audiometric configuration, and degraded for listeners with steep audiometric configuration. In a similar investigation Gabrielsson *et al*. (1988), used five frequency responses: one flat and others combinations of reductions at lower frequencies and/or increases at higher frequencies. The most preferred scheme was characterized by flat response at low frequencies with 6 dB/octave increase thereafter. The signal level, type of speech, SNR, and audiometric configuration all may interact and influence the effects of spectral shape on speech recognition performance of hearing impaired listeners.

In an attempt to reduce the effect of upward spread of masking, Summers and Leek (1997) attenuated the energy of the first formant (F1) of vowel in six synthetic consonant-vowel (CV) stimuli /b*a*/, /d*a*/, /g*a*/, /be/, /de/, and /ge/. Normal hearing and hearing impaired listeners were asked to label the stimuli presented in quiet and in broadband noise sufficient to mask the initial burst release. The performance improved for F1 attenuation up to 18 dB particularly in the /*a*/ vowel context. Consistent improvement for hearing impaired listeners, particularly for those with steep increase in hearing threshold between first and second formant regions, is reported.

2.6.4 Processing to improve the perception of temporal cues

Speech produced with an effort to be highly intelligible, in comparison with conversational speech in which clarity may be compromised, is known as clear speech. In conversational speech, duration of vowels is often reduced and bursts of stops occurring in word-final positions are not released. Clear speech becomes acoustically more distinct (Kent and Read, 1992). In comparison with conversational speech, clear speech is slower due to longer pauses between words and lengthening of some speech segments. Also vowels are not reduced or modified and stop consonants are released.

Some researchers have dealt with the improvement in intelligibility associated with clear speech for hearing impaired listeners (Payton *et al.*, 1994; Picheny *et al*., 1985; 1986; 1989; Uchanski *et al*., 1996). Picheny *et al.* (1985) reported that significant improvement in intelligibility is obtained, when speech is produced with an effort to make it clear. The improvement was roughly independent of the listener, presentation level and the frequency gain characteristics of the amplifying system used.

Picheny *et al.* (1985; 1986; 1989) have reported that uniform adjustment of duration of conversational speech degraded the intelligibility. Nejime and Moore (1998) have commented that uniform time expansion makes the speech signal sound unnatural. Further non-uniform speech-rate reduction, to make the speech sound similar to natural slow speech, was evaluated in listeners with either flat or sloping sensorineural hearing losses and in normal hearing listeners in noisy background (Uchanski *et al*., 1996). Intelligibility increased compared to uniform time expansion, but it was still less than that for original speech. Even though slow (clear) speech is more intelligible compared to normal speech, artificial time expansion has helped in improving the intelligibility of speech only in few hearing impaired listeners (Nejime and Moore, 1998).

In the studies conducted by Gordon-Salent (1986; 1987), two properties were modified in conversational speech, on the basis of known improved characteristics of clear speech. The consonant-to-vowel intensity ratio (CVR) was increased by 10 dB and the consonantal duration was increased 100 % for a set of 19 consonants conversationally spoken in CV syllables. Elderly normal hearing listeners in a background of babble and hearing impaired listeners benefited from the increase in CVR especially for consonant place, manner, and voicing features, but not from the increase in consonant duration.

Thomas (1996) and Thomas *et al*. (1996) studied the effect of increase in CVR and increase in the duration of certain acoustic segments of the synthesized speech stimuli. The CVR was increased by +3, +6, +9, and +12 dB. Four sets of synthesized speech were used as test material, involving three stop consonants / p, t, k / in consonant-vowel (CV) and vowel-consonant (VC) context with vowel /*a*, i, u/ and six stop consonants / p, t, k, b, d, g / in CV and VC context with vowel /*a*/. Listening tests conducted on normal hearing subjects with simulated hearing loss showed significant increase in recognition score for all test material sets with increase in CVR. Higher scores were obtained for VC position than for CV position, indicating that an increase in CVR is more effective in reducing forward masking than backward masking. At high CVR, few vowel confusions occurred in the VC context. In the above study, the duration of acoustic segments like bursts, closure, formant transitions, and VOT was increased in synthesized speech stimuli involving six stop consonants / p, t, k, b, d, g / in consonant-vowel (CV) context with vowel /*a*/. The overall duration was kept constant at 300 ms, by adjusting the vowel duration. Listening tests were conducted on four normal hearing subjects with simulated hearing loss under three SNR conditions of ∞, 12 and 6 dB. It was reported that, under adverse listening conditions, better performance could be obtained for a formant transition duration modification of up to 50 % combined with burst duration modification. VOT modification provided improvement for only one subject.

Kennedy *et al*. (1998) investigated the effect of CVR on consonant recognition with 48 VC nonsense syllables adjusted in steps of 3 – 6 dB depending on the subject's dynamic range on 18 sensorineural hearing impaired listeners. CVR for maximum consonant recognition was found to have significant relation to consonant type and a small relation to vowel environment and audiogram configuration of the subject. Substantial improvements in consonant recognition were obtained due to individual adjustments made for each subject for each consonant-vowel combination. Such adjustments may be difficult to incorporate in practical hearing aids.

Enhancement of speech cue by altering the vowel duration led to substantial improvement in the perception of voicing in final fricatives for hearing impaired listeners (Revoile *et al*., 1986). The improvement in voicing perception was observed when the vowels preceding voiced fricative was on an average 73 % longer than that preceding an unvoiced fricative.

2.6.5 Binaural hearing aids

In the previous sub-sections, speech processing schemes for improving the speech perception of persons with sensorineural hearing have been reviewed. Many persons have some hearing in both the ears and benefit by using binaural hearing aids. In binaural hearing, presenting same signal to both ears is termed as diotic presentation and presenting different signals to the two ears is termed as dichotic presentation. Dichotic presentation schemes, in which the speech signal is split into two to form two complementary signals, have helped in reducing the effect on increased masking in persons with moderate bilateral sensorineural hearing impairment. Chapter 3 reviews these schemes and gives an outline of the proposed study.

Chapter 3

BINAURAL DICHOTIC PRESENTATION SCHEMES

3.1 Introduction

Sensorineural hearing impairment results from the malfunctioning of the cochlea or auditory nerve. It is characterized by increased threshold of hearing, loudness recruitment, and increased spectral and temporal masking in the auditory system. This impairment cannot be treated medically, but its effects can be reduced, by processing the acoustic signal before its presentation to the auditory system, taking into consideration the different characteristics of sensorineural loss. Some of these processing techniques are reviewed in the last chapter. In case of persons who can use binaural hearing aids, speech signal can be processed to take advantage of binaural presentation for reducing the effects of masking and thereby improving speech perception. In this chapter, first the earlier studies on binaural dichotic presentation schemes for reducing the effect of increased spectral and temporal masking are reviewed. Sections to follow give an outline of the proposed study.

3.2 Binaural dichotic presentation

Listening through the two ears enhances aural discrimination for both normal hearing and hearing impaired persons, especially in noisy environment. The advantages of binaural hearing include, increased frequency and intensity discrimination, loudness summation, improved intelligibility of speech in noise, and better source localization (Hall *et al.*, 1984; Helfer, 1994; Jesteadt and Wier, 1977; Marks, 1978). Binaural hearing could be with diotic or dichotic presentation. In diotic presentation, same signal reaches both the ears. In dichotic presentation, different signals reach the two ears (Gelfand, 1998; Moore, 1997). Masking takes place at the peripheral level of the auditory system. Speech perception at higher auditory levels involves the integration of information received from both the ears. Hence splitting of speech signal in a complementary manner and presenting it to the two ears can be used to reduce the effect of increased masking. Speech processing schemes, using splitting of speech for dichotic presentation, are reviewed in the following sections.

3.3 Speech processing schemes using spectral splitting

Spectral masking reduces spectral contrast and frequency components lying close to each other become less distinguishable. A component having higher level will mask a band of frequencies lying close to it. Such bands of frequencies are represented in the form of the auditory critical bands. Spectral masking increases with sensorineural loss. The auditory filters act as though the bandwidth is broader than normal. Speech processing schemes making use of binaural dichotic presentation have been investigated, to reduce the effect of spectral masking. Such schemes split the speech signal into two and present it to the two ears, with the objective of separating the frequency components that are likely to mask or get masked, making use of the fact that masking takes place at the peripheral level of the auditory system.

Upward spread of masking may be responsible for the reduction in detection of F1 and F2 formants. In an investigation by Turek *et al.* (1980), impaired listeners were asked to identify a continuum of three-formant synthetic stimuli in which place of articulation was cued only by the second and third formants transitions as /b*a*/, /d*a*/, or /g*a*/. There was a great variability in the perception reported by listeners as (i) all the formants were presented to one ear and (ii) F1 was presented to one ear while F2 and F3 to the other. Identification was consistently better for dichotic presentation for only one or two of the nine hearing impaired listeners.

Rand (1974) investigated a dichotic method of listening with synthetic CV speech syllables, /b*a*/, /d*a*/, and /g*a*/. The experiments were conducted in two parts. In the first part, formant F1 was presented to one ear and formants F2 and F3 with different levels of attenuation were presented to the other ear. In the second part, F2 and F3 transitions with different levels of attenuation were presented to one ear and the remainder of the syllable to the other. In the two experiments conducted on four normal hearing subjects, at each presentation level the dichotic presentation was compared with binaural presentation in which both the signals were mixed and presented to both the ears (diotic). For 90% recognition score, the attenuation in the level of formants F2 and F3 for dichotic presentation could be 20 dB less than that for diotic presentation. For the second experiment, the attenuation in transitions could be 5 dB less in dichotic presentation than in diotic.

Lyregaard (1982) used an analog delay for realizing complementary comb filter for the purpose of splitting speech signal. The delayed signal was added to and subtracted from the original signal to obtain two separate signals for binaural dichotic presentation. The delay was varied to obtain comb filters with different bandwidths. Two lists, each with 25 words, were used as test material for diotic and dichotic presentation, at signal-to-noise ratios (SNR) of 12 dB and 4 dB. But, the improvement of dichotic over diotic was insignificant. Response of the filter used in this study is discussed later in Chapter 4.

An 8-channel digital filter bank in monaural, diotic and dichotic modes was tested by Lunner *et al*. (1993). It was designed as complementary interpolated linear phase FIR filters, with a constant bandwidth of approximately 700 Hz. For dichotic presentation, the odd bands were presented to one ear and even bands to the other ear. An overall improvement of 2 dB in SNR was observed for dichotic condition with respect to diotic. Focus in the design was on efficiency, and not on separation of bands and undistorted perception for spectral components at band crossovers. In Chapter 4 these filters are dealt in more detail.

Chaudhari and Pandey (1998a, b) implemented an audio-signal processing scheme for splitting the speech into two, with two complementary comb filters with magnitude response based on eighteen critical bands, corresponding to the auditory pattern described by Zwicker (1961). The bandwidths of these filters were constant at 100 Hz for center frequencies below 500 Hz and increased for further increase in center frequencies up to 5 kHz, reaching a maximum bandwidth of approximately 1 kHz. For dichotic presentation, speech signal filtered by the alternate bands were presented to the two ears. A real time implementation and testing of the scheme was conducted using two TI/TMS320C50 DSP processors. For real time processing, the two comb filters were realized as 128-coefficients FIR filters, with filters designed by frequency sampling technique, to have sharp transition between pass and stop bands. They investigated two types of comb filters: (a) with constant gain in each of the pass bands and (b) with pass band gain adjusted between ±3 dB, to partly compensate the frequency dependent shifts in hearing threshold. Chapter 4 deals with the comb filters in more detail.

Experimental evaluation was done on normal hearing subjects with simulated hearing loss and on subjects with bilateral sensorineural hearing loss. The test material consisted of twelve consonants /p, t, k, b, d, g, m, n, s, z, f, v/ in vowel-consonant-vowel (VCV) and consonant-vowel (CV) contexts with vowel /*a*/. Hearing impaired subjects showed significant improvement in recognition scores. The percentage relative improvement in recognition score was in the range of 9.2 – 23.6 in VCV context and was between 14.4 – 19.2 in CV context, for hearing impaired subjects with bilateral sensorineural impairment. The contribution of each feature (duration, frication, nasality, manner, place and voicing) was obtained from information transmission analysis. The improvement obtained was contributed by all features, with maximum contribution from place feature. The capacity to distinguish place feature is associated with the frequency resolution capacity of the auditory system. Hence maximum contribution from place feature indicates the usefulness of the scheme in reducing the effect of spectral masking. With adjustment of pass band gains, the percentage relative improvement in recognition score further increased, by 2.2 – 6.4 and 1.6 – 7.8 in CVC and CV contexts respectively. All the features contributed to the overall improvement, but there was no distinct contribution from place feature. The gain adjustment to compensate the hearing loss at different frequencies did not cause any further decrease in the effect of spectral masking.

In a recent investigation, Murase *et al*., (2004) divided the speech stimuli into two frequency bands, using a low pass filter and a high pass filter, considering the formant frequencies of Japanese vowels. Diotic listening with and without 6 dB attenuation and dichotic listening with filtering done at two different crossover frequencies at 0.8 kHz (between formant frequencies F1 and F2) and 1.6 kHz (middle of F2) were compared. To simulate different listening conditions, speech spectral noise and road noise recorded on highway were added to the speech stimuli at SNRs of 4 and 0 dB. When tested on four sensorineural hearing impaired persons, speech intelligibility was improved for dichotic listening with crossover of 0.8 kHz, particularly in no noise condition. Further sound localization tests were conducted on six normal hearing and three elderly hearing impaired persons. For hearing impaired persons sound was appeared to be located on the side where low frequency portion was presented and for normal hearing subjects the sound appeared to be from two different locations.

3.4 Speech processing schemes using temporal splitting

In temporal masking, low energy subsegments of speech get suppressed by adjacent high energy segments. In an investigation conducted in our lab (Jangamashetti, 2003; Jangamashetti and Pandey, 2000a, b) to reduce the effect of temporal masking, the speech signal was temporally segmented and the alternate segments were presented to the two ears so that successive segments that are likely to mask or get masked by each ear are presented to different ears. In this scheme, step and trapezoidal fading functions with inter-aural switching period of 20 ms were used to split the speech, to be presented to the two ears. A schematic representation of the scheme is shown in Fig. 3.1 (a). Figure 3.1 (b) shows the trapezoidal fading functions with inter-aural switching period T_c. This scheme was tested, by conducting listening test on normal hearing subjects with simulated hearing loss. Sensorineural hearing loss was simulated by adding broadband noise with short-time SNR. Listening tests involved a closed set evaluation of twelve nonsense syllables in vowel-consonant-vowel (VCV) context, with the consonants /p, t, k, b, d, g, m, n, s, z, f, v/ and the vowel /*a*/.

The scheme was evaluated using step fading function, with different duty cycles (> 50 %) to study the effect of overlap between segments. It was found that temporal splitting of speech helps in improving the response time, consonant recognition scores, and relative information transmitted for overall and for different speech features under all test conditions analyzed. The improvement was highest for 70 % duty cycle. The overlap between segments must have helped by reducing the perception of temporal gaps. As step fading function introduced severe spectral distortion, use of trapezoidal fading function with different transition durations, was evaluated for 70 % duty cycle. The intelligibility of speech was further improved, response times decreased, recognition scores and information transmission of features improved. The scores were higher for transition durations of 2 and 3 ms. Relative improvements were higher under adverse listening conditions. From the information transmission analysis, it was observed that the maximum improvement was for duration feature (244 %) at −6 dB SNR, indicating that the scheme helped in reducing the effect of temporal masking. The relative improvement for place feature was 71 % at −6 dB SNR. Processing with trapezoidal fading function with slower variation resulted in significant additional improvement in the place feature at lower SNR conditions. The relaxation period provided between the alternate segments helped in improving the temporal resolution of the ears and this resulted in increased perception of temporal cues.

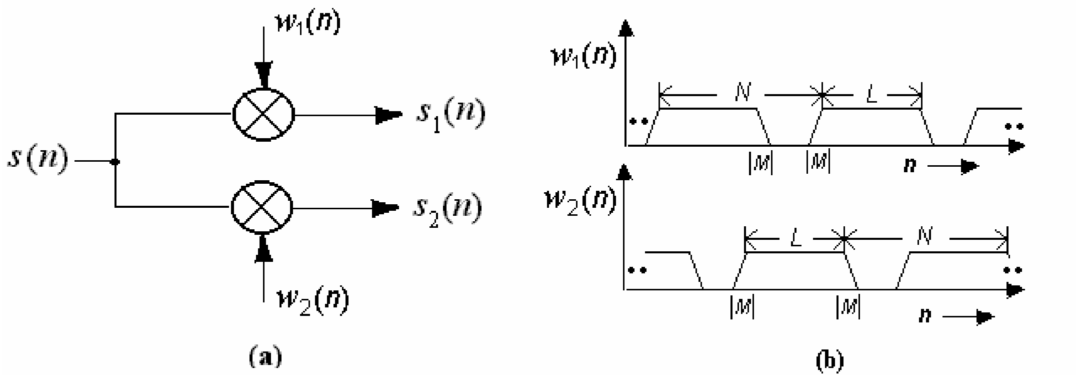


FIG. 3.1. A schematic representation of the temporal splitting scheme and the trapezoidal fading function with sample frequency of f_s, inter-aural switching period $T_c = N/f_s$, transition duration = M/f_s, $L= N–2M$, duty cycle = L/N. (Jangamashetti, 2003).

3.5 Speech processing schemes using combined splitting

Previous two sections have reviewed speech processing schemes involving binaural dichotic presentation based on spectral splitting and temporal splitting. In the scheme of spectral splitting, the sensory cells corresponding to alternate bands of the basilar membrane are always stimulated, whereas sensory cells of the other bands are always relaxing. In the temporal splitting scheme, all the sensory cells of the ears get relaxed alternately for some time. A combined splitting scheme can be suitably devised to provide all the sensory cells of the basilar membrane periodic relaxation from stimulation, and thereby achieve a simultaneous reduction in the adverse effects of increased spectral and temporal masking.

In the study by Lunner *et al.* (1993), an 8-channel digital filter bank with constant bandwidth of approximately 700 Hz was implemented for spectral splitting of speech and 2 dB improvement in SNR was obtained (reviewed in section 3.2). For combined spectral and temporal splitting, a scheme was developed in which the comb filters were alternated between the two ears after every 10 ms. No further improvement over spectral splitting could be obtained, and switching of bands resulted in poor sound quality. The lack of improvement and deterioration of speech quality may have been caused due to abrupt transitions during switching of odd and even bands between the two ears.

In an investigation in our lab, combined splitting was done using a pair of time varying comb filters with pre-calculated sets of coefficients (Jangamashetti, 2003; Jangamashetti *et al*, 2001; Pandey *et al.*, 2001). For an implementation with *m* shiftings, each of the time-varying comb filter contained pre-calculated sets of coefficients of *m* perceptually balanced comb filters, which have magnitude responses such that the pass bands of each of these comb filter pairs are shifted in a complementary manner along the frequency axis. The sets of coefficients were selected in steps to process the speech such that a cyclic sweeping of magnitude responses occurs. At any time a pair of comb filters, with complementary magnitude responses, processed the speech and provided spectral splitting. Temporal splitting was obtained by the sweeping of bands. The scheme was implemented for off-line processing for 2, 4, 8, and 16 shiftings, for a constant sweep cycle of 20 ms. Experimental evaluation of the scheme was done, by conducting listening test on normal hearing subjects with simulated hearing loss. Different levels of sensorineural hearing loss (∞, 6, 3, 0, −3, −6, −9, −12, −15 dB), was simulated by adding Gaussian noise with short-time SNR to the test stimuli. Listening tests involved a closed set evaluation of twelve nonsense syllables (/ *a*p*a*, *ata*, *a*k*a*, *a*b*a*, *a*d*a*, *aga*, *a*m*a*, *a*n*a*, *asa*, *aza*, *a*f*a*, *ava* /).

The scheme provided improvement in response times, recognition scores, and transmission of features with an increasing trend from high SNR to low SNRs for all shiftings. The relative improvement in recognition score was maximum at −15 dB SNR and was 32 and 29 % for 4 and 8 shiftings respectively. Improvements were statistically significant for higher levels of noise. It was observed from the information transmission analysis that more improvement was obtained for duration and place features. The relative improvements for 4 and 8 shiftings were 177 and 149 % for place feature and 365 and 318 % for duration feature at −15 dB SNR conditions. Perception of frication and manner features also improved for adverse listening conditions. The scheme was successful in simultaneously reducing the effects of increased temporal and spectral masking.

Further an overall comparison of the schemes of temporal, spectral, and combined splitting was carried out by conducting listening tests, on five subjects with moderate to severe bilateral sensorineural hearing loss (Jangamashetti, 2003). All the schemes provided different degrees of improvement to subjects, depending on the individual’s hearing loss configuration. Persons with high frequency hearing loss preferred temporal splitting, while persons with low frequency and gradual sloping symmetrical loss preferred combined splitting.

3.6 Proposed study

Dichotic presentation schemes to reduce the effect of spectral masking and temporal masking separately and simultaneously, have been reviewed in the previous sections of this chapter. The present investigation aims to explore the possibilities of further improving speech intelligibility by persons with moderate bilateral sensorineural hearing impairment and by normal hearing persons under adverse listening conditions. This is carried out in two phases. The first phase of investigation, includes design of perceptually balanced comb filters for spectral splitting, to reduce the effect of increased spectral masking. The second phase is to implement and evaluate the three speech processing schemes namely, spectral splitting, temporal splitting, and combined splitting for different processing parameters.

Chaudhari and Pandey (1998a, b) investigated spectral splitting by using a pair of comb filters having complementary magnitude responses based on auditory critical bands, as reviewed earlier in Section 3.3. The odd and even bands which form a pair of comb filters split the speech into two, such that the frequency components in the adjacent bands are presented to different ears. The comb filters used were 128-coefficient linear phase FIR filters, designed using frequency sampling technique. The filters had sharp transition between pass and stop bands. While listening with slowly sweeping pure tones through these filters, an imbalance in loudness was perceived at different frequencies and also at the crossovers between adjacent bands. This can be due to the high ripple in the pass band (maximum of 4 dB), low attenuation in the stop band (minimum of 10 dB), and finite crossovers in the magnitude response, which causes the spectral components lying in the transition regions to be presented to both the ears at different levels.

Hence, it is proposed to design perceptually balanced comb filters with three parameters in consideration, adjustment in magnitude response at transition crossovers to minimize the changes in intensity perception, reduction in passband ripple, and increase in stop-band attenuation. Chaudhari and Pandey (1998b) had provided compensation for frequency dependent shift in hearing threshold, by adjusting the pass band gains of the comb filter over ± 3 dB. In our investigation, it is proposed to cascade a filter with an audiogram dependent magnitude response with gain adjusted between ± 3 dB to partly compensate for frequency dependent shifts in hearing threshold, along with the perceptually balanced comb filters for the left and right ear independently.

In another investigation, Jangamashetti and Pandey (2000a, b), implemented a scheme of temporal splitting to reduce the effect of temporal masking, in which the speech was split into time segments and the alternate segments were presented to the two ears so that the adjacent segments that are likely to mask or get masked are presented to different ears, as reviewed in Section 3.4. The scheme was implemented for step and trapezoidal “fading functions” with inter-aural switching period of 20 ms and for different duty cycles and transition durations, and evaluated by conducting listening test on normal hearing subjects with simulated hearing loss. Maximum improvement was obtained for trapezoidal fading functions with 70 % duty cycle for 2 and 3 ms transition durations. Effect of inter-aural switching period on speech perception needs to be investigated.

In the scheme of combined spectral and temporal splitting, a pair of time-varying comb filters with magnitude responses cyclically swept, provided the spectral and temporal splitting of speech simultaneously, as reviewed in Section 3.5. At a given time, two comb filters with complementary magnitude responses processed the signal, providing the spectral splitting of speech. Temporal splitting was obtained by the sweeping of bands in the ears. The implementation and evaluation of combined temporal and spectral splitting scheme using time-varying comb filters, was carried out for a constant sweep cycle of 20 ms (Jangamashetti, 2003; Jangamashetti *et al.*, 2001; Pandey *et al.*, 2001). It is proposed to optimize the scheme of time-varying comb filters for two parameters namely, the time duration for a sweep cycle and the number of shiftings. Since in combined splitting the filter bandwidths are swept cyclically, the stimulation on the sensory cells moves slowly along the length of the basilar membrane. The time duration of the sweep cycle selected, determines the time for which the sensory cells are stimulated in one cycle. Hence the optimized value for sweep cycle duration will depend on temporal integration of the auditory system.

Further it is proposed to carry out an overall evaluation of the three schemes: spectral splitting, temporal splitting, and combined splitting with different parameters, initially on normal hearing subjects with simulated hearing loss and then on subjects with bilateral sensorineural hearing impairment.

Investigations for the design and evaluation of perceptually balanced comb filters are referred as first phase. Optimizing the different parameters of the time-varying comb filters for combined splitting and overall evaluation of the three splitting schemes are referred as the second phase.

3.7 Evaluation methods

The perception of speech by hearing impaired persons, is evaluated by speech intelligibility tests, which determine the ability to understand speech presented at comfortable listening levels. Various researchers have used different types of material, measurement techniques, and presentation methods. In the following sections, some of these are reviewed. Along with this, the test material, presentation methods and measurement techniques used in the present research work are also discussed.

3.7.1 Listening test material and techniques

Different researchers have used variety of speech materials to conduct listening tests for the investigation of perception by normal hearing and hearing impaired listeners under various conditions. These present a range of acoustic, phonetic and lingustic, and lexical variables, and vary in complexity from vowels (ter Keurs *et al*., 1992), nonsense syllables (Dubno and Schaefer, 1992; Gutnick, 1982; Hickson *et al*., 1994; Humes *et al*., 1987; Pandey, 1987), single words (Monotgomery and Edge, 1988; Tyler *et al*., 1983), spondee (Dubno *et al*., 1984; Nilson *et al*., 1994), to sentences (Moore *et al*., 1998; ter Keurs *et al*., 1993). Synthesized speech stimuli are attractive, since these provide independent control over spectral, temporal and intensity characteristics (Dorman and Hannley, 1985; Turner and Robb, 1987; Turner *et al*., 1992).

Most word recognition tests consist of about fifty monosyllables representing familiar words that are equally difficult to understand. The word lists are also phonetically balanced (PB) to represent the frequency of occurrence of sounds in everyday life (Lippmann *et al.*, 1981). Nonsense syllables in different contexts like consonant-vowel-consonant (CVC), vowel-consonant (VC), consonant-vowel (CV) and vowel-consonant-vowel (VCV) have also been used for evaluation in many researches (Miller and Nicely, 1955; Wang and Bilger, 1973; Wilson *et al*., 1984). The responses from 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. Some researchers have added noise to the processed speech and evaluated signal-to-noise ratios for 50 % correct recognition scores (Dubno *et al*., 1984; Gabrielsson *et al*., 1988; Lunner *et al*., 1993; Lyregaard 1992; Van Dijkhuizen *et. al*., 1991).

There are different approaches to analyze the subject responses. Recognition scores are used very often, which uses a binary criterion i.e. correct or incorrect (Dorman and Hannely, 1985; Humes *et al*., 1987). In another approach, stimulus-response confusion matrix is constructed for response analysis (Gutnick, 1982; Kates, 1994; ter Keurs *et al*., 1992; Tyler and Moore, 1992). This helps to find the pattern of phoneme confusions. Information transmission analysis provides a measure of covariance between stimuli and responses (Miller and Nicely, 1955). It takes care of chance scoring and pattern errors. Information transmission analysis can be carried out for different feature grouping of stimulus and it has been used for evaluation in many studies (Apoux *et al*., 2000a, b; Chaudhari and Pandey, 1999b, Hou and Pavlovic, 1994; Jangamashetti, 2003; Miller and Nicely, 1955; Pandey, 1987; Revoile *et al*., 1991a; Thomas, 1996; Tyler and Moore, 1992; Wang and Bilger, 1973). For assessing speech processing schemes, some researchers have used response time statistics (Apoux, 2001; Chaudhari and Pandey, 1999b; Gatehouse and Gordon, 1990; Jangamashetti, 2003; Meftah and Boudelaa, 1996; Pandey, 1987; Thomas *et al*., 1996). Decrease in response time is an indicator of decreased load on perception process.

Earlier studies have used CV, VC, CVC, and VCV syllables. It has been reported earlier (Revoile *et al*., 1991a) that greater masking takes place in intervocalic consonants due to the presence of vowels on both sides. Since our primary objective is to study improvement in consonantal identification due to reduction in the effect of masking, VCV syllables was considered as the most appropriate test material.

In the first phase of evaluation, a closed set evaluation of twelve nonsense syllables in VCV context was carried out with consonants / p, b, t, d, k, g, m, n, s, z, f, v / and vowel /*a*/ as in f<u>a</u>ther. 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 information transmitted. Further the consonants were grouped according to the articulatory features (Miller and Nicely, 1955) and the contribution of different features was analyzed. The features selected for study were voicing (voiced: / b d g m n z v / and unvoiced: / p t k s f /), place (front: / p b m f v /, middle: / t d n s z /, and back: / k g /), manner (oral stop: / p b t d k g /, fricative: / s z f v /, and nasals: / m n /), nasality (oral: / p b t d k g s z f v /, nasal: /m n /), frication (stop: / p b t d k g m n /, fricative: / s z f v /), and duration (short: / p b t d k g m n f v / and long: / s z /).

The assessment by closed set evaluation using nonsense syllables may overestimate the ability of hearing impaired persons to perceive conversational speech. Presentation of the twelve VCV nonsense syllables monotonously, causes irritation and fatigue in the subjects. Hence for the second phase of evaluations, an open set evaluation with sets of phonetically balanced monosyllabic words in three languages (Hindi, Marathi, and English) were used. In each language there were 50 to 60 monosyllabic words. The word lists were obtained from Ali Yaver Jung National Institute for Hearing Handicapped (AYJNIHH), Mumbai, where these monosyllables are used to evaluate the discrimination capacity of hearing impaired persons. The lists of monosyllables are shown in Tables F.1, F.2 and F.3. Each subject chose one of these languages for the listening tests. In addition to subject response, response time was also recorded. Recognition score and average response time were calculated for the evaluation of the schemes.

Speech discrimination capacity reduces at very low (< 30 dB) and at very high (> 90 dB) presentation levels (Dorman and Dougherty, 1981; Gabrielsson *et al*., 1988; Gorden-Salant, 1987; Simon, 1978). Mean value for comfortable listening levels are 83 dB SPL for normal hearing subjects and 16 – 20 dB more for hearing impaired subjects with moderate loss (Gabrielsson *et al*., 1988). In this investigation, listening tests were carried out at the most comfortable level of the individual subject, both normal hearing and hearing impaired. For a subject, the same level was maintained for all the tests in a set of experiments.

3.7.2 Simulation of hearing loss

For the evaluation of different speech processing schemes, listening tests involving recognition of speech syllables is necessary, to understand the benefit of the processing scheme and to optimize the different parameters. This is time consuming and many times causes fatigue in hearing impaired subjects undergoing listening tests, even though the presentation is done at the most comfortable hearing levels. Hence it becomes difficult to get sufficient number of bilateral sensorineural hearing impaired subjects, ready to participate in the listening tests for many hours. In our experimental evaluations, initial round of investigation was carried on normal hearing subjects with different levels of simulated hearing loss. For the second round of investigations, involving a smaller set of processing parameters, listening tests were conducted on subjects with sensorineural hearing loss.

Different types of simulation have been used to characterize the different aspects of impairment, including loudness recruitment (Moore and Glasberg, 1993; Nejime and Moore 1997; Villchur, 1974) and increased masking (Bear and Moore, 1993; 1994; ter Keurs *et al*., 1992; ter Keurs *et al*., 1993; Villchur, 1977). Villchur (1974) simulated loudness recruitment by splitting the speech signal into three frequency bands and applying dynamic range expansion with different ratios, in each band. The simulation was tested in the normal ears of four subjects with severe acquired unilateral hearing loss. Subjects judged the simulated stimuli presented to the normal ear, to be similar to unprocessed stimuli presented to the impaired ear. Moore and Glasberg (1993), in a similar approach split the input signal into thirteen bands and processed the envelope in each band to simulate loudness recruitment.

The reduced frequency resolution of the auditory system was simulated by smoothing the envelope of the squared short-time fast Fourier transform (FFT) by convolving it with a Gaussian-shaped filter (ter Keurs *et al*., 1992; ter Keurs *et al*., 1993). The effects of reduced frequency selectivity was simulated by spectral smearing, using the overlap-add method (Bear and Moore, 1993; 1994). The smearing of the spectra of the stimuli evoked similar response in normal hearing persons as the broadened auditory filters of the hearing impaired persons.

Nejime and Moore (1997; 1998), investigated a scheme which simulates the combined effects of elevated threshold, loudness recruitment, and reduced frequency selectivity. Loudness recruitment was simulated by filtering the speech stimuli into number of frequency bands, and raising the temporal envelope of the waveforms at the output of each filter to a power greater than one. The effect of reduced frequency selectivity is simulated by smearing the short-term power spectrum of the stimuli in such a way that the excitation pattern produced in a normal ear, calculated over a short period of time, resemble unprocessed stimuli excitation of an impaired ear.

Simulation of sensorineural loss has often been carried out, by employing different types of masking noise. In a study conducted by Dubno and Schaefer (1992), hearing loss was simulated using spectrally shaped broadband noise and hearing threshold of normal hearing subjects was matched with hearing impaired subjects. Although in consonant recognition the results were similar, frequency selectivity of hearing impaired listeners was poorer than normal hearing subjects with simulated hearing loss. The results of frequency resolution, temporal resolution and speech recognition obtained from hearing impaired persons were used to predict the results on noise-masked normal listeners (Humes *et al*., 1988), the prediction was accurate for frequency resolution and speech recognition. Elevated thresholds can be simulated by adding broadband noise (Jesteadt, 1997). In a study to determine the minimum spectral contrast required for vowel identification, Leek *et al*. 1987, used broadband noise to simulate elevated thresholds in the range of 72 – 75 dB in normal hearing subjects.

In our investigations, we are concerned with the effects of increased spectral and temporal masking and not the compression of the dynamic range. Out of the various methods reported for simulation of masking, addition of noise is the simplest and has been shown to simulate elevated thresholds as well as increased masking, with degradation in speech perception being related to SNR. In both the phases of our investigation, evaluation was first carried on normal hearing subjects. For simulating hearing loss, broadband Gaussian noise band limited to speech frequency range, was added to the speech signal. Noise addition was carried out at specific SNR with respect to short time (≈10 ms) energy. During the segments of no signal, the background noise also will not be present. Different levels of loss were simulated by varying the SNR.

The effect of adding broadband Gaussian noise, with a specified short time SNR, to simulate sensorineural hearing loss was evaluated earlier in our lab (Jangamashetti, 2003), with VCV monosyllables, on five normal hearing subjects. The average response time of the subjects increased from 1.89 s at no-noise condition to 2.61 s at –15 dB SNR. This indicates increased difficulty in auditory perception. A plot of recognition scores and relative information transmission of consonantal features averaged across the subjects, is given in Fig. 3.2. The average recognition scores decreased from 100 % at no-noise condition to 64.5 % at –15 dB SNR condition. Corresponding decrease in overall information transmitted was from 100 % to 70 %. Relative information transmission analysis was carried out for different feature grouping of the consonants. Reception of voicing and nasality features was nearly unaffected with simulation. These features are known to be less susceptible to the adverse effects of temporal and spectral masking. Reception of the other features decreased from 100 % with decrease in SNR. The values were 29 %, 24 % 38 %, and 58 % for place, duration, frication, and manner respectively at −15 dB SNR. Further the results were matched with results obtained from five persons with moderate sensorineural hearing impairment. The values of the recognition score and the information transmitted averaged across five hearing impaired subjects corresponded to simulation level of −9 dB and −8 dB SNR conditions respectively. The relative information transmitted due to features of place, duration, and frication were approximately equal to that obtained from simulation levels of −7, −9 and −11 dB SNR conditions respectively. From this it can be inferred that the addition of broadband noise simulates the effects of increased temporal and spectral masking, and as the SNR degrades, the severity of masking effect increases.

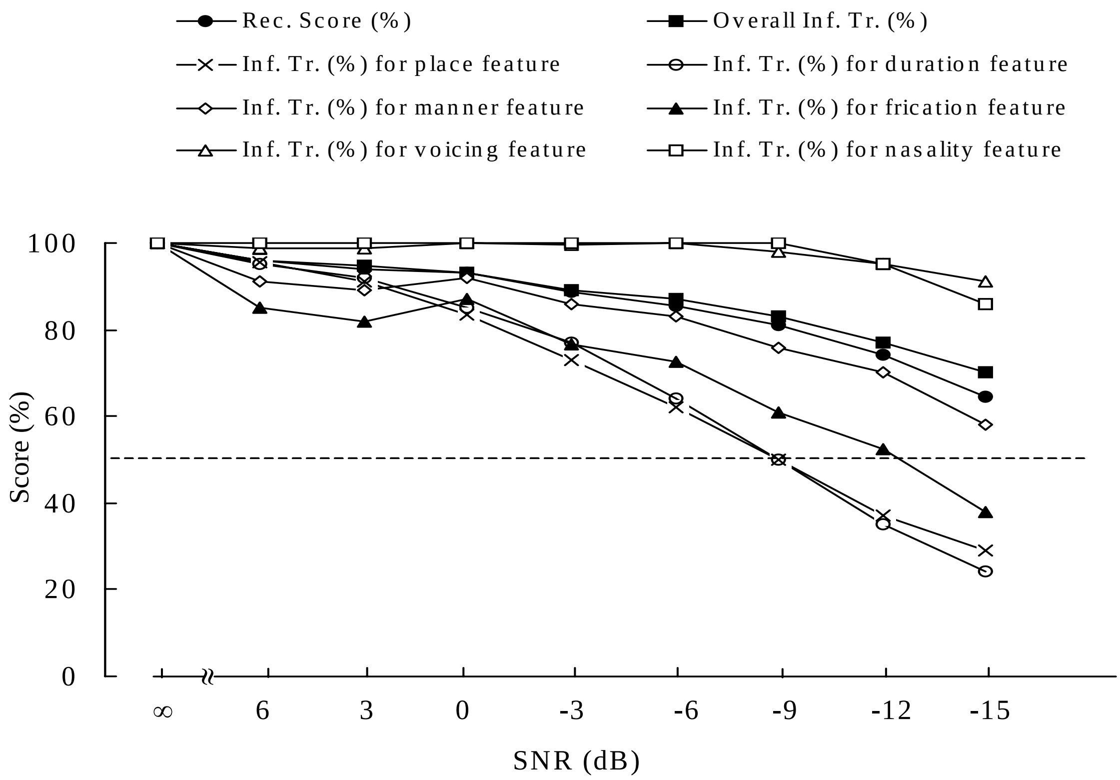


FIG. 3.2. The effect of simulation by adding broadband Gaussian noise, with constant short-time SNR, on recognition score and relative information transmission for different features (Jangamashetti, 2003).

This chapter has provided a review of binaural dichotic presentation schemes, and an outline of the proposed investigations scheme and the evaluation methods. The implementation of the speech processing schemes and the evaluation are presented in the next three chapters.

Chapter 4

SPECTRAL SPLITTING WITH PERCEPTUALLY BALANCED COMB FILTERS

4.1 Introduction

Binaural dichotic presentation of speech signal by using comb filters having complementary magnitude response, has helped in improving the perception for persons with moderate bilateral sensorineural hearing loss. Earlier work in our lab (Chaudhari, 2000; Chaudhari and Pandey, 1998a, b), reviewed in Section 3.3, involved pair of comb filters with 18 bands based on auditory critical bandwidths lying in the range of 0 to 5 kHz. The filter design based on frequency sampling technique had sharp transition between pass and stop bands. The focus of the current investigation is to design and test comb filters with perceptual balance; i.e., to obtain minimum perceived spectral distortion in the pass bands and perceptual balance of intensity at band crossovers. Evaluation of processing with perceptually balanced comb filters was performed on normal hearing subjects with simulated loss and on subjects with moderate bilateral sensorineural hearing loss. While conducting evaluation on hearing impaired subjects, adjustable magnitude response filters were cascaded with perceptually balanced comb filters, to partly compensate for the frequency dependent hearing thresholds. Response time, recognition scores, and relative information transmission were used for performance evaluation. This chapter gives the design of perceptually balanced comb filters and its evaluation on normal hearing subjects with simulated loss and on subjects with moderate bilateral sensorineural hearing impairment.

4.2 Review of previous work

In the investigation on splitting of speech spectrally for binaural dichotic presentation, Lyregaard (1982) used an analog delay to obtain two complementary comb filter magnitude responses with constant bandwidth pass and stop bands. Three bandwidths of 200, 500, and 800 Hz were used. Listening tests were conducted on three hearing impaired subjects, all with binaural hearing loss of approximately 50 dB HL and degraded frequency selectivity, and two normal hearing subjects. The test material, consisting of two lists of 25 words each, was presented at a signal-to-noise ratio (SNR) of +12 and +4 dB for hearing impaired subjects. For normal hearing subjects the word list was presented at +6 and –2 dB SNR conditions. The improvement of dichotic over diotic was insignificant. Figure 4.1 shows a representation of the comb filter with a constant bandwidth of 800 Hz used by Lyregarrd (1982). It is to be noted that the comb filters are implemented very efficiently, but they do not provide band separation. There are no distinct pass and stop bands, and band crossovers occur at 3 dB.

In spectral splitting scheme reported by Lunner *et al.* (1993), a pair of comb filters with complementary magnitude responses was used. The comb filters with eight-channel filter bank of constant bandwidth of 700 Hz, were realized using complementary interpolated linear phase FIR filters. The pass band gain was adjusted according to the threshold of the hearing impaired subject. For dichotic listening, odd numbered channels were presented to one ear and even numbered channels to the other ear. In diotic listening, all the 8 channels were presented to both the ears. Speech material, consisting of a list of five-word sentences, was presented in the presence of background noise. Listening tests were conducted on bilateral sensorineural hearing impaired subjects with moderate bilateral losses in the age group of 38 to 69 years. An overall improvement of 2 dB in speech-to-noise ratio for dichotic with respect to diotic for 50% correct recognition was reported. Gain adjustments to compensate for frequency dependent shifts in hearing thresholds provided additional improvement. Figure 4.2 shows a representation of the magnitude response of the pair of comb filters without considering gain adjustments in the pass bands. Focus in the design was on efficiency, and not on separation of bands and undistorted perception for spectral components. Here the inter-band crossovers were lying at different levels, which may have led to spectral distortion due to perceptual imbalance. The components lying in the transition region are presented to both the ears, whereas the components lying in the pass bands are presented to either of the ears. The gain in the crossover region needs to be adjusted for perceptual balance.

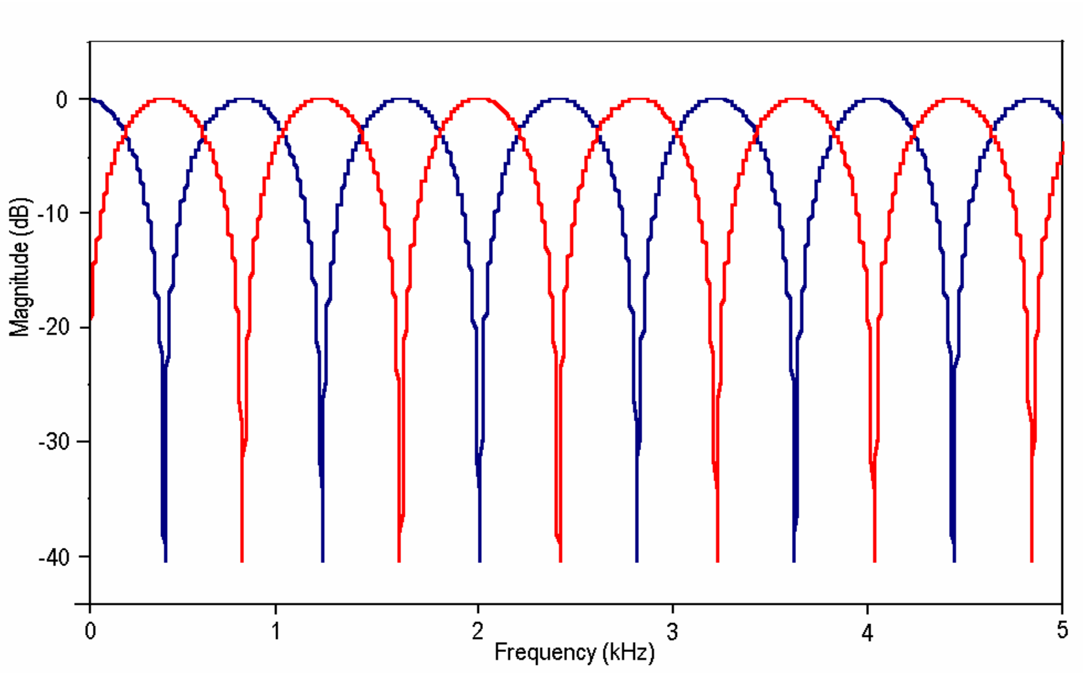


FIG. 4.1. Representation of the magnitude response of the pair of comb filters with 800 Hz used by Lyregaard (1982).

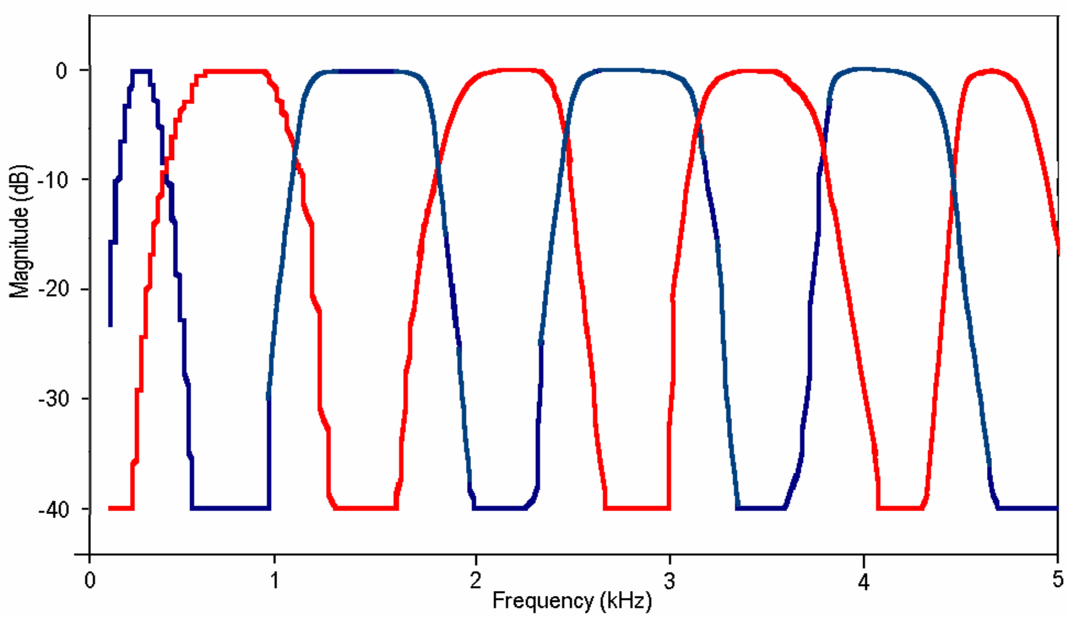


FIG. 4.2. Representation of the magnitude response of the pair of comb filters with constant bandwidth of 700 Hz used by Lunner *et al.* (1993).

The above two investigations used comb filters with constant bandwidths, which was exploited for design and implementation efficiency. However, the auditory system acts as a bank of overlapping band pass filters, with bandwidths increasing from low to high frequencies. Hence use of comb filters with bandwidths increasing with center frequency may be more effective in reducing the effect of spectral masking and improving speech perception.

Chaudhari and Pandey (1998a, b) investigated the use of a pair of comb filters with complementary magnitude responses for spectral splitting for binaural dichotic presentation. Each of these comb filters had 9 pass bands based on auditory critical bands, as described by Zwicker (1961). The bandwidths were approximately 100 Hz for center frequencies below 500 Hz and were 15 − 17 % of the center frequency in the range of 1 − 5 kHz. The linear phase FIR filters were designed using frequency sampling techniques with 128 coefficients, having sharp transitions between pass and stop bands. Evaluation of the scheme on subjects with moderate bilateral loss showed improvement in recognition scores, transmission of information of all consonantal features (voicing, place, manner, nasality, frication, and duration) with maximum contribution from place feature. The percentage relative improvements ranged from 9.2 to 23.6 in VCV context and 14.4 to 19.2 in CV context. Further, magnitude response of the comb filters was adjusted for each band within +3 dB and −3 dB, to partly match the hearing thresholds of the hearing impaired subjects (Chaudhari, 2000; Chaudhari and Pandey, 1999b). Additional improvement was reported but there was no specific contribution from place feature.

Chaudhari and Pandey (1998c) designed the comb filters with the consideration of relatively flat response in the pass band and sharp inter-band transitions. The comb filters were designed as 128-coefficient filters, for real time implementation on TI/TMS320C50 based DSP boards. Magnitude response of the pair of filters is shown in Fig. 4.3. The response has ripple of up to 4 dB in pass band and minimum attenuation of 10 dB in stop bands. Band transition width was 78 Hz. The design has no control over inter-band crossovers, and these vary over 0 −10 dB. An analysis of these filters was carried out to find perceptual distortion due to passband ripple, stop-band attenuation, and difference in crossover points at the various bands. The focus of present work is to design and test comb filters that will improve the performance of spectral splitting based dichotic presentation by improving the magnitude response of the comb filters.

4.3 Comb filters with sharp transition between pass and stop bands

The relative phase between two components separated by more than a critical bandwidth may not be of importance in speech perception (Hartmann, 1997; Moore, 1997). The possible cues for sound location are inter-aural intensity difference, inter-aural time difference, and the spectrum of sound reaching the eardrum. The inter-aural time difference cue may get affected by changes in wave shapes due to nonlinear phase response of the filter. Hence we have designed comb filters with linear phase characteristics, as in the case of previous design (Chaudhari and Pandey, 1998a, b). We may obtain more efficient comb filters due to reduction in number of coefficients if the linear phase characteristics is relaxed. Filters were designed using frequency sampling technique (Ifeachor and Jevis, 1997; Oppenheim *et al*., 1999; Proakis and Manolakis, 1997; Rabiner and Gold, 1998), as described in Appendix C. As a first step towards designing better filters, comb filters were designed with 256 coefficients, taking into consideration the advances in DSP processors. The magnitude response of the pair of comb filters is shown in Fig. 4.4. The filters have sharp transition between bands, passband ripple of 4 dB, and stop-band attenuation of 11 dB. Doubling of filter coefficients halved the transition width (Δf = 39 Hz), but it did not result in significant improvements in passband ripple and stop-band attenuation. As in the earlier design, there was no control on inter-band crossover points, and these ranged over 1 –10 dB.

Relatively small stop-band attenuation may lead to inadequate spectral separation and may not help in reducing the effect of spectral masking. Hence we need to have larger stop-band attenuation. Another problem may be caused by perceived spectral distortion due to ripple of up to 4 dB in pass bands, particularly for the hearing impaired persons with loudness recruitment. Perceptual distortion was noticed when a slowly sweeping sinusoidal tone with frequency variation limited to lie in a pass band with ripple of 4 dB was perceived. The just noticeable difference (JND) for overall intensity in case of synthetic vowel is reported to be 1.5 dB, while for first and second formants it is 1.5 dB and 3 dB respectively (Flanagan, 1972). The JND remains very near to 1 dB for a wide range of levels for wide band of noise (Yost, 1994). Hence it was decided to limit the maximum passband ripple to 1 dB.

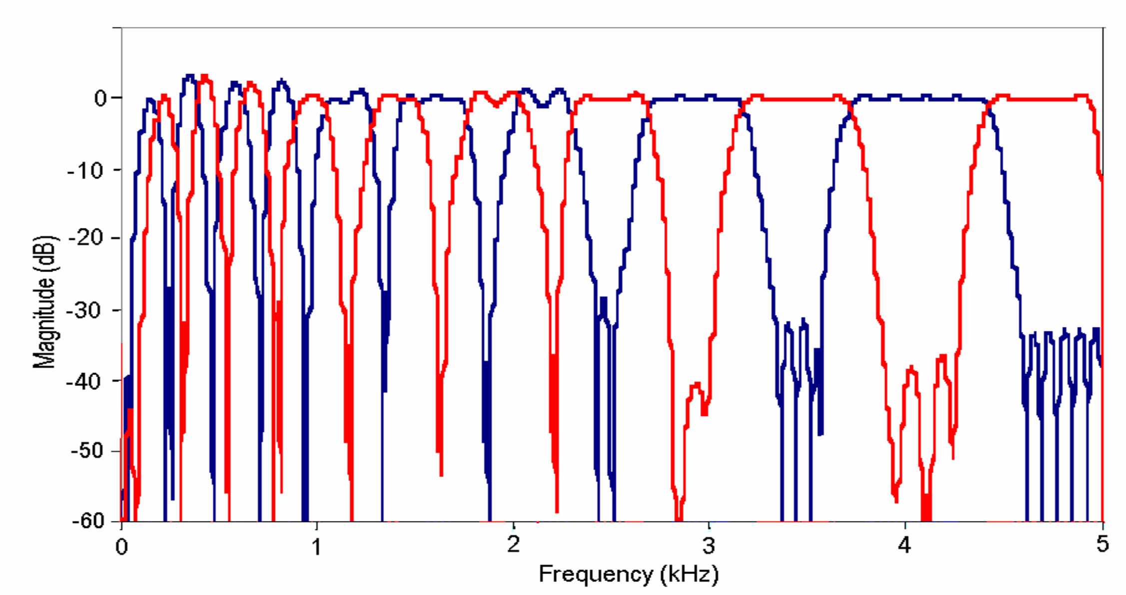


FIG. 4.3. Magnitude response of the pair of 128-coefficients comb filters based on auditory critical bandwidths and with sharp transition between bands, S.R. = 10 kSa/s, passband ripple < 4 dB, stop-band attenuation > 10 dB, transition width = 78 Hz, and band crossover points = 0 – 10 dB.

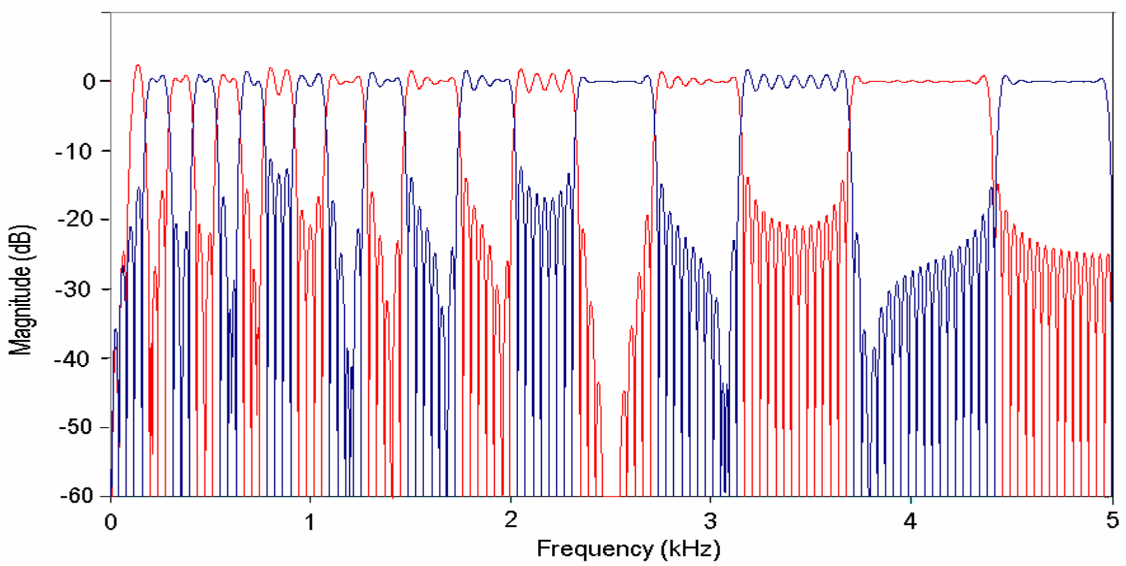


FIG. 4.4. Magnitude response of the pair of 256-coefficients comb filters with sharp transition between bands. S.R. = 10 kSa/s, passband ripple < 4 dB, stop-band attenuation > 10 dB, transition width = 39 Hz, and band crossover points = 1 –10 dB.

In ideal splitting, any spectral components would be presented to one ear. However, with the filters with finite crossover in magnitude response, the components lying in the pass band are presented to one ear, whereas those lying in the transition region are presented to both ears. Another serious problem with these filters was that the crossovers between adjacent bands were at different levels with respect to the pass band response, and this may result in perceptual distortions due to decrease or increase in the perceived intensity of spectral components lying in the transition region. If the magnitude response is not properly adjusted at the transitions, the components lying in the overlapped region will be perceived with different loudness and will reduce the speech quality and may degrade speech perception. When a sine wave with frequency sweeping slowly between 0 and 5 kHz was processed with the pair of comb filters and presented binaurally, a change in intensity was clearly perceived at the transition between bands.

In a study by Scharf (1969), the perceived loudness of tones presented dichotically at the same level did not change with the separation of frequency between the tones over a wide frequency range, when two distinct auditory images were perceived. The critical bandwidths did not play a role, which can be related to the fact that the interaction of tones between critical bands takes place at the peripheral auditory level. Dichotic presentation of tones with same loudness at different frequencies was compared with binaural presentation of tones with frequency equal to the mean value of frequencies used in dichotic presentation, for perceived similarity. The perceived loudness was found to be the same, when the tones presented in both ways were at the same intensity level. With the same intensity, binaurally presented components will be louder than monaurally presented components, however the loudness is generally less than double (Hellman and Zwislocki, 1963; Marks, 1987; Scharf, 1969). The ratio of monaural to binaural loudness at moderate sound levels, lies closer to 1.5 than 2 (Hellman and Zwislocki, 1963). According to another study (Marks, 1978; 1987), loudness of binaural summation is double of monaural loudness when measured in sone scale, which is a power function of sound pressure with exponent 0.6 (Hartmann, 1998; Moore, 1997). Accordingly monaural intensity needs to be 10 dB SPL more than binaural, for perceptual similarity.

In our investigation, the design of perceptually balanced comb filters was carried out in two phases. In the first phase, the focus was to reduce the passband ripple and increase stop-band attenuation. This is presented in Section 4.4. In the second phase, the magnitude responses were further modified to minimize the perceptual distortion at the band crossovers. To this extent, loudness evaluation test was conducted to determine the difference in intensity for same perception in monaural and binaural presentations. Comb filters were designed with different magnitudes at crossovers between adjacent bands. These tests and results are presented in Section 4.5.

4.4 Comb filters with improved passband ripple and stop-band attenuation

Rabiner *et al.* (1970) reported a method, involving frequency sampling technique for FIR filter design, for increasing the side lobe attenuation of prototype filters by trading sharp transition between bands. The magnitude of the transition samples was considered unconstrained and was adjusted to make the required adjustments in the response. They used linear programming technique to find the optimal magnitude for the transition samples. For the design of the critical bandwidth based comb filters, use of automated design techniques based on optimization criteria was not found suitable, and hence the filters were designed by using an iterative procedure, as described in Appendix C.

For adjusting the magnitude response of the comb filters, the samples in the pass band are considered as constrained samples taking a magnitude of 1. Samples lying close to the edge of the pass band are taken as unconstrained transition samples, and remaining samples of the stop band are constrained with value 0. The magnitude of the unconstrained samples lying in the transition region was varied to modify the magnitude response and the interpolated response was observed. This iterative process was continued for each transition sample, until optimization was obtained for parameters under consideration. The number of transition samples (0, 1, 2) was dependent on the available stop bandwidth, increasing from low to high frequencies.

The transition samples in the magnitude response were adjusted to limit the passband ripple to 1 dB and to maximize the stop-band attenuation. The 256-coefficient comb filters provided stop-band attenuation of 38 dB with passband ripple constrained to 1 dB. With sampling rate of 10 kSa/s, the filters have a transition band of 78 Hz at lower frequencies and 117 Hz at higher frequencies. The crossover points were between 1 and 4 dB. The magnitude response of the pair of comb filters is shown in Fig. 4.5. Listening tests did not show any change in intensity perception due to passband ripple, when the frequency of a sinusoidal tone was swept over the frequency range of the pass band with maximum ripple. When a sinusoidal wave with its frequency slowly swept over 0 to 5 kHz was processed with these comb filters and was presented binaurally, a change in intensity was perceived, for frequencies in the transition region. Thus, even though this comb filter provided relatively flat pass bands and adequate band separation, it was found necessary to modify the magnitude response at the transitions to obtain perceptual balance at all frequencies.

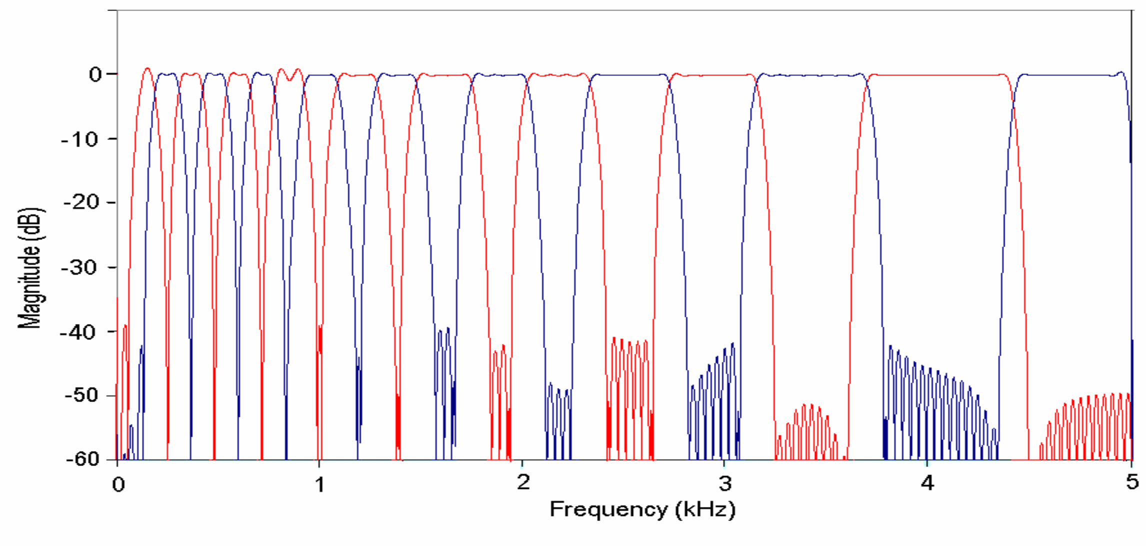


FIG. 4.5. Magnitude response of the pair of 256-coefficients comb filters with improved passband ripple and stop-band attenuation. S.R. = 10 kSa/s, passband ripple < 1 dB, stop-band attenuation > 38 dB, transition width = 78 – 117 Hz, and band crossover points = 1 – 4 dB. .

4.5 Perceptually balanced comb filters

For comb filters with perceptual balance, the magnitude responses need to be properly adjusted at the transitions, so that the components lying in the overlapped region will be perceived with equal loudness. Loudness evaluation test was conducted to determine the difference in intensity for the same perception in monaural and binaural presentations. Comb filters were designed with different magnitudes at crossovers between adjacent bands. These tests and results are presented in the following subsections.

4.5.1 Perceptual balance of monaural and binaural intensity levels.

Listening tests were conducted to determine the difference in the intensity in monaural and binaural presentations, such that they evoke the same perceived loudness level. Stimuli used were pure tones of four different frequencies (0.25, 1, 2, 4 kHz), sustained vowel /*a*/, and broadband noise. Five normal hearing subjects participated in the tests.

The stimuli of 1 s duration were presented through headphones monaurally and binaurally one after the other, with an inter-stimulus interval of 1 s. Monaural intensity was kept constant at 85 dB and binaural intensity was varied from 84 dB down to 70 dB in steps of 1 dB, to establish the monaural versus binaural intensity balance, using a 2-step matching procedure (described in Appendix C, and depicted in Fig. C.1). The subject was instructed to label the binaural loudness as "high", "same", or “low” as compared to the monaural loudness. For a particular binaural level, the pair of monaural and binaural stimuli of 1 s duration was presented respectively with an inter-presentation gap of 1 s. The presentations were repeated with a time gap of 2 s, until the subject was able to respond. This process was continued with different binaural levels. From these observations, the median value of the binaural intensity for "same" response was obtained. For obtaining a more refined estimate of binaural intensity for "same" loudness, listening tests were conducted for binaural intensities spread over 4 dB on both sides of this intensity level. If the subject response was "same" or "high", the presentation was repeated with the binaural level decreased by 2 dB. If the response was "less", the level was increased by 1 dB. This process was continued until similar response was obtained consistently for at least 3 times. This intensity levels with the "same" response was taken as binaural intensity for perceptual balance for a monaural level of 85 dB.

Table 4.1 gives the listening test results for different stimuli on the five subjects, and average and standard deviation across subjects. The perceived levels match when the binaural level was 4 – 12 dB lower than monaural level. An interesting observation here is that for tones of 0.25 and 2 kHz, there is a significant inter-subject variation. At 1 kHz, the average monaural binaural difference is 6.4 dB with a relatively small inter-subject variation. For vowel /*a*/ and broadband noise, the inter-subject variations are small and the average difference is approximately 9 dB.

4.5.2 Comb filters with crossovers adjusted for perceptual balance

Based on the results obtained from the listening tests for determining intensity level difference for perceptual balance between monaural and binaural presentation, pairs of comb filters were designed with different crossovers. Comb filters were designed with different crossover points varying between –3 dB and –9 dB at the crossover region of the first two auditory critical bands, i.e. crossover frequency of 200 Hz. Sine wave with frequency swept between 100 Hz to 300 Hz over an interval of 30 s was processed with these comb filters and listening tests were conducted. For swept sine waves processed with comb filter pairs with crossover points between –4 and –6 dB, change in intensity perception was not noticeable as the swept sine wave moved from one ear to the other. To verify the effect in high frequency range, comb filter pairs were designed with crossover points varying between –3 dB and –9 dB at the overlapping of the pass bands of the 15th and 16th auditory critical bands, i.e. crossover frequency of 3.15 kHz. Listening tests were conducted using sine wave with frequency swept between 3 kHz and 3.5 kHz, so as to cover the overlapping region of these bands. Test results were the same as those obtained for lower frequency bands.

TABLE 4.1. The intensity difference (in dB) between monaural and binaural presentations for same loudness perception.

S	Stimuli						
	Pure Tone (kHz)					Vowel /a/	Noise
	0.25	1	2	4	Avg.		
AC	8	7	4	7	7.2	8	9
DJ	8	6	5	7	6.8	8	7
VK	12	7	9	7	9.0	10	9
AJ	5	5	6	6	6.7	9	9
MD	12	7	9	9	9.2	9	9
Avg.	9.0	6.4	6.6	7.2	7.3	8.8	8.6
s.d.	3.0	0.9	2.3	1.1	1.2	0.8	0.9

Next the transition regions of the pair of comb filters were modified to obtain all the crossovers to lie between –4 dB and –6 dB with respect to the pass band response. The passband ripple was constrained to 1 dB and stop-band attenuation was maximized. The comb filters were designed with 256 coefficients at sampling frequency of 10 kSa/s. The transition width was 78 to 117 Hz at lower end to upper end of the 0 – 5 kHz. The adjustment of magnitude response was done, by adjusting the magnitude of the different transition samples iteratively. Figure 4.6 shows the magnitude response of the pair of comb filters. The minimum stop-band attenuation in this case is 30 dB. A swept sine tone, processed with the pair of complementary comb filters, resulted in no perceived variation in loudness, due to passband ripple or band crossovers.

4.5.3 Filters with adjustable magnitude response to partly compensate for shifts in hearing thresholds

It has been earlier established (Chaudhari, 2000; Chaudhari and Pandey, 1999b; Lunner *et al.*, 1993) that adjustment of pass band gains, to partly compensate for frequency dependent shifts in hearing thresholds, contributes to improvement in speech perception and does not affect the reduction in spectral masking because of spectral splitting. In our investigation, filters with adjustable magnitude response were cascaded with perceptually balanced comb filters. This helped in retaining the same comb filters for all the listeners. Hearing threshold values from the audiogram of the two ears of the listener were used for designing the two filters with adjustable magnitude responses. The frequency compensation was restricted between –3 and +3 dB, keeping in view the limited dynamic range of the hearing impaired subjects. The magnitude response in dB at audiometric test frequencies is given as

$$A_a(f) = -3 + 6\,\frac{\alpha(f) - \alpha_{\min}}{\alpha_{\max} - \alpha_{\min}} \tag{4.1}$$

where α (f) is the value of hearing loss in dB, with the minimum value of α_{min} and the maximum value of α_{max} over 125 Hz to 5 kHz frequency range. The magnitude response in dB as a function of frequency was obtained by linearly interpolating the response at test frequencies. Each filter is a 256-coefficient linear phase filter, designed using frequency sampling technique (as described in Appendix C). Figure 4.7 shows the pure tone audiogram of one of the subjects (SM) and the corresponding magnitude responses (desired and actual) of the filter used to partly compensate for the frequency dependent shift in threshold of hearing for the right and left ears. Audiograms and corresponding compensating filter responses for all the hearing impaired subjects, who participated in the listening tests, are given in Appendix G.

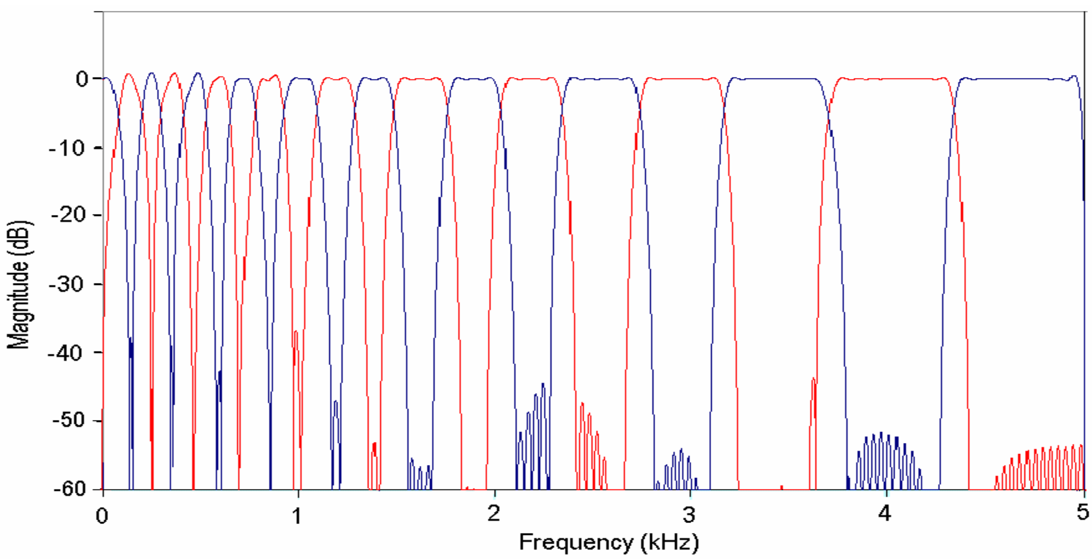


FIG. 4.6. Magnitude response of the pair of 256-coefficients perceptually balanced comb filters. S.R. = 10 kSa/s, passband ripple < 1 dB, stop-band attenuation > 30 dB, transition width = 78 – 117 Hz and band crossover points = 4 – 6 dB.

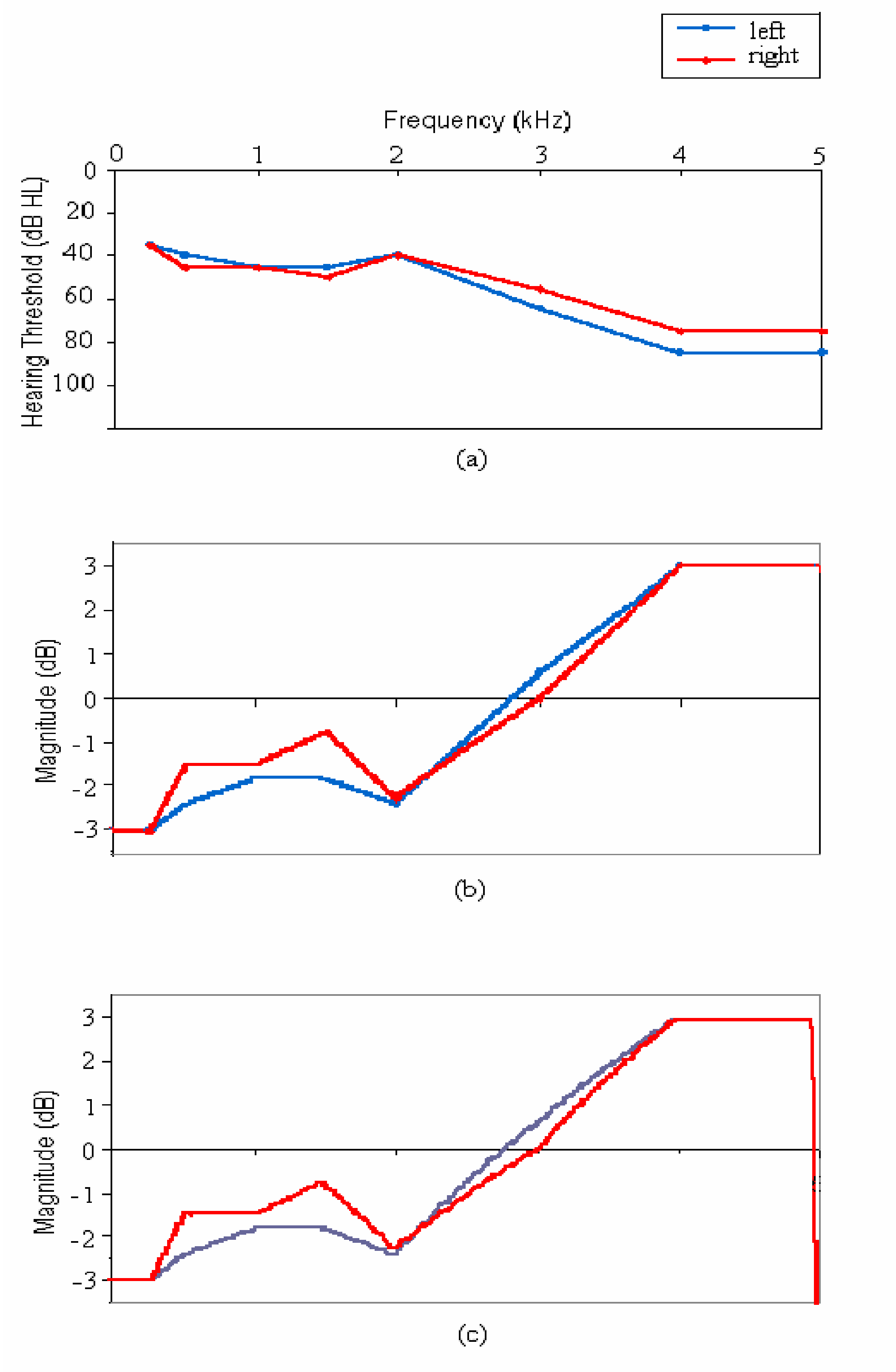


FIG. 4.7. Example of adjustable magnitude response filter. (a) pure tone audiogram for subject SM, (b) desired magnitude response of the two compensating filters (c) 256-coefficient filter magnitude response for the two ears.

4.6. Evaluation of the comb filters

The processing schemes for spectral splitting were implemented for evaluation through listening tests. All the comb filters were 256-coefficient linear phase FIR filters, implemented with the sampling frequency of 10 kSa/s. The schemes evaluated were:

(i) SpA: Comb filters with sharp transition between bands, passband ripple < 4 dB, stop-band atten. > 10 dB, transition width = 39 Hz, crossover points = 1 –10 dB. The magnitude response is shown earlier in Fig. 4.4.

(ii) SpB: Perceptually balanced comb filters, passband ripple < 1 dB, stop-band atten. > 30 dB, transition width = 78 – 117 Hz, crossover points = 2 – 6 dB. The magnitude response is shown in Fig. 4.6.

(iii) SpC: 256-coefficient linear phase filters with magnitude response adjusted within ±3 dB, to partly match the audiogram of the respective ear, cascaded with perceptually balanced comb filters. The magnitude response of the adjustable magnitude response filter for one subject is shown in Fig. 4.7

4.6.1 Listening tests

A schematic representation of the scheme of spectral splitting using comb filters (schemes SpA and SpB) is shown in Fig. 4.8. Figure 4.9 shows the schematic representation for scheme SpC. The bandwidth of the pass bands of the pair of comb filters for the left and right ear are given in Table 4.2.

As described later, evaluation has been carried out in two phases. In the first phase, evaluation was carried on five normal hearing subjects with simulated hearing loss, for comparing schemes SpA and SpB with unprocessed speech Su. This will be referred as Experiment I. The second phase of evaluation was carried out after obtaining favorable results on normal hearing subjects with simulated loss. In the second phase, which will be referred as Experiment II, evaluation was carried out for comparing unprocessed speech Su with speech signals processed with SpA, SpB, and SpC, on five subjects with moderate bilateral sensorineural hearing impairment.

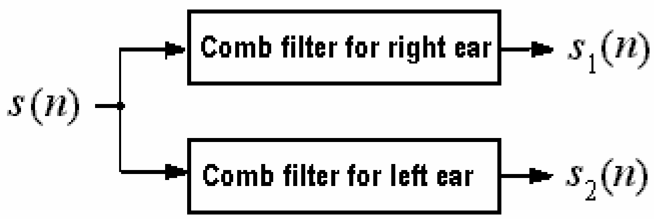


FIG. 4.8. A schematic representation of the scheme of spectral splitting using a pair of comb filters. Input signal: $s(n)$, and outputs to the two ears: $s_1(n)$ and $s_2(n)$.

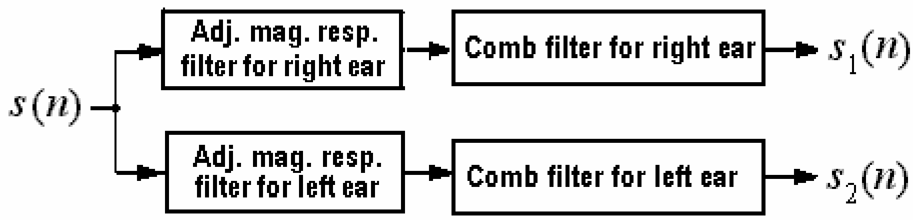


FIG. 4.9. A schematic representation of the scheme SpC. Adjustable magnitude response filters to partly compensate for frequency dependent shift in hearing threshold cascaded with perceptually balanced comb filters. Input signal: $s(n)$, and outputs to the two ears: $s_1(n)$ and $s_2(n)$.

TABLE 4.2. Eighteen critical bands for spectral splitting.

Filter for left ear			Filter for right ear		
Band	Center frequency (kHz)	Passband frequency (kHz)	Band	Center frequency (kHz)	Passband frequency (kHz)
1	0.15	0.10-0.20	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

4.6.2 Test material

Listening tests involved a closed set identification of 12 English consonants /p, b, t, d, k, g, m, n, s, z, f, v/ in a vowel-consonant-vowel (VCV) context with vowel /a/ as in "father". Nonsense syllables were used to reduce the effect of linguistic factors. As discussed earlier in Section 3.7.1, VCV syllables were considered specifically to test the effect of masking, since greater masking takes place in intervocalic consonants due to the presence of vowels on both sides. These speech stimuli were acquired using a setup consisting of microphone, amplifier, and anti-aliasing filter with low pass cut-off frequency of 4.6 kHz, and were digitized with 16-bit resolution at 10 kSa/s (details in Appendix D). The acquired signals were equalized for energy and processed offline by the three schemes.

4.6.3 Experimental set-up

A computerized test administration system, consisting of a PC interfaced through RS-232 serial port to the subject terminal placed in an acoustically isolated chamber, was used. The subject sitting in the acoustically isolated chamber listened to the test material presented through headphones (Telephonics TDH-39P) and responded through the keyboard of the terminal placed there. Presentation of the speech to the headphones was done through two DAC ports of a data acquisition card, smoothening low pass filter, and audio amplifiers, with gain adjusted for most comfortable listening level for each ear. The experimental set-up is detailed in Appendix D. The presentation level was kept constant throughout the series of tests for each of the subject.

4.6.4 Experimental method

The details about the test and the method to respond were explained to each of the subjects. There was a provision to listen each of the 12 VCV syllables as many times as required, at the start of each test run. The 12 speech stimuli were presented 5 times in a random order in one test run (60 presentations). During the listening test session, all the 12 choices of the stimuli and the corresponding key to be pressed were displayed on the monitor of the subject terminal. The choice order was also randomized to avoid any response bias. In the computerized test administration system, updated information about the stimulus presented, response and the number of correct response, along with response time was displayed on the operator screen kept outside the acoustic chamber. For each test run, the responses were recorded in the form of stimulus-response confusion matrices. Each cell in the matrix represents the frequency of occurrence of a stimulus response pair. The numbers along the diagonal of the confusion matrix corresponds to correct response. The percentage of correct recognitions (recognition score) and average response time were displayed along with the confusion matrix after every test run, and stored for further analysis. Response time was used to compare the load on the perception process.

For each of the experimental conditions, first test run was with feedback about the correct response. Further test runs were without feedback, which were repeated until the recognition scores become consistent (variation in the scores not exceeding 5 %) over at least 5 test runs. One test run normally required 5 – 8 minutes. Hearing impaired subjects required more time for each test run, ranging up to 20 minutes. Scheduling of the listening tests was done as per the convenience of the subjects. Test sessions for normal hearing subjects were spread over three months, followed by another four months for hearing impaired subjects.

To determine the level of significance of relative improvement of the processing scheme with respect to unprocessed speech, paired t-test (Bailey, 1981; Snedector and Cochran, 1980; Veerarajan, 2003) was carried out. Information transmission analysis (Miller and Nicely, 1955), described in Appendix B, which gives a measure that is independent of subject’s response bias, was performed on confusion matrix. For each of the processing conditions, combined confusion matrices for five tests were obtained and were subjected to information transmission analysis. The combined confusion matrix was for 25 presentations for each of the 12 stimuli (5 presentation in each test × 5 tests). For studying the contribution of specific consonantal features on speech reception, the consonants were grouped according to the articulatory features (Ladefoged, 1982; Miller and Nicely, 1955). The features selected for study were

voicing: voiced / b d g m n z v /, unvoiced / p t k s f /
place: front: / p b m f v /, middle: / t d n s z /, back: / k g /
manner: oral stop: / p b t d k g /, fricative: / s z f v /, nasal: / m n /
nasality: oral: / p b t d k g s z f v /, nasal: /m n /
frication: stop: / p b t d k g m n /, fricative: / s z f v /
duration short: / p b t d k g m n f v /, long: / s z /

4.7 Experiment I: Listening tests on normal hearing subjects with simulated loss

Listening tests were initially carried out on normal hearing subjects with simulated loss. The tests involved diotic presentation of the unprocessed speech (Su) and binaural dichotic presentation of speech processed by (i) comb filters with sharp transition between bands (denoted as SpA) and (ii) with perceptually balanced comb filters (denoted as SpB). The filters were implemented for off-line processing and tests were conducted on normal hearing subjects with hearing loss simulated by adding broadband noise with constant short-time signal-to-noise ratio (SNR). SNR conditions used were ∞, 6, 3, 0, –3, –6, –9, –12, and –15 dB. Simulation method used here has been discussed earlier in Section 3.7.2. Thus there were 27 test conditions, 3 types of speech signals (unprocessed speech: Su, processed speech: SpA, processed speech: SpB) × 9 SNR values.

Five normal hearing subjects (MP: M 22, RJ: M 28, VK: M 27, JK: M 25, AC: F: 38) participated in the listening tests. All subjects had pure tone thresholds less than 20 dB HL in the frequency range of 125 Hz to 6 kHz. Further, confusion matrices were grouped according to features of voicing, place, manner, nasality, frication, and duration, and were subjected to information transmission analysis. Response times, percentage correct recognition scores, relative improvements in response times and recognition scores, information transmission analysis results for unprocessed and processed speech for each subject and averaged across the subjects are discussed in the following subsections. To determine the significance of processing, t-tests were carried out.

4.7.1 Response time

Table 4.3 gives the response time for unprocessed (Su) and the two processed (SpA, SpB) schemes for all experimental conditions. The average over 5 subjects, standard deviation, relative improvement (decrease) in response time, and significance of relative improvement *p* from one-tailed paired t-test are also shown in the table. Figure 4.10 shows the response times and the percentage relative improvement in response time (averaged across 5 subjects) respectively. The response times for unprocessed speech increased as SNR decreased. The response averaged across 5 subjects under no-noise condition was 1.8 s. The relative increase in response times at 6, 3, 0, –3, –6, –9, –12, and –15 dB SNR conditions were 13.8, 8.8, 10, 13.9, 7.7, 14.4, 24.4, and 25.6 % respectively. The percentage relative increase in response times with decrease in SNR conditions shows that, the load on perception increases with increase in masking noise.

Averaged across 5 subjects, the relative decrease (%) in response time with processing scheme SpA with respect to unprocessed were 10.5, 5.1, 6.7, 12.7, 3.64, 7, 1.9 and –0.3 for SNR conditions of 6, 3, 0, −3, −6, −9, −12, and −15 dB respectively. The corresponding values for processing scheme SpB were 13.7, 6.2, 8, 15.5, 3.7, 13.6, 8.3, and –2.0 % at SNR conditions of 6, 3, 0, −3, −6, −9, −12, and −15 dB respectively. The improvements are not significant at all the SNR values. The improvement because of SpB is marginally better than that of SpA. However the difference is not statistically significant.

TABLE 4.3. Experiment I. Response time (s) for Su: unprocessed speech, SpA: processed speech with comb filters with sharp transitions, SpB: processed speech with perceptually balanced comb filters. S: Subject, Avg.: averaged recognition scores, s.d. = standard deviation, R.I._u: average of relative improvement in % with respect to Su and *p*: significance level (one-tailed) for paired t-test (n = 5, df = 4), R.I._A: average of relative improvement in % with respect to SpA and *p*: significance level (one-tailed) for paired t-test (n = 5, df = 4).

S	∞ SNR			6 dB SNR			3 dB SNR			0 dB SNR			-3 dB SNR		
	Su	SpA	SpB	Su	SpA	SpB	Su	SpA	SpB	Su	SpA	SpB	Su	SpA	SpB
MP	2.02	2.05	2.03	2.43	1.77	1.73	1.71	1.70	1.84	2.06	1.91	1.73	2.05	1.62	1.56
RJ	1.90	2.36	2.05	2.28	1.68	1.54	1.93	2.20	1.41	1.83	1.65	1.79	1.81	1.44	1.36
VK	1.94	2.24	2.07	1.97	1.83	2.12	2.10	1.87	2.36	1.97	2.21	2.09	2.28	2.04	2.00
JK	1.29	1.30	1.45	1.50	1.58	1.31	1.79	1.48	1.48	1.68	1.58	1.46	1.79	1.83	1.55
AC	1.81	2.12	2.15	2.09	2.15	2.04	2.25	2.01	2.10	2.37	1.83	2.02	2.34	2.02	2.27
Avg.	1.80	2.01	1.95	2.05	1.80	1.75	1.96	1.85	1.84	1.98	1.84	1.82	2.05	1.79	1.75
s.d	0.29	0.41	0.28	0.36	0.22	0.34	0.22	0.28	0.40	0.26	0.25	0.25	0.26	0.26	0.38
R.I._u (p)		-11.8 (0.03)	-9.3 (0.02)		10.5 (0.09)	13.7 (0.8)		5.1 (0.2)	6.2 (0.2)		6.7 (0.2)	8.0 (0.07)		12.7 (0.2)	15.5 (0.008)
R.I._A (p)			-1.7 (0.2)			3.4 (0.3)			-0.6 (0.5)			0.7 (0.4)			2.8 (0.3)

S	∞ SNR			-6 dB SNR			-9 dB SNR			-12 dB SNR			-15 dB SNR		
	Su	SpA	SpB	Su	SpA	SpB	Su	SpA	SpB	Su	SpA	SpB	Su	SpA	SpB
MP	2.02	2.05	2.03	1.96	1.61	1.63	1.77	1.78	1.62	2.17	1.61	1.85	1.84	1.66	1.76
RJ	1.90	2.36	2.05	1.75	1.68	1.48	1.95	1.65	1.22	2.29	1.84	1.40	2.24	2.60	2.40
VK	1.94	2.24	2.07	2.21	2.24	2.33	2.34	1.99	2.03	2.54	2.91	2.50	2.41	1.96	2.21
JK	1.29	1.30	1.45	1.35	1.48	1.57	1.73	1.80	1.63	1.85	1.96	1.91	1.87	2.06	2.07
AC	1.81	2.12	2.15	2.45	2.27	2.26	2.50	2.27	2.42	2.37	2.73	2.62	2.92	3.03	3.06
Avg.	1.80	2.01	1.95	1.94	1.86	1.86	2.06	1.90	1.78	2.24	2.21	2.06	2.26	2.26	2.30
s.d	0.29	0.41	0.28	0.423	0.37	0.41	0.35	0.24	0.46	0.26	0.57	0.5	0.44	0.55	0.49
R.I._u (p)		-11.8 (0.03)	-9.3 (0.02)		3.6 (0.18)	3.7 (0.23)		7.0 (0.07)	13.6 (0.04)		2.0 (0.4)	8.3 (0.2)		-0.3 (0.5)	-2.0 (0.3)
R.I._A (p)			-1.7 (0.2)			0.2 (0.5)			7.2 (0.2)			5.9 (0.1)			-2.5 (0.3)

4.7.2 Recognition score

The recognition scores for each experimental condition for each subject at different SNR condition are given in Table. 4.4. Average scores for five subjects, standard deviation, average relative improvement with processing, its paired t-test (one-tailed) values are also shown. Figure 4.11 shows the recognition scores (averaged across five subjects) for unprocessed and processed speech and the relative improvement with processing. The average recognition score for unprocessed speech decreased from 98.7% at no-noise condition to 92.2, 90.6, 89.6, 81.8, 78.9, 76, 68.7, and 63.3% for 6, 3, 0, –3, –6, –9, –12, and –15 dB SNR conditions respectively. Averaged across the five subjects, the relative increase in recognition scores with processing were 5.9, 4.5, 7.5, 18.8, 19.1, 18.3, 24, and 21.7 % for scheme SpA and 7.7, 7.6, 9.3, 20.7, 21.5, 21.4, 29.7 and 27.5 % for scheme SpB for SNR conditions of 6, 3, 0, –3, –6, –9, –12, and –15 dB respectively.

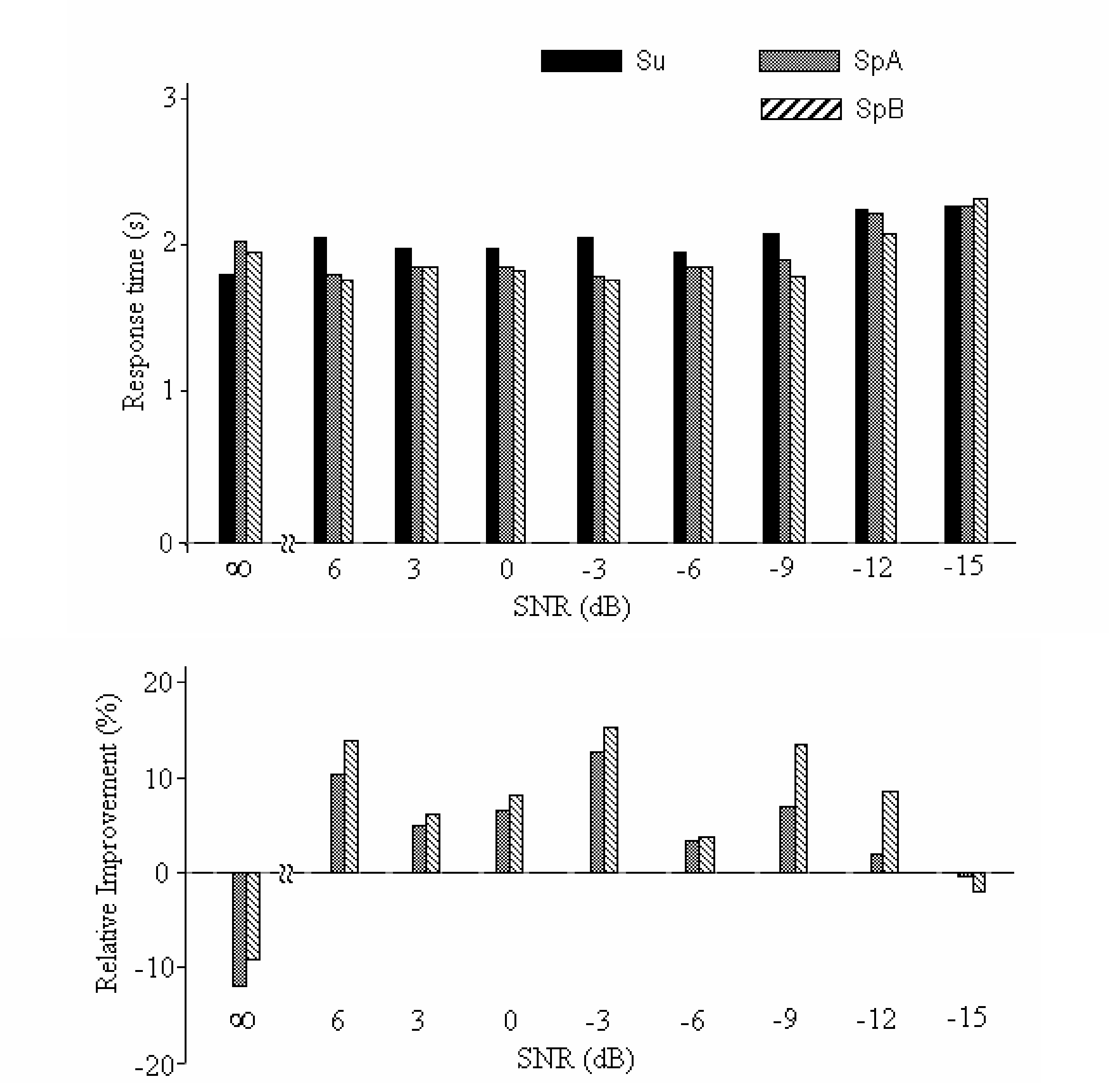


FIG. 4.10. Experiment I. Averaged response time (s) and average relative improvement (%). Su: unprocessed speech, SpA: processed speech with comb filters with sharp transitions, and SpB: processed with perceptually balanced comb filter.

TABLE 4.4. Experiment I. Recognition scores (%) for Su: unprocessed speech, SpA: processed speech with comb filters with sharp transitions, SpB: processed speech with perceptually balanced comb filters. S: Subject, Avg.: averaged recognition scores, s.d. = standard deviation, R.I._u: average of relative improvement in % with respect to Su and *p*: significance level (one-tailed) for paired t-test (*n* = 5, df = 4), R.I._A: average of relative improvement in % with respect to SpA and *p*: significance level (one-tailed) for paired t-test (*n* = 5, df = 4).

S	∞ SNR			6 dB SNR			3 dB SNR			0 dB SNR			-3 dB SNR		
	Su	SpA	SpB	Su	SpA	SpB	Su	SpA	SpB	Su	SpA	SpB	Su	SpA	SpB
MP	97.4	92.0	94.3	86.3	96.6	97.0	88.0	90.3	91.7	87.3	89.0	94.3	77.3	92.7	95.0
RJ	96.0	91.7	87.0	79.3	90.3	96.3	70.3	81.0	90.7	70.3	89.7	91.3	64.3	91.3	93.7
VK	100.0	98.7	100.0	96.0	100.0	100.0	98.7	100.0	99.7	96.7	99.7	100.0	90.3	99.0	99.3
JK	100.0	100.0	100.0	100.0	100.0	100.0	98.7	100.0	100.0	99.0	99.7	100.0	91.0	98.0	100.0
AC	100.0	99.7	100.0	99.3	99.0	100.0	97.3	99.3	99.7	94.7	98.7	98.7	86.3	99.0	99.3
Avg.	98.7	96.4	96.3	92.2	97.2	98.7	90.6	94.1	96.4	89.6	95.4	96.9	81.8	96.0	97.5
s.d	1.9	4.2	5.7	9.0	4.1	1.9	12.2	8.4	4.7	11.6	5.5	3.9	11.2	3.7	2.9
R.I._u		-2.3	-2.5		5.9	7.7		4.5	7.6		7.5	9.3		18.8	20.7
(*p*)		(0.05)	(0.1)		(0.05)	(0.06)		(0.06)	(0.1)		(0.09)	(0.06)		(0.008)	(0.007)
R.I._A			-0.2			1.6			2.7			1.7			1.6
(*p*)			(0.5)			(0.1)			(0.2)			(0.1)			(0.02)

S	∞ SNR			-6 dB SNR			-9 dB SNR			-12 dB SNR			-15 dB SNR		
	Su	SpA	SpB	Su	SpA	SpB	Su	SpA	SpB	Su	SpA	SpB	Su	SpA	SpB
MP	97.4	92.0	94.3	71.3	90.3	90.7	77.7	85.3	87.3	76.0	84.0	85.3	69.0	80.9	81.3
RJ	96.0	91.7	87.0	61.0	86.0	86.0	60.3	84.7	84.3	47.0	73.1	80.0	43.0	62.4	66.4
VK	100.0	98.7	100.0	92.7	98.3	99.0	82.3	99.0	98.7	73.3	87.3	90.7	75.7	94.3	94.0
JK	100.0	100.0	100.0	93.0	98.3	99.7	87.7	87.3	96.7	83.3	87.7	89.0	71.7	73.3	77.7
AC	100.0	99.7	100.0	76.7	89.0	96.0	72.0	87.3	89.7	64.0	83.0	86.7	57.0	68.0	75.7
Avg.	98.7	96.4	96.3	78.9	92.4	94.3	76.0	88.7	91.30	68.7	83.0	86.3	63.3	75.8	79.0
s.d	1.9	4.2	5.7	13.9	5.6	5.8	10.5	5.9	6.2	14.0	5.9	4.1	13.3	12.4	10.0
R.I._u		-2.3	-2.5		19.1	21.5		18.3	21.4		24.0	29.7		21.7	27.5
(*p*)		(0.05)	(0.1)		(0.01)	(0.008)		(0.02)	(0.003)		(0.01)	(0.01)		(0.009)	(0.003)
R.I._A			-0.2			2.1			3.0			4.2			4.8
(*p*)			(0.5)			(0.1)			(0.1)			(0.02)			(0.05)

The improvements in the processing with both the schemes (SpA and SpB) were statistically highly significant ($p \leq 0.01$) for SNR conditions of –3, −6, −9, −12, and –15 dB. At 6, 3, 0, −3, –6, −9, −12, and −15 dB SNR conditions, the relative improvement (%) for processing scheme SpB over SpA was 1.6, 2.7, 1.7, 1.6, 2.1, 3.0, 4.2 and 4.8 respectively. Thus splitting with perceptually balanced comb filters provided higher improvements in recognition scores at low SNR conditions.

4.7.3 Information transmission analysis

For each experimental condition, combined confusion matrix for five test runs with consistent scores were subjected to information transmission analysis. Overall relative information transmitted for unprocessed (Su) and processed (SpA and SpB) for all subjects with the average over five subjects are given in Table 4.5. Figure 4.12 shows the overall relative information transmitted, averaged across five subjects.

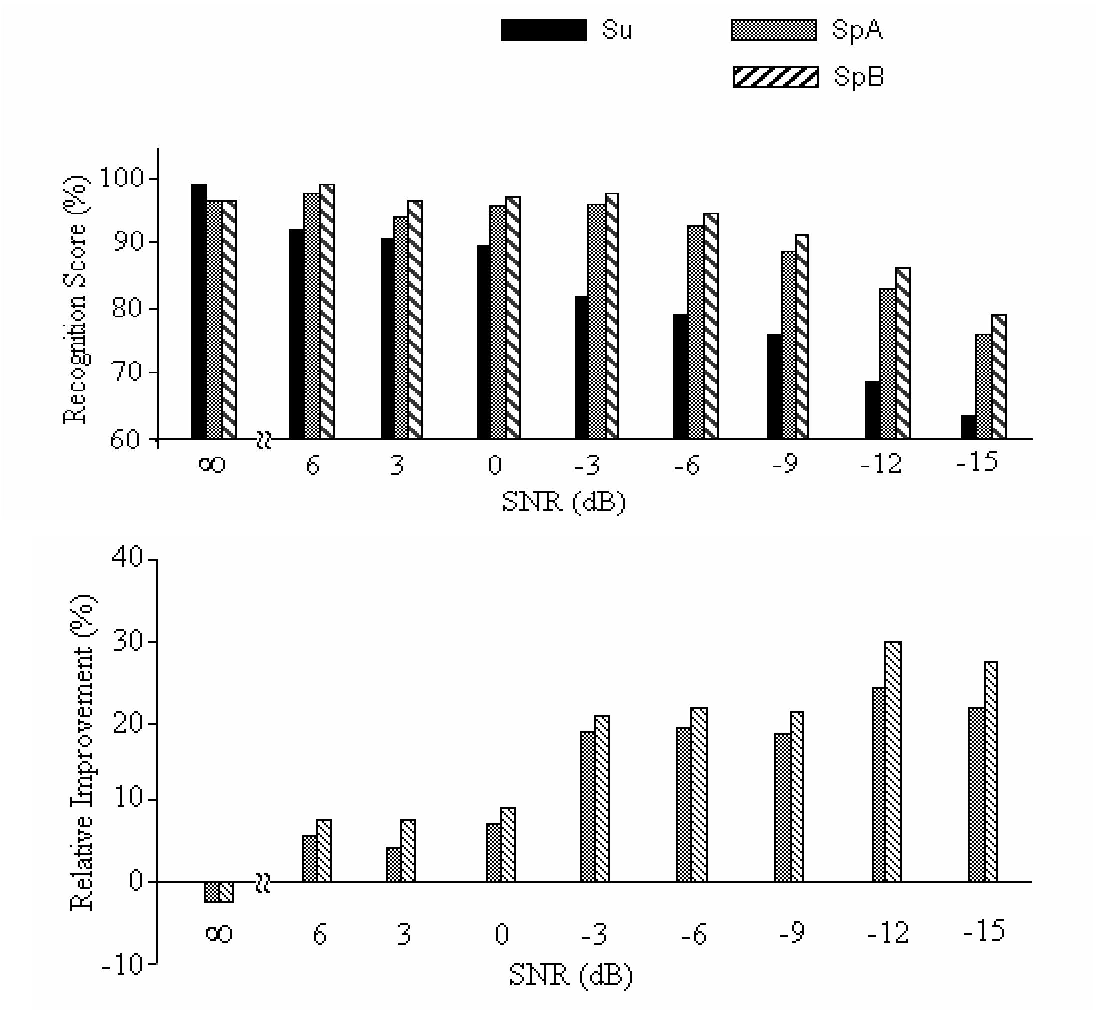


FIG. 4.11. Experiment I. Averaged recognition score (%) and average relative improvement (%). Su: unprocessed speech, SpA: processed speech with comb filters with sharp transitions, and SpB: processed with perceptually balanced comb filter.

The overall information transmitted was degraded with decrease in SNR conditions for unprocessed speech. The values were 98.6 % for no-noise and 94, 91.2, 92.8, 86.4, 84, 81, 74, and 69.4 % for SNR conditions of 6, 3, 0, –3, –6, –9, –12, and –15 dB respectively. With processing scheme SpA, the overall information transmitted were 96.8, 94.8, 95.2, 95.2, 91.8, 89.6, 85, and 76 % for 6, 3, 0, –3, –6, –9, –12, and –15 dB SNR conditions respectively. The corresponding values with SpB were 98.2, 97, 96.6, 96.8, 94, 91.4, 86, and 79.8 %. The results of information transmission analysis for various groupings are also given in Table 4.5.

Figures 4.12 shows the relative information transmitted for individual features of voicing, place, manner, nasality, duration, and frication, averaged across five subjects.

Voicing: The relative information transmitted for voicing feature in no-noise condition for unprocessed speech was 99.4 %. With decrease in SNR, the information transmitted by this feature decreases to 83.4 % at −15 dB SNR condition. The relative increase in information transmitted by this feature at −12 and −15 dB SNR conditions were 6.8 and 9.7 % for SpA and 4.6 and 12.3 % for SpB respectively.

Place: Place feature is severely affected by the addition of noise. The relative information transmitted for unprocessed speech under no-noise condition is 99.2 %, which was decreased to 29 % at −15 dB SNR. The relative improvement in information transmitted were 14.1, 14.3, 16.2, 58.5, 69.0, 57.9, 45.8, and 73.1 % for 3, 6, 0, –3, –6, –9, –12, and −15 dB SNR conditions for scheme SpA. The corresponding values for SpB are 19.6, 21.3, 20.8, 66.8, 79.7, 59.5, 68.9, and 92.6 %. The relative improvement increases with increase in masking noise, which is maximum at −15 dB SNR for scheme SpB. The relative improvement in place feature for SpB with respect SpA was 15.8 and 13.2 % at −12 and −15 dB SNR conditions.

Manner: The relative information transmitted for manner feature decrease from 95.4 % under no-noise condition to 51.8 % for −15 dB SNR for unprocessed speech. The relative information transmitted for manner feature for SNR conditions of 6, 3, 0, –3, –6, –9, –12, and –15 dB were 5.7, 10.2, 11.4, 12.0, 12.8, 10.3, 35.9, and 25.0 % and 8.3, 13.3, 14.6, 15.0, 18.6, 19.1, 36.7 and 47.2 % for SpA and SpB respectively. The improvement of SpB with respect SpA at −12 and −15 dB SNR conditions were 1.3 and 17.4 % respectively.

Nasality: Relative information transmitted for unprocessed speech under no-noise condition was almost perfect. The corresponding value at −12 and −15 dB SNR conditions were 76.2 and 63.2 %. Up to –9 dB SNR condition, there was not much decrease in the information transmitted for unprocessed speech. The relative improvements at −12 and −15 dB SNR conditions for SpA were 34.6 and 32.5 % and for SpB were 36.3 and 53.5 %.

Frication: Frication was affected badly by the addition of noise in unprocessed speech. At no-noise condition, the relative information transmitted was 93.2 %, which decreased to 38.6 % at −15 dB. The relative increase in information transmitted for frication at −12 and −15 dB SNR conditions were 62.4 and 38.9 % for SpA and 59.5 and 56.5 % for SpB.

TABLE 4.5. Experiment I. Relative information transmitted (%) for Su: unprocessed speech, SpA: processed speech with comb filters with sharp transitions, SpB: processed speech with perceptually balanced comb filters. (a) overall, and feature groupings: (b) voicing, (c) place, (d) manner, (e) nasality (f) frication and (g) duration. Avg.: average S: Subject, Avg.: averaged relative information transmitted (%), R.I.: average of relative improvement in % with respect to Su and *p*: significance level (one-tailed) for paired t-test (n = 5, df = 4).

(a) Overall

S	∞ SNR			6 dB SNR			3 dB SNR			0 dB SNR			-3 dB SNR		
	Su	SpA	SpB	Su	SpA	SpB	Su	SpA	SpB	Su	SpA	SpB	Su	SpA	SpB
MP	97	95	96	90	95	95	85	92	94	88	90	94	82	92	93
RJ	96	92	89	84	91	96	79	84	93	86	90	91	79	91	93
VK	100	98	100	97	100	100	98	100	99	97	99	100	90	98	99
JK	100	100	100	100	100	100	98	100	100	98	99	100	93	97	100
AC	100	99	100	99	98	100	96	98	99	95	98	98	88	98	99
Avg.	98.6	96.8	97.0	94.0	96.8	98.2	91.2	94.8	97.0	92.8	95.2	96.6	86.4	95.2	96.8
R.I.		-1.8	-1.7		3.2	4.8		4.2	6.9		2.6	4.2		10.4	12.2
p		0.03	0.15		0.07	0.06		0.01	0.04		0.005	0.003		0.001	0.0004

S	∞ SNR			-6 dB SNR			-9 dB SNR			-12 dB SNR			-15 dB SNR		
	Su	SpA	SpB	Su	SpA	SpB	Su	SpA	SpB	Su	SpA	SpB	Su	SpA	SpB
MP	97	95	96	77	89	90	81	87	89	80	85	85	77	80	84
RJ	96	92	89	75	85	87	75	85	86	56	76	83	55	65	71
VK	100	98	100	93	98	99	84	98	98	76	85	89	76	92	91
JK	100	100	100	93	98	99	88	90	96	85	87	88	73	70	78
AC	100	99	100	82	89	95	77	88	88	73	82	85	66	73	75
Avg.	98.6	96.8	97.0	84.0	91.8	94.0	81.0	89.6	91.4	74.0	85.0	86.0	69.4	76.0	79.8
R.I.		-1.8	-1.7		9.6	12.3		10.8	12.9		13.7	18.3		9.9	15.7
p		0.03	0.15		0.002	0.002		0.007	0.0004		0.02	0.02		0.05	0.004

(b) Feature: voicing

S	∞ SNR			6 dB SNR			3 dB SNR			0 dB SNR			-3 dB SNR		
	Su	SpA	SpB	Su	SpA	SpB	Su	SpA	SpB	Su	SpA	SpB	Su	SpA	SpB
MP	97	100	100	100	91	86	90	73	71	88	87	83	73	95	93
RJ	100	97	100	90	94	100	100	92	97	100	95	100	97	97	100
VK	100	100	100	100	100	100	100	100	100	100	100	100	100	97	100
JK	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
AC	100	100	100	100	97	100	97	97	100.	97	100	100	97	100	100
Avg.	99.4	99.4	100.0	98.0	96.4	97.2	97.4	92.4	93.6	97.0	96.4	96.6	93.4	97.8	98.6
R.I.		0.0	0.6		-1.5	-0.6		-5.4	-4.2		-0.6	-0.5		2.2	3.1
p		0.5	0.2		0.2	0.4		0.1	0.2		0.3	0.4		0.2	0.1

S	∞ SNR			-6 dB SNR			-9 dB SNR			-12 dB SNR			-15 dB SNR		
	Su	SpA	SpB	Su	SpA	SpB	Su	SpA	SpB	Su	SpA	SpB	Su	SpA	SpB
MP	97	100	100	76	97	92	87	87	97	90	94	90	86	88	95
RJ	100	97	100	97	97	100	86	97	100	71	100	100	72	85	84
VK	100	100	100	100	100	100	97	100	100	100	95	100	100	97	100
JK	100	100	100	100	100	100	100	78	100	90	87	92	68	70	69
AC	100	100	100	100	100	97	97	97	100	97	83	97	91	86	94
Avg.	99.4	99.4	100.0	94.6	98.8	97.8	93.4	91.8	99.4	89.6	91.8	95.8	83.4	85.2	88.4
R.I.		0.0	0.6		1.6	3.5		9.2	1.3		6.8	4.6		9.7	12.3
p		0.5	0.2		0.2	0.2		0.3	0.04		0.4	0.2		0.3	0.05

TABLE 4.5. (Contd.)

(c) Feature: place

S	∞ SNR			6 dB SNR			3 dB SNR			0 dB SNR			-3 dB SNR		
	Su	SpA	SpB	Su	SpA	SpB	Su	SpA	SpB	Su	SpA	SpB	Su	SpA	SpB
MP	98	98	100	79	91	88	68	76	78	67	70	85	46	76	83
RJ	98	85	76	52.	72.	87	46	59	76	46	71	72	31	74	78
VK	100	95.	100	87	100	100	87	100	98	97	98	100	76	96	97
JK	100	100	100	100	100	100	94	100	100	95	98	100	76	91	100
AC	100	98	100	96	98	100	91.	100	98	83	98	93	68	97	97
Avg.	99.2	95.2	95.2	82.8	92.2	95.0	77.2	87.0	90.0	77.6	87.0	90.0	59.4	86.8	91.0
R.I.		-4.1	-4.1		14.1	19.6		14.3	21.3		16.2	20.8		58.5	66.8
p		0.09	0.2		0.03	0.06		0.001	0.02		0.06	0.02		0.002	0.001

S	∞ SNR			-6 dB SNR			-9 dB SNR			-12 dB SNR			-15 dB SNR		
	Su	SpA	SpB	Su	SpA	SpB	Su	SpA	SpB	Su	SpA	SpB	Su	SpA	SpB
MP	98	98	100	32	70	70	45	62	66	48	59	63	37	47	52
RJ	98	85	76	25	58	65	36	60	59	24	35	54	12	27	31
VK	100	95.	100	85	92	95	59	97	94	40	58	70	37	82	82
JK	100	100	100	76	94	98	64	72	87	60	69	66	37	39	51
AC	100	98	100	48	78	86	36	75	69	30	60	61	22	41	45
Avg.	99.2	95.2	95.2	53.2	78.4	82.8	48.0	73.2	75.0	40.4	56.2	62.8	29.0	47.2	52.2
R.I.		-4.1	-4.1		69.0	79.7		57.9	59.5		45.8	68.9		73.1	92.6
p		0.09	0.2		0.005	0.004		0.007	0.0004		0.008	0.006		0.03	0.008

(d) Feature: manner

S	∞ SNR			6 dB SNR			3 dB SNR			0 dB SNR			-3 dB SNR		
	Su	SpA	SpB	Su	SpA	SpB	Su	SpA	SpB	Su	SpA	SpB	Su	SpA	SpB
MP	90	79	84	83	94	90	86	95	100	84	90	96	87	88	94
RJ	87	87	85	81	95	100	67	92	100	63	90	95	73	95	96
VK	100	96	100	100	100	100	97	100	98	91	100	100	89	98	100
JK	100	100	100	100	100	100	100	100	100	100	100	100	96	98	100
AC	100	100	100	100	98	100	98	98	98	98	95	96	84	98	100
Avg.	95.4	92.4	93.8	92.8	97.4	98.0	89.6	97.0	99.2	87.2	95.0	97.4	85.8	95.4	98.0
R.I.		-3.2	-1.8		5.7	8.3		10.2	13.3		11.4	14.6		12.0	15.0
p		0.1	0.1		0.1	0.09		0.1	0.1		0.1	0.08		0.03	0.01

S	∞ SNR			-6 dB SNR			-9 dB SNR			-12 dB SNR			-15 dB SNR		
	Su	SpA	SpB	Su	SpA	SpB	Su	SpA	SpB	Su	SpA	SpB	Su	SpA	SpB
MP	90	79	84	71	82	81	79	73	82	72	68	69	66	68	73
RJ	87	87	85	66	79	83	68	88	72	34	68	65	32	47	59
VK	100	96	100	88	100	100	80	96	100	65	87	90	69	91	87
JK	100	100	100	94	98	100	88	82	98	71	82	88	54	56	72
AC	100	100	100	75	83	100	67	78	100	59	80	79	38	53	69
Avg.	95.4	92.4	93.8	78.8	88.4	92.8	76.4	83.4	90.4	60.2	77.0	78.2	51.8	63.0	72.0
R.I.		-3.2	-1.8		12.8	18.6		10.3	19.1		35.9	36.7		25.0	47.2
p		0.1	0.1		0.002	0.006		0.1	0.03		0.03	0.02		0.02	0.004

TABLE 4.5. (Contd.)

(e) Feature: nasality

S	∞ SNR			6 dB SNR			3 dB SNR			0 dB SNR			-3 dB SNR		
	Su	SpA	SpB	Su	SpA	SpB	Su	SpA	SpB	Su	SpA	SpB	Su	SpA	SpB
MP	96	100	100	100	95	85	100	100	100	100.0	96	96	100	100	100
RJ	100	100	100	88	96	100	100	96	100	95	100	100	100	100	100
VK	100	100	100	100	100	100	100	100	100	100	100	100	100	96	100
JK	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
AC	100	100	100	100	100	100	100	100	100	96	100	100	95	100	100
Avg.	99.2	100.0	100.0	97.6	98.2	97.0	100.0	99.2	100.0	98.2	99.2	99.2	99.0	99.2	100.0
R.I.		0.8	0.8		0.8	-0.3		-0.8	0.0		1.1	1.1		0.3	1.1
p		0.2	0.2		0.4	0.4		0.2	-		0.3	0.3		0.4	0.2

S	∞ SNR			-6 dB SNR			-9 dB SNR			-12 dB SNR			-15 dB SNR		
	Su	SpA	SpB	Su	SpA	SpB	Su	SpA	SpB	Su	SpA	SpB	Su	SpA	SpB
MP	96	100	100	100	100	100	100	100	100	88	95	91	96	91	100
RJ	100	100	100	95	100	96	84	100	100	38	92	91	46	48	80
VK	100	100	100	100	100	100	95	100	100	82	100	100	92	100	100
JK	100	100	100	100	100	100	100	59	100	78	78	91	28	50	61
AC	100	100	100	95	100	100	96	100	100	95	96	95	54	95	88
Avg.	99.2	100.0	100.0	98.0	100.0	99.2	95.0	91.8	100.0	76.2	92.0	93.6	63.2	76.8	85.8
R.I.		0.8	0.8		2.1	1.3		-2.5	5.7		34.6	36.3		32.5	53.5
p		0.2	0.2		0.09	0.1		0.4	0.08		0.09	0.07		0.09	0.01

(f) Feature: frication

S	∞ SNR			6 dB SNR			3 dB SNR			0 dB SNR			-3 dB SNR		
	Su	SpA	SpB	Su	SpA	SpB	Su	SpA	SpB	Su	SpA	SpB	Su	SpA	SpB
MP	86	66	75	74	92	87	77	92	100	73	83	97	79	79	90
RJ	80	79	76	75	94	100	44	87	100	40	86	92	55	91	94
VK	100	94	100	100	100	100	95	100	97	85	100	100	82	100	100
JK	100	100	100	100	100	100	100	100	100	100	100	100	94	97	100
AC	100	100	100	100	97	100	97	97	97	100	92	93	78	97	100
Avg.	93.2	87.8	90.2	89.8	96.6	97.4	82.6	95.2	98.8	79.6	92.2	96.4	77.6	92.8	96.8
R.I.		-6.1	-3.6		9.3	10.2		24.5	31.9		27.7	34.7		23.0	28.3
p		0.1	0.1		0.1	0.1		0.1	0.1		0.1	0.09		0.04	0.01

S	∞ SNR			-6 dB SNR			-9 dB SNR			-12 dB SNR			-15 dB SNR		
	Su	SpA	SpB	Su	SpA	SpB	Su	SpA	SpB	Su	SpA	SpB	Su	SpA	SpB
MP	86	66	75	53	69	69	65	54	69	59	48	52	44	52	53
RJ	80	79	76	46	64	74	56	81	52	20	51	46	23	47	40
VK	100	94	100	80	100.	100.	72	94	100	55	79	83	54	85	79
JK	100	100	100	90	97	100	80	97	97	56	83	81	48	57	62
AC	100	100	100	63	73	100	49	63	84	37	68	68	24	23	51
Avg.	93.2	87.8	90.2	66.4	80.6	88.6	64.4	77.8	80.4	45.4	65.8	66.0	38.6	52.8	57.0
R.I.		-6.1	-3.6		23.6	37.2		21.6	26.1		62.4	59.5		38.9	56.5
p		0.1	0.1		0.002	0.005		0.05	0.05		0.03	0.02		0.04	0.003

TABLE 4.5. (Contd.)

(g) Feature: duration

S	∞ SNR			6 dB SNR			3 dB SNR			0 dB SNR			-3 dB SNR		
	Su	SpA	SpB	Su	SpA	SpB	Su	SpA	SpB	Su	SpA	SpB	Su	SpA	SpB
MP	100.	100	100	72	78	78	47	33.0	38	51	31	57	13	67	70
RJ	96	74	65	33	60	80	32	52	62	24	62	69	26	64	73
VK	100	94	100	64	100	100	91	100	100	91	95	100	78	92	91
JK	100	100	100	100	100	100	96	100	100	96	100	100	80	85	100
AC	100	100	100	100	95	100	92	100	100	88	95	89	67	100	100
Avg.	99.2	93.6	93.0	73.8	86.6	91.6	71.6	77.0	60.0	70.0	76.6	83.0	52.8	81.6	86.8
R.I.		-5.8	-6.5		28.3	41.4		11.1	19.5		27.1	42.9		127.0	142.0
p		0.1	0.2		0.09	0.07		0.2	0.1		0.3	0.09		0.01	0.007

S	∞ SNR			-6 dB SNR			-9 dB SNR			-12 dB SNR			-15 dB SNR		
	Su	SpA	SpB	Su	SpA	SpB	Su	SpA	SpB	Su	SpA	SpB	Su	SpA	SpB
MP	100.	100	100	13	71	70	40	65	73	48	84	85	32	58	96
RJ	96	74	65	22	63	65	18	49	83	13	31	83	8	29	54
VK	100	94	100	96	85	91	72	100	100	43	59	65	49	79	94
JK	100	100	100	70	92	100	68	79	100	72	63	57	43	33	28
AC	100	100	100	55	75	91	44	70	65	30	63	61	30	12	25
Avg.	99.2	93.6	93.0	51.2	77.2	83.4	48.4	72.6	84.2	41.2	60.0	70.2	32.4	42.2	59.4
R.I.		-5.8	-6.5		137.8	147.4		69.8	115.5		69.6	149.8		64.3	163.1
p		0.1	0.2		0.04	0.02		0.001	0.005		0.04	0.05		0.2	0.08

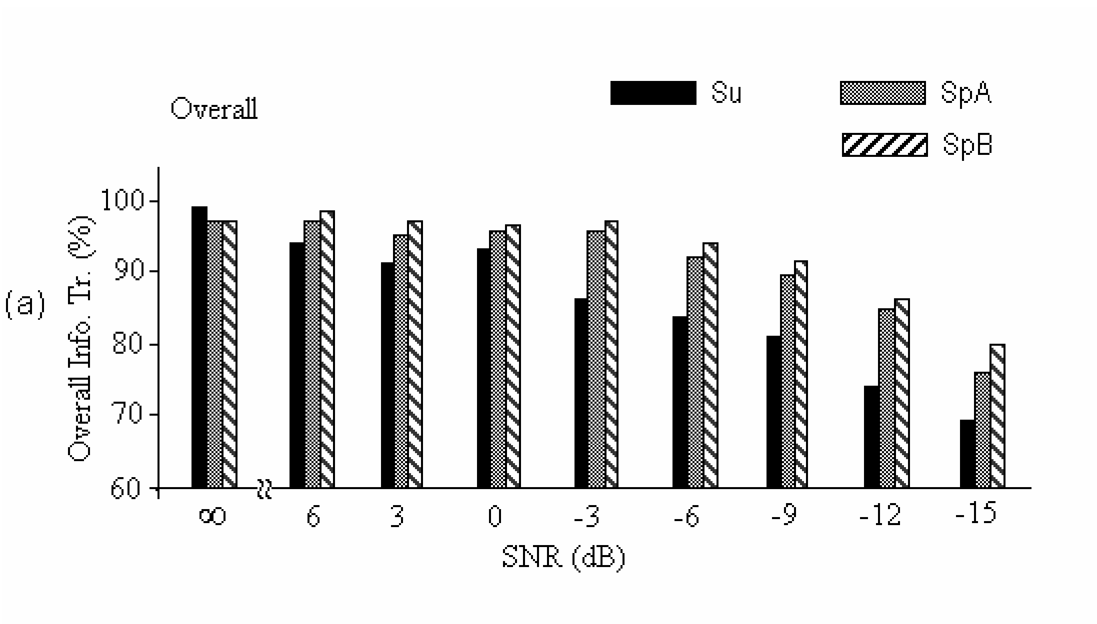


FIG. 4.12. Experiment I. Averaged relative information transmitted (%) (a) overall, (b) voicing, (c) place, (d) manner, (e) nasality, (d) frication, and (e) duration. Su: unprocessed speech, SpA: processed speech with comb filters with sharp transitions, and SpB: processed with perceptually balanced comb filter.

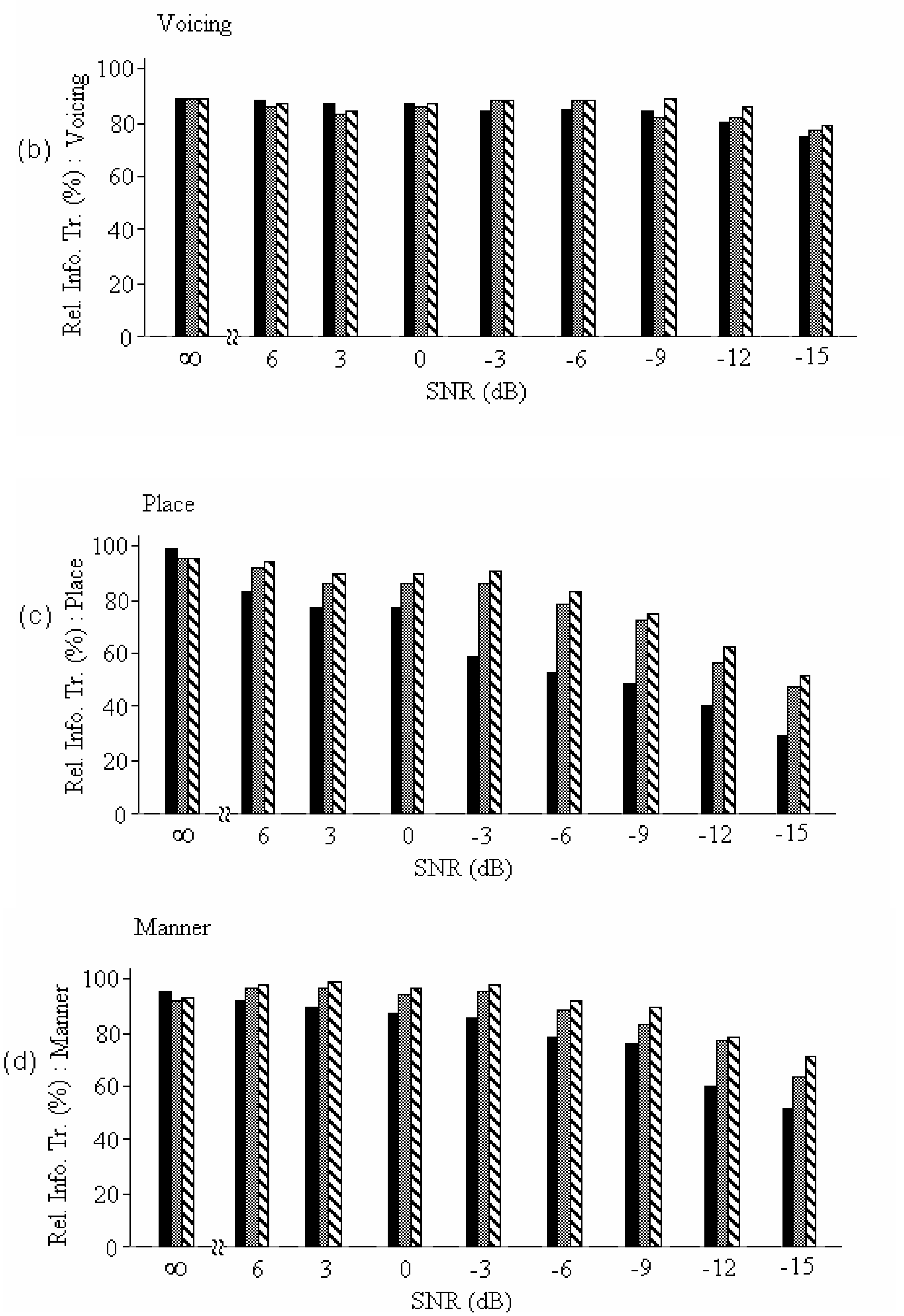


FIG. 4.12. (Contd.)

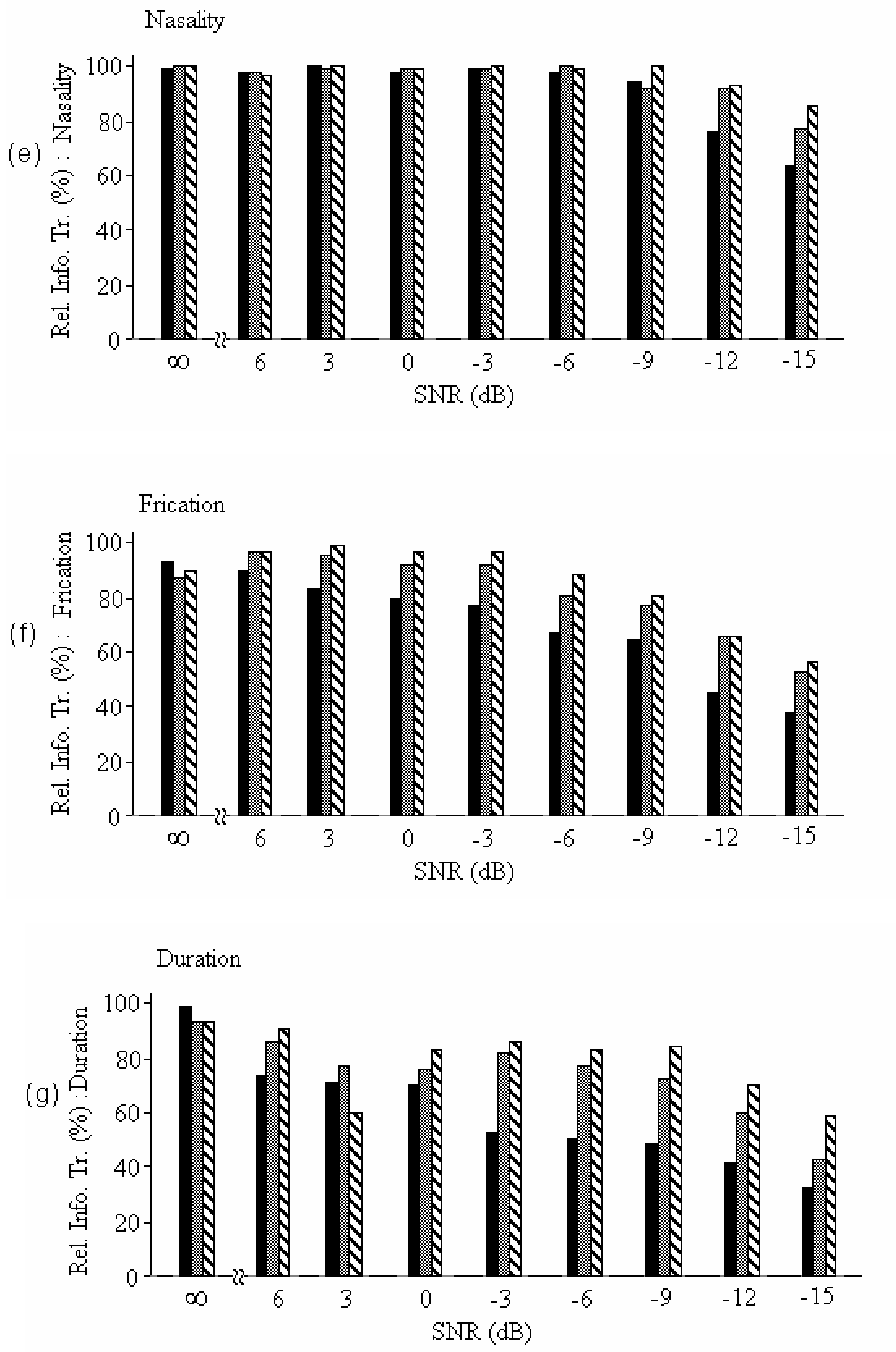


FIG. 4.12. (Contd.)

Duration: The adverse effect of the addition of masking noise on the reception of duration feature was next to that for the place feature. The value varies from 99.2 % under no-noise condition to 32.4 % under −15 dB SNR. At −12 and −15 dB SNR conditions, the percentage relative increase for SpA was 69.6 and 64.3 % respectively. The corresponding values for SpB were 149.8 and 163.1 %. Of all the features more relative improvement for scheme SpB over SpA was obtained for duration. The values at −12 and −15 dB were 33.3 and 52.8 %.

From the analysis of the listening tests conducted on normal hearing subjects with simulated sensorineural hearing loss, it is clear that processing helps in improving the perception of all consonantal features. The features of place, duration and frication were very much degraded by low SNR conditions. The average relative improvement was maximum for duration feature, then place and next frication. Manner feature was affected at low SNR conditions, which improved after processing to some extent. Perceptually balanced filters (SpB) showed higher improvement in information transmission of consonantal features compared to comb filters with sharp transition (SpA). The relative improvement was maximum for duration feature. Thus the listening tests on normal hearing subjects with simulated hearing loss showed that perceptually balanced comb filters are suitable for improving the perception under adverse listening conditions.

4.8 Experiment II: Listening tests on hearing impaired subjects

After the effectiveness of perceptually balanced filters was observed in Experiment I involving listening tests on normal hearing subjects with simulated loss, listening tests were subsequently carried out on five persons with moderate bilateral sensorineural loss. For these subjects, effects of cascading a linear phase filter with adjustable magnitude response to partly match the audiogram as a way of partial compensation for frequency dependent shifts in hearing thresholds was also investigated. The processing with these filters is referred to as SpC.

For the five subjects who participated in the listening tests, the degree of bilateral sensorineural hearing loss varied from moderate to severe. Table 4.6 shows the hearing thresholds of these subjects. Subject AB was a 61-year old male, having low frequency hearing loss. Subjects KS and BS having high frequency loss, were females with age of 31 and 38 years respectively. The losses were symmetrical for subject BS, whereas a small asymmetry was noticed for subject KS for frequencies below 1 kHz. Subject SK a male of 42 years had symmetrical loss, progressively increasing from 65 dB HL at low frequency 105 dB HL at high frequency. Subject SM (M, 59 years) had a moderate symmetrical loss of about 40 dB HL up to 2 kHz and progressively increasing to 85 dB HL after that. No measurements on the extent of spectral masking was carried out. During the listening tests, hearing impaired listeners were not using hearing aids and speech was presented through headphones. This procedure was followed to avoid differences in effectiveness of different types of hearing aids, and to facilitate binaural presentation.

TABLE 4.6. Experiment II. Hearing thresholds of the subjects with bilateral hearing impairment. PTA: average pure tone thresholds in dB HL, taken 0.5, 1, 2 kHz.

Subject Code (Sex, Age)	Ear L=left R=right	Hearing threshold (dB HL) Frequency (kHz)						PTA (dB)
		0.25	0.5	1	2	4	6	
AB (M, 61)	L	55	65	50	40	45	45	52
	R	60	60	45	35	35	35	47
KS (F, 32)	L	60	45	100	110	120	120	85
	R	65	85	100	115	120	120	100
BS (F, 38)	L	75	80	105	110	120	120	98
	R	70	65	85	110	120	120	87
SK (M, 41)	L	75	75	85	90	100	100	83
	R	75	85	90	95	110	105	90
SM (M, 59)	L	35	40	45	40	85	85	42
	R	35	45	45	40	75	75	42

4.8.1 Response time

The response time for unprocessed and processed test conditions are shown in Table 4.7 for five subjects AB, KS, BS, SK and SM. The relative decrease in response time (processed vs. unprocessed), significance level p from one-tailed paired t-test are also shown in the same table. Figure 4.13 shows the response time and its relative improvement for the five subjects. For unprocessed speech, the response time varied from 3.8 to 9.0 s. The range of percentage relative decrease in response times for the three schemes, SpA, SpB, and SpC were 2.4 – 35.1 s, 0 – 40.4 s, and 0 – 23.4 s respectively. Averaged across the subjects, the percentage relative improvements in response times for the three schemes, SpA, SpB, and SpC were 14.4, 21.5, and 16.2 s respectively.

TABLE 4.7. Experiment II. Response times for listening tests with hearing impaired subjects for Su: unprocessed speech, SpA: processed speech with comb filters with sharp transitions, SpB: processed speech with perceptually balanced comb filters and SpC: processed speech with adjustable magnitude response filter cascaded with perceptually balanced comb filter. RT = response time in s, S: subject, s.d. = standard deviation, R.I. = relative improvement in % with respect to unprocessed. *p*: significance level for one-tailed t-test (processed vs. unprocessed)

S		Su	SpA	SpB	SpC
AB	RT	3.83	3.74	2.9	3.16
	s.d	0.28	0.8	0.28	0.61
	R.I		2.4	24.2	17.6
	p		0.4	0.0004	0.03
KS	RT	4.72	4.33	4.49	3.61
	s.d	0.42	0.48	0.7	0.17
	R.I		8.2	4.8	23.4
	p		0.1	0.3	0.0003
BS	RT	9.02	5.86	5.56	7.2
	s.d	1.16	0.26	0.87	0.58
	R.I		35.1	38.4	20.2
	p		0.0002	0.0004	0.007
SK	RT	8.69	6.82	5.18	6.94
	s.d	1.13	0.65	0.64	0.86
	R.I		21.5	40.4	20.2
	p		0.006	0.0002	0.01
SM	RT	6.34	6.03	6.38	6.37
	s.d	0.68	0.41	0.69	0.4
	R.I		4.9	-0.5	-0.5
	p		0.2	0.5	0.5
Avg. RT		6.52	5.36	4.9	5.46
Avg. R.I.			14.4	21.5	16.2
p(paired)			0.06	0.1	0.2

Among the five subjects, subject AB had the least response time for unprocessed speech. He showed a modest (2.4 %) improvement with SpA, but had a large statistically significant improvement with perceptually balanced comb filters: 24.2 % for SpB and 17.6 % for SpC. Subject KS had highly statistically significant improvement in response time for SpC, with relative improvement of 23.4 %. In this case, the relative decrease in response time was 8.2 and 4.8 % for schemes SpA and SpB. The response time for unprocessed speech was maximum for subject BS. This subject showed highly statistically significant decrease response time for all the three processing schemes; with percentage relative improvements for SpA, SpB, and SpC of 35.1, 38.4, and 20.2 respectively. Subject SK had statistically significant improvement in response times of 21.3, 40.4 and 20.2 % for SpA, SpB, and SpC respectively. The subject SM had almost the same response time for unprocessed as well as all the processed schemes. The maximum improvement in response time for processing scheme SpC with respect to SpB was for subject KS who has asymmetry in hearing loss at low frequencies.

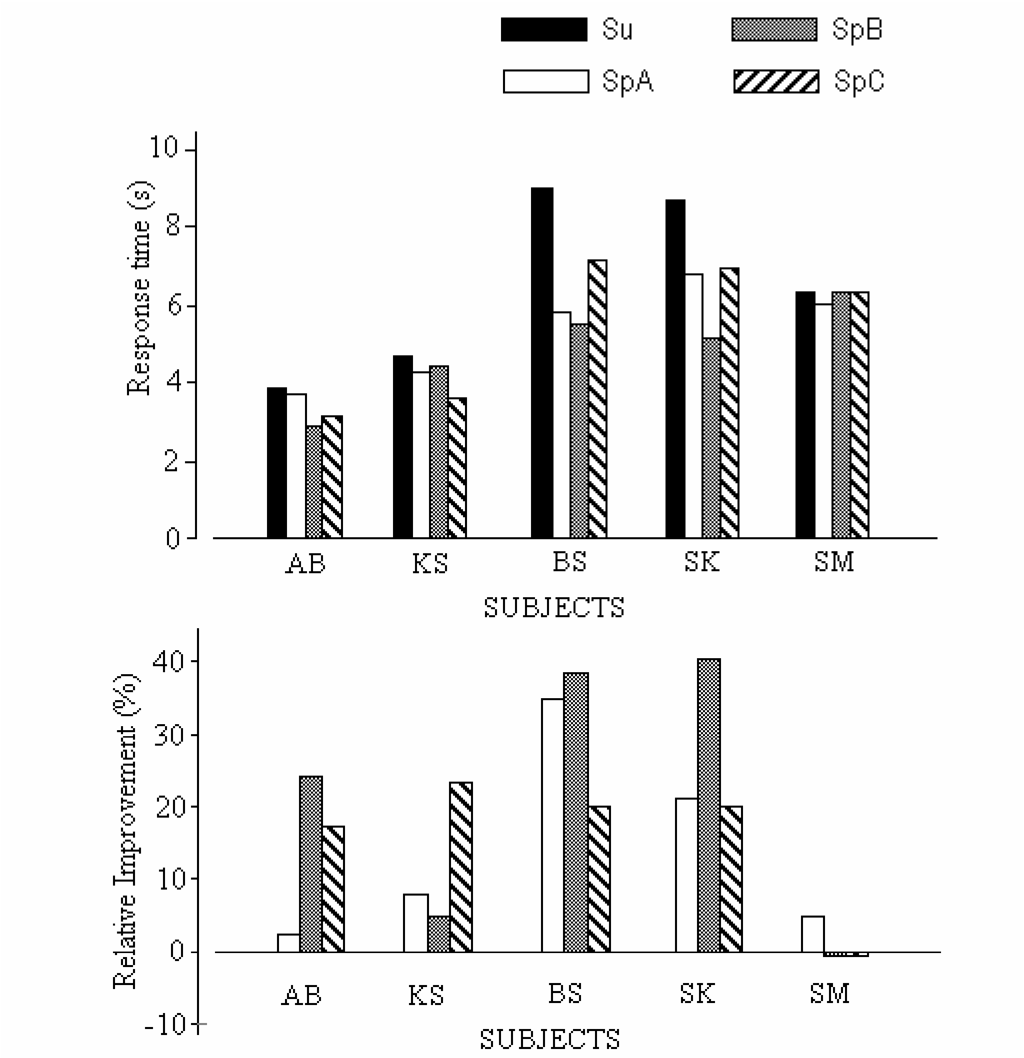


FIG. 4.13. Experiment II. Response time (s) and relative improvement (%) for five hearing impaired subjects. Su unprocessed speech, SpA: processed speech with comb filters with sharp transitions, SpB: processed speech with perceptually balanced comb filter, and SpC: processed speech with adjustable magnitude response filter cascaded with perceptually balanced comb filter.

Thus we see that across the subjects, spectral splitting helped in reducing the response time, i.e. in reducing the load on the perception process. The improvements were generally higher with perceptually balanced comb filters.

4.8.2 Recognition scores

Table 4.8 shows the recognition scores for unprocessed and processed speech, relative improvement for processed speech over unprocessed, relative improvement (%), and significance level p from one-tailed paired (processed vs. unprocessed). The recognition scores and the relative improvements in recognition scores are shown in Fig. 4.14. The recognition scores for unprocessed speech varied from 51.0 to 88.7 %. The percentage relative improvement range for SpA was from 1.3 to 27.4 with an average of 12.4. For SpB and SpC the relative improvement (%) varied from 8.7 to 39.2 with an average of 19.3 and from 12.0 to 47.7 with an average of 25.2 respectively.

TABLE 4.8. Experiment II. Recognition scores (%) for listening tests with hearing impaired subjects for Su: unprocessed speech, SpA: processed speech with comb filters with sharp transitions, SpB: processed speech with perceptually balanced comb filters and SpC: processed speech with adjustable magnitude response filter cascaded with perceptually balanced comb filter. RS = % recognition score, S: subject, s.d. = standard deviation, R.I. =relative improvement in % with respect to unprocessed. *p*: significance level (one-tailed) for t-test (processed vs. unprocessed). Avg. average.

S		Su	SpA	SpB	SpC
AB	RS	88.7	97.0	98.3	99.3
	s.d	4.2	1.4	0.0	0.9
	R.I		9.4	10.9	12.0
	p		0.001	0.004	0.0003
KS	RS	76.0	77.0	82.7	87.36
	s.d	4.3	3.2	2.8	2.5
	R.I		1.3	8.7	14.9
	p		0.4	0.01	0.0005
BS	RS	51.0	65.0	71.0	75.3
	s.d	4.3	2.0	3.8	3.8
	R.I		27.4	39.2	47.7
	p		0.0001	0.0000	0.0000
SK	RS	75.7	82.3	87.4	89.0
	s.d	0.9	3.5	2.5	2.2
	R.I		8.9	15.4	17.6
	p		0.002	0.0000	0.0000
SM	RS	67.7	77.8	82.7	90.3
	s.d	3.6	4.7	1.9	1.4
	R.I		14.9	22.2	33.5
	p		0.003	0.0000	0.0000
Avg. RS		71.8	79.8	84.4	88.3
Avg. R.I.			12.4	19.3	25.2
p(paired)			0.01	0.003	0.02

4.8.3 Information transmission analysis

For each of the processing conditions, the confusion matrices involving 25 presentations of each of the 12 stimuli, were subjected to information transmission analysis. Table 4.9 gives the relative information transmitted (%) for all consonantal features. Figure 4.15 shows the overall relative information transmitted and the relative information transmitted (%) by individual features of voicing, place, manner, nasality, duration, and frication, for the five hearing impaired subjects.

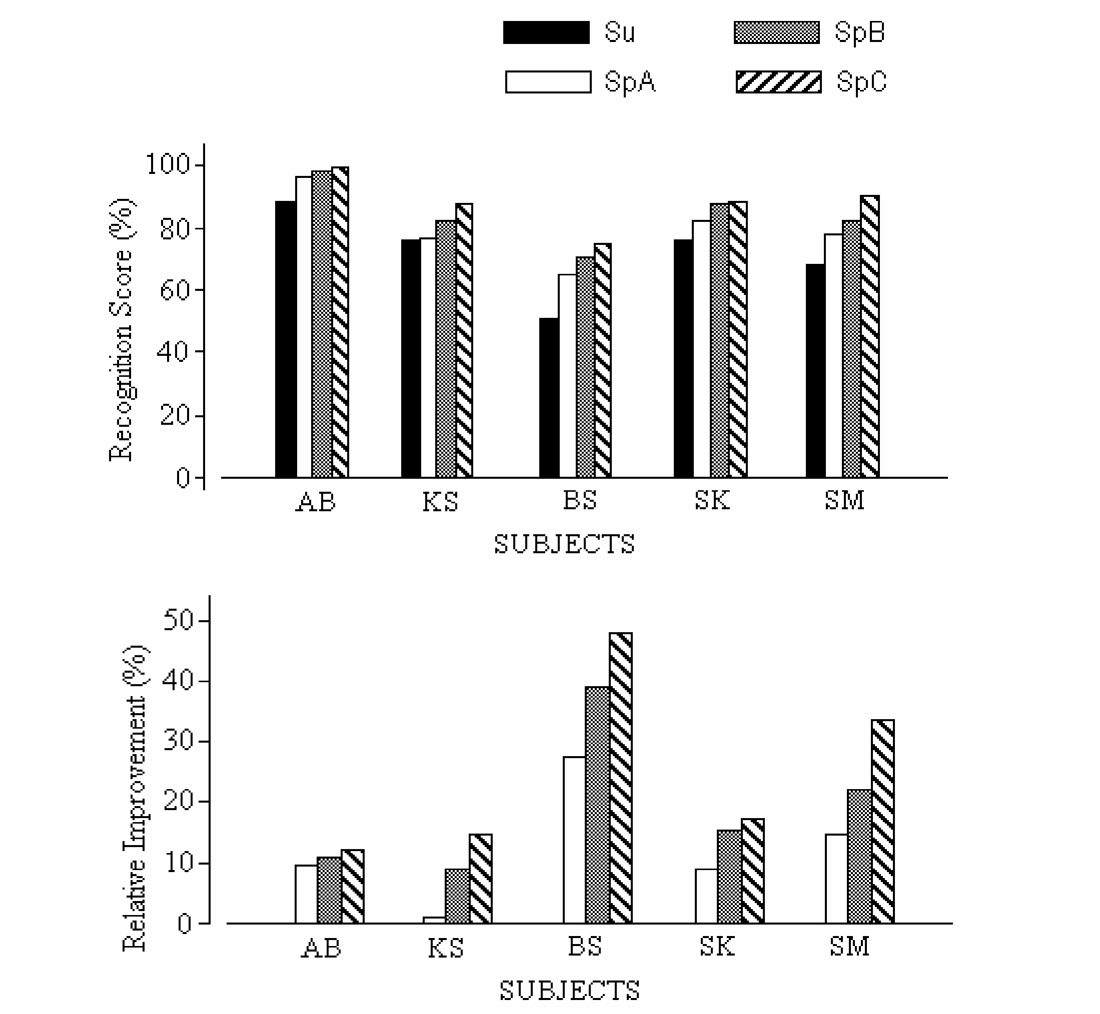


FIG. 4.14. Experiment II. Recognition scores (%) and the relative improvement (%) for five hearing impaired subjects Su: unprocessed speech. SpA: processed speech with comb filters with sharp transitions, SpB: processed speech with perceptually balanced comb filter, and SpC: processed speech with adjustable magnitude response filter cascaded with perceptually balanced comb filter.

Overall: The overall relative information transmitted varied from 61 to 85 % for unprocessed speech. The same for processed speech with different schemes SpA, SpB, and SpC were between 71 – 95 %, 75 – 98 %, 81 – 99 % respectively. The relative improvements varied from –2.5 to 23.0 %, 4.9 to 32.8 %, and 7.4 to 54.1 % for processed speech with SpA, SpB, and SpC respectively. For the three processing conditions SpA, SpB, and SpC, maximum relative improvement was for subject SM. He obtained more relative overall information transmitted for processing scheme SpC over SpB among the five subjects. Interestingly this subject did not show any improvement in response time (as described in Section 4.8.1). Subject BS also obtained good percentage relative information transmitted for SpC over SpB.

TABLE 4.9. Experiment II. Relative information transmitted (%) for listening tests with hearing impaired subjects for Su: unprocessed speech, SpA: processed speech with comb filters with sharp transitions, SpB: processed speech with perceptually balanced comb filters and SpC: processed speech with adjustable magnitude response filter cascaded with perceptually balanced comb filter. S: Subject, Avg. = average, R.I. average relative improvement, *p* significance level (one tailed) for paired t-test (unprocessed vs processed)

(a) Overall

S	Su	SpA	SpB	SpC
AB	85	95	98	99
KS	81	79	85	87
BS	62	71	75	81
SK	78	83	86	88
SM	61	75	81	94
Avg.	73.4	80.6	85.0	89.8
R.I.		10.6	16.8	24.3
p		0.03	0.006	0.01

(b) Feature: voicing

S	Su	SpA	SpB	SpC
AB	80	97	97	100
KS	82	79	92	97
BS	59	89	89	100
SK	58	88	86	88
SM	37	63	77	92
Avg.	63.2	83.2	88.2	95.4
R.I.		38.1	48.1	62.6
p		0.02	0.004	0.006

(c) Feature: place

S	Su	SpA	SpB	SpC
AB	74	94	95	98
KS	42	45	55	65
BS	8	23	33	34
SK	63	61	67	77
SM	46	64	70	95
Avg.	46.6	57.4	64.0	73.8
R.I.		51.5	86.1	108.2
p		0.03	0.006	0.005

(d) Feature: manner

S	Su	SpA	SpB	SpC
AB	80	91	100	98
KS	73	82	78	91
BS	60	58	56	69
SK	54	73	74	78
SM	46	52	63	78
Avg.	62.5	71.2	74.2	82.8
R.I.		14.2	19.8	35.2
p		0.03	0.04	0.003

(e) Feature: nasality

S	Su	SpA	SpB	SpC
AB	81	96	100	100
KS	100	92	100	100
BS	100	100	100	100
SK	86	100	94	96
SM	64	65	75	96
Avg.	86.2	90.6	93.8	98.4
R.I.		5.7	10.0	17.0
p		0.2	0.05	0.06

(f) Feature: frication

S	Su	SpA	SpB	SpC
AB	79	88	100	97
KS	55	74	65	85
BS	33	29	25	50
SK	31	55	60	64
SM	34	45	56	64
Avg.	46.4	58.2	61.2	72.0
R.I.		28.7	35.8	64.7
p		0.03	0.04	0.0008

(g) Feature: duration

S	Su	SpA	SpB	SpC
AB	35	87	94	96
KS	48	72	55	78
BS	36	48	48	67
SK	42	58	61	74
SM	43	76	73	89
Avg.	40.8	68.2	66.2	80.8
R.I.		69.4	66.3	101.2
p		0.009	0.03	0.001

Voicing: The relative information transmitted for consonantal voicing feature for unprocessed speech varied from 37 to 82 %. Relative improvement with processed speech varied in the range of −3.7 – 70.3 %, 12.2 – 108.1 %. and 18.3 – 148.7 % for SpA, SpB, and SpC respectively. Highest relative percentage improvement in voicing was obtained for subject SM in all the processing schemes.

Place: For place feature, the relative information transmitted for unprocessed speech varied from 8 to 74 %. The relative improvement percentage varied from −3.2 to 187.5 % for SpA, from 6.4 to 312.5 % for SpB and from 22.2 to 325 % for SpC. The subject BS who has severe high frequency loss, had more relative improvement (SpA: 187.5 %, SpB: 312.5 %, SpC: 325 %) compared to other subjects. Subject SM had high relative improvement for SpC with respect to SpB of 35.7 %.

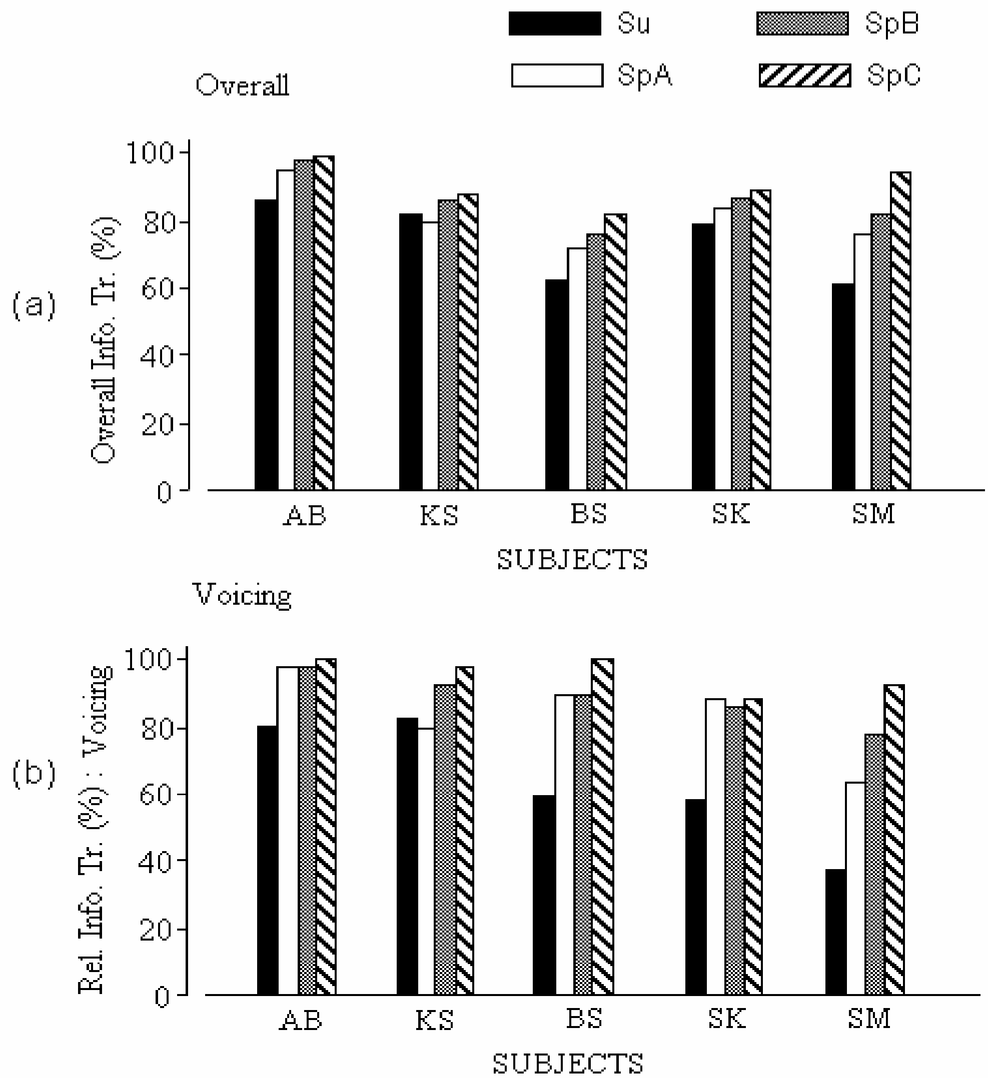


FIG. 4.15. Experiment II. Relative information transmitted (%) for (a) overall, (b) voicing, (c) place, (d) manner, (e) nasality, (d) frication, and (e) duration, for five hearing impaired subjects Su: unprocessed speech. SpA: processed speech with comb filters with sharp transitions, SpB: processed speech with perceptually balanced comb filter, and SpC: processed speech with adjustable magnitude response filter cascaded with perceptually balanced comb filter.

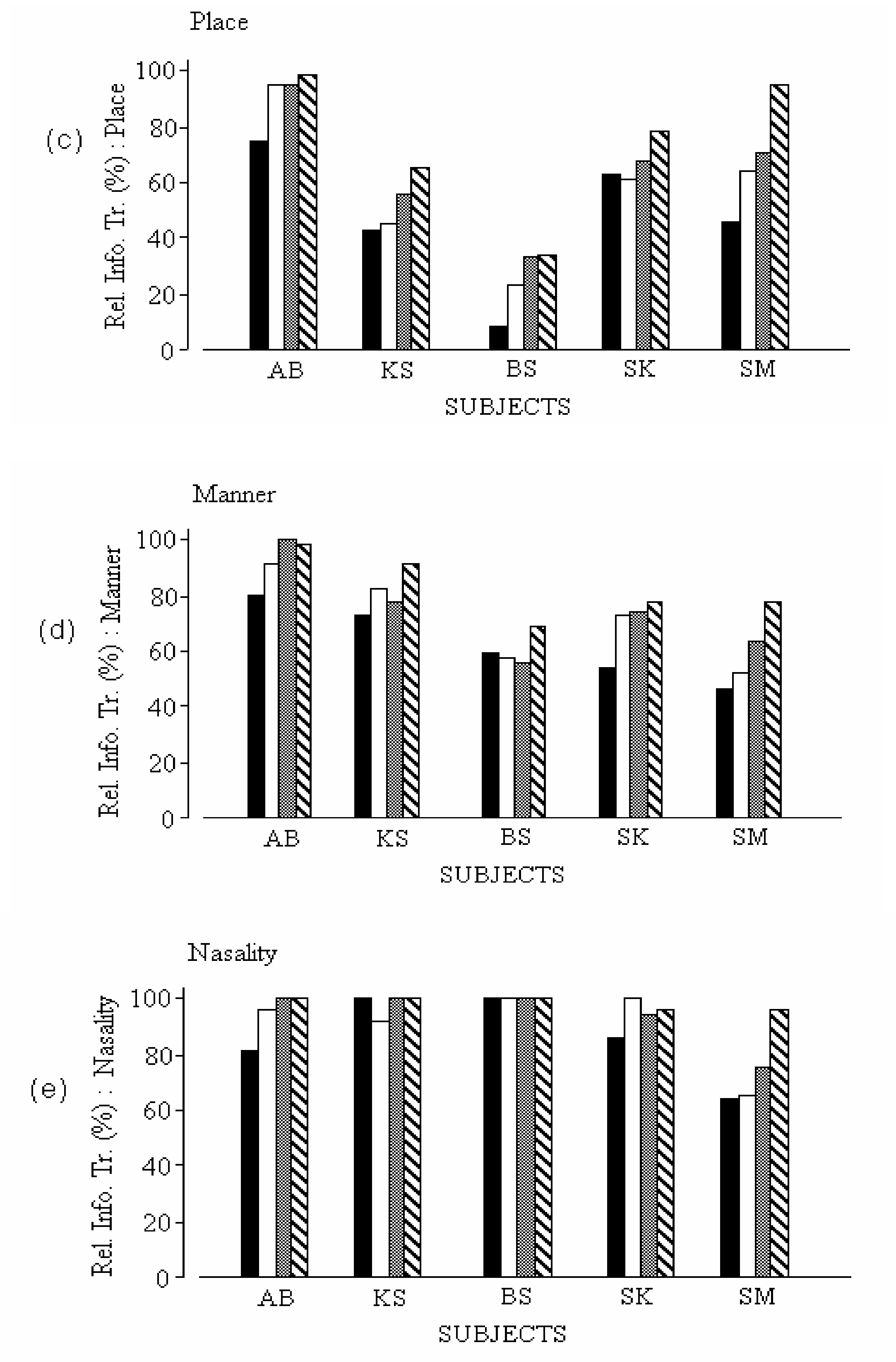
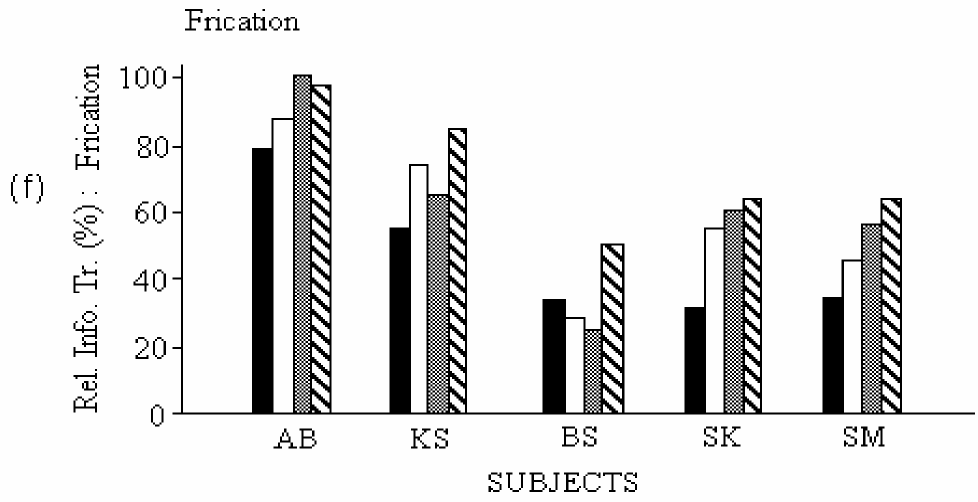


FIG. 4.15. (Contd.)



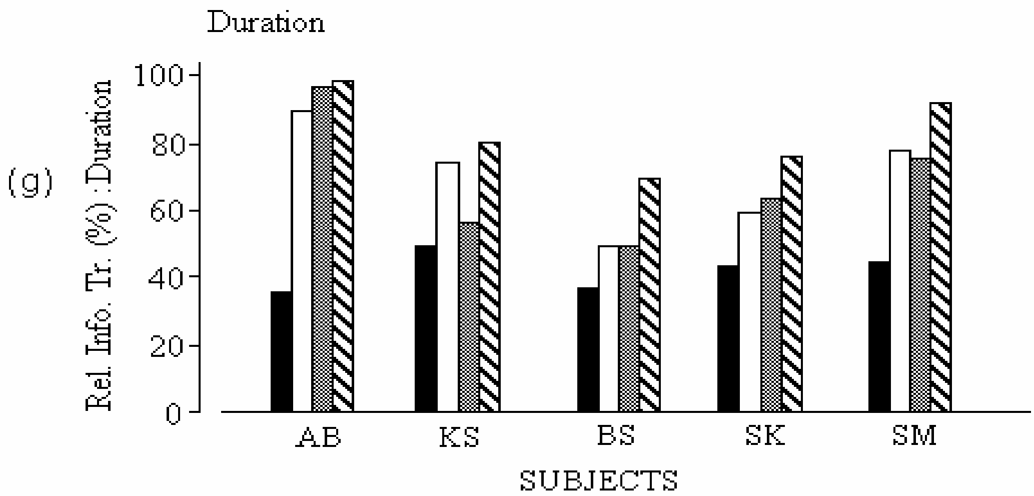


FIG. 4.15. (Contd.)

Manner: With unprocessed speech the relative information transmitted for manner speech varied from 46 to 80 %. The relative percentage improvements were between –3.3 and 35.2, –6.7 and 37, and 15 and 69.6 for SpA, SpB, and SpC respectively. The relative percentage improvement was more for subjects SM (69.6) and SK (44.4) for SpC, both with almost constant hearing loss.

Nasality: This feature was not affected for subjects with severe high frequency loss BS and KS. For other three subjects relative improvement increased with processing from SpA to SpC.

Frication: The relative information transmitted for frication with unprocessed speech varied from 31 to 79 %. The percentage relative improvements with processing increased for subject SK and SM from SpA to SpC. The subjects BS and KS with severe high frequency loss showed higher relative improvement for SpC with respect SpB.

Duration: For duration feature, the relative information transmitted for unprocessed speech varied from 35 to 48 %. Subject AB with low frequency loss showed more relative improvement increasing from SpA to SpC. Subjects with severe high frequency loss BS and KS had more relative improvement for SpC with respect to SpB.

The main conclusion that can be derived from the information transmission analysis is that spectral splitting helps in improving the speech perception of sensorineural hearing impaired subjects. Perceptually balanced comb filters provided more improvement compared to comb filters with sharp transitions. Further improvement was obtained with the inclusion of adjustable magnitude response filter to partly compensate for frequency dependent shifts in hearing thresholds. Subject with relatively constant hearing loss (particularly below 2 kHz) had the maximum additional benefit. The reason for this may be the relatively small variation in gain (6 dB) provided by the variable gain filter. The subjects with high frequency losses were also benefited by the scheme SpC with respect to SpB. The subjects who had low frequency losses showed no additional benefit with SpC.

4.9 Discussion

From the analysis of the listening tests conducted on normal hearing subjects with simulated sensorineural hearing loss, it is seen that processing helps in improving the response time, recognition scores, and the perception of all consonantal features. For the perceptually balanced comb filter (SpB), the improvements in recognition scores were 20.7, 21.5, 21.4, 29.7, and 27.5 % for the SNR conditions of –3, –6, –9, –12, and –15 dB respectively. In terms of relative information transmission, the relative improvement was maximum for duration feature, followed by place and frication. The improvements for both the schemes SpA and SpB were highly significant ($p \leq 0.01$) for these SNR conditions. The relative percentage improvement for SpB over SpA was 1.6. 2.1, 3.0, 4.2 and 4.8 % for SNR conditions –3, –6, –9, –12, and –15 dB respectively. Perceptually balanced filters showed higher improvement in information transmission of consonantal features compared to comb filters with sharp transition. The relative improvement was maximum for duration feature. Thus the listening tests on normal hearing subjects with simulated loss showed that spectral splitting by comb filtering improved speech perception, by reducing the effect of spectral masking, under poor SNR conditions. Further the improvements were higher with perceptually balanced comb filters. There was a reduction in subjects' response time, indicating a reduction in load on perception process, but these improvements were statistically not very significant.

To find the SNR advantage due to processing, the scores were plotted with second order polynomial fit and the improvement at 80 % score was determined. The schemes SpA and SpB provided SNR advantages in recognition scores of approximately 7.5 and 9 dB respectively. The advantages in information transmission were 8 and 9 dB respectively. For specific consonantal features the advantages for the schemes, SpA and SpB, were approximately 7 and 8 dB for place, 7.5 and 9.5 dB for frication and 9 and 14 dB for duration.

The listening tests on bilateral sensorineural hearing impaired subjects included a filter with magnitude response adjusted to partly compensate for the frequency dependent shifts in hearing thresholds, along with spectral splitting with perceptually balanced comb filter. All the three processing schemes evaluated in this investigation provided better speech reception in listening tests for persons with mild to severe bilateral sensorineural hearing impairment. Processing with perceptually balanced comb filters provided better results compared to processing with comb filters with sharp transitions. The highest improvement was for the scheme of adjustable magnitude response filter cascaded with perceptually balanced comb filters. The recognition scores for unprocessed speech varied from 51.0 to 88.7 %. The percentage relative improvements for SpA, SpB and SpC, ranged between 1.3 – 27.4, 8.7 – 39.2 and 12.0 – 47.7 respectively. Across the subjects, spectral splitting helped in reducing the response time. The relative improvements were more for perceptually balanced comb filters. Cascading of the filter with adjustable magnitude response further reduced the response time of subject with asymmetrical loss.

The scheme of compensating for frequency dependent shifts in hearing threshold cascaded with perceptually balanced comb filters provided better intelligibility for all subjects but with varying degrees of improvement. Subject with almost flat loss (particularly below 2 kHz) showed higher improvement in recognition score and relative information transmission of features (overall and place). The subjects with high frequency loss are benefited in the information transmission of features of frication and duration. Subject with high loss at low frequency is least benefited by all the schemes.

The relative improvements in recognition score and information transmission for consonantal features (particularly place feature) is likely to be related to the extent of spread of masking for individual subjects. Hence psychoacoustic experimentation for characterizing the auditory bands and relating these to the improvements because of spectral splitting is likely to give useful insight into the suitability of this scheme as well as selection of appropriate bandwidth for spectral splitting.

Chapter 5

COMBINED SPLITTING USING TIME-VARYING COMB FILTERS

5.1 Introduction

Spectral splitting with perceptually balanced comb filter has helped in reducing the effect of spectral splitting. The design, implementation, and evaluation on normal hearing subjects with simulated hearing loss and on subjects with bilateral sensorineural hearing impairment, are presented in Chapter 4. Earlier work (Jangamashetti, 2003; Jangamashetti and Pandey, 2000b), reviewed in Section 3.4, has shown that temporal splitting using trapezoidal fading functions with inter-aural switching period of 20 ms provided improvement by reducing the effect of temporal masking. In spectral splitting, sensory cells corresponding to alternate bands are always relaxed; while in temporal splitting, all the sensory cells of the two ears are alternately relaxed. A combined splitting scheme (Jangamashetti *et al.*, 2001; Pandey *et al*., 2001) is devised to have spectral and temporal splitting simultaneously, so as to provide periodic relaxation from stimulation to all the sensory cells of the basilar membrane, and hence reduce the effects of spectral and temporal masking. The implementation of the combined splitting scheme was done, using a pair of time-varying comb filters with pre-calculated sets of coefficients (each set corresponds to a perceptually balanced comb filter), which were selected in steps for a cyclic sweeping of magnitude responses. This chapter presents investigations carried out to select appropriate values for the sweep cycle and the number of steps in sweeping the magnitude responses.

5.2 Implementation of time-varying comb filters

A schematic representation of combined splitting using time-varying comb filters is shown in Fig. 5.1 (a). The digitized input signal $s(n)$ is passed through a pair of time-varying comb filters to produce two digitized outputs, $s_1(n)$ and $s_2(n)$, for binaural dichotic presentation. Each of these filters was realized by stepping through m comb filters, which have magnitude responses such that the pass bands and stop bands of each of these filters get slightly shifted along the frequency axis with respect to the pass and stop bands of the previous comb filter, as shown in Fig. 5.1 (b). If the filters in the time-varying comb filter for the left ear were numbered in the order of sweeping as [1], [2], …, [m/2], [m/2 +1], …, [m −1], [m], then the same for the right will be [m/2 +1], [m/2 +2], …, [m], [1], [2], …, [m/2 −1], [m/2]. At a given time, two complementary comb filter responses corresponding to odd and even bands process the speech, and provide spectral splitting. Filters [1] and [m/2 +1], [2] and [m/2 +2], …, and [m/2] and [m] form these complementary pairs. We refer to the number of steppings in the filter response as the number of shiftings m. In actual realization, we have m set of pre-calculated filter coefficients, and these were cyclically selected as shown in Fig. 5.1.

Eighteen critical bands, spread over the frequency range of 100 Hz to 5 kHz, were used for spectral splitting, with odd numbered bands forming one comb filter and even numbered bands forming the other comb filter of the pair. The range between center frequencies of any two neighboring bands of the comb filters used for spectral splitting was uniformly divided into m, forming the center frequencies of the pass bands in the comb filters in the combined splitting scheme with m shiftings. The bandwidths corresponding to these center frequencies were obtained from the auditory critical bandwidths, as described by Zwicker (1961) shown in Fig 2.5. The center frequencies and bandwidths of the different pass bands for a pair of time-varying comb filters with four shiftings is given in Table 5.1. The comb filters were designed as 256-coefficient linear phase FIR filters, using frequency-sampling technique with sampling rate of 10 kSa/s. Each pair was designed separately as perceptually balanced comb filters, similar to the one used for spectral splitting described in Chapter 4. To minimize the perceived spectral distortion at the transitions between any two adjacent bands of a pair used simultaneously, the crossovers were adjusted to lie between 4 and 6 dB with respect to passband gain. This was obtained by iteratively adjusting the magnitude of the transition samples. The transition bandwidth of the comb filters was 78 Hz at the lower frequencies and 117 Hz at higher frequencies. The stop-band attenuation is greater than 30 dB with passband ripple constrained to 1 dB. The pairs of comb filter responses for 4-shiftings (center frequencies as given in Table 5.1) are shown in Fig 5.2.

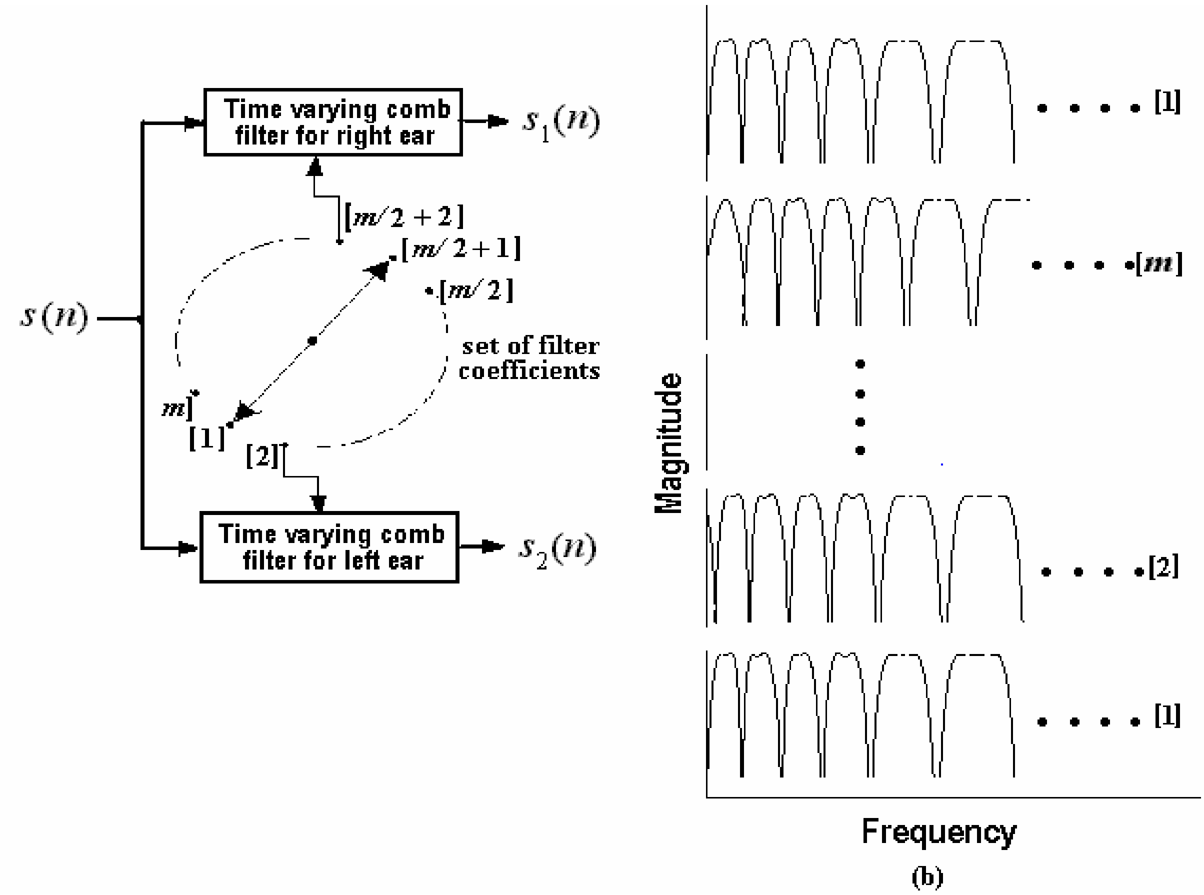


FIG. 5.1. (a) A schematic representation of the scheme of combined splitting using time-varying comb filters. Input signal: $s(n)$, and outputs to the two ears: $s_1(n)$ and $s_2(n)$. (b) Magnitude responses of comb filters stepped through cyclically

A swept sine tone, processed with each of these complementary pair of comb filters that are used together, resulted in no perceived variation in loudness, due to passband ripple or band crossovers. For a time-varying comb filter with T_c sweep cycle and m shiftings, each successive pair of comb filter coefficients were used for T_c/m in every cycle. As the number of shiftings increases, the sweeping becomes smoother and the stimulation of the sensory cells in the basilar membrane moves smoothly. But this increases the set of filter coefficients to be stored, and may lead to difficulties in real time implementation. Sweeping of bands provides temporal splitting and helps in reducing the effect of increased forward and backward masking. For a particular sweep cycle, the stimulation time of each of the sensory cells of the basilar membrane is independent of the number of shiftings. For an acoustic signal with non-time varying spectrum, variation in the stimulation of sensory cells becomes more smooth with increase in the number of shiftings. As the sweep cycle increases, each of the sensory

TABLE 5.1. Bandwidths for 18 critical band filters with 4 shiftings between odd-even-odd bands

Filter for left ear		
Band	Center frequency kHz	Bandwidth (kHz)
1	0.15	0.10-0.20
	0.20	0.15-0.25
	0.25	0.20-0.30
	0.30	0.25-0.35
3	0.35	0.30-0.40
	0.40	0.35-0.46
	0.45	0.40-0.51
	0.51	0.45-0.57
5	0.57	0.51-0.63
	0.63	0.57-0.69
	0.70	0.63-0.77
	0.77	0.70-0.84
7	0.84	0.77-0.92
	0.91	0.84-0.99
	1.00	0.92-1.08
	1.08	1.00-1.17
9	1.17	1.08-1.27
	1.27	1.18-1.37
	1.37	1.27-1.48
	1.48	1.37-1.59
11	1.60	1.48-1.72
	1.72	1.60-1.85
	1.86	1.72-2.00
	2.00	1.85-2.15
13	2.16	2.00-2.32
	2.33	2.15-2.51
	2.51	2.32-2.70
	2.71	2.50-2.92
15	2.92	2.70-3.15
	3.16	2.91-3.42
	3.42	3.15-3.70
	3.72	3.41-4.03
17	4.05	3.70-4.40
	4.39	4.01-4.77
	4.70	4.40-5.00
	5.07	4.62-5.52

Filter for right ear		
Band	Center frequency (kHz)	Bandwidth (kHz)
2	0.25	0.20-0.30
	0.30	0.25-0.35
	0.13	0.07-0.20
	0.20	0.15-0.25
4	0.45	0.40-0.51
	0.51	0.45-0.57
	0.35	0.30-0.40
	0.40	0.35-0.46
6	0.70	0.63-0.77
	0.77	0.70-0.84
	0.57	0.51-0.63
	0.63	0.57-0.69
8	1.00	0.92-1.08
	1.08	1.00-1.17
	0.84	0.77-0.92
	0.91	0.84-0.99
10	1.37	1.27-1.48
	1.48	1.37-1.59
	1.17	1.08-1.27
	1.27	1.18-1.37
12	1.86	1.72-2.00
	2.00	1.85-2.15
	1.60	1.48-1.72
	1.72	1.60-1.85
14	2.51	2.32-2.70
	2.71	2.50-2.92
	2.16	2.00-2.32
	2.33	2.15-2.51
16	3.42	3.15-3.70
	3.72	3.41-4.03
	2.92	2.70-3.15
	3.16	2.91-3.42
18	4.70	4.40-5.00
	5.07	4.62-5.52
	4.05	3.70-4.40
	4.39	4.01-4.77

(d)
pair
4-2

(c)
pair
3-1

(b)
pair
2-4

(a)
pair
1-3

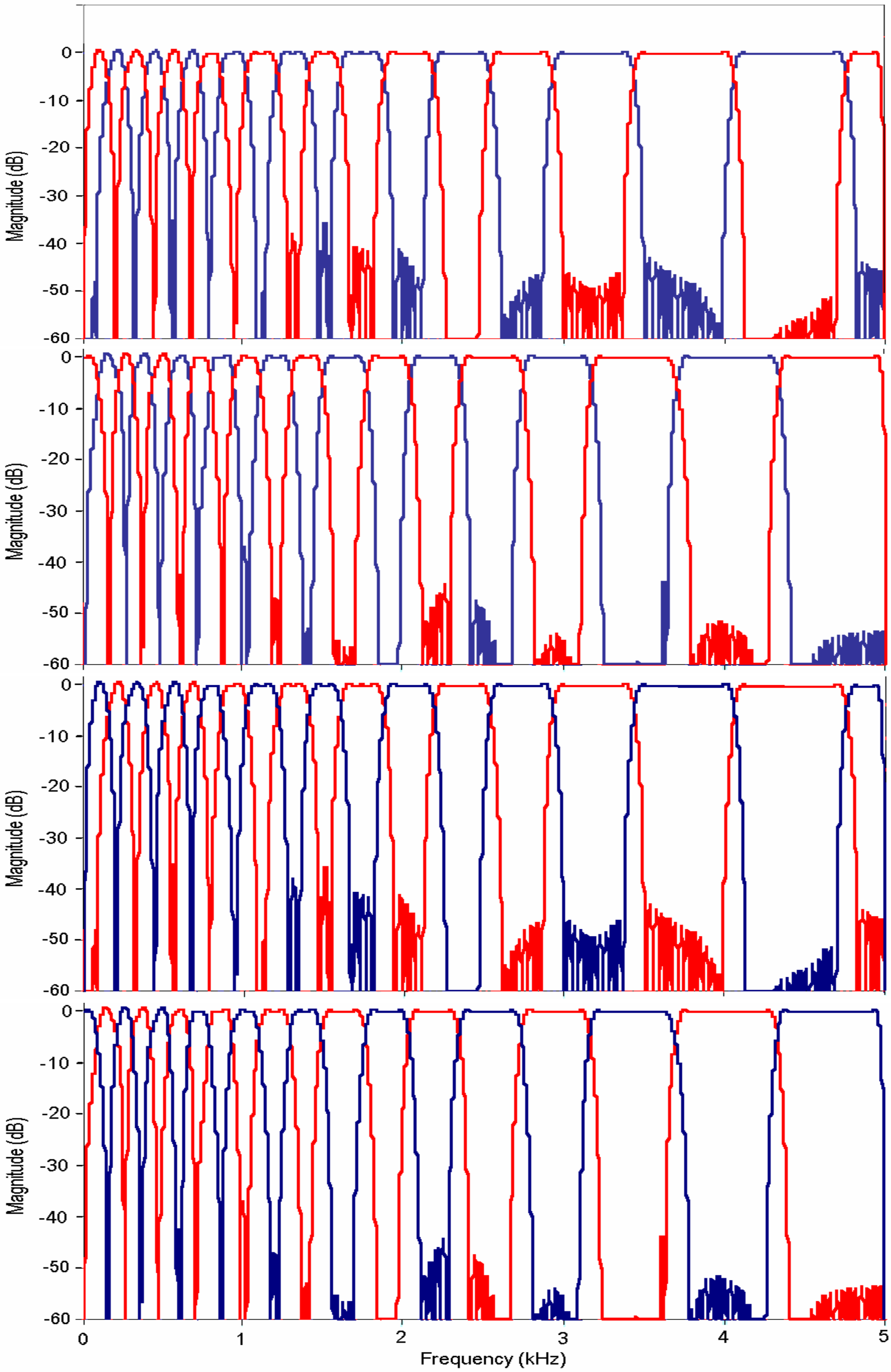


FIG. 5.2. Magnitude response of the time-varying comb filter with four shiftings. Four pairs of complementary perceptually balanced comb filter [1-3], [2-4], [3-1] and [4-2]

cells gets stimulated and relaxed for a longer time. Certain minimum time of stimulation may be required for perceiving some of the temporal cues. On the other hand, a large stimulation time may not help in reducing the effect of increased forward and backward masking in persons with sensorineural hearing impairment. Hence there was a need to establish appropriate values for the sweep cycle time T_c and number of shiftings m.

5.3 Experiment III: Listening tests for selection of parameters

The implementation of the combined splitting scheme was done, using a pair of time-varying comb filters with pre-calculated sets of coefficients, which were selected in steps for a cyclic sweeping of magnitude responses. In an earlier investigation (Jangamashetti, 2003; Jangamashetti *et al.*, 2001; Pandey *et al.*, 2001), reviewed in Chapter 3, experimental evaluation on normal hearing subjects with simulated hearing loss was conducted for a sweep cycle of 20 ms. The scheme provided improvement in response times and recognition scores. Information transmission analysis showed increase in the perception of all features especially the place and duration features. The scheme was successful in simultaneously reducing the effects of increased temporal and spectral masking.

With a given sweep cycle, the time duration of stimulation of the sensory cells remains constant, irrespective of the number of shiftings. For the perception of consonants, temporal cues related to rapid formant transitions, voice-onset time, etc. are important. The duration of formant transitions in stop consonants is 15 – 30 ms (Turner, *et al.*, 1997). Effect of forward masking is more if the signal occurs within 10 ms of the masker and decreases with time extending up to 200 – 300 ms. Significant backward masking occurs up to a temporal gap of 20 ms and with decreasing effect up to 100 ms (Fitzgibbons and Wightman, 1982; O'Shaughnessy, 1987). Hence an optimal value for sweep cycle duration may lie between 10 and 100 ms.

It was decided to investigate the appropriate values of sweep cycle time T_c by experimentation. For each sweep cycle, shiftings m of 2, 4, 8, and 16 were considered. With T_c = 10, 20, 50, 80, and 100 ms along with number of shiftings m = 2, 4, 8, and 16, we have a total of 20 conditions. Hence, we used quality assessment of processed speech for preliminary selection of appropriate values, for at least narrowing down the range of values. The listening tests were carried out in two phases with different test material with the same experimental set-up and listeners. In the first phase, which will be referred as Experiment III-A, a sine wave linearly swept from 100 Hz to 5 kHz and back with a sweep period of 1 minute was used as the test material. Results from these listening tests were analyzed to narrow down the range of processing conditions for listening tests of the second phase, referred as Experiment III-B. The speech material in this case was a speech passage of about 15 minute duration, processed for dichotic presentation with the selected processing conditions. Further these tests involved the use of simulated hearing loss to study the interaction between the processing conditions and degree of simulated loss.

Three normal hearing subjects (AC: 40, DJ: 37, and PL: 33), with pure tone thresholds less than 20 dB HL in the range of 125 Hz to 6 kHz, participated in both the listening tests. The test material was processed off-line. For binaural listening, the test material was outputted at 10 kSa/s through the line-out of the PC sound card to a 2-channel audio amplifier and presented through headphones.

5.3.1 Results of Experiment III-A

Listening tests were carried out for rank ordering the processing conditions on the basis of perceived similarity of the dichotically presented processed signal with diotically presented unprocessed signal. A sine wave linearly swept from 100 Hz to 5 kHz and back with a sweep period of 1 minute was used as the test material. The 20 processing conditions formed by the sweep cycle time T_c = 10, 20, 50, 80, and 100 ms and m = 2, 4, 8, and 16 were labeled as Sp_CS-T_c/m. For example Sp_CS-50/16 represents combined splitting scheme with sweep cycle duration of T_c = 50 ms and number of shiftings m = 16. A different randomized list of processing conditions was prepared for each subject. In each presentation, the unprocessed sound was followed by two processed sounds. The subject ranked the perceptual similarity of the processed sounds as one of the two being higher, or both being similar. The test started with the first two processed conditions in the randomized list. In the next presentation, the higher ranked sound (or either one in case of similar ranking) was paired with the sound with the next processing conditions from the list. This process was continued until all the conditions were exhausted. Based on the ranking comparison, the conditions were ranked 1 onwards, with 1 being the highest in similarity. The condition whose perceptual rank could be resolved was taken out of the list, and the procedure was continued for ranking of the other conditions. This was carried out until all the conditions were ranked. Because of many test conditions being ranked perceptually similar by the listener, the ranking ranged from 1 to 5.

Table 5.2 gives the final ranking obtained for each processed test material, for the three subjects, and also the medians and means. Table 5.3 shows the mean value of the rankings for processing conditions tabulated according to T_c and m. All processing conditions with sweep cycle duration T_c of 10 ms, and the condition with m = 2 and T_c = 20 ms had the lowest ranking. Among the number of shiftings, m = 8 and 16 provided better results compared to shiftings, m = 2 and 4 for sweep cycles of T_c = 50, 80, and 100 ms and the difference in mean ranking for m = 8 and 16 was negligible. Out of these processing conditions, Sp_CS-50/16 had the highest mean rank. Conditions Sp_CS-50/8 and Sp_CS-80/16 were ranked next. From these results, we can say that the preferred conditions are $T_c \approx$ 50 ms and m = 16

TABLE 5.2. Experiment III-A. Ranking of processing conditions for combined splitting, with swept sine wave. Combined splitting schemes are denoted as Sp_CS-T /ms, T_c = 10, 20, 50, 80, 100 ms, number of shiftings m = 2, 4, 8, 16, S: Subject.

S	Sp_ CS-10/2	Sp_ CS-10/4	Sp_ CS-10/8	Sp_ CS-10/16	Sp_ CS-20/2	Sp_ CS-20/4	Sp_ CS-20/8	Sp_ CS-20/16	Sp_ CS-50/2	Sp_ CS-50/4
AC	5	5	5	5	5	4	4	4	3	2
DJ	5	5	5	5	5	5	5	5	3	3
PL	5	5	5	5	5	5	4	5	5	4
Median	5	5	5	5	5	5	4	5	3	3
Mean	5.0	5.0	5.0	5.0	5.0	4.7	4.3	4.7	3.7	3.0

S	Sp_ CS-50/8	Sp_ CS-50/16	Sp_ CS-80/2	Sp_ CS-80/4	Sp_ CS-80/8	Sp_ CS-80/16	Sp_ CS-100/2	Sp_ CS-100/4	Sp_ CS-100/8	Sp_ CS-100/16
AC	1	1	3	2	2	2	4	3	3	3
DJ	1	1	3	3	2	1	4	4	2	2
PL	2	1	4	3	1	1	5	4	3	3
Median	1	1	3	3	2	1	4	4	3	3
Mean	1.3	1.0	3.3	2.7	1.7	1.3	4.3	3.7	2.7	2.7

Further to find the difference in similarity for small variations in sweep cycle duration, evaluation was carried out with swept sine wave processed for sweep cycles T_c = 40, 50, and 60 ms, with shiftings m = 8, 16. Listening tests were conducted for these 6 processing conditions, for perceptual similarity with the unprocessed test signal. The same procedure as described above was used for conducting the listening tests, and the same subjects participated. Any distinct ranking of these processing conditions could not be obtained.

TABLE 5.3. Experiment III-A.- Mean value of the rankings for processing conditions tabulated according to T_c and m

m	T_c				
	10	20	50	80	100
2	5.0	5.0	3.7	3.3	4.3
4	5.0	4.7	3.0	2.7	3.7
8	5.0	4.3	1.3	1.7	2.7
16	5.0	4.7	1.0	1.3	2.7

5.3.2 Results of Experiment III-B

In the second phase, quality assessment was carried out with a speech passage in English of approximately 15 minute duration. The passage was recorded from a male and a female speaker with considerations for clarity of speech. The two speech signals were processed by the time-varying comb filters with sweep cycle time periods $T_{c,}$ = 20, 50, 80, and 100 ms, and shiftings m = 8 and 16. Perceptual quality ranking of these processing conditions was carried out by conducting listening tests on three subjects with normal hearing. These subjects were the same ones as in Experiment III-A. Bilateral sensorineural hearing impairment was simulated by adding broadband Gaussian noise with constant short-time SNR, as described in Section 3.7.2 and used in Chapter 4. For simulating different levels of hearing impairment, SNR conditions of ∞, 6, 3, 0, and −3 were used. The total number of stimuli used were

(unprocessed + 8 processed conditions) × (2 speakers) × (5 SNR values) = 90.

The ranking was obtained through paired comparison for various processing conditions for a given SNR and speaker. Stimuli with two processing conditions were selected randomly to form a pair and presented. Subject had the choice of listening only a part of the passage and could listen to the two speech signals of the pair any number of times, to establish perceptual ranking. Subjects generally did not listen to the speech beyond 1−2 minutes. The subject ranked the perceptual quality of these two sounds as one of them being higher or both being similar. In the next presentation, the higher ranked sound (or either one in case of similar ranking) was paired with the next stimulus from the list. This process was continued until a ranking for all the 9 stimuli (unprocessed + 8 processed conditions) was obtained, with highest quality being ranked as 1. Because of some of the stimuli being ranked perceptually similar, the ranking ranged from 1 to 5.

The quality rankings for the 5 SNR conditions and two speakers (male and female) by three subjects are given in Table 5.4. There is a general agreement among the subjects on perceptual ranking of processing conditions. The rankings occasionally differ for male and female speakers. Perceptual quality rankings by the three subjects were averaged and these values are given in Table 5.5.

Under no-noise condition (SNR = ∞), the unprocessed speech was obviously ranked highest, T_c = 100 ms was ranked distinctly lower than the other three conditions. At lower SNR conditions, ranking of processed speech versus unprocessed improved for both the speakers. At 0 and –3 dB, the unprocessed speech was ranked the lowest. This is in conformity with results obtained from earlier experiments (Jangamashetti, 2003; Jangamashetti *et al.*, 2001; Pandey *et al*., 2001), that the improvements of dichotic listening become apparent under adverse listening conditions. For SNR = 0 and –3 dB, T_c = 50 ms ranked the highest for both the speakers, with T_c = 80 ms being close. There is not much change in perceptual quality ranking for m = 8 and 16.

5.4 Discussion

The scheme of combined splitting was implemented using time-varying comb filters to reduce the effects of spectral and temporal masking simultaneously. Perceptually balanced comb filters were cyclically swept in the time-varying comb filter to obtain spectral and temporal splitting together. Two sets of listening test were conducted to select the most appropriate processing conditions, based on perceptual ranking of processed signals, by three normal hearing subjects. In the first listening test (Experiment III-A), input signal was a sine wave linearly swept from 100 Hz to 5 kHz and back over 1 minute duration. Processing conditions tested were combinations of T_c = 10, 20, 50, 80, and 100 ms and m = 2, 4, 8, and 16. The average ranking from three subjects clearly established that preferred conditions are T_c = 40 – 60 ms and with m = 8 or 16.

TABLE 5.4 Experiment III-B Perceptual ranking of processing conditions for combined splitting, for male and female speaker. Combined splitting schemes are denoted as Sp_CS-T_c/m, where T_c = 20, 50, 80, 100 ms, number of shiftings m = 8, at SNR conditions of ∞, 6, 3, 0, –3.dB S: Subject.

SNR dB	Speaker M/F	S	Su	Sp_CS-20/4	Sp_CS-20/8	Sp_CS-50/8	Sp_CS-50/16	Sp_CS-80/8	Sp_CS-80/16	Sp_CS-100/8	Sp_CS-100/16
∞	M	AC	1	2	2	1	1	1	2	3	3
		DJ	1	1	2	1	1	2	1	3	3
		PL	1	2	2	2	2	2	1	2	2
		Mean	1	1.7	2	1.3	1.3	1.7	1.3	2.7	2.7
	F	AC	1	2	2	2	2	2	2	3	3
		DJ	1	2	2	2	2	2	1	3	3
		PL	1	2	1	2	1	2	2	3	3
		Mean	1	2	1.7	2	1.7	2	1.7	3	3
6	M	AC	1	2	1	2	1	2	1	3	3
		DJ	2	2	4	3	3	1	1	2	3
		PL	1	3	2	3	3	2	2	2	2
		Mean	1.3	2.3	2.3	2.3	2.3	1.7	1.3	2.3	2.7
	F	AC	2	3	3	1	1	1	2	2	2
		DJ	2	2	2	1	2	2	1	2	2
		PL	1	3	2	2	1	1	2	3	2
		Mean	1.7	2.7	2.3	1.3	1.3	1.3	1.7	2.3	2
3	M	AC	4	3	3	1	1	2	2	2	2
		DJ	4	4	3	2	2	1	1	1	2
		PL	3	2	2	2	1	3	3	2	2
		Mean	3.7	3.0	2.7	1.7	1.3	2.0	2.0	1.7	2
	F	AC	5	3	3	1	1	2	2	3	3
		DJ	4	4	3	3	1	3	3	3	4
		PL	4	3	3	1	2	2	2	3	3
		Mean	4.3	3.3	3.0	1.7	1.3	2.3	2.3	3.0	3.3
0	M	AC	5	5	2	1	1	2	3	3	3
		DJ	4	3	2	1	1	2	3	4	3
		PL	5	3	2	2	1	2	1	2	2
		Mean	4.7	3.7	2.0	1.3	1.0	2.0	2.3	3.0	2.7
	F	AC	5	2	3	2	1	1	2	3	3
		DJ	5	4	3	1	2	2	2	2	3
		PL	5	2	3	1	1	2	1	3	3
		Mean	5.0	2.7	3.0	1.3	1.7	1.7	1.7	2.7	3.0
-3	M	AC	5	3	3	2	1	2	2	2	2
		DJ	5	2	3	1	1	2	1	3	3
		PL	5	2	2	1	1	2	2	2	2
		Mean	5.0	2.3	2.7	1.3	1.0	2.0	1.7	2.3	2.3
	F	AC	5	3	3	1	2	3	2	4	3
		DJ	5	4	4	2	1	2	2	2	2
		PL	5	2	3	1	1	2	1	3	3
		Mean	5.0	3.0	3.3	1.3	1.3	2.3	1.7	3.0	2.7

TABLE 5.5. Experiment III-B: Mean value of perceptual ranking of speech processed by combined splitting

SNR dB	Speaker M/F	Su	Sp_CS-T_c/m				
			T_c =20	T_c =50	T_c =80	T_c =100	×
∞	M	1.0	1.7	1.3	1.7	2.7	m =8
			2.0	1.3	1.3	2.7	m =16
	F	1.0	2.0	2.0	2.0	3.0	m =8
			1.7	1.7	1.7	3.0	m =16
6	M	1.3	2.3	2.3	1.7	2.3	m =8
			2.3	2.3	1.3	2.7	m =16
	F	1.7	2.7	1.3	1.3	2.3	m =8
			2.3	1.3	1.7	2.0	m =16
3	M	3.7	3.0	1.7	2.0	1.7	m =8
			2.7	1.3	2.0	2.0	m =16
	F	4.3	3.3	1.7	2.3	3.0	m =8
			3.0	1.3	2.3	3.3	m =16
0	M	4.7	3.7	1.3	2.0	3.0	m =8
			2.0	1.0	2.3	2.7	m =16
	F	5.0	2.7	1.3	1.7	2.7	m =8
			3.0	1.7	1.7	3.0	m =16
-3	M	5.0	2.3	1.3	2.0	2.3	m =8
			2.7	1.0	1.7	2.3	m =16
	F	5.0	3.0	1.3	2.3	3.0	m =8
			3.3	1.3	1.7	2.7	m =16

In the second set of listening tests, a speech passage, recorded from a male and a female speaker, was used as test material. Processing conditions with combinations of T_c = 20, 50, 80, 100 ms and m = 8, 16 were used. Perceptual quality ranking was obtained through paired comparisons. Further hearing loss was simulated by adding broadband noise with constant short-time SNR. At low SNR values, dichotic listening under all processing conditions were ranked higher than diotic listening of unprocessed speech. For both the speakers, the highest ranking was obtained for T_c = 50 ms with not much difference between rankings for m = 8 and 16. Thus we can conclude that under adverse listening conditions the perceptual quality of dichotic presentation with combined splitting was ranked much higher over the unprocessed speech, and T_c = 50 – 80 ms with m = 8 or 16 may be considered as the most appropriate processing condition.

An investigation for evaluation of dichotic presentation with the three types of splitting, i.e. spectral, temporal, and combined splitting with appropriate set of processing parameters was taken up next and this is presented in the following chapter.

Chapter 6

EVALUATION OF THE DICHOTIC PRESENTATION SCHEMES

6.1 Introduction

The implementation and evaluation of spectral splitting scheme with perceptually balanced comb filters based on auditory critical bands has been presented in Chapter 4. Results from listening tests demonstrate that spectral splitting scheme improved speech reception for persons with moderate bilateral loss. Cascading of adjustable magnitude response filter with gain adjustable between +3 dB and −3 dB to partly compensate for the frequency dependent shifts in hearing threshold provided additional improvement. Earlier work (Jangamashetti, 2003; Jangamashetti and Pandey, 2000b), reviewed in Section 3.4, has shown that temporal splitting using trapezoidal fading functions with inter-aural switching period of 20 ms provided improvement by reducing the effect of temporal masking. In spectral splitting, sensory cells corresponding to alternate bands are always relaxed; while in temporal splitting, all the sensory cells of the two ears are alternately relaxed. A combined splitting scheme (Jangamashetti *et al.*, 2001; Pandey *et al.*, 2001) is devised to have spectral and temporal splitting simultaneously, so as to provide periodic relaxation from stimulation to all the sensory cells of the basilar membrane, and hence reduce the effects of spectral and temporal masking. The implementation of the combined splitting scheme was done, using a pair of time-varying comb filters with pre-calculated sets of coefficients, which were selected in steps for a cyclic sweeping of magnitude responses. An evaluation carried out to select an appropriate value for the sweep cycle and the number of steps in sweeping the magnitude responses, is presented in Chapter 5.

This chapter deals with an overall evaluation of the three schemes with different processing parameters, which was carried out in two phases. In the first phase of the evaluation, listening tests were conducted on normal hearing subjects with simulated hearing loss. In the second phase, listening tests were conducted on subjects with bilateral sensorineural hearing impairment. These two sets of listening tests are referred to as Experiment IV and V respectively.

6.2 Speech processing schemes

In the investigation presented in Chapter 4, the evaluation of spectral splitting scheme using comb filters with sharp transitions and perceptually balanced comb filters, on normal hearing and bilateral sensorineural hearing impaired subjects, has established that speech perception with perceptually balanced comb filters, are superior to comb filters with sharp transitions in all respects. Hence, in the present evaluation, spectral splitting was considered only with perceptually balanced comb filters.

In the earlier investigation on temporal splitting scheme (Jangamashetti, 2003, Jangamashetti and Pandey, 2000b), reviewed in Chapter 3, evaluation was carried out only for inter-aural switching period of 20 ms. Inter-aural switching period of 20 ms was decided, since speech subsegments with important acoustic cues are of the order of about 20 ms. Trapezoidal fading functions with 70 % duty cycle, with transition durations of 2 and 3 ms resulted in higher improvement. Significant temporal masking effects extend over 10 –100 ms, as described earlier in Sections 2.4.3 and 5.3. Further, evaluation of combined splitting scheme with different sweep cycles, presented in Chapter 5, showed best perceptual quality for speech processed with sweep cycle duration T_c = 50 – 80 ms. This emphasized the need for investigating the effect of different inter-aural switching periods in temporal splitting. Hence in the present investigation, temporal splitting scheme with 70 % duty cycle and 3 ms transition duration was evaluated for different inter-aural switching periods in the range of 20 – 80 ms.

The combined splitting scheme using time-varying comb filters, evaluated for sweep cycle of 20 ms earlier (Jangamashtti, 2003, Jangamashetti *et al.*, 2001; Pandey *et al.*, 2001), showed more improvement for 4 and 8 shiftings. Further our investigation, presented in Chapter 5, carried out to select the appropriate values for the sweep cycle and number of shiftings, showed best results for sweep cycles T_c = 50 – 80 ms and number of shiftings m = 8 and 16. In this investigation it was decided to carry out the evaluation for combined splitting scheme for different sweep cycle durations of T_c = 20 – 160 ms, along with number of shiftings m = 4, 8, and 16.

The three speech processing schemes evaluated in Experiment IV by conducting listening tests on persons with normal hearing are:

(i) *Spectral splitting scheme with perceptually balanced comb filters (referred as Sp_SS).* Figure 4.8 shows a schematic diagram of the spectral splitting scheme. The magnitude response of the perceptually balanced comb filter designed with 256 coefficients, passband ripple < 1 dB, stop-band attenuation > 30 dB, transition width 78 – 117 Hz, and band crossovers lying between –4 and –6 dB, is shown in Fig 4.6. A spectrographic representation of magnitude response of the comb filters is given in Fig. 6.1. This can be visualized as the spectrogram of the two output signals, for an input signal with uniform power spectral density.

(ii) *Temporal splitting scheme with trapezoidal fading functions of 70 % duty cycle and 3 ms transition duration for different inter-aural switching periods "T_c" (referred as Sp_TS-T_c).* A schematic representation of the temporal splitting scheme and trapezoidal fading functions with inter-aural switching period of T_c are shown in Fig 3.1. A spectrographic representation of the scheme using trapezoidal fading function with inter-aural switching period of 20 ms, duty cycle of 70 % and transition duration of 3 ms, is given in Fig. 6.2.

(iii) *Combined splitting using time-varying comb filters with different sweep cycle durations T_c and number of shiftings m (referred as Sp_CS-T_c/m).* A schematic diagram of the combined splitting scheme and the magnitude response of the time-varying comb filter are shown in Fig. 5.1 (a) and (b) respectively. Figure 6.3 shows a spectrographic representation of magnitude response of the pair of time-varying comb filters with 16 shiftings in a sweep cycle of 50 ms, for a duration of 1.5 sweep cycle.

The wideband spectrograms ($\Delta f \approx$ 300 Hz) of swept sine wave as input signal for the above three processing schemes (with a specific set of parameters) are shown in Figs 6.4, 6.5, and 6.6 respectively. Spectrograms for random noise and speech syllable /*asa*/ for the same set of parameters are shown in Figs 6.7 – 6.9 and 6.10 – 6.12 respectively.

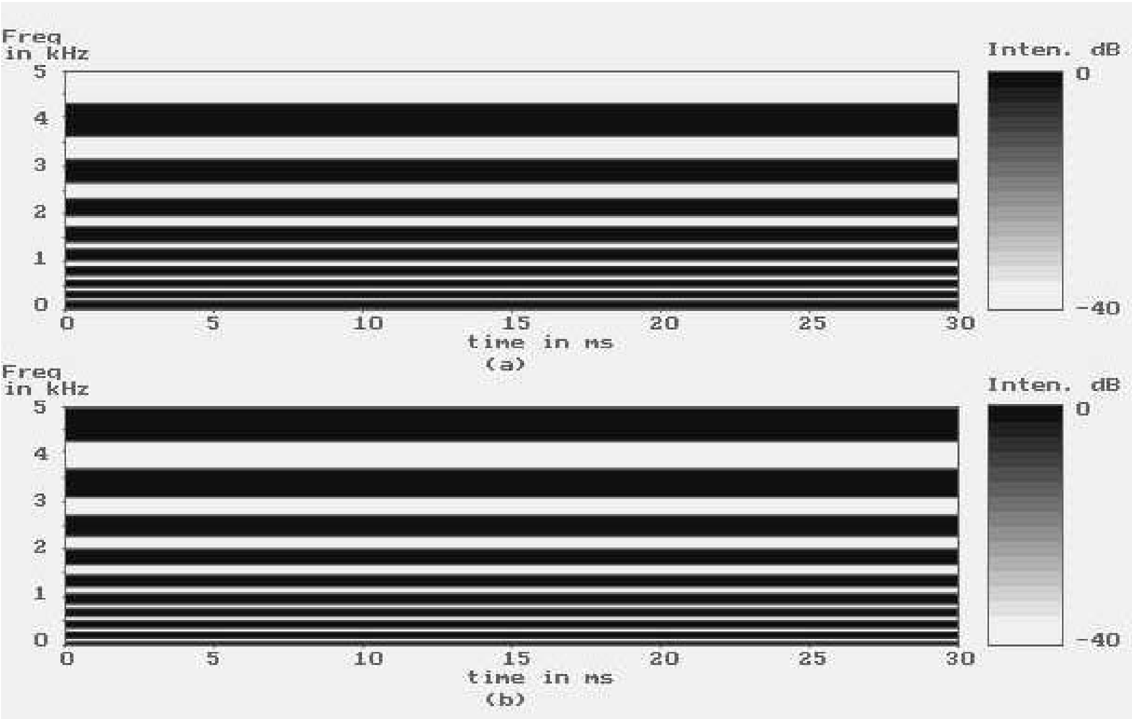


FIG. 6.1. A spectrographic representation of magnitude response of the comb filters using in spectral splitting (a) for the left ear and (b) for right ear.

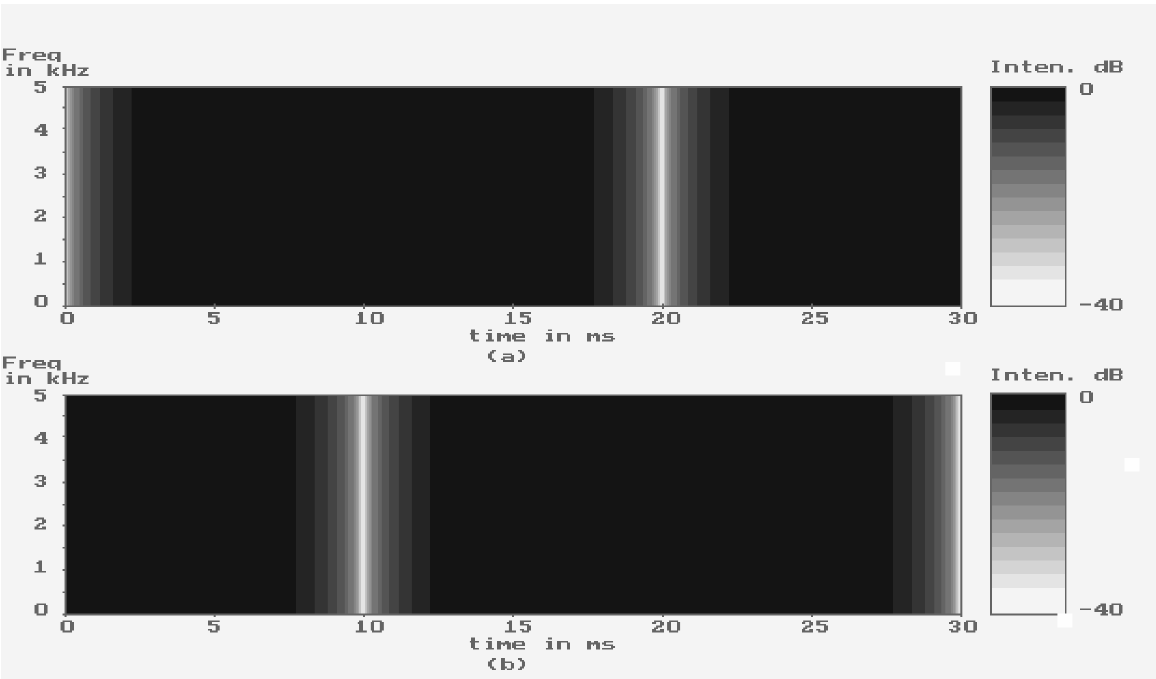


FIG. 6.2. A spectrographic representation of the temporal splitting scheme using trapezoidal fading function with inter-aural switching period of 20 ms, duty cycle of 70 % and transition duration of 3 ms.

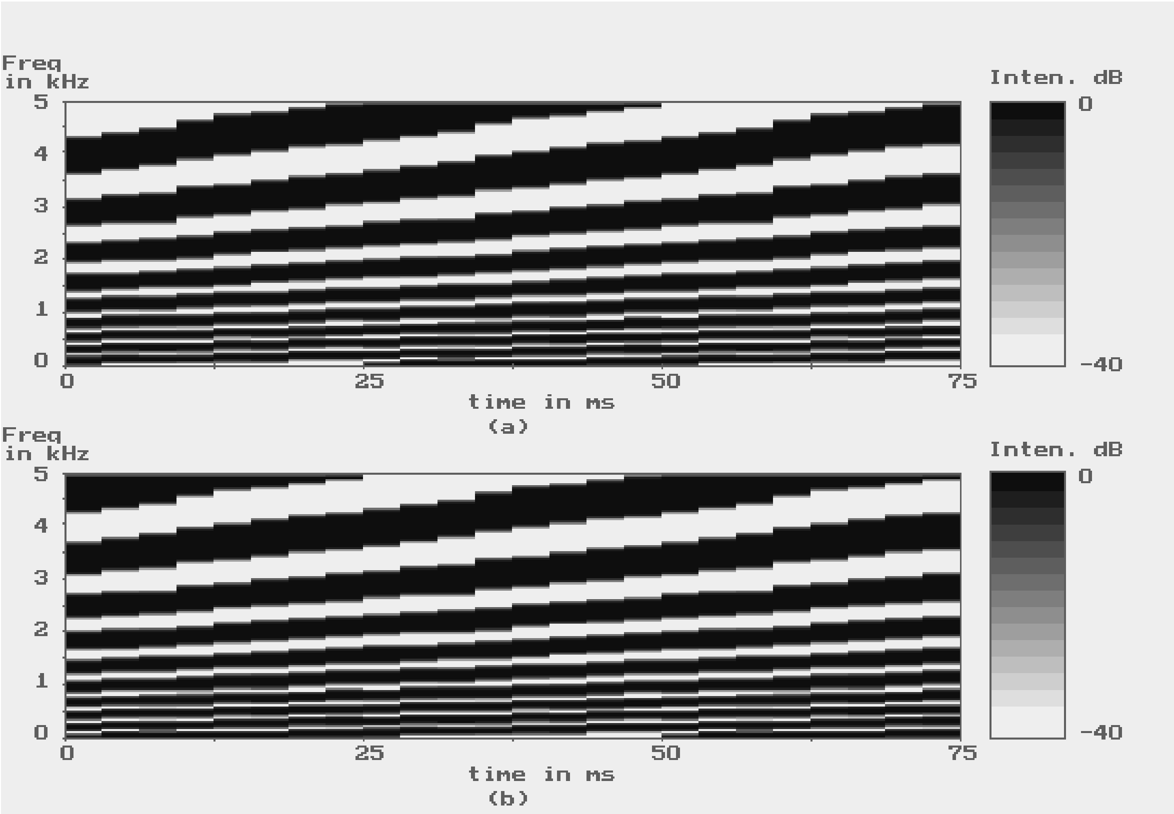


FIG. 6.3. A spectrographic representation of magnitude response of the pair of time-varying comb filters with 16 shiftings and 50 ms sweep cycle (a) for the left ear and (b) for right ear.

It has been established that use of adjustable magnitude response filters contributes improvement in speech perception and does not affect the reduction in spectral masking (Chaudhari, 2000; Chaudhari and Pandey, 1999b; Lunner *et al.*, 1993). In our evaluation in Experiment II, presented in Chapter 4, it has been found that cascading of adjustable magnitude response filter with perceptually balanced comb filters provided more improvement over perceptually balanced comb filters used alone. It was decided to also carry out the evaluation of adjustable magnitude response filter to partly compensate for the frequency dependent shifts in hearing thresholds. The frequency compensation was restricted between +3 dB and −3 dB, as given in Section 4.5.3, used in Experiment II with perceptually balanced comb filters, keeping in view the limited dynamic range of the hearing impaired subjects.

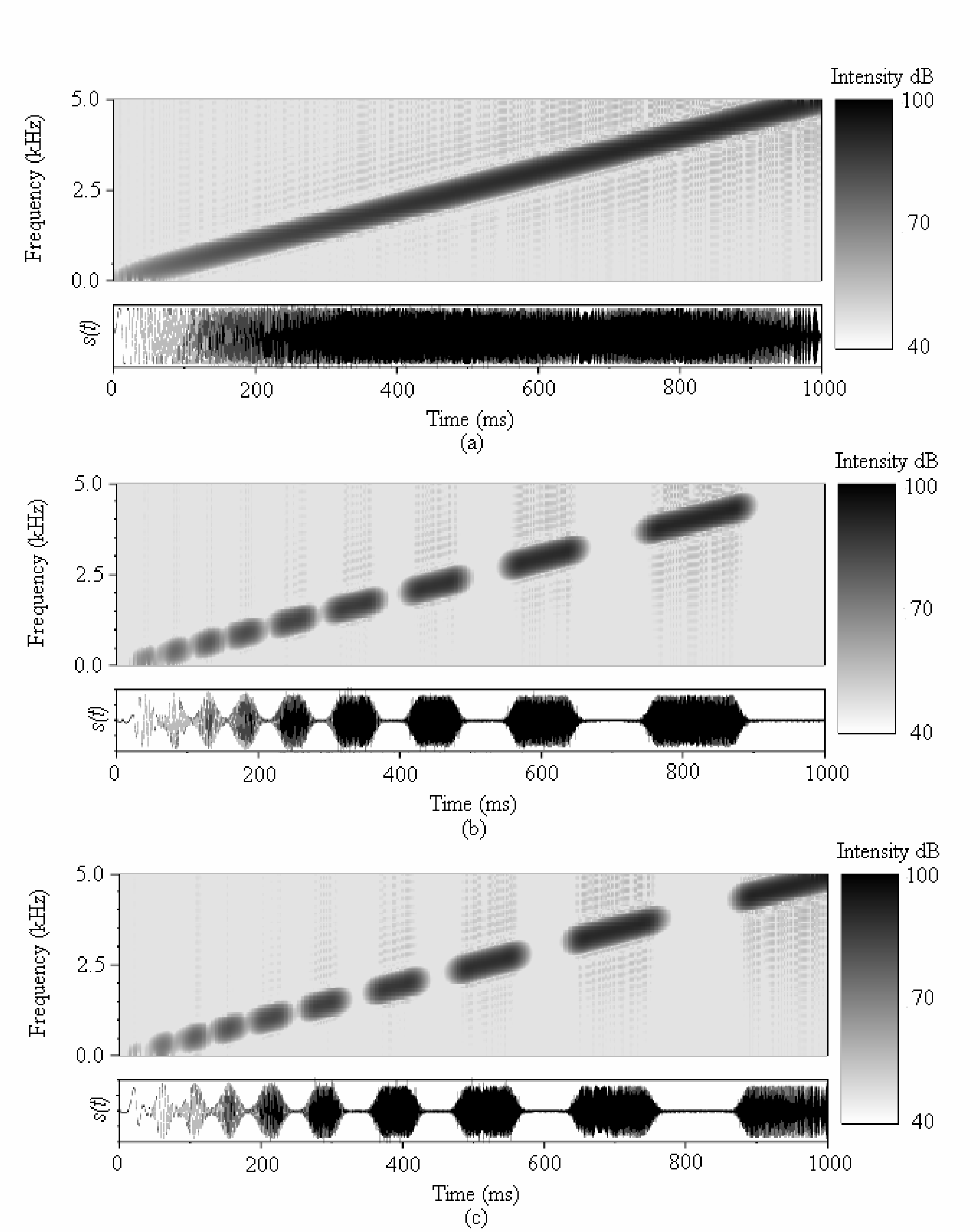


FIG. 6.4. Spectral splitting with perceptually balanced comb filters (Sp_SS). Wideband spectrogram ($\Delta f \approx 300$ Hz) of swept sine wave. (a) unprocessed, (b) processed right ear, and (c) processed left ear.

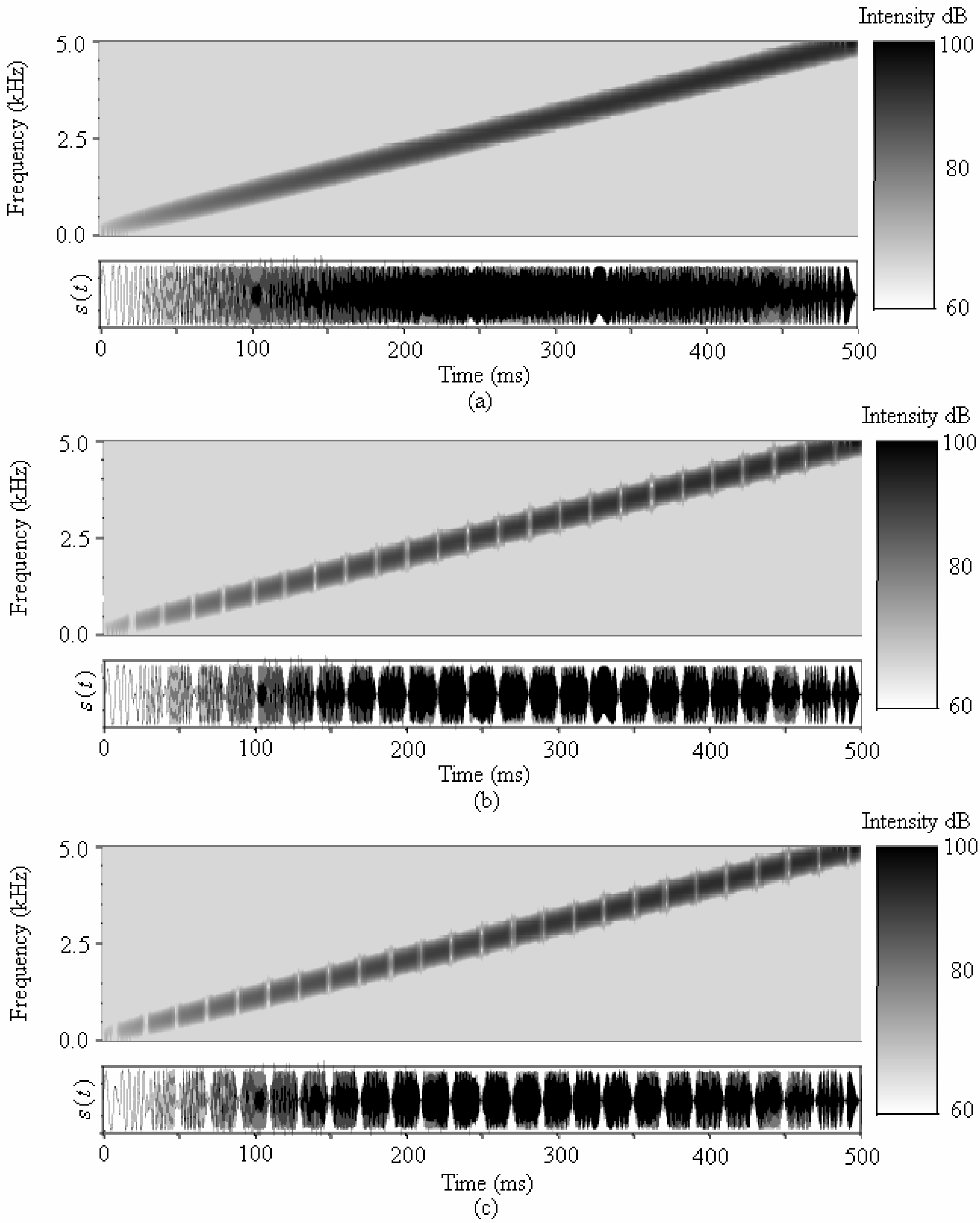


FIG. 6.5. Temporal splitting with trapezoidal fading functions, inter-aural switching period = 20 ms, duty cycle = 70 %, and transition duration = 3 ms (Sp_TS-20). Wideband spectrogram ($\Delta f \approx 300$ Hz) of swept sine wave. (a) unprocessed, (b) processed right ear, and (c) processed left ear.

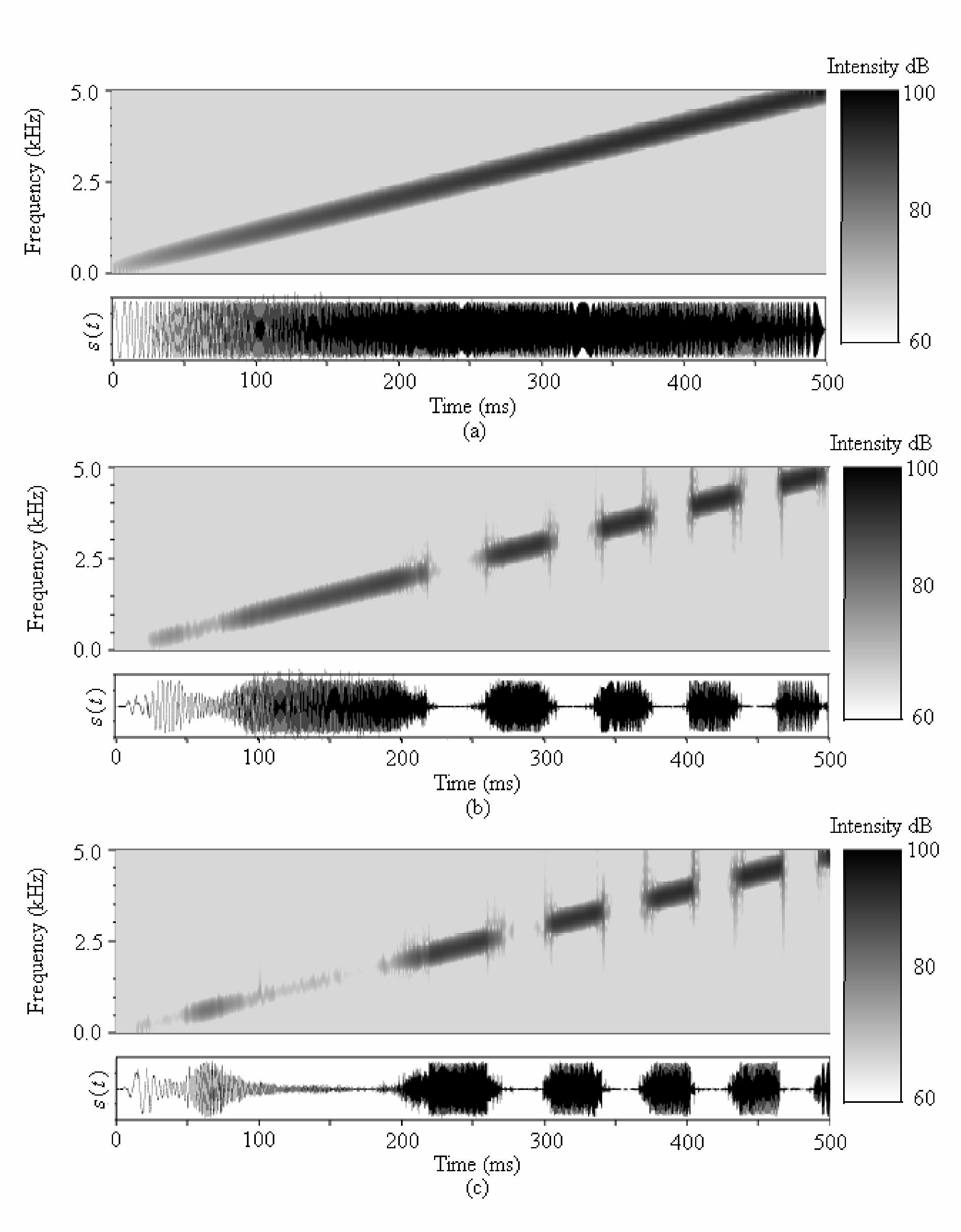


FIG. 6.6. Combined splitting with time-varying comb filters, sweep cycle duration = 40 ms and number of shiftings = 16 (Sp_CS-40/16). Wideband spectrogram ($\Delta f \approx 300$ Hz) of swept sine wave. (a) unprocessed, (b) processed right ear, and (c) processed left ear.

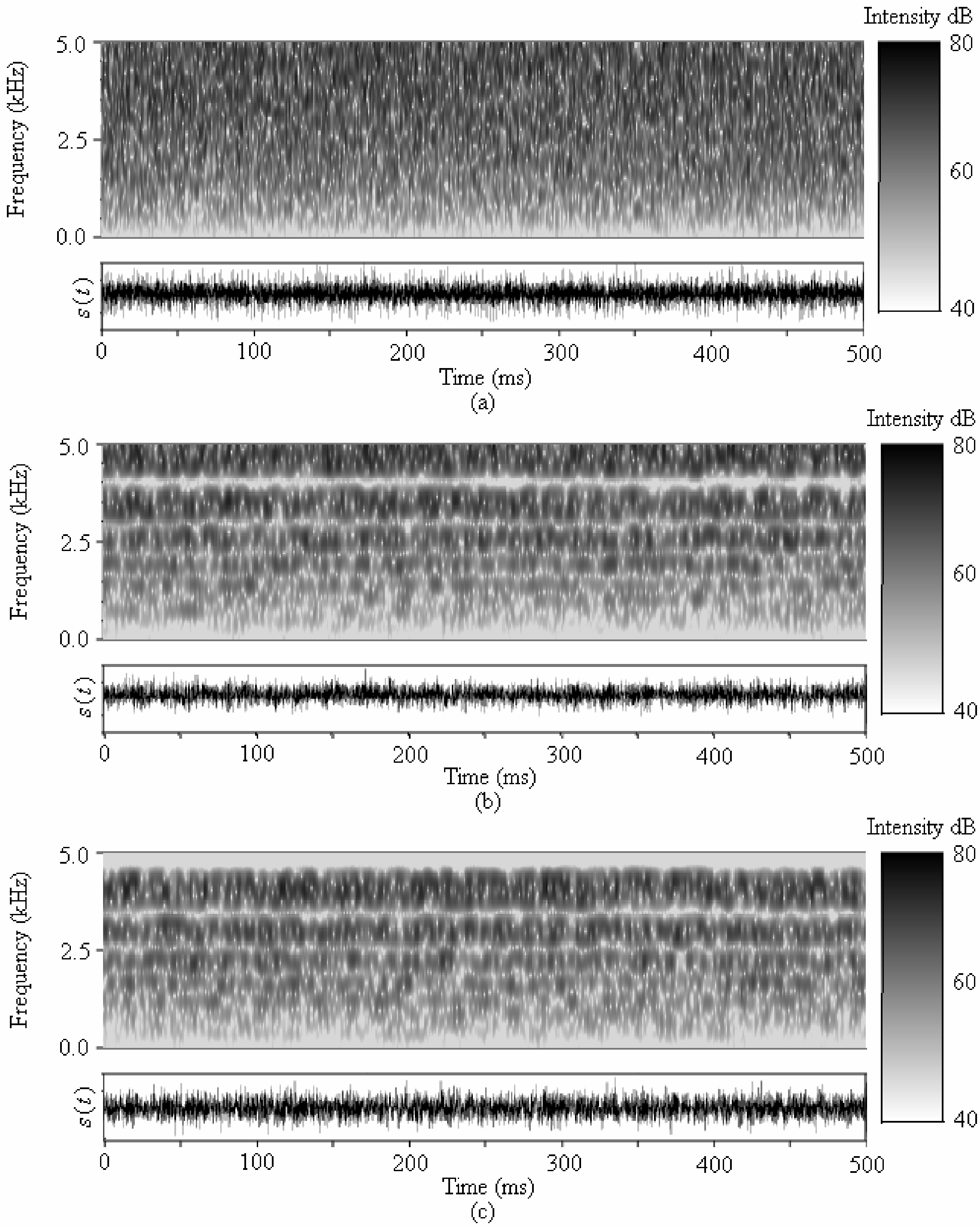


FIG. 6.7. Spectral splitting with perceptually balanced comb filters (Sp_SS). Wideband spectrogram ($\Delta f \approx 300$ Hz) of random noise (a) unprocessed, (b) processed right ear, and (c) processed left ear.

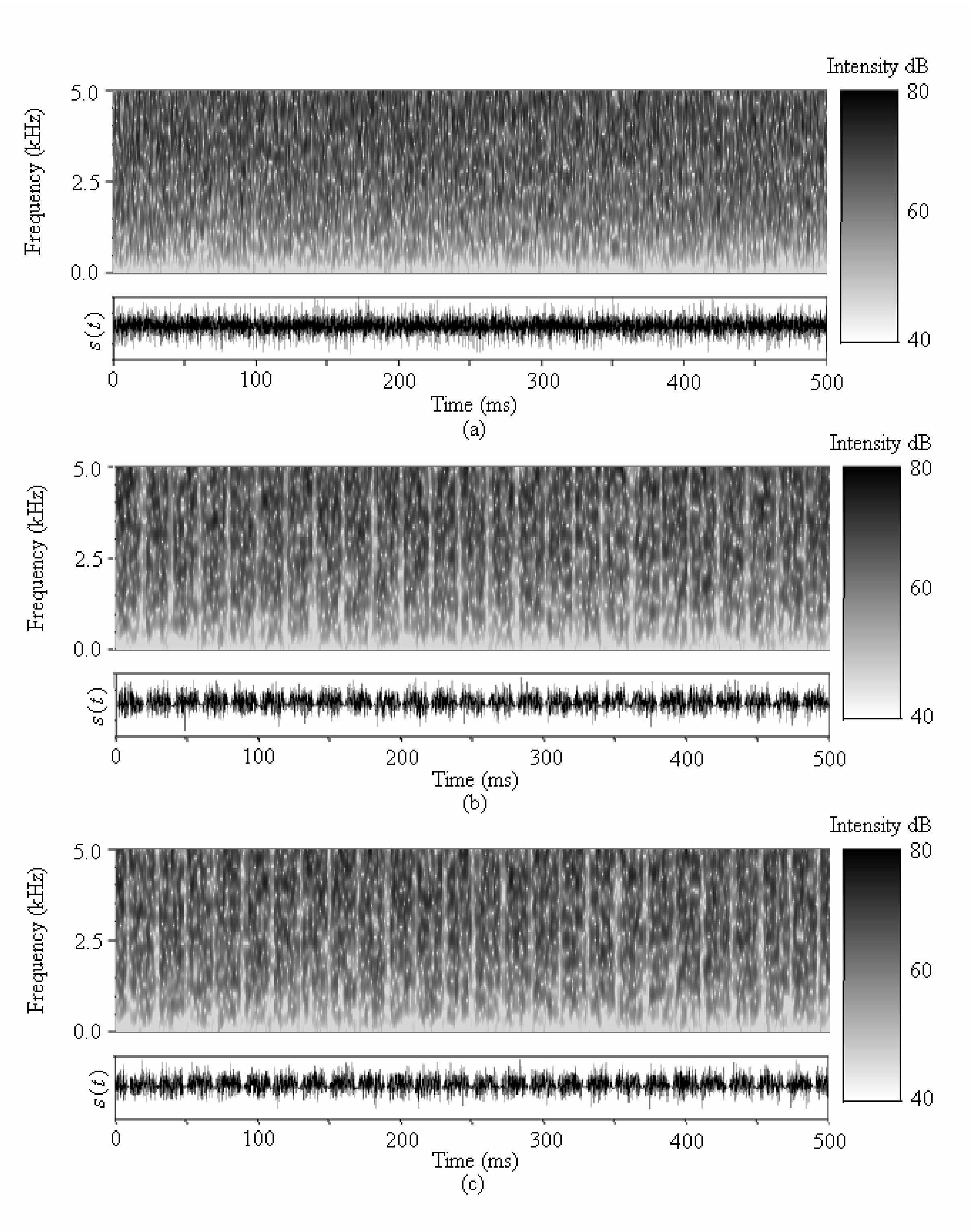


FIG. 6.8. Temporal splitting with trapezoidal fading functions, inter-aural switching period = 20 ms, duty cycle = 70 %, and transition duration = 3 ms (Sp_TS-20). Wideband spectrogram ($\Delta f \approx 300$ Hz) of random noise (a) unprocessed, (b) processed right ear, and (c) processed left ear.

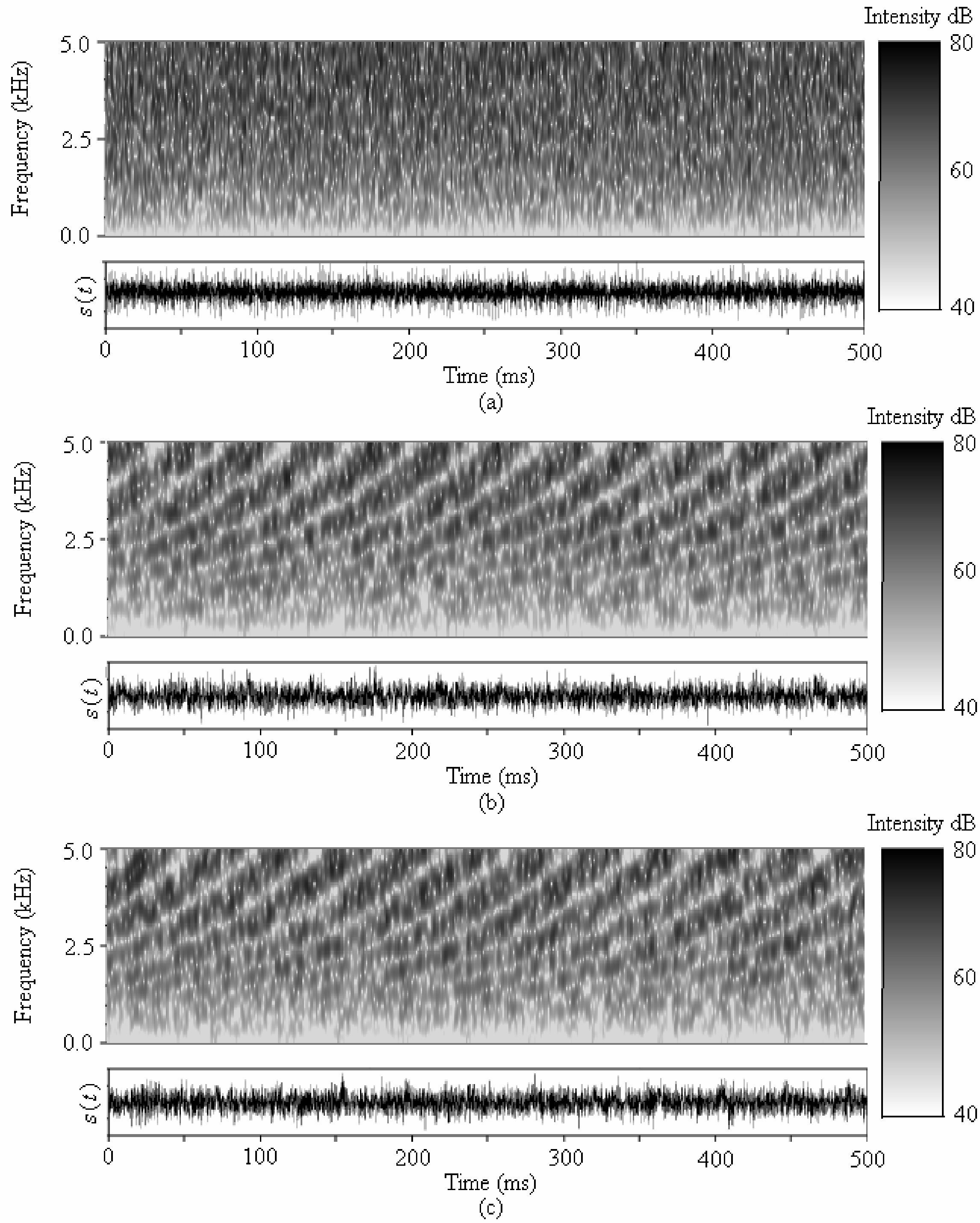


FIG. 6.9. Combined splitting with time-varying comb filters, sweep cycle duration = 40 ms and number of shiftings = 16 (Sp_CS-40/16). Wideband spectrogram ($\Delta f \approx$ 300 Hz) of random noise (a) unprocessed, (b) processed right ear, and (c) processed left ear.

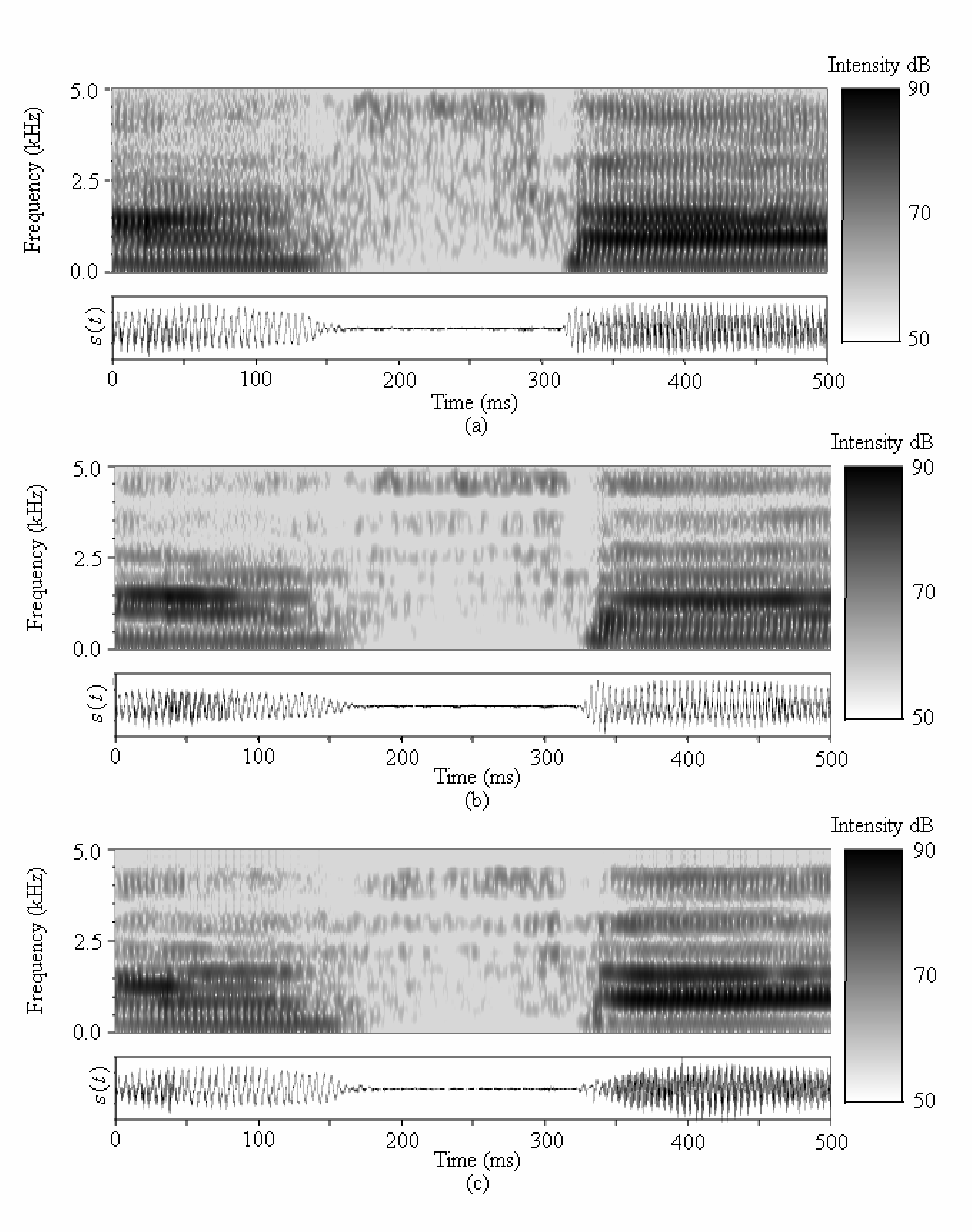


FIG. 6.10. Spectral splitting with perceptually balanced comb filters (Sp_SS). Wideband spectrogram ($\Delta f \approx 300$ Hz) of speech syllable /*asa*/ (a) unprocessed, (b) processed right ear, and (c) processed left ear.

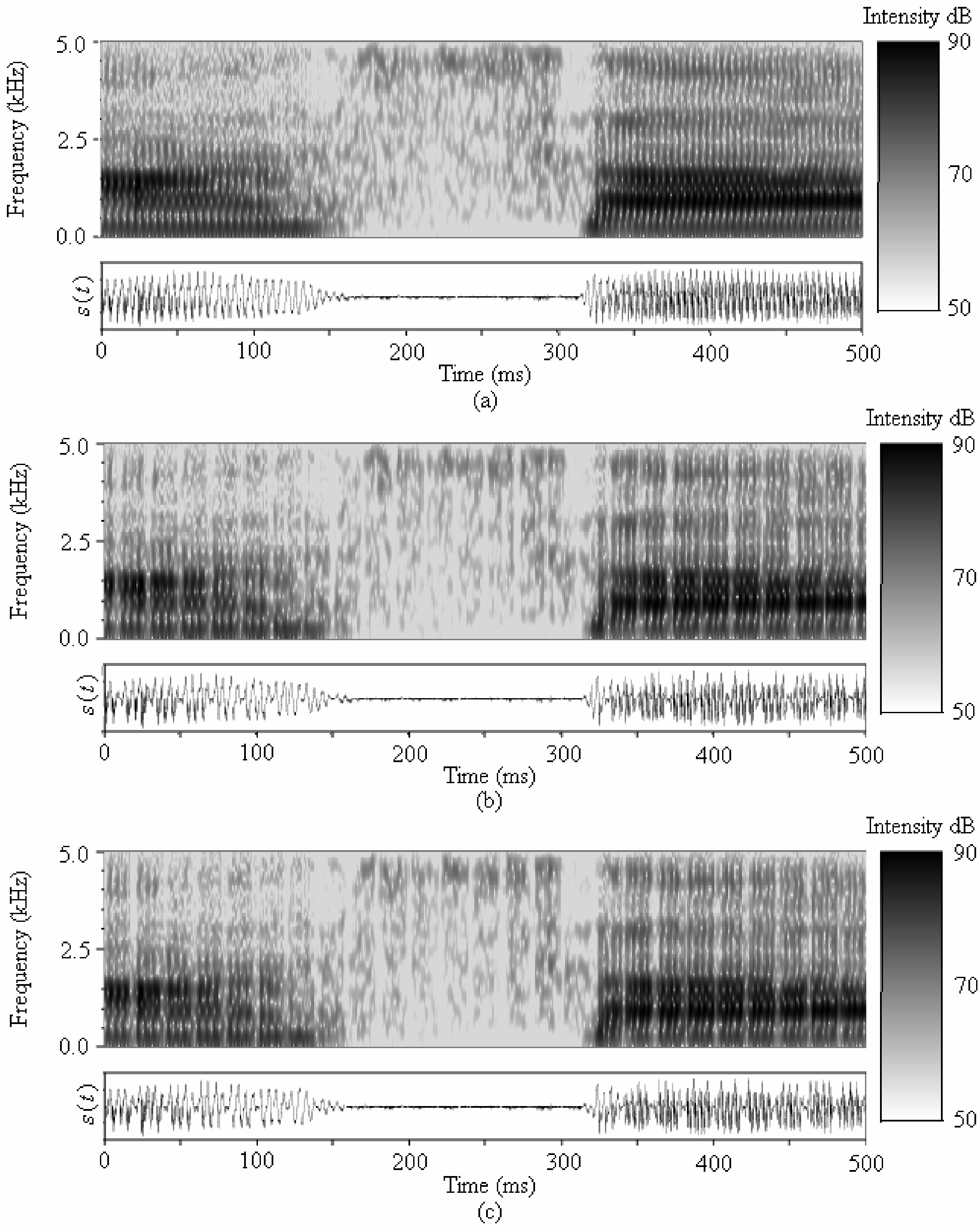


FIG. 6.11. Temporal splitting with trapezoidal fading functions, inter-aural switching period = 20 ms, duty cycle = 70 %, and transition duration = 3 ms (Sp_TS-20). Wideband spectrogram ($\Delta f \approx 300$ Hz) of speech syllable /*asa*/ (a) unprocessed, (b) processed right ear, and (c) processed left ear.

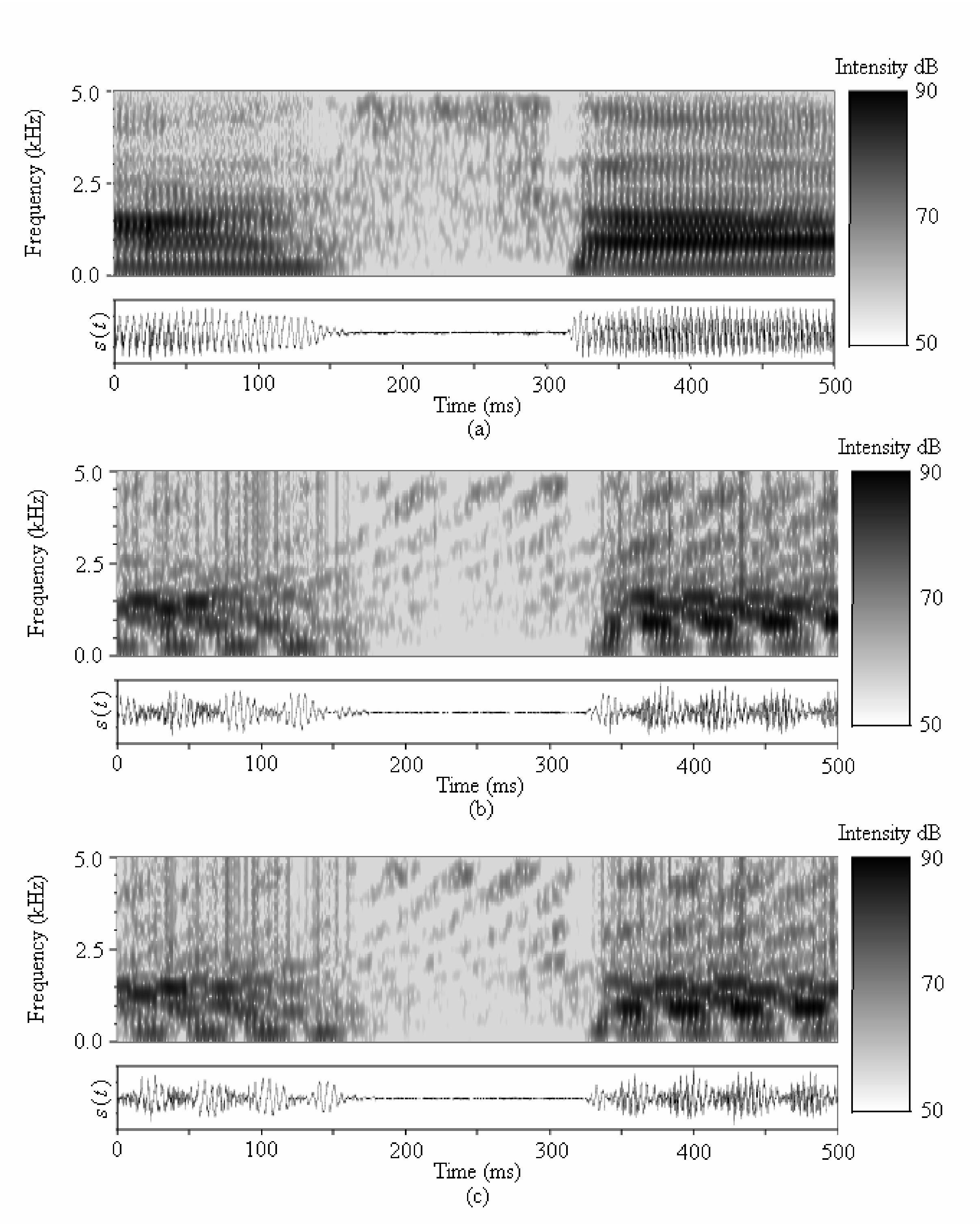


FIG. 6.12. Combined splitting with time-varying comb filters, sweep cycle duration = 40 ms and number of shiftings = 16 (Sp_CS-40/16). Wideband spectrogram ($\Delta f \approx 300$ Hz) of speech syllable /*asa*/ (a) unprocessed, (b) processed right ear, and (c) processed left ear.

In the listening tests with subjects with bilateral loss, the filters with adjustable magnitude response were used by themselves and also in cascade with the three dichotic schemes. Thus the speech processing schemes evaluated in Experiment V are:

(i) *Filters with adjustable magnitude response to partly compensate for the shifts in hearing thresholds, for right and left ear separately (referred as Sp_AG).* Figure 6.13 shows the schematic diagram of the scheme. Each filter is a 256-coefficient linear phase FIR filter, designed using frequency sampling techniques, as described in Appendix C. An example of one subject's pure tone audiogram and the desired and actual magnitude response used for binaural dichotic presentation are shown earlier in Fig. 4.7. Audiograms and corresponding compensating filter responses for various hearing impaired subjects are given in Appendix G.

(ii) *Filters with adjustable magnitude response cascaded with spectral splitting scheme using perceptually balanced comb filters (referred as Sp_AG-SS).* Figure 6.14 shows a schematic diagram of the cascaded scheme. Spectral splitting is the same as that used in Experiment IV.

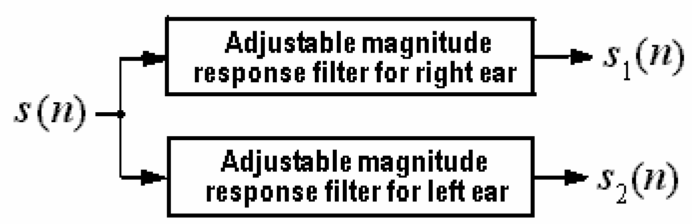


FIG. 6.13. A Schematic representation of adjustable magnitude response filter (Sp_AG), used to partly compensate for the frequency dependent shifts in hearing thresholds. Input signal: $s(n)$, and outputs to the two ears: $s_1(n)$ and $s_2(n)$.

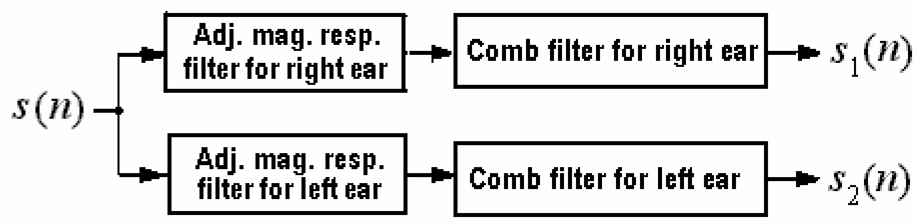


FIG. 6.14. A Schematic representation of scheme of adjustable magnitude response filter cascaded with spectral splitting (Sp_AG-SS). Input signal: $s(n)$, and outputs to the two ears: $s_1(n)$ and $s_2(n)$.

(iii) *Filters with adjustable magnitude response cascaded with temporal splitting scheme with trapezoidal fading functions for different inter-aural switching periods T_c (referred as Sp_AG-TS-T_c).* Figure 6.15 shows a schematic diagram of the cascaded scheme. Temporal splitting is the same as that used in Experiment IV.

(iv) *Filters with adjustable magnitude response cascaded with combined splitting scheme with time-varying comb filters for different sweep cycle durations T_c and number of shiftings m (referred as Sp_AG-TS-T_c/m).* Figure 6.16 shows a schematic diagram of the cascaded scheme. Combined splitting is the same as that used in Experiment IV.

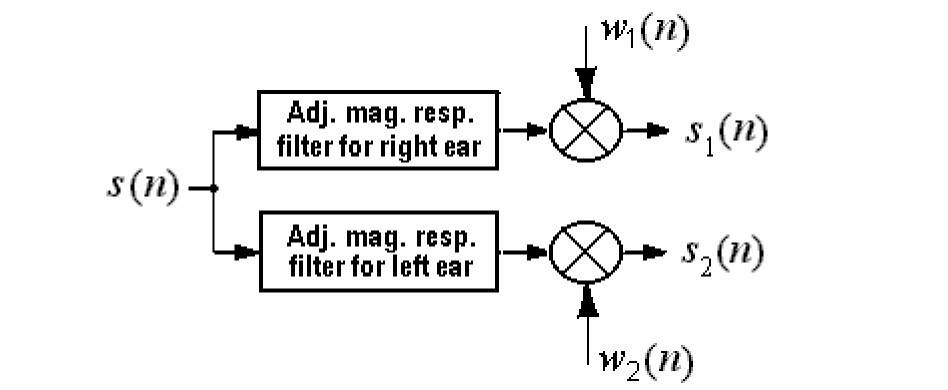


FIG. 6.15. A Schematic representation of scheme of adjustable magnitude response filter cascaded with temporal splitting (Sp_AG-TS-T_c). Input signal: $s(n)$, and outputs to the two ears: $s_1(n)$ and $s_2(n)$.

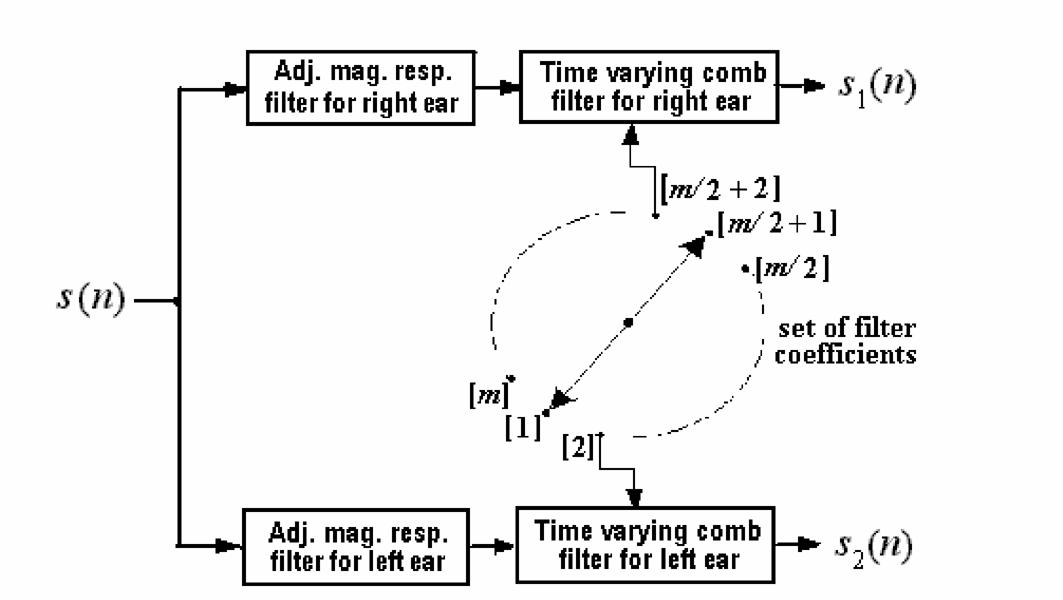


FIG. 6.16. A Schematic representation of scheme of adjustable magnitude response filter cascaded with combined splitting (Sp_AG-CS-T_c/m). Input signal: $s(n)$, and outputs to the two ears: $s_1(n)$ and $s_2(n)$.

6.2.1 Listening tests

Evaluation of the schemes was carried out in two phases. The first phase, referred as Experiment IV, involved listening tests conducted on normal hearing subjects with simulated hearing loss for the schemes of spectral, temporal, and combined splitting, for the different processing parameters for comparing with unprocessed speech. The scheme of spectral splitting was evaluated with perceptually balanced comb filters, and this is labeled as Sp_SS. The temporal splitting scheme with 70 % duty cycle and 3 ms transition duration was evaluated for inter-aural switching periods T_c = 20, 40, 80 ms, and these are termed as Sp_TS-20, Sp_TS-40 and Sp_TS-80 respectively. The scheme of combined splitting was carried out for sweep cycle T_c = 20, 40, 80 ms, each with shiftings m = 4, 8, 16 (Sp_CS-20/4, Sp_CS-20/8, Sp_CS-20/16, Sp_CS-40/4, Sp_CS-40/8, Sp_CS-40/16, Sp_CS-80/4, Sp_CS-80/8, Sp_CS-80/16) and T_c = 120, 180 ms with m = 8, 16 (Sp_CS-120/8, Sp_CS-120/16, Sp_CS-160/8, Sp_CS-160/16).

Second phase of evaluation, referred as Experiment V, involved listening tests with hearing impaired subjects. Since the usefulness of adjustable magnitude response filter of the type mentioned above, has been established already, it was decided to cascade the adjustable magnitude response filter with the schemes of spectral, temporal and combined splitting schemes, while carrying out listening tests on hearing impaired subjects. After the evaluation on normal hearing subjects (Experiment IV), the processing parameters with no or very small improvements were omitted. The processing parameters considered in Experiment V were spectral splitting with perceptually balanced comb filters (Sp_AG-SS), temporal splitting with T_c = 20, 40 ms (Sp_AG-TS-20, Sp_AG-TS-40), and combined splitting with T_c = 20, 40, 80 ms, each with shiftings m = 8, 16 (Sp_AG-CS-20/8, Sp_AG-CS-20/16, Sp_AG-CS-40/8, Sp_AG-CS-40/16, Sp_AG-CS-80/8, Sp_AG-CS-80/16). Scheme of adjustable magnitude response filter Sp_AG, also was evaluated.

6.2.2. Test material

For evaluation of spectral splitting schemes, presented in Chapter 4, listening tests were conducted using 12 nonsense syllables in VCV context. The assessment by closed set evaluation using nonsense syllables may overestimate the ability of hearing impaired persons to perceive conversational speech. Listening tests involving presentation of the nonsense syllables with a large number of processing conditions cause irritation and fatigue in the subjects. Hence for the present investigation, an open set evaluation with sets of phonetically balanced (PB) monosyllabic words in three languages (Hindi, Marathi, and English) were used. In each of these word lists there were 50 to 60 monosyllabic words. The lists were obtained from Ali Yaver Jung National Institute for Hearing Handicapped (AYJNIHH), Mumbai, where these monosyllables are used to evaluate the discrimination capacity of hearing impaired persons. The lists of monosyllables used are given in Appendix F (Tables F.1, F.2, and F.3). The subjects were allowed to choose one of these languages for carrying out the listening tests. In addition to subject response, response time was also recorded. Recognition score and average response time, and their relative improvements were used for the evaluation of the schemes.

For recording each set of nonsense syllables, the speaker was instructed to speak at normal conversational level, and the distance of the microphone was adjusted such that the sound level meter indicated 70 – 75 dB SPL. The analog signal (obtained using the microphone and ac output of sound level meter B&K 2235 which has an in-built preamplifier, input attenuator, weighting filter, and buffer amplifier for ac output), was passed through an amplifier, anti-aliasing filter with low pass cut-off frequency of 4.6 kHz, connected to the line-input of the PC sound card, and was digitized at 10 kSa/s with 16-bit resolution. All the words had approximately the same intensity. The details of the speech acquisition set-up are given in Appendix D.

6.2.3 Experimental set-up

For listening tests, we have used a set-up based on two audio line outputs of a Notebook PC. The line-outputs of the sound card were connected to the headphones via two audio amplifiers. The set-up details are given in Appendix D. The subject was seated in an acoustically isolated room and a talk-back microphone was used to get the verbal response from the subject to the experimenter, sitting outside. The subject repeated what he/she has heard and the experimenter compared the response with the word presented and keyed in ‘y’ or ‘n’. Thus the response was recorded as correct/incorrect, and reaction time was also recorded. It is to be noted that recorded reaction time included the reaction time of the experimenter, in addition to that of the subject.

6.2.4 Experimental method

Before the commencement of presentations, the subject was given a copy of the list of PB word list to be presented. Initially the words were presented one after the other in the order given in the list with a fixed inter-presentation interval decided by the subject, for listening practice. Listening practice was carried out as many times as opted by the subject. The subject could also listen to specific words repeatedly. Once the subject was familiar with the words, the listening tests involving randomized presentation of words commenced.

One listening test consisted of presentation of each word of the list three times. The number of consecutive presentations for a word was restricted to two. For maintaining mid-range uniformity, in M consecutive presentations, any stimulus should not occur more than $M/n+2$ times, where n is the number of words. The result of the test along with the percentage recognition scores and response times (average and standard deviation separately for correct and incorrect responses) were stored. The presentation and response recording was done using program “`stest.c`” described in Appendix E.

6.3 Experiment IV: Listening tests on normal hearing subjects with simulated loss

Seven normal hearing subjects participated in the listening tests (AU: M 18, JK: M 30, SM: M 24, VJ: M 24, ST: M 28, VS: M 21, and GA: M 21). The subjects were having pure-tone hearing threshold less than 20 dB in the hearing range of 125 to 6 kHz. Hearing loss was simulated by adding broadband noise to the speech stimuli, at constant short-time SNR of 10 ms, in the same manner as in Experiment I and described in Section 3.7.2. The simulation was done at noise levels of ∞, 3, 0, –3, –6, and –9 dB. Minimum level of SNR was decided so as to have less than 50 % recognition score for unprocessed speech for all the subjects.

Results of the listening tests conducted on normal hearing subjects with hearing loss simulated at different levels are shown in Tables 6.1 and 6.2, for response time and recognition score respectively.

6.3.1 Response time

The response times for unprocessed and processed speech under different processing conditions for the different SNR values are shown in Table 6.1. The average across seven subjects and standard deviation, percentage relative improvement in response time, and paired t-test significance levels (one-tailed) are also given. Figure 6.17 (a) and (b) show the average response times (s) and average relative improvements (%), for all the schemes under different processing conditions and SNR levels. For unprocessed speech, the average response time increased with decrease in SNR, from 2.09 s for no noise to 2.83 s for –9 dB SNR. For spectral splitting, there was an improvement in response time, and relative improvement increased with decreasing SNR. For temporal splitting, the response time degraded at high SNR levels but improved at low SNR levels. Combined splitting with cycle time of 20 and 40 ms has the same pattern as temporal splitting, while splitting with cycle time of 80 ms and higher showed a pattern similar to that of spectral splitting.

At –9 dB SNR, the average response times for all processing conditions were less than that for unprocessed speech. The relative improvement in response time at –9 dB SNR, was maximum for spectral splitting (14 %) and it was statistically significant (p = 0.05). Next to spectral splitting was combined splitting with sweep cycle of 40 ms and more. The highest improvement of 13.2 % ($p < 0.05$) was for 80 ms sweep cycle with 16 shiftings. At –6 dB SNR, the relative improvement in response time was more for combined splitting under various processing conditions and for spectral splitting, with the maximum for combined splitting for sweep cycle of 80 ms with 8 and 16 shiftings. The improvements, ranging 13-18 %, were statistically significant for combined splitting for sweep cycles of 80 ms and above for 8 and 16 shiftings, sweep cycle of 40 ms with 4 shiftings, and spectral splitting.

Decrease in response time shows decrease in the load on perception. Spectral splitting scheme provides maximum improvement. Temporal splitting at different inter-aural switching cycles provided less improvement compared to other processing schemes. Among the three inter-aural switching periods considered, 20 ms provided higher improvements at high SNR conditions whereas 40 ms was better at low SNR conditions. Combined splitting with sweep cycles of 80 ms and more provided statistically significant improvement in response time. At low SNR conditions sweep cycle of 80 ms provided higher improvements, for number of shiftings m = 8 and 16.

6.3.2 Recognition scores

Table 6.2 shows the recognition scores and its averages for unprocessed and processed speech for the three schemes for all the processing conditions tested on seven normal hearing subjects. The standard deviation, the relative improvement in recognition scores, and paired t-test significance levels (one-tailed) are also given. Figure 6.18 (a) and (b) show the average recognition scores and the average relative improvements for all the schemes under different processing conditions and SNR levels. For unprocessed speech at no noise condition, all the subjects had perfect or near perfect scores. The scores decreased with decreasing SNR, with increasing inter-subject variability. The average recognition score for unprocessed speech reduced from 99.8 % at no noise condition to 23.9 % at –9 dB SNR. At all SNR conditions, spectral splitting generally showed highest percentage relative improvement, with values of 10.9, 15.7, 30.1, 93.2, and 142.8 % for 3, 0, –3, –6, and –9 dB respectively. Combined splitting with sweep cycle T_c = 80 ms with 8 and 16 shiftings resulted in recognition scores close to spectral splitting. The percentage relative improvements for 8 and 16 shiftings were respectively, 10.7 and 9.2 % at 3 dB, 12.7 and 13.8 at 0 dB, 26.3 and 27.4 at –3 dB, 75.4 and 81.0 at –6 dB, and 130.2 and 118.2 at –9 dB. Some other conditions also gave good results: 3 dB SNR: T_c = 120, 160 ms with 8 and 16 shiftings (11 –12 %), 0 dB SNR: T_c = 120 ms with 16 shiftings (14.4 %), and –6 dB SNR: T_c = 160 ms with 8 shiftings (84.7 %). The results showed almost similar recognition scores for processing with 8 and 16 shiftings, both having higher scores than 4 shiftings in most of the cases.

Figure 6.19 shows the average recognition scores for unprocessed speech and for the three schemes with processing conditions with higher scores. For 60% recognition score, spectral splitting provided an advantage of approximately 5 dB in SNR. The advantage for temporal splitting with T_c = 20 and 40 ms was approximately 1.5 dB. Combined splitting resulted in an advantage of approximately 4 dB for T_c = 40 ms and 4.5 dB for T_c = 80 ms for both m = 8 and 16.

A two-way analysis of variance (ANOVA) on relative improvements (with reference to unprocessed speech) averaged across subjects, for SNR conditions and processing schemes and conditions (results given in Table H.3 in Appendix H), showed significant effect (p < 0.01) of processing and SNR conditions when all the processing schemes and conditions were considered. ANOVA was also carried out for temporal splitting with its 3 switching durations

TABLE 6.1 Experiment IV Response time (s) for Unprocessed speech (Su) and processed speech: spectral splitting Sp_SS, temporal splitting Sp_TS-T_c with inter-aural switching interval of T_c ms, and combined splitting Sp_CS-T_c/m with cycle time T_c ms and m shiftings for SNRs of (a) ∞, (b) 3, (c) 0, (d) −3, (e) −6, and (f) −9 dB. S: Subject, Avg. = average recognition scores, s.d. = standard deviation, R.I. = Average relative improvements (%) with respect to unprocessed. p: significance level (one-tailed) for paired t-test (unprocessed vs processed, averaged across the subjects, n =7, df =6).

(a) ∞ SNR

S	Su	Sp_SS	Sp_TS-			Sp_CS-													
			20	40	80	20/4	20/8	20/16	40/4	40/8	40/16	80/4	80/8	80/16	120/8	120/16	160/8	160/16
AU	2.69	2.53	2.56	2.46	2.48	3.04	2.95	2.73	2.49	2.60	2.45	2.30	2.26	2.09	1.78	1.69	1.83	1.76
JK	2.60	2.71	2.51	2.41	2.47	2.99	2.92	2.68	2.77	2.66	2.52	2.44	2.44	2.45	1.86	1.84	1.86	1.94
SM	1.85	1.98	1.91	1.80	1.95	2.14	2.13	2.13	2.01	2.04	1.91	2.01	1.83	1.82	1.88	1.87	1.82	1.80
VJ	1.75	1.86	1.80	1.79	1.77	2.08	1.94	2.20	1.99	1.93	1.97	1.92	1.86	1.78	1.95	1.78	1.75	1.88
ST	1.72	1.73	1.75	1.78	1.79	2.27	2.11	2.08	1.92	1.94	1.99	1.90	2.02	1.80	2.00	1.89	1.81	1.74
VS	1.77	1.75	1.81	1.96	1.98	2.13	2.56	2.07	1.96	2.07	1.87	1.85	1.82	1.91	1.73	1.72	1.76	1.90
GA	2.26	2.17	2.01	2.46	2.52	2.32	2.17	2.05	2.06	2.01	2.09	1.96	2.16	1.75	1.95	1.96	1.96	1.86
Avg.	2.09	2.10	2.05	2.09	2.14	2.42	2.40	2.28	2.17	2.18	2.11	2.05	2.06	1.94	1.88	1.82	1.83	1.84
s.d.	0.42	0.39	0.34	0.33	0.34	0.41	0.41	0.30	0.33	0.31	0.26	0.22	0.24	0.25	0.10	0.10	0.07	0.07
R.I.		-1.0	1.3	-1.0	-3.0	-16.8	-15.9	-10.6	-5.0	-5.5	-2.5	0.1	0.2	5.4	7.0	10.0	10.1	9.2
p		0.48	0.42	0.49	0.41	0.08	0.10	0.18	0.35	0.33	0.45	0.42	0.42	0.22	0.11	0.06	0.06	0.07

(b) 3 dB SNR

S	Su	Sp_ SS	Sp_TS-			Sp_CS-													
			20	40	80	20/4	20/8	20/16	40/4	40/8	40/16	80/4	80/8	80/16	120/8	120/16	160/8	160/16
AU	2.57	2.43	2.85	2.76	3.00	3.20	2.69	2.58	2.46	2.41	2.28	2.34	2.34	2.49	2.16	2.02	1.97	1.99
JK	2.46	2.40	2.46	2.49	2.70	2.42	2.52	2.45	2.42	2.60	2.42	2.56	2.42	2.62	1.93	1.89	1.93	1.87
SM	2.10	1.94	2.06	2.11	2.17	2.26	2.13	2.45	2.03	2.08	2.08	2.03	2.07	1.88	1.99	2.10	2.09	2.06
VJ	2.03	1.84	1.88	2.03	2.23	2.18	2.08	2.12	1.90	1.98	1.99	2.01	1.96	1.94	1.99	2.06	1.84	1.85
ST	1.99	1.82	2.02	2.10	2.22	2.26	2.16	2.09	1.95	1.95	2.01	1.93	1.85	1.81	1.88	1.91	1.89	1.89
VS	2.05	1.87	1.93	2.14	2.22	2.00	1.94	1.90	1.99	2.02	1.99	1.96	1.93	1.85	1.89	1.86	1.99	1.90
GA	1.94	1.86	1.91	2.05	2.13	2.14	2.14	2.08	1.83	2.07	1.93	1.83	1.85	1.83	1.98	1.96	1.93	2.03
Avg.	2.16	2.02	2.16	2.24	2.38	2.35	2.24	2.24	2.08	2.16	2.10	2.09	2.06	2.06	1.97	1.97	1.95	1.94
s.d.	0.25	0.27	0.36	0.28	0.33	0.40	0.27	0.25	0.25	0.25	0.18	0.26	0.23	0.34	0.09	0.09	0.08	0.08
R.I.		6.6	0.6	-3.5	-9.9	-8.5	-3.5	-3.7	3.8	0.1	2.6	3.2	4.7	5.1	8.0	7.9	9.0	9.3
p		0.166	0.49	0.296	0.095	0.153	0.299	0.291	0.28	0.487	0.299	0.311	0.219	0.266	0.042	0.039	0.025	0.022

(c) 0 dB SNR

S	Su	Sp_ SS	Sp_TS-			Sp_CS-													
			20	40	80	20/4	20/8	20/16	40/4	40/8	40/16	80/4	80/8	80/16	120/8	120/16	160/8	160/16
AU	2.96	2.60	2.69	2.82	3.21	2.91	2.95	2.95	2.61	2.94	2.39	2.72	2.43	2.55	2.12	2.05	2.06	2.04
JK	2.45	2.38	2.25	2.29	2.27	2.24	2.21	2.16	2.28	2.13	2.19	2.06	2.10	2.10	2.03	2.06	2.07	2.03
SM	2.23	2.14	2.17	2.12	2.32	2.18	2.30	2.24	2.26	2.51	2.17	2.05	2.01	2.00	2.22	2.06	2.22	2.32
VJ	2.12	1.89	2.29	2.22	2.33	2.09	2.15	2.10	2.03	1.95	1.84	2.03	1.93	1.87	1.95	1.86	1.90	1.90
ST	2.07	2.05	2.05	2.16	2.35	2.12	2.05	1.94	1.97	2.06	2.22	2.00	2.03	1.95	2.07	1.98	1.99	1.93
VS	2.25	2.08	2.03	2.04	2.12	2.01	2.03	2.17	2.05	2.04	2.19	1.88	2.11	1.98	1.96	1.93	2.01	2.06
GA	1.98	1.82	1.90	1.99	2.11	2.02	1.91	1.85	1.86	2.02	1.86	1.86	1.89	1.90	1.78	1.76	1.88	1.85
Avg.	2.29	2.14	2.20	2.23	2.50	2.36	2.40	2.36	2.30	2.38	2.15	2.22	2.12	2.13	2.08	2.01	2.06	2.07
s.d.	0.33	0.27	0.26	0.28	0.38	0.31	0.34	0.36	0.25	0.36	0.20	0.29	0.18	0.23	0.14	0.11	0.11	0.15
R.I.		6.6	3.8	2.3	-4.2	2.9	2.8	4.2	5.9	2.4	6.7	8.9	9.1	10.3	11.0	13.7	10.9	10.9
p		0.175	0.275	0.36	0.316	0.346	0.361	0.311	0.19	0.378	0.131	0.117	0.071	0.068	0.032	0.013	0.029	0.034

(d) -3 dB SNR

S	Su	Sp_ SS	Sp_TS-			Sp_CS-													
			20	40	80	20/4	20/8	20/16	40/4	40/8	40/16	80/4	80/8	80/16	120/8	120/16	160/8	160/16
AU	2.78	2.43	2.81	2.71	2.94	2.44	2.68	2.60	2.50	2.61	2.08	2.27	2.18	2.27	2.46	2.52	2.34	2.45
JK	2.29	2.35	2.43	2.41	2.51	2.37	2.34	2.43	2.37	2.35	2.21	2.23	2.09	2.22	2.28	2.19	2.32	2.27
SM	2.57	2.45	2.53	2.69	2.20	2.37	2.42	2.28	2.54	2.49	2.53	2.27	2.50	2.61	2.21	2.34	2.51	2.57
VJ	2.38	2.05	1.96	2.16	2.51	2.49	2.26	2.40	1.91	1.92	2.07	1.97	1.85	1.92	2.05	2.11	1.96	1.94
ST	2.26	1.90	2.00	2.16	2.34	2.35	2.17	2.04	2.08	2.11	2.08	2.07	2.09	2.01	2.05	1.98	2.01	2.14
VS	2.31	2.08	2.29	2.39	2.67	2.02	1.98	2.03	2.03	2.05	1.90	1.96	1.95	1.96	2.01	2.03	2.01	2.03
GA	2.10	1.80	1.94	1.96	2.16	2.09	1.99	1.92	1.87	1.93	1.84	1.93	1.97	2.15	1.83	1.88	2.13	2.05
Avg.	2.38	2.15	2.28	2.35	2.48	2.30	2.26	2.24	2.19	2.21	2.10	2.10	2.09	2.16	2.13	2.15	2.18	2.21
s.d.	0.22	0.26	0.33	0.28	0.27	0.18	0.25	0.25	0.28	0.28	0.23	0.15	0.21	0.24	0.21	0.22	0.21	0.23
R.I.		9.8	4.6	1.4	-4.1	3.0	5.1	5.9	8.4	7.4	11.6	11.6	12.1	9.0	10.7	9.9	8.2	7.3
p		0.05	0.25	0.41	0.25	0.24	0.18	0.14	0.08	0.11	0.02	0.01	0.01	0.05	0.02	0.04	0.05	0.09

(e) -6 dB SNR

S	Su	Sp_ SS	Sp_TS-			Sp_CS-													
			20	40	80	20/4	20/8	20/16	40/4	40/8	40/16	80/4	80/8	80/16	120/8	120/16	160/8	160/16
AU	3.16	2.32	3.08	3.13	3.25	2.82	2.77	3.05	2.63	2.85	2.64	2.32	2.34	2.21	2.39	2.74	2.45	2.59
JK	2.48	2.28	2.39	2.50	2.54	2.38	2.24	2.31	2.38	2.40	2.30	2.21	2.18	2.19	2.21	2.28	2.17	2.17
SM	3.04	2.55	2.99	3.24	3.49	2.78	2.89	3.01	2.60	2.84	2.72	2.91	2.46	2.46	2.73	2.88	2.65	2.54
VJ	2.72	2.01	2.49	2.41	2.85	2.35	2.15	2.07	2.05	2.03	2.12	2.05	1.99	2.09	2.05	2.26	2.18	2.32
ST	2.60	2.07	2.61	2.44	2.38	2.24	2.18	2.26	2.17	2.27	2.22	2.27	2.17	2.14	2.31	2.24	2.16	2.28
VS	2.36	2.17	2.41	2.26	2.44	2.18	2.45	2.44	2.34	2.45	2.52	2.38	2.16	2.33	2.23	2.18	2.31	2.27
GA	2.29	2.14	2.48	2.28	2.72	2.14	2.03	2.02	2.04	2.04	2.13	2.04	1.98	1.95	2.03	2.03	1.92	2.02
Avg.	2.66	2.22	2.64	2.61	2.81	2.41	2.39	2.45	2.32	2.41	2.38	2.31	2.18	2.20	2.28	2.37	2.26	2.31
s.d.	0.33	0.18	0.28	0.40	0.42	0.28	0.33	0.42	0.24	0.34	0.25	0.29	0.17	0.16	0.24	0.31	0.23	0.20
R.I.		16.0	0.8	2.3	-5.5	9.3	10.2	8.1	12.6	9.3	10.2	12.7	17.5	16.8	14.0	10.9	14.7	12.8
p		0	0.43	0.39	0.24	0.08	0.07	0.16	0.02	0.09	0.05	0.03	0	0	0.01	0.06	0.01	0.02

(f) -9 dB SNR

S	Su	Sp_ SS	Sp_TS-			Sp_CS-													
			20	40	80	20/4	20/8	20/16	40/4	40/8	40/16	80/4	80/8	80/16	120/8	120/16	160/8	160/16
AU	3.56	3.12	3.35	2.88	2.79	2.89	3.01	3.15	2.96	3.09	2.81	2.78	2.82	2.73	3.13	3.16	3.04	3.01
JK	2.52	2.36	2.54	2.57	2.58	2.48	2.52	2.38	2.32	2.39	2.41	2.51	2.53	2.20	2.34	2.32	2.43	2.33
SM	3.34	2.82	3.52	3.78	3.61	3.23	3.09	3.34	2.93	2.91	3.20	3.14	2.91	3.03	2.93	2.88	2.95	3.05
VJ	2.74	2.33	2.53	2.67	2.67	2.50	2.75	2.91	2.75	2.63	2.38	2.31	2.30	2.47	2.33	2.46	2.56	2.55
ST	2.42	2.17	2.23	2.31	2.59	2.32	2.29	2.62	2.51	2.40	2.50	2.57	2.48	2.28	2.39	2.32	2.38	2.31
VS	2.79	2.09	2.38	2.49	2.32	2.15	2.28	2.23	2.18	2.18	2.16	2.16	2.14	2.10	2.08	2.07	2.23	2.07
GA	2.45	2.15	2.47	2.25	2.39	2.29	2.24	2.37	2.16	2.23	2.09	2.30	2.19	2.28	2.10	2.25	2.26	2.30
Avg.	2.83	2.43	2.72	2.68	2.69	2.55	2.60	2.71	2.54	2.55	2.51	2.54	2.48	2.44	2.47	2.49	2.55	2.52
s.d.	0.45	0.39	0.50	0.56	0.47	0.38	0.36	0.43	0.34	0.35	0.39	0.34	0.30	0.33	0.40	0.39	0.32	0.38
R.I.		13.8	4.2	4.3	3.7	9.4	7.8	3.7	9.5	9.6	11.0	9.5	11.5	13.2	12.5	11.6	9.4	10.7
p		0.05	0.33	0.3	0.3	0.12	0.15	0.31	0.1	0.1	0.09	0.1	0.06	0.04	0.07	0.08	0.1	0.09

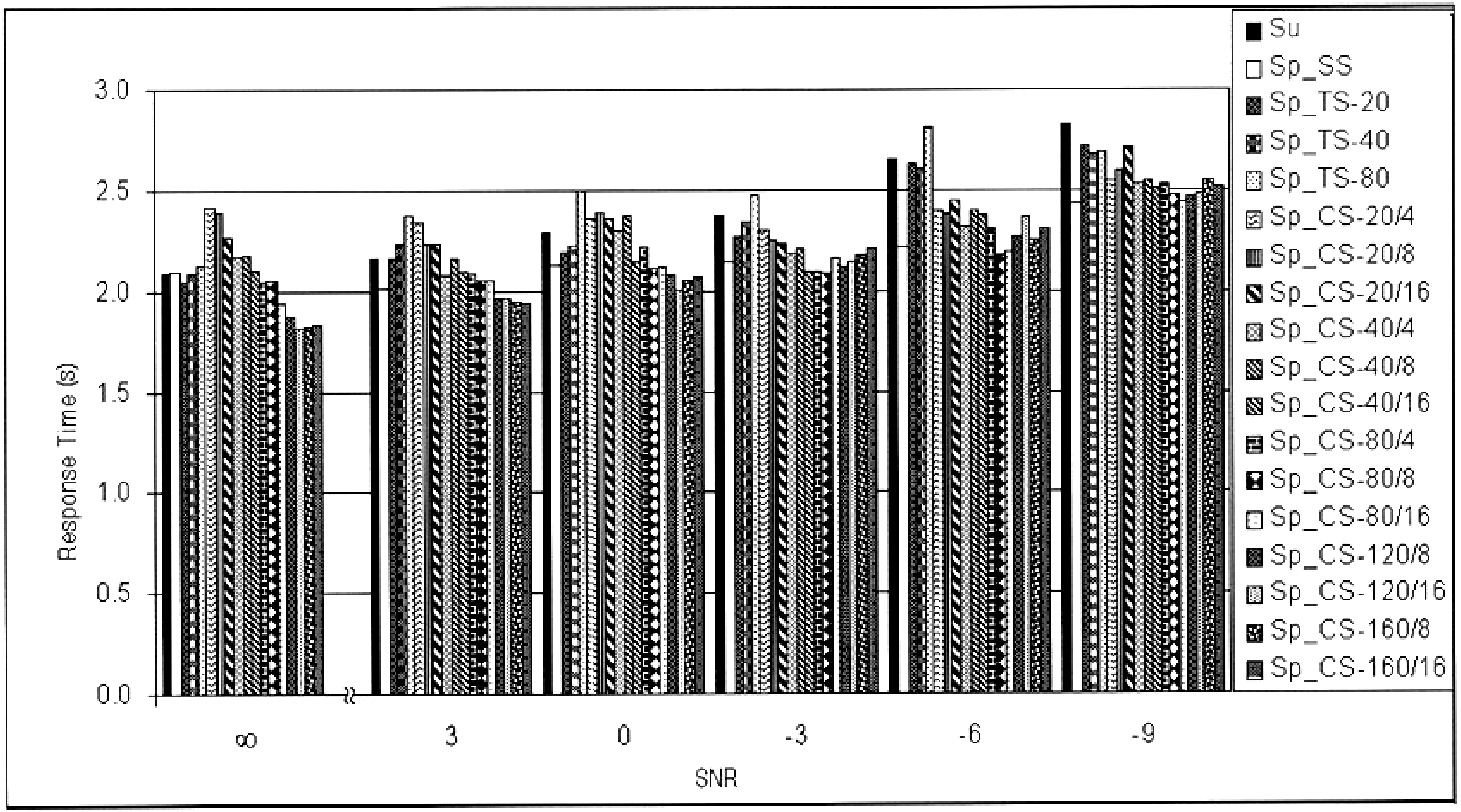


FIG 6.17 (a). Experiment IV. (a) Response time for unprocessed speech (Su) and processed speech: spectral splitting Sp_SS, temporal splitting Sp_TS-T_c with inter-aural switching interval of T_c ms, and combined splitting Sp_CS-T_c/m with cycle time T_c ms and m shiftings for SNRs of ∞, 3, 0, -3, -6, and -9 dB.

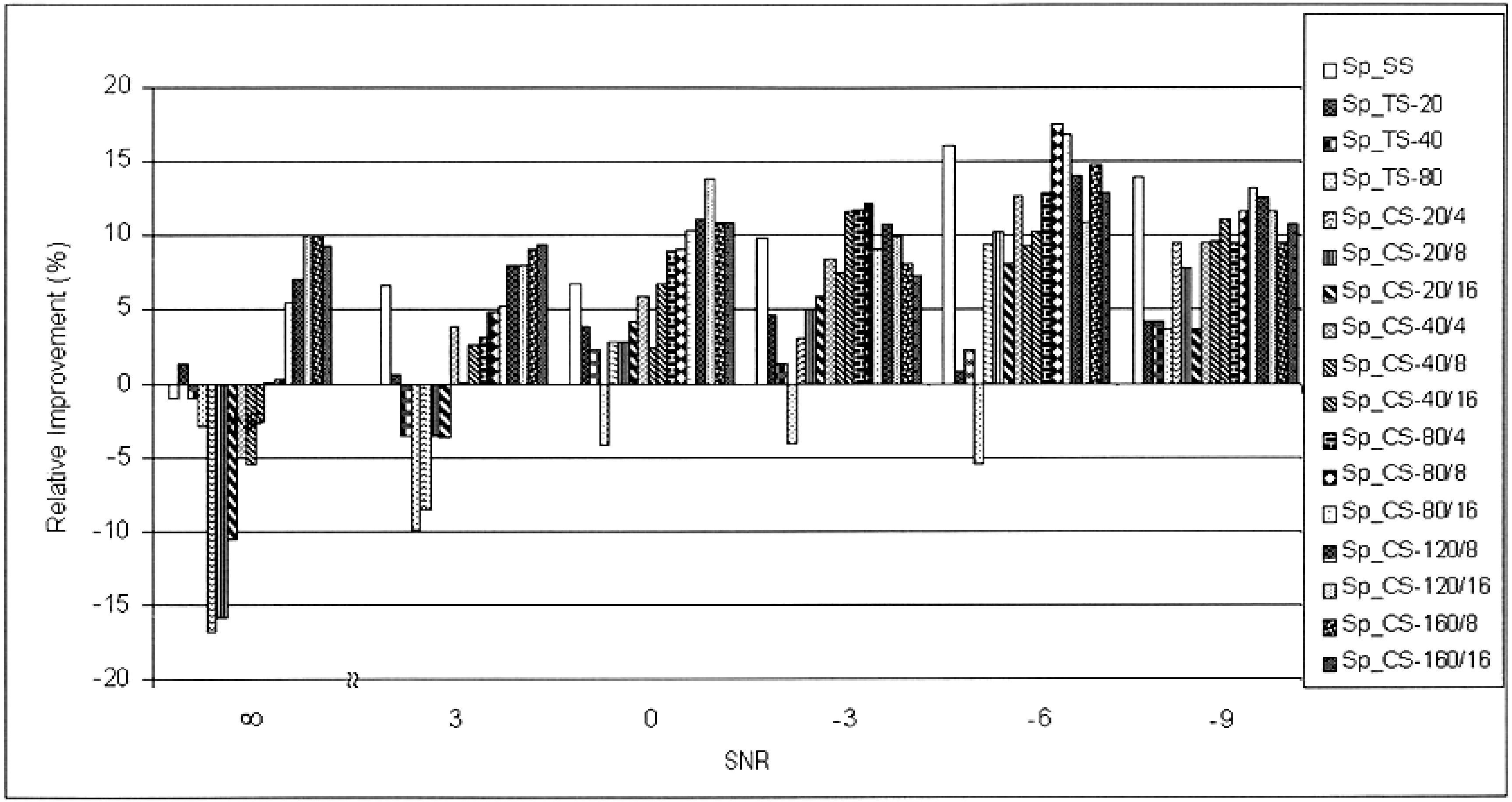


FIG 6.17 (b). Experiment IV. Average relative improvements in response time for different processing schemes and conditions with reference to Su spectral splitting Sp_SS, temporal splitting Sp_TS-T_c with inter-aural switching interval of T_c ms, and combined splitting Sp_CS-T_c/m with cycle time T_c ms and m shiftings for SNRs of ∞, 3, 0, −3, −6, and −9 dB.

TABLE 6.2. Experiment IV. Recognition scores (%) for unprocessed speech (Su) and processed speech: spectral splitting Sp_SS, temporal splitting Sp_TS-T_c with inter-aural switching interval of T_c ms, and combined splitting Sp_CS-T_c/m with cycle time T_c ms and m shiftings for SNRs of (a) ∞, (b) 3, (c) 0, (d) −3, (e) −6, and (f) −9 dB. S: Subject, avg. = average recognition scores, s.d. = standard deviation, R.I. = Average relative improvement (%) with respect to unprocessed. p: significance level (one-tailed) for paired t-test (unprocessed vs processed, averaged across the subjects, n =7, df =6).

(a) ∞ SNR

S	Su	Sp_SS	Sp_TS-			Sp_CS-													
			20	40	80	20/4	20/8	20/16	40/4	40/8	40/16	80/4	80/8	80/16	120/8	120/16	160/8	160/16
AU	100.0	100.0	100.0	100.0	98.6	85.1	85.8	87.9	92.2	88.6	95.0	97.2	97.9	97.9	98.6	99.3	99.3	98.6
JK	100.0	98.6	98.6	99.3	98.6	80.9	80.1	75.9	80.9	83.7	90.8	91.5	96.5	95.7	95.7	94.3	96.5	98.6
SM	100.0	98.6	98.6	99.3	100.0	83.0	87.2	90.1	88.7	91.5	94.3	95.7	100.0	99.3	99.3	98.6	99.3	99.3
VJ	100.0	100.0	100.0	100.0	100.0	90.8	91.5	87.9	93.6	97.2	93.6	97.9	98.5	100.0	100.0	100.0	100.0	100.0
ST	100.0	100.0	100.0	100.0	100.0	75.9	78.7	85.8	96.5	97.2	94.3	100.0	100.0	100.0	93.6	95.7	97.9	98.6
VS	100.0	98.6	98.6	96.5	95.0	88.7	86.5	91.5	95.0	98.6	97.9	97.9	100.0	97.2	96.5	97.2	97.9	97.2
GA	98.6	98.6	98.6	100.0	97.9	83.0	85.1	88.7	89.4	90.1	94.3	94.3	98.6	98.6	98.6	100.0	98.6	97.9
Avg.	99.8	99.2	99.2	99.3	98.6	83.9	85.0	86.8	90.9	92.4	94.3	96.4	98.8	98.4	97.5	97.9	98.5	98.6
s.d.	0.5	0.8	0.8	1.3	1.8	4.9	4.3	5.1	5.2	5.5	2.1	2.8	1.3	1.6	2.3	2.2	1.2	0.9
R.I.		-0.6	-0.6	-0.5	-1.2	-15.9	-14.8	-13.0	-8.9	-7.4	-5.5	-3.5	-1.0	-1.5	-2.3	-1.9	-1.3	-1.2
p		0.05	0.05	0.18	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.02	0.01	0.02	0.01	0.01

(b) 3 dB SNR

S	Su	Sp_ SS	Sp_TS-			Sp_CS-													
			20	40	80	20/4	20/8	20/16	40/4	40/8	40/16	80/4	80/8	80/16	120/8	120/16	160/8	160/16
AU	84.4	92.2	85.8	82.3	78.0	68.0	80.9	78.0	83.6	85.1	90.8	89.4	93.6	89.4	94.3	96.5	97.9	93.6
JK	77.3	85.1	78.7	78.7	64.5	76.6	78.7	76.6	76.6	86.5	83.7	76.6	84.4	78.0	90.1	90.8	87.9	91.5
SM	90.8	97.9	87.9	91.5	91.5	92.2	89.4	85.1	93.6	88.7	90.1	90.8	95.0	98.6	97.2	95.0	91.5	91.5
VJ	86.5	95.0	86.5	84.4	78.7	80.1	84.4	86.5	94.3	91.5	93.6	93.6	95.7	94.3	97.2	95.7	93.6	92.2
ST	86.5	95.0	91.5	90.1	85.1	78.7	87.2	82.3	92.2	89.4	87.2	93.6	94.3	97.2	95.7	96.5	94.3	95.0
VS	73.8	87.2	82.2	76.6	70.2	76.6	88.7	77.3	83.0	85.8	80.9	86.5	87.2	85.1	85.1	85.8	88.7	88.7
GA	85.8	96.5	91.5	91.5	89.4	85.1	75.9	89.4	90.1	86.5	91.5	94.3	97.2	96.5	92.2	95.0	97.9	96.5
Avg.	83.6	92.7	86.3	85.0	79.6	79.6	83.6	82.2	87.6	87.6	88.2	89.3	92.5	91.3	93.1	93.6	93.1	92.7
s.d.	5.9	4.8	4.7	6.2	9.9	7.6	5.2	5.0	6.7	2.3	4.6	6.2	4.8	7.5	4.4	3.9	4.0	2.6
R.I.		11.1	3.5	1.7	-4.9	-4.6	0.4	-1.5	4.9	5.2	5.8	6.9	10.8	8.3	11.6	12.2	11.7	11.3
p		0.00	0.18	0.33	0.19	0.15	0.5	0.32	0.13	0.06	0.06	0.05	0.00	0.03	0.00	0.00	0.00	0.00

(c) 0 dB SNR

S	Su	Sp_ SS	Sp_TS-			Sp_CS-													
			20	40	80	20/4	20/8	20/16	40/4	40/8	40/16	80/4	80/8	80/16	120/8	120/16	160/8	160/16
AU	81.6	91.5	83.7	78.0	71.6	73.1	80.1	75.9	83.0	86.5	88.7	84.4	85.8	90.0	92.9	94.3	91.5	90.1
JK	68.1	83.0	76.6	78.7	68.1	71.6	75.9	72.3	78.0	72.3	78.7	75.9	77.3	73.8	78.7	81.6	76.6	75.9
SM	86.5	94.3	90.1	87.2	83.0	90.8	80.5	83.7	85.1	83.0	87.2	92.2	94.3	96.5	92.2	92.9	93.6	92.2
VJ	90.8	95.0	95.0	93.6	85.1	92.9	87.9	85.1	90.1	90.8	95.0	92.9	97.9	98.6	92.2	95.7	96.5	95.7
ST	76.6	92.2	81.6	73.8	67.4	78.7	75.9	85.1	82.3	85.1	83.0	91.5	90.8	91.5	84.4	87.9	88.7	82.3
VS	67.4	87.2	73.1	70.2	70.9	78.7	75.2	65.3	79.4	74.5	68.1	81.6	83.0	93.6	82.3	80.9	77.3	77.3
GA	80.9	95.0	83.7	87.2	75.9	81.6	84.4	90.8	90.3	85.8	90.8	89.4	88.7	94.3	94.3	95.0	88.7	91.5
Avg.	78.8	91.2	83.4	81.3	75.0	82.1	81.1	79.3	84.0	83.2	87.4	86.3	88.8	89.7	89.0	91.1	89.5	88.5
s.d.	8.8	4.5	7.5	8.4	7.1	8.2	4.8	8.9	4.8	6.7	8.9	6.4	6.9	8.2	6.2	6.4	7.7	7.9
R.I.		16.4	6.0	3.4	-5.1	3.2	2.1	1.4	7.3	5.1	7.4	10.7	12.4	13.4	12.4	14.4	11.3	9.9
p		0.00	0.16	0.3	0.17	0.32	0.38	0.42	0.1	0.19	0.13	0.04	0.02	0.01	0.02	0.01	0.04	0.06

(d) -3 dB SNR

S	Su	Sp_ SS	Sp_TS-			Sp_CS-													
			20	40	80	20/4	20/8	20/16	40/4	40/8	40/16	80/4	80/8	80/16	120/8	120/16	160/8	160/16
AU	73.8	86.5	69.5	73.1	71.6	78.0	80.9	82.2	83.0	80.1	85.8	89.4	89.4	89.4	88.7	85.8	86.5	86.5
JK	54.6	80.1	66.0	68.1	51.8	68.1	70.2	63.1	68.8	68.8	74.5	74.5	70.9	67.4	66.7	74.5	73.1	71.6
SM	73.8	89.4	77.3	69.5	60.3	77.3	73.8	85.1	83.0	80.1	83.0	90.1	85.1	87.2	85.1	75.2	84.4	87.9
VJ	83.7	93.6	90.8	86.5	78.7	76.6	81.6	92.9	86.5	88.0	90.8	89.4	95.7	92.9	91.5	95.0	94.3	92.9
ST	56.0	83.0	68.8	62.4	55.3	68.1	73.8	71.6	75.9	72.3	72.3	87.2	79.9	85.1	78.7	83.0	81.6	71.6
VS	55.3	79.4	53.9	51.8	41.1	65.3	65.3	68.1	67.4	73.1	74.5	77.3	78.0	80.1	74.5	78.0	71.6	71.6
GA	66.7	91.5	75.2	73.8	61.0	79.4	75.9	81.6	80.1	77.3	83.0	81.6	86.5	88.7	83.7	86.5	87.2	87.2
Avg.	66.3	86.2	71.6	69.3	60.0	73.2	74.5	77.8	77.8	77.1	80.5	84.2	83.7	84.4	81.3	82.6	82.7	81.4
s.d.	11.4	5.6	11.3	10.7	12.5	5.9	5.7	10.5	7.4	6.4	6.9	6.4	8.1	8.5	8.6	7.3	8.1	9.3
R.I.		32.3	8.8	5.3	-9.7	12.2	14.2	18.1	18.8	17.9	23.1	29.2	27.8	21.7	23.9	26.7	26.3	23.8
p		0.00	0.2	0.31	0.17	0.09	0.06	0.04	0.02	0.02	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.01

(e) -6 dB SNR

S	Su	Sp_ SS	Sp_TS-			Sp_CS-													
			20	40	80	20/4	20/8	20/16	40/4	40/8	40/16	80/4	80/8	80/16	120/8	120/16	160/8	160/16
AU	58.1	85.1	62.4	56.7	45.4	68.8	73.8	62.4	71.0	70.2	74.5	80.9	85.8	86.5	78.7	69.5	80.1	74.5
JK	29.1	69.5	36.9	38.3	24.1	46.8	51.8	50.4	51.8	55.3	59.6	55.3	61.0	60.3	56.7	51.1	54.6	60.3
SM	44.0	80.1	56.0	48.2	31.2	72.3	68.8	71.6	73.1	67.4	63.8	64.5	68.8	79.4	70.2	70.9	80.9	78.0
VJ	48.9	86.5	62.4	75.9	36.2	73.1	73.8	71.6	75.9	78.7	78.7	80.9	83.0	72.3	78.7	78.7	78.0	78.0
ST	31.2	70.2	43.3	48.9	31.9	56.0	53.9	54.6	63.8	58.2	61.7	61.7	58.9	61.0	55.3	63.1	66.7	58.9
VS	33.3	76.6	43.3	37.6	24.1	63.1	64.5	56.7	65.3	58.9	57.5	65.3	64.5	75.2	67.4	61.0	68.1	67.4
GA	34.8	71.6	50.4	44.7	27.0	51.5	64.5	63.1	72.3	69.5	73.8	70.9	68.1	70.9	68.1	66.7	71.6	66.7
Avg.	39.9	77.1	50.7	50.0	31.4	61.7	64.4	61.5	67.6	65.4	67.1	68.5	70.0	72.2	67.9	65.9	71.4	69.1
s.d.	10.7	7.0	10.0	13.2	7.6	10.4	8.8	8.2	8.2	8.3	8.4	9.6	10.5	9.4	9.3	8.7	9.4	7.9
R.I.		100.8	28.9	27.5	-20.3	58.6	66.3	59.5	75.7	69.7	74.5	77.0	80.2	45.3	75.2	70.6	84.7	79.3
p		0.00	0.04	0.07	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

(f) -9 dB SNR

S	Su	Sp_ SS	Sp_TS-			Sp_CS-													
			20	40	80	20/4	20/8	20/16	40/4	40/8	40/16	80/4	80/8	80/16	120/8	120/16	160/8	160/16
AU	32.6	60.3	38.3	29.9	15.8	54.6	51.8	53.2	54.6	51.1	60.3	64.5	66.0	61.7	57.5	58.2	56.0	57.5
JK	17.7	45.4	22.0	20.6	10.6	27.7	27.7	37.6	43.3	24.8	39.7	36.9	53.2	46.8	33.3	34.8	32.6	44.0
SM	24.8	63.8	33.3	28.4	17.7	51.8	46.1	44.0	53.2	48.2	50.4	51.1	58.9	53.9	51.8	56.0	51.8	51.8
VJ	28.4	70.2	35.4	31.2	19.1	55.3	52.5	54.6	53.2	54.6	56.0	55.3	58.9	52.5	56.0	61.0	63.1	55.3
ST	15.6	55.3	23.4	22.7	14.9	34.0	39.0	38.3	37.6	41.1	44.0	41.9	44.7	46.1	43.3	40.4	42.6	41.1
VS	24.1	53.2	34.8	34.8	17.7	38.3	49.4	48.2	43.3	50.4	52.5	48.2	57.6	52.5	48.2	49.6	45.4	43.3
GA	24.1	58.2	33.3	39.0	20.6	43.3	41.1	42.6	45.4	45.4	49.7	48.2	46.1	51.8	46.1	48.9	50.4	47.5
Avg.	23.9	58.1	31.5	29.5	16.6	43.6	43.9	45.5	47.2	45.1	50.4	49.4	55.0	52.2	48.0	49.8	48.8	48.6
s.d.	5.8	7.9	6.3	6.4	3.3	10.8	8.8	6.8	6.5	9.9	6.9	9.0	7.6	5.2	8.3	9.6	9.8	6.3
R.I.		151.7	33.3	26.2	-28.3	83.4	87.3	95.3	103.1	92.0	116.4	110.7	137.6	126.2	105.5	111.9	108.1	109.7
p		0.00	0.02	0.06	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

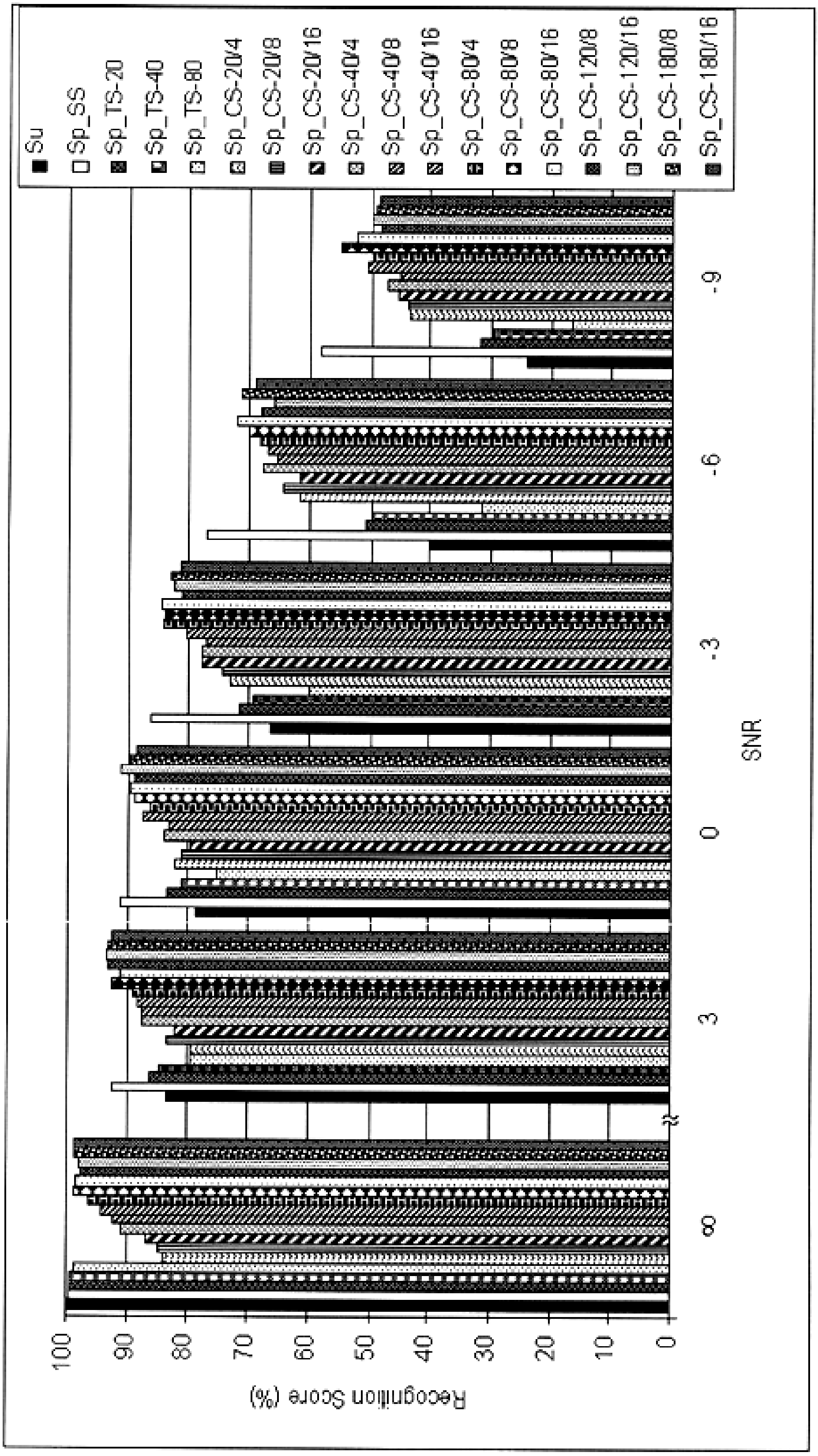


FIG 6.18. (a) Experiment IV. Recognition scores for unprocessed speech (Su) and processed speech: spectral splitting Sp_SS, temporal splitting Sp_TS-T_c with inter-aural switching interval of T_c ms, and combined splitting Sp_CS-T_c/m with cycle time T_c ms and m shiftings for SNRs of ∞, 3, 0, −3, −6, and −9 dB.

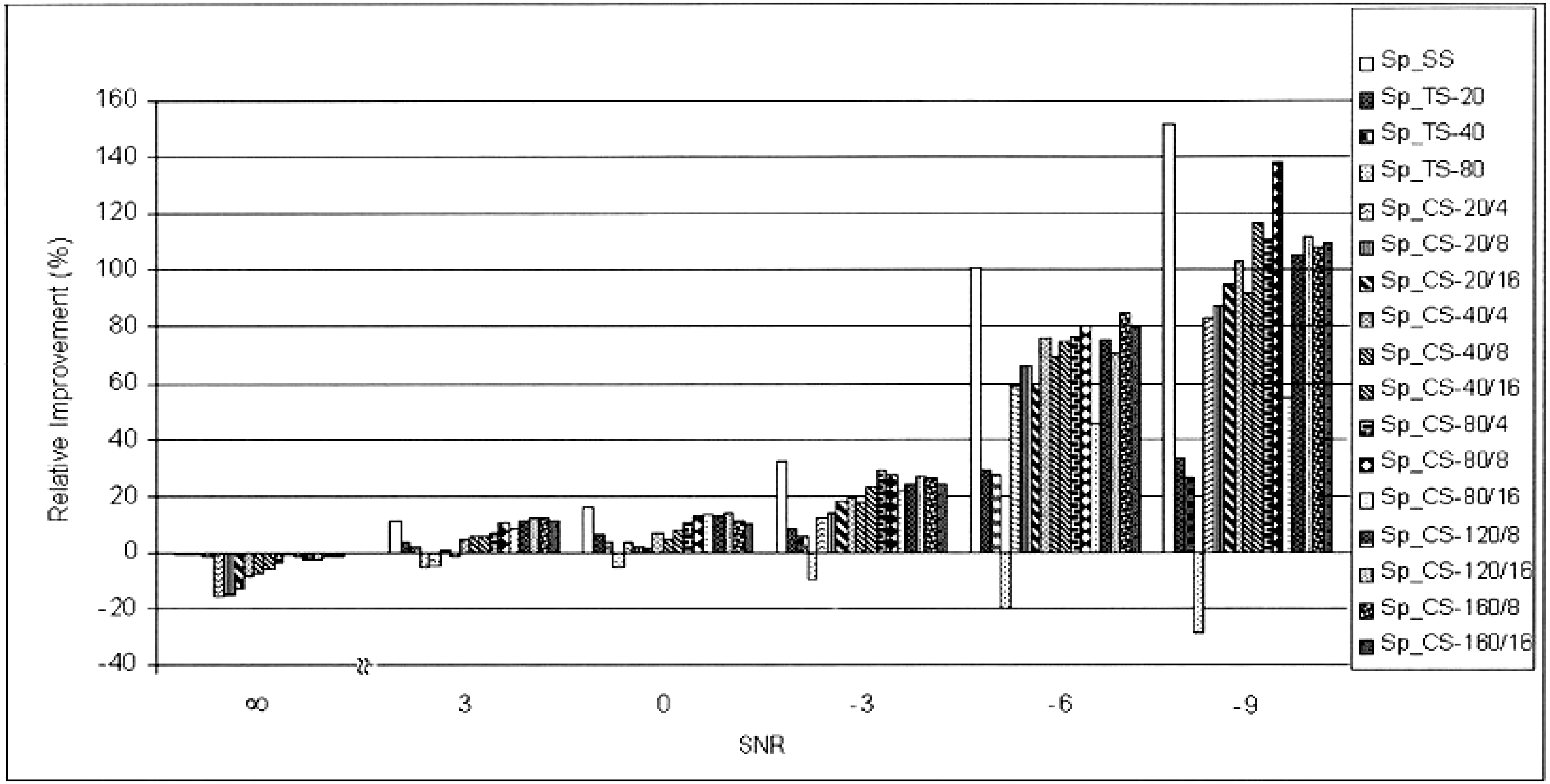


FIG 6.18. (b) Experiment IV. Average relative improvements in recognition scores for different processing schemes and conditions with reference to unprocessed speech (Su): spectral splitting Sp_SS, temporal splitting Sp_TS-T_c with inter-aural switching interval of T_c ms, and combined splitting Sp_CS-T_c/m with cycle time T_c ms and m shiftings for SNRs of ∞, 3, 0, −3, −6, and −9 dB.

and different SNR levels. The effects of switching duration and SNR were not significant. Similar analysis for combined splitting with its processing conditions (13 combinations of sweep cycle time and number of shiftings) showed significant effect ($p < 0.01$) for both the processing conditions as well as the SNR levels. Based on the relative improvements across SNR levels, T_c = 20 ms is optimal for temporal splitting and T_c = 80 ms with 16 shiftings may be taken as the optimal condition for combined splitting. ANOVA for the spectral, temporal, and combined splitting, with the appropriate optimal conditions, showed a significant effect of SNR ($p < 0.01$) and processing scheme ($p < 0.05$).

Since the effects of processing are more at lower SNR levels, a repeated measure ANOVA was carried out, with the processing conditions as the main effect, separately for different SNR levels (results given in Table H.4). It showed significant effects ($p < 0.01$) at all SNR levels. For SNR = –6 dB and –9 dB, similar analysis was carried out for (a) temporal splitting with the 3 switching durations, (b) combined splitting with 13 combinations of sweep cycle duration and number of shiftings, and (c) the three splitting schemes with the corresponding optimal conditions (results given in Table H.5). The effect of processing is significant ($p < 0.01$) in all the three cases, at both the SNR levels.

Further as the improvements are higher at lower SNR levels, particularly at SNR = –6 dB and –9 dB, a comparison of relative improvements of the various processing schemes and conditions was carried out at these two SNR levels, by calculating the averaged relative improvements and paired t-test significance levels between the processing conditions (results given in Table H.6). For temporal splitting, highest improvements are with T_c = 20 ms, 29 % for -6 dB and 33 % for –9 dB, and both are significant ($p < 0.01$). These improvements are only slightly higher over T_c = 40 ms and not statistically significant. However the improvements over T_c = 80 ms are large and significant.

For combined splitting, the improvements due to processing are statistically significant ($p < 0.01$) under all processing conditions, and are in the range of 59-87 % and 83 –138 % at –6 and –9 dB SNR levels respectively. The highest and statistically significant ($p < 0.01$) improvements are obtained for T_c = 80 ms and m = 8 and 16: 80-87 % for –6 dB SNR and 126 –138 % for –9 dB SNR. These improvements are significant only with respect to T_c = 20 ms.

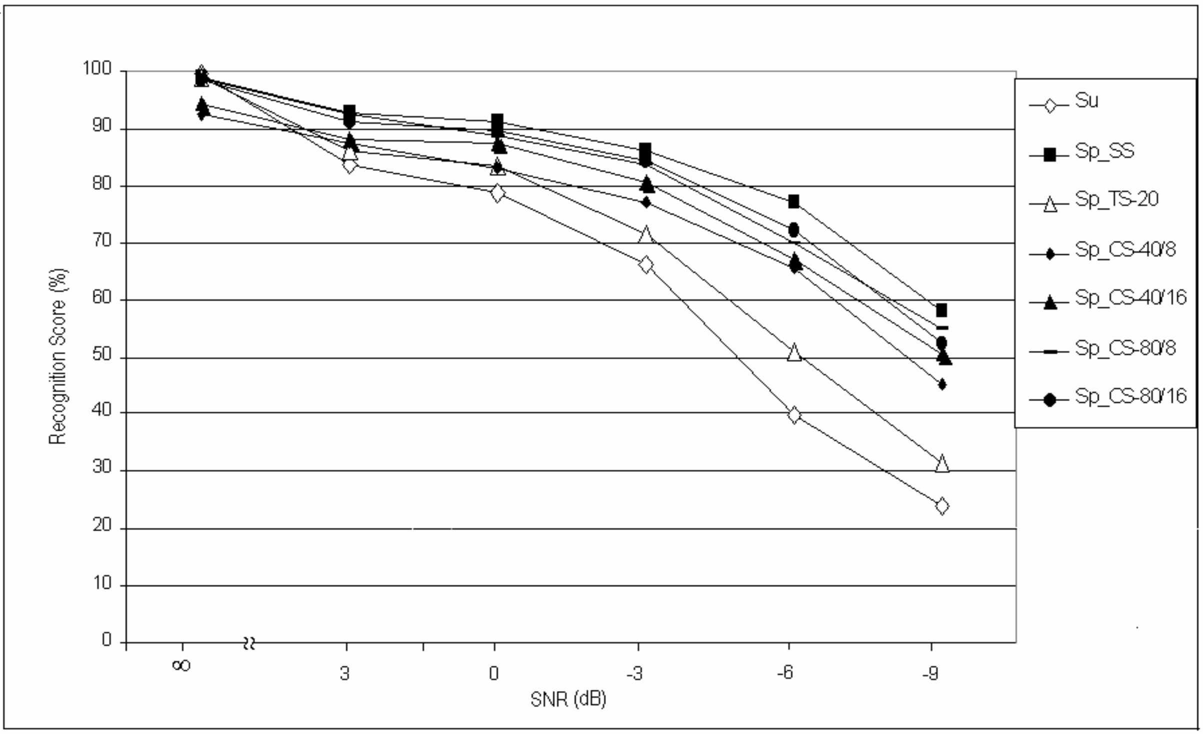


FIG 6.19. Experiment IV. Average recognition score (%) for unprocessed speech and processed speech: spectral splitting (Sp_SS), temporal splitting for T_c = 20 ms (Sp_TS-20) and combined splitting for T_c = 40 and 80 ms with m = 8 and 16 (Sp_CS-40/8, Sp_CS-40/16, Sp_CS-80/8, and Sp_CS-80/16).

A comparison of the three splitting schemes with the corresponding optimal conditions, showed that the splitting gave statistically significant ($p < 0.05$) improvements; 29 % and 33 % for temporal splitting, 87 % and 138 % for combined splitting, and 101 % and 152 % for spectral splitting at SNR of –6 dB and –9 dB respectively. The relative improvements of combined splitting over temporal splitting at both the SNR levels are statistically significant ($p < 0.01$). The improvements of spectral splitting over combined splitting are statistically not significant at both the SNR levels.

6.4 Experiment V: Listening tests on hearing impaired subjects

Results of Experiment IV, involving listening tests on normal subjects with simulated hearing loss, showed maximum improvement for spectral splitting with perceptually balanced comb filters. Temporal splitting provided improvement for inter-aural switching periods of 20 and 40 ms. Inter-aural switching period of 80 ms did not provide any improvement, and hence was not considered for listening tests on hearing impaired persons. Improvements in recognition scores for combined splitting, for sweep cycle of 40 and 80 ms were generally higher as compared to 20 ms, while higher values of sweep cycles did not show any further improvements. Almost similar improvements were obtained for 8 and 16 shiftings, which was higher compared to 4 shiftings. While conducting listening tests on hearing impaired subjects, processing with 4 shiftings and sweep cycles more than 80 ms was not considered with combined splitting scheme.

Experimental evaluation involved diotic presentation of unprocessed speech (Su) and dichotic presentation of processed speech: adjustable magnitude response filter (Sp_AG), adjustable magnitude response filter cascaded with spectral splitting (Sp_AG-SS), adjustable magnitude response filter cascaded with temporal splitting (Sp_AG-TS-20 and Sp_AG-TS-40) and adjustable magnitude response filter cascaded with combined splitting (Sp_AG-CS-20/8, Sp_AG-CS-20/16, Sp_AG-CS-40/8, Sp_AG-CS-40/16, Sp_AG-CS-80/8, and Sp_AG-CS-80/16).

The listening tests were conducted with sets of phonetically balanced monosyllabic words in three languages (Hindi, Marathi, and English), the same as used for normal subjects, in Experiment IV, described in Section 6.2.3. Thirteen bilateral sensorineural hearing impaired subjects with mild to severe loss, participated in the listening tests (SM: M 59, RM: M 52, AB: M 61, DV: M 59, KJ: M 20, WK: M 19, BS: M 38, PP: M 23, NA: M 42, SK: M 42, SS: F 27, PK: M 44, and TT: M 35). The hearing loss of these subjects, were categorized by considering their pure tone audiograms. Table 6.3 gives the hearing thresholds of these subjects. While participating in the listening tests, hearing impaired listeners were not using hearing aids and speech was presented through headphones

Subject SM with moderate hearing loss was having almost constant loss up to 2 kHz, slowly increasing afterwards. RM had moderately severe hearing loss, almost constant up to 750 Hz, and increasing afterwards. The subjects AB and DV were having moderate and moderately severe loss respectively. Both had symmetrical low frequency loss with maximum loss at 500 Hz. The subjects KJ and WK with moderately severe hearing loss, had symmetrical sloping high frequency loss, steeply sloping up to 1 kHz, further on the slope reduced. BS and PP were having severe high frequency loss, steeply sloping up to 1.5 kHz and 2 kHz respectively. Subject NA had sloping moderately severe high frequency loss. SK had severe symmetrical slowly sloping high frequency hearing loss. SS had symmetrical

TABLE 6.3. Hearing thresholds of the subjects with bilateral sensorineural hearing impairment. PTA: average pure tone thresholds in dB HL, taken at test frequencies 0.5, 1, 2 kHz.

Subject Code (Sex, Age)	Ear L=left R=right	Hearing threshold (dB HL) Frequency (kHz) 0.25	0.5	1	2	4	6	PTA (dB)
SM (M, 59)	L	35	40	45	40	85	85	42
	R	35	45	45	40	75	75	42
RM (M, 52)	L	50	50	55	65	85	85	57
	R	35	35	65	90	85	85	63
AB (M, 61)	L	55	65	50	40	45	45	52
	R	60	60	45	35	35	35	47
DV (M, 59)	L	70	80	60	65	70	60	68
	R	85	90	75	60	65	90	75
KJ (M, 20)	L	40	45	65	75	75	90	62
	R	45	50	65	70	75	75	62
WK (M, 19)	L	45	55	75	85	95	95	72
	R	50	50	75	80	85	85	68
SK (M, 42)	L	75	75	85	90	100	100	83
	R	75	85	90	95	105	105	90
SS (F, 27)	L	45	60	90	85	80	75	78
	R	45	60	70	85	75	70	72
PP (M, 23)	L	25	45	65	115	105	100	75
	R	55	60	80	105	110	100	82
PK (M, 44)	L	80	80	85	80	100	120	82
	R	60	65	45	40	70	80	50
TT (M, 35)	L	95	100	115	120	120	115	112
	R	60	65	65	80	85	80	70
NA (M, 42)	L	30	25	75	80	120	115	60
	R	40	45	70	80	95	100	65
BS (F, 38)	L	75	80	105	110	120	120	98
	R	70	65	85	110	120	120	87

severe loss more in the mid frequency range between 1 and 2 kHz. The subjects PK and TT had asymmetrical loss. PK had less loss at mid frequencies and TT had more loss in the frequency range of 3 - 4 kHz. The pure tone audiogram and the magnitude response of the adjustable gain filter used to partly compensate for the shifts in hearing thresholds for each of the subjects are given in Appendix G.

The listening test set-up and testing procedure is same as that used for normal subjects in Experiment IV, described in Section 6.2.4 and 6.2.5. The results of the listening tests conducted on the thirteen bilateral sensorineural hearing impaired subjects, are given for individual subjects in Tables 6.4 and 6.5 for response times and recognition scores respectively.

6.4.1 Response time

The response times for unprocessed and processed speech under different schemes for the different processing conditions for all the thirteen subjects, are shown in Table 6.4. The relative improvements in response times for the different schemes for the different processing conditions, and the average values and paired t-test significance levels (one-tailed) for the different processing schemes and conditions are also tabulated.

Figure 6.20 gives the response time of the thirteen subjects for the unprocessed (Su) and the different processed schemes under various processing conditions. The response time varied from 2.1 to 6.6 s for unprocessed speech. For processing with adjustable magnitude response filter (Sp_AG) and spectral splitting (Sp_AG-SS), the relative decrease in response times ranged over –4.6 to 38.7 % and 0.5 to 41.6 % respectively. The relative improvement in response time for temporal splitting with inter-aural switching of 20 and 40 ms (Sp_AG-TS-20 and Sp_AG-TS-40) ranged over –1 to 41.1 % and –2.8 to 31.8 % respectively. For the different processing conditions of combined splitting scheme, the ranges of values for relative improvements in response times were: for Sp_AG-CS-20/8 from –88.7 to 24.2 %, for Sp_AG-CS-20/16 from –82.0 to 29.2 %, for Sp_AG-CS-40/8 from –89.2 to 27.5 %, for Sp_AG-CS-40/16 from –22.7 to 34.4 %, for Sp_AG-CS-80/8 from –71.2 to 36.5 %, and for Sp_AG-CS-80/16 from –87.8 to 33.1 %. Subjects AB and DV, having symmetrical low frequency loss, showed increase in response time for combined splitting scheme for all processing conditions. They had some improvements under the other three processing conditions.

TABLE 6.4. Experiment V. Response time (s) for unprocessed speech (Su) and processed speech: adjustable magnitude response filter (Sp_AG), spectral splitting (Sp_AG-SS), temporal splitting (Sp_AG-TS-T_c) for T_c = 20, 40 ms, and combined splitting (Sp_AG-CS-T_c/m) for T_c = 20, 40, 80 and m = 8, 16. S: Subject, Avg. = average, RT = response time, R.I. = relative improvement (%). p: significance level (one-tailed) for paired t-test (unprocessed vs processed, averaged across the subjects, n =13, df =12).

S	RT/ R.I.	Su	Sp_ AG	Sp_AG -SS	Sp_AG-TS- 20	Sp_AG-TS- 40	Sp_AG-CS- 20/8	Sp_AG-CS- 20/16	Sp_AG-CS- 40/8	Sp_AG-CS- 40/16	Sp_AG-CS- 80/8	Sp_AG-CS- 80/16
SM	RT	2.60	2.38	2.11	2.26	2.37	2.50	2.39	2.10	2.08	2.23	2.18
	R.I.		8.5	18.8	13.1	8.8	3.8	8.1	19.2	20.0	14.2	16.2
RM	RT	2.82	2.95	2.56	2.58	2.90	3.13	2.88	2.79	3.46	3.00	3.01
	R.I.		-4.6	9.2	8.5	-2.8	-11.0	-2.1	1.1	-22.7	-6.4	-6.7
AB	RT	2.44	2.16	2.06	2.24	2.37	2.69	2.78	2.86	2.81	3.28	3.02
	R.I.		11.5	15.6	8.2	2.9	-10.2	-13.9	-17.2	-15.2	-34.4	-23.8
DV	RT	2.22	2.00	2.02	1.96	1.97	4.19	4.04	4.20	2.47	3.80	4.17
	R.I.		9.9	9.0	11.7	11.3	-88.7	-82.0	-89.2	-11.3	-71.2	-87.8
KJ	RT	2.99	2.39	2.05	1.99	2.04	2.73	2.78	2.06	2.16	2.14	2.12
	R.I.		20.1	31.4	33.4	31.8	8.7	7.0	31.1	27.8	28.4	29.1
WK	RT	2.40	2.36	2.16	2.12	2.21	2.46	2.41	2.46	2.39	2.38	2.23
	R.I.		1.7	10.0	11.7	7.9	-2.5	-0.4	-2.5	0.4	0.8	7.1
SK	RT	2.40	2.36	2.16	2.12	2.21	2.46	2.41	2.46	2.39	2.38	2.23
	R.I.		1.7	10.0	11.7	7.9	-2.5	-0.4	-2.5	0.4	0.8	7.1
SS	RT	2.98	2.71	2.85	2.41	2.38	2.61	2.48	2.65	2.34	2.39	2.33
	R.I.		9.1	4.4	19.1	20.1	12.4	16.8	11.1	21.5	19.8	21.8
PP	RT	4.18	3.94	3.66	2.77	3.26	3.17	2.96	3.05	2.90	3.10	2.87
	R.I.		5.7	12.4	33.7	22.0	24.2	29.2	27.0	30.6	25.8	31.3
PK	RT	2.09	2.09	2.08	2.11	2.14	2.54	2.12	2.19	2.41	1.90	1.87
	R.I.		0.0	0.5	-1.0	-2.4	-21.5	-1.4	-4.8	-15.3	9.1	10.5
TT	RT	2.65	2.36	2.33	2.26	2.33	2.13	2.19	2.02	2.05	2.06	2.02
	R.I.		10.9	12.1	14.7	12.1	19.6	17.4	23.8	22.6	22.3	23.8
NA	RT	3.26	2.23	2.35	2.06	2.31	2.82	2.97	2.39	2.14	2.07	2.18
	R.I.		31.6	27.9	36.8	29.1	13.5	8.9	26.7	34.4	36.5	33.1
BS	RT	6.59	4.04	3.85	3.88	8.99	5.53	5.63	4.78	4.39	4.30	5.40
	R.I.		38.7	41.6	41.1	-36.4	16.1	14.6	27.5	33.4	34.7	18.1
Avg. RT		3.05	2.61	2.48	2.37	2.88	3.00	2.93	2.77	2.61	2.69	2.74
s.d. RT		2.82	1.17	1.23	1.15	4.68	2.14	2.29	1.90	1.63	1.46	2.28
Avg. R.I.		11.1	15.6	18.7	8.6	-2.9	0.1	3.9	9.7	6.2	6.1	11.1
p			0.1	0.07	0.04	0.4	0.5	0.4	0.2	0.1	0.2	0.2

It is seen that for unprocessed speech, there is a large variation in response time, with a range of 2.1 – 6.65 s, average = 3.1 s, s.d. = 2.8 s. Hence for comparison of the processing conditions across the subjects, we have considered the relative improvement of particular processing condition with reference to the unprocessed speech. The averaged relative improvement is highest, 18.7 %, for spectral splitting and it is statistically significant ($p <$ 0.1). Other statistically significant improvements are: 15.6 % for Sp_AG and 8.5 % for Sp_AG-TS-20.

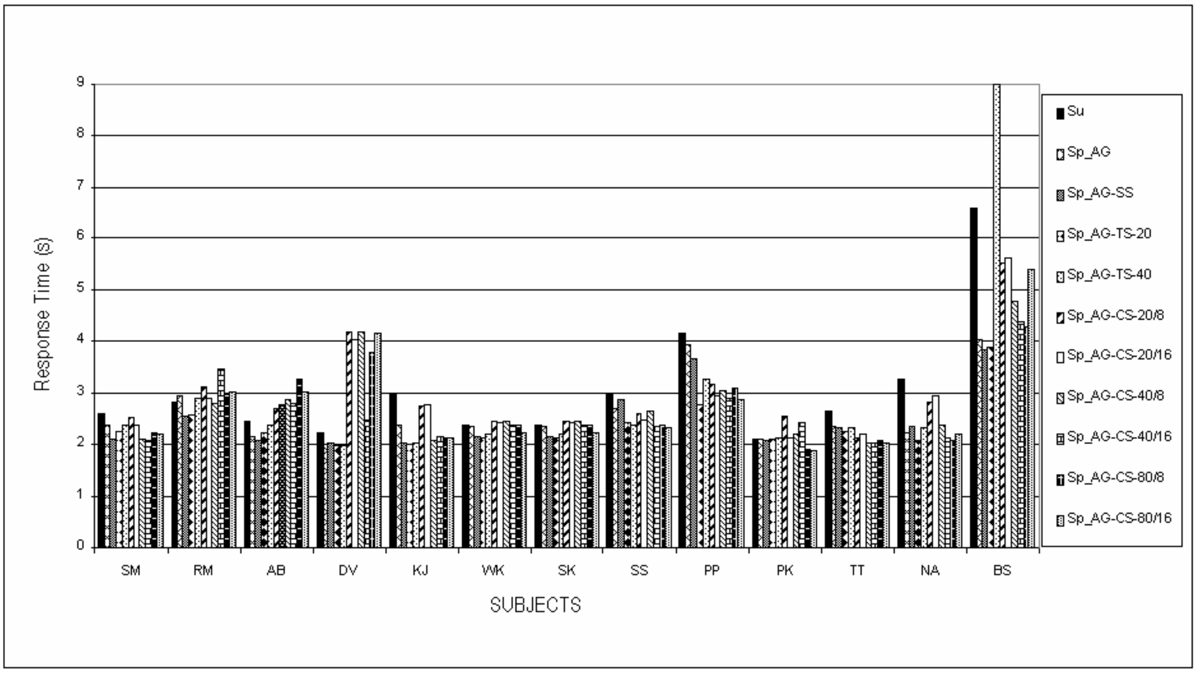


FIG. 6.20. Experiment V. Average response time (s) for unprocessed speech (Su) and processed speech: adjustable magnitude response filter (Sp_AG), spectral splitting (Sp_AG-SS), temporal splitting (Sp_AG-TS-T_c) for T_c = 20, 40 ms, and combined splitting (Sp_AG-CS-T_c/m) for T_c = 20, 40, 80 and m = 8, 16.

6.4.2 Recognition scores

Table 6.5 shows the recognition scores for unprocessed and processed speech for the different schemes for all processing conditions tested on thirteen hearing impaired subjects. The percentage relative improvement, the average recognition scores and percentage relative improvement in recognition scores, and the paired t-test significant levels (one-tailed) are also given. Figure 6.21 shows the recognition score for the different schemes at the various processing conditions for the thirteen subjects.

The scores for unprocessed speech ranged over 20.6 – 90.1 %. The percentage relative improvements because of the various processing schemes and parameters varied across the subjects: 1.2 – 65.6 for Sp_AG, 5.7 – 120.8 for Sp_AG-SS, 2.4 – 110.4 for Sp_AG-TS-20, 0.8 – 79.4 for Sp_AG-TS-40, -87.7 – 55.2 for Sp_AG-CS-20/8, -87.6 – 69.0 for Sp_AG-CS-

20/16, -65.3 - 107 for Sp_AG-CS-40/8, and -37.7 – 138.0 for Sp_AG-CS-40/16, -57.0 – 110.4 for Sp_AG-CS-80/8, -53.7 – 117.3 for Sp_AG-CS-80/16. The relative improvements in recognition scores were the least for the combined splitting scheme with 20 ms cycle time (Sp_AG-CS-20/8 and Sp_AG-CS-20/16). For most subjects, this processing even reduced the recognition scores.

TABLE 6.5. Experiment V. Recognition scores (%) for unprocessed speech (Su) and processed speech: adjustable magnitude response filter (Sp_AG), spectral splitting (Sp_AG-SS), temporal splitting (Sp_AG-TS-T_c) for T_c = 20, 40 ms, and combined splitting (Sp_AG-CS-T_c/m) for T_c = 20, 40, 80 and m = 8, 16. Subject, Avg. = average RS = recognition scores, R.I. = relative improvement (%). p: significance level for one-tailed paired t-test (unprocessed vs processed, averaged across the subjects, n =13, df =12).

S		Su	Sp_AG	Sp_AG-SS	Sp_AG-TS-		Sp_AG-CS-					
					20	40	20/8	20/16	40/8	40/16	80/8	80/16
SM	RS	76.6	84.4	87.9	84.4	83.7	66.7	66.0	82.3	87.2	83.7	85.8
	R.I.		10.2	14.8	10.2	9.2	-13.0	-13.9	7.4	13.9	9.2	12.0
RM	RS	46.1	63.8	71.6	66.0	63.1	56.0	61.0	65.3	63.1	63.8	63.2
	R.I.		38.5	55.4	43.1	36.9	21.5	33.8	41.5	36.9	38.5	36.9
AB	RS	86.5	92.2	91.5	90.1	87.2	10.6	14.2	50.4	53.9	58.9	63.2
	R.I.		6.6	5.7	4.1	0.8	-87.7	-83.6	-41.8	-37.7	-32.0	-27.0
DV	RS	85.8	90.8	90.8	90.1	90.8	12.8	10.6	29.8	60.3	36.9	39.7
	R.I.		5.8	5.8	5.0	5.8	-85.1	-87.6	-65.3	-29.8	-57.0	-53.7
KJ	RS	78.0	93.6	95.0	94.3	91.5	75.2	67.4	90.8	86.5	92.2	92.2
	R.I.		20.0	21.8	20.9	17.3	-3.6	-13.6	16.4	10.9	18.2	18.2
WK	RS	63.1	77.3	83.0	84.4	78.7	54.6	54.6	74.5	73.8	78.0	79.4
	R.I.		22.5	31.5	33.7	24.7	-13.5	-13.5	18.0	16.8	23.6	25.8
SK	RS	20.6	34.0	45.4	43.3	36.9	31.9	34.8	42.6	48.9	43.3	44.7
	R.I.		65.6	120.8	110.4	79.4	55.2	69.0	107.0	138.0	110.4	117.3
SS	RS	52.5	61.0	65.3	69.5	68.1	41.1	43.3	56.0	58.9	56.8	67.4
	R.I.		16.2	24.3	32.4	29.7	-21.6	-17.6	6.8	12.2	8.2	28.4
PP	RS	59.6	60.3	70.2	71.6	64.5	53.2	53.9	57.5	68.8	58.7	68.1
	R.I.		1.2	17.9	20.2	8.3	-10.7	-9.5	-3.6	15.5	-1.2	14.3
PK	RS	90.1	92.9	95.8	92.2	90.1	75.9	67.4	82.3	87.9	93.6	94.3
	R.I.		3.2	6.3	2.4	0.0	-15.8	-25.2	-8.7	-2.4	3.9	4.7
TT	RS	53.9	72.3	76.6	75.9	72.3	54.6	57.5	74.5	76.6	77.3	78.0
	R.I.		34.2	42.1	40.8	34.2	1.3	6.6	38.2	42.1	43.4	44.7
NA	RS	61.7	83.0	84.4	84.4	80.1	53.2	58.2	75.2	73.1	87.2	85.1
	R.I.		34.5	36.8	36.8	29.9	-13.8	-5.7	21.8	18.4	41.4	37.9
BS	RS	40.4	50.4	50.4	53.9	42.6	28.4	34.6	33.3	31.9	32.6	37.6
	R.I.		24.6	24.6	33.3	5.3	-29.8	-14.0	-17.5	-21.1	-19.4	-7.0
Avg.RS		62.68	73.5	77.5	76.9	73.1	47.2	48.0	62.6	67.0	66.4	69.1
s.d. RS		25.58	24.08	26.58	21.56	29.08	27.09	22.07	34.61	39.12	36.13	34.1
Avg.R.I.			21.8	31.4	30.3	21.7	-16.7	-13.4	9.2	24.1	14.4	19.4
p			0.09	0.03	0.03	0.09	0.04	0.04	0.5	0.3	0.3	0.2

Repeated measure analysis of variance (ANOVA) on relative improvements (with reference to unprocessed speech), for the 13 subjects and processing schemes and conditions (results given in Table H.5), showed significant effect ($p < 0.01$) of processing when all the processing schemes and conditions were considered. ANOVA was also carried out for temporal splitting with its 2 switching durations and the effects was significant ($p < 0.01$). Similar analysis for combined splitting with its processing conditions (6 combinations of sweep cycle time and number of shiftings) showed significant effect ($p < 0.01$) for processing conditions. Based on the average relative improvements, T_c = 20 ms is optimal for temporal splitting, and T_c = 40 ms with 16 shiftings may be taken as the optimal condition for combined splitting. ANOVA for adjustable magnitude response filter and the three splitting schemes with the appropriate optimal conditions showed a significant effect of subjects ($p < 0.01$), but the effect of the schemes was not significant.

Tukey's HSD test for multiple comparisons (Daniel, 1999; Myers and Well, 2003) was carried out for different paring of processing scheme and conditions (results given in Table H.9). The improvements were significant ($p \leq 0.05$) for processing schemes of Sp_AG-SS and Sp_AG-TS-20. For combined splitting, improvements of processing with sweep cycle duration of 40 and 80 ms both with 8 and 16 shiftings were significant when compared with processing with sweep cycle duration of 20 ms.

Further a comparison of the various processing schemes and conditions was carried out, by calculating the averaged relative improvements and paired t-test significance levels between the processing conditions (results given in Table H.6). For temporal splitting, the highest improvements is with T_c = 20 ms and it is statistically significant ($p < 0.05$). The improvements are only slightly higher compared to these for T_c = 40 ms and not significant. The improvements due to processing are in the range of 9 – 24 % for combined splitting for T_c = 40 and 80 ms for 8 and 16 shiftings and are not statistically significant. The highest improvement is obtained for T_c = 80 ms and m = 16. These improvements are significant only with respect to T_c = 20 ms. A comparison of the adjustable magnitude response filter and the three splitting schemes with the corresponding optimal conditions, showed that spectral and temporal splitting gave statistically significant ($p < 0.05$) improvements; of 31 and 30 % respectively. The improvements for the adjustable magnitude response filtering and combined splitting are 22 % and 16 % respectively and are not statistically significant. Spectral splitting provided the highest improvement, but the improvements over other schemes are not statistically significant. The results of t-test comparisons are in agreement with Tukey's HSD test for multiple comparisons.

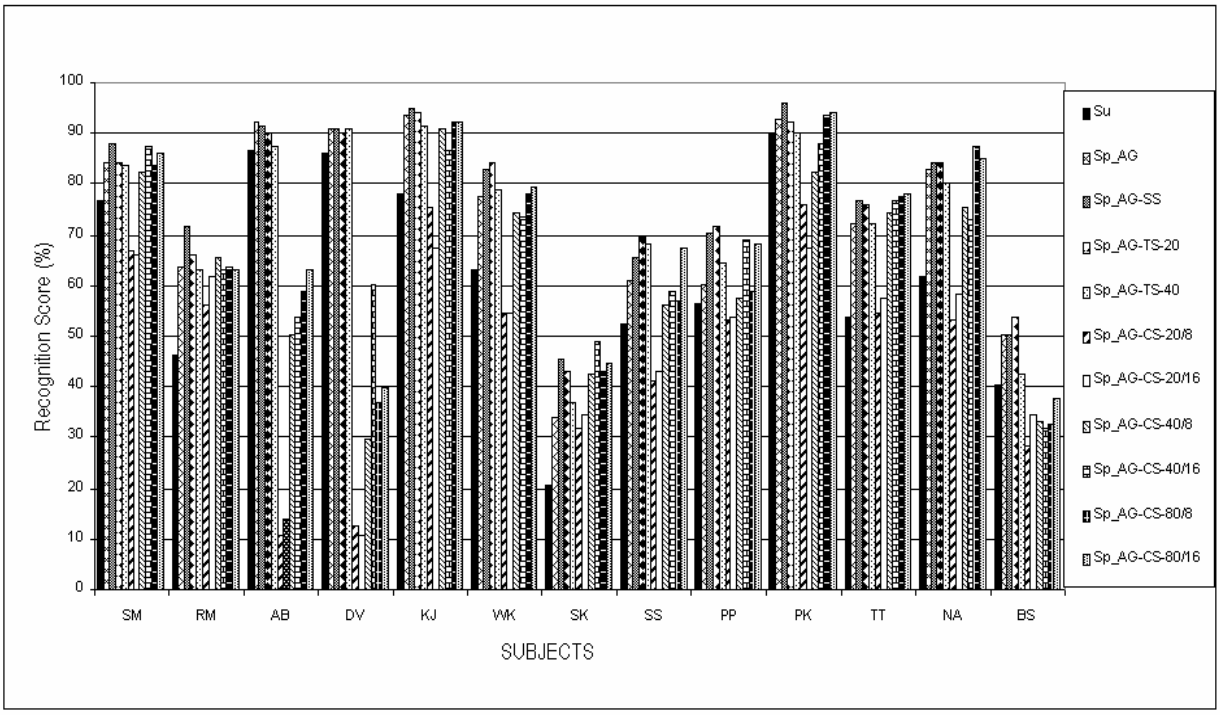


FIG. 6.21. Experiment V. Average recognition score (%) for unprocessed speech (Su) and processed speech: adjustable magnitude response filter (Sp_AG), spectral splitting (Sp_AG-SS), temporal splitting (Sp_AG-TS-T_c) for T_c = 20, 40 ms, and combined splitting (Sp_AG-CS-T_c/m) for T_c = 20, 40, 80 and m = 8, 16.

Since a large variation was observed in relative improvements for the processing schemes and conditions across the subjects, results were studied for optimal processing condition for each of the subjects. These results are summarized in Table 6.6. Subjects SM and RM had almost flat loss at low frequencies, which slowly increased at higher frequencies. For unprocessed speech, their recognition scores were 76.6 and 46.1 % respectively. The relative improvement in recognition scores, were maximum for spectral splitting, 14.8 and 55.4 % respectively. Combined splitting with sweep cycle of 40 ms provided almost similiar improvement: 13.9 % for SM under Sp_AG-CS-40/16 and 41.5 % for RM under Sp_AG-CS-40/8.

Subjects AB and DV, had symmetrical low frequency loss with maximum loss at 500 Hz. The recognition scores for unprocessed speech were 86.5 and 85.8 % for subjects AB and DV respectively. The percentage relative improvement in recognition scores, was maximum for processing with adjustable magnitude response filter (Sp_AG): 6.6 for AB and 5.8 for DV. The subject DV had the same improvement for spectral splitting and temporal splitting with T_c = 40 ms. For these two subjects, the recognition scores for combined splitting at all the different values of sweep cycle and number of shiftings, were lower than the score for unprocessed speech. It has been earlier observed that these two subjects had no improvement in response time for combined splitting.

TABLE 6.6. Experiment V. Subject-wise relative improvements with reference to unprocessed speech. In case of temporal and combined splitting the maximum improvements with corresponding parameters are given. The overall maximum improvement is given in bold.

S	Rec.score for Su	Relative improvement for proc. Cond			
		Sp_AG	Sp_AG_SS	Sp_AG_TS	Sp_AG_CS
SM	76.6	10.2	**14.8**	10.2 T_c = 20	13.9 T_c = 40, m = 16
RM	46.1	38.5	**55.4**	43.1 T_c = 20	41.5 T_c = 40, m = 8
AB	86.5	**6.6**	5.7	4.1 T_c = 20	-27.0 T_c = 80, m = 16
DV	85.8	**5.8**	**5.8**	**5.8** T_c = 40	-29.8 T_c = 40, m = 16
KJ	78.0	20.0	**21.8**	20.9 T_c = 20	18.2 T_c = 80, m = 16
WK	63.1	22.5	31.5	**33.7** T_c = 20	25.8 T_c = 80, m = 16
SK	20.6	65.6	120.8	110.4 T_c = 20	**138.0** T_c = 40, m = 16
SS	52.5	16.2	24.3	**32.4** T_c = 20	28.4 T_c = 80, m = 16
PP	59.6	1.2	17.9	**20.2** T_c = 20	14.3 T_c = 80, m = 16
PK	90.1	3.2	**6.3**	2.4 T_c = 20	4.7 T_c = 80, m = 16
TT	53.9	34.2	42.1	40.8 T_c = 20	**44.7** T_c = 80, m = 16
NA	61.7	34.5	36.8	36.8 T_c = 20	**41.4** T_c = 80, m = 8
BS	40.4	24.6	24.6	**33.3** T_c = 20	-7.0 T_c = 80, m = 16

The recognition scores for subjects KJ and WK having symmetrical sloping high frequency loss, steeply sloping up to 1 kHz, were similar for all speech processed with different schemes. The recognition scores for unprocessed speech was 78.0 and 63.1 % for subjects KJ and WK respectively. Maximum relative improvements were obtained for Sp_AG-SS and Sp_AG-TS-20 for both the subjects: 21.8 and 20.9 % for KJ, and 31.46 and 33.7 % for WK for processing with Sp_AG-SS and Sp_AG-TS-20 respectively. The relative improvements for Sp_AG are 20.0 and 22.5 % respectively for KJ and WK. Among the combined splitting schemes, processing with sweep cycle of 80 ms with 16 shiftings provided better results: 18.2 % for KJ and 25.8 % for WK.

Subjects BS and PP, with severe steeply sloping high frequency loss, had recognition scores for unprocessed speech of 40.4 and 59.6 % respectively, for unprocessed speech. Both subjects had maximum relative improvement for Sp_AG-TS-20: 33.3 % for BS and 20.2 % for PP. For Sp_AG-SS, the relative improvements were 24.6 and 17.9 % respectively. Subject BS showed decrease in recognition scores for all the processing parameters in combined splitting, whereas subject PP had improvement for Sp_AG-CS-40/16 and Sp_AG-CS-80/16.

Subject NA with sloping moderately severe high frequency loss, had 61.7 % recognition scores for unprocessed speech. Relative improvements were 41.4 % for combined splitting Sp_AG-CS-80/4, 36.8 % for Sp_AG-SS and Sp_AG-TS-20, and 34.5 % for Sp_AG. Subject SK, with severe symmetrical hearing loss, had very low recognition score for unprocessed speech (20.6 %). Higher relative improvements were 138 % for combined splitting Sp_AG-CS-40/16, 121 % for Sp_AG-SS and 110 % for SP_AG-TS-20.

Subject SS with symmetrical severe mid frequency loss, has recognition score of 52.5 % for unprocessed speech. The relative improvements were the highest for temporal splitting: 32.4 % for Sp_AG-TS-20 and 29.7 % for Sp_AG-TS-40. Combined splitting Sp_AG-CS-80/16 and spectral splitting Sp_AG-SS provided relative improvements of 28.4 and 24.3 % respectively.

The subjects PK and TT had asymmetry in the loss (≈ 40 dB) for the two ears. PK had less loss at mid-frequencies in the better ear. The recognition scores for unprocessed speech were 90.1 and 53.9 % for PK and TT respectively. Subject PK had maximum relative improvement for Sp_AG-SS (6.3 %), whereas subject TT has maximum improvement for Sp_AG-CS-80/16 (44.7 %).

With the optimal processing scheme and condition for each of the thirteen subjects, the relative improvements in recognition scores ranged over 6 – 138 %, with an average of 35 % and median of 32 %. A comparison of relative improvements in recognition scores and response time showed that the optimal conditions based on these two measures were almost same.

6.5 Discussion

An overall evaluation of the three splitting schemes: spectral, temporal and combined for different processing parameters was carried out with combinations of parameters, initially on normal hearing subjects with simulated hearing loss (Experiment IV) and further with a limited set of parameters on subjects with moderate bilateral loss (Experiment V). Listening tests included an open-set evaluation of phonetically balanced monosyllabic words. Response time and recognition scores were used for comparing the processing schemes and processing conditions.

In Experiment IV involving seven normal hearing subjects with simulated loss, response time increased and recognition scores decreased with decrease in SNR. The relative improvements in both were higher at low SNR levels, indicating that binaural dichotic presentation was more effective under adverse listening conditions. The improvements for the response time and recognition scores had a similar pattern across the processing conditions. The highest improvements in response time as well as in recognition score were generally provided by spectral splitting. At SNR levels of –6 and –9 dB, the spectral splitting provided improvements of 16.0 and 13.8 % in response time and 101 and 152 % in recognition score. This indicates that processing resulted in a modest reduction in load on the perception process and a large improvement in speech perception. Combined splitting with sweep cycle of 80 ms and with 8 and 16 shiftings resulted in recognition scores close to spectral splitting. The highest improvements in recognition scores for temporal splitting were for T_c = 20 and 40 ms, and these were lower than that for combined splitting. Another comparison between the three schemes was made on the basis of SNR advantage for the same recognition score. For 60% score, temporal splitting with T_c = 20 and 40 ms provided an SNR advantage of approximately 1.5 dB. Combined splitting gave an advantage of approximately 4 dB for T_c =

40 ms and 4.5 dB for T_c = 80 ms for both m = 8 and 16. Spectral splitting with perceptually balanced comb filters provided an advantage of approximately 5 dB in SNR.

Thus from these experiments on normal hearing subjects with simulated loss, it may be concluded that spectral splitting is most effective and combined splitting with sweep cycle $\approx$ 80 ms with 8 or more shiftings gives improvements which are very close to spectral splitting.

In Experiment V involving listening tests on hearing impaired subjects, the processing conditions which provided very low or no improvements in Experiment IV were omitted. Adjustable magnitude response filter was cascaded with the splitting schemes and filtering with the adjustable magnitude response filter alone, also was evaluated. Thirteen subjects with moderate to severe bilateral sensorineural hearing loss participated in these listening tests. Subjects' response times and recognition scores were analyzed. The optimal processing condition based on maximum improvement in recognition score varied across subjects, and could be related to the individual's audiogram. Further, the optimal conditions based on improvements in recognition scores and response time were generally in agreement. Under optimal processing conditions, the relative improvements ranged from 6 to 138 % with an average of 35 % and median of 32 %. Adjustable magnitude response filtering, spectral splitting and temporal splitting provided improvement in recognition scores for all the hearing impaired subjects. Five subjects had highest improvement for spectral splitting. For temporal splitting scheme, inter-aural switching interval of 20 ms provided highest improvement for four subjects. Three subjects had highest improvement for combined splitting, with different combination of processing conditions (Sp_AG-CS-40/16, Sp_AG-CS-80/8, and Sp_AG-CS-80/16). For the two subjects with low frequency hearing loss, the improvement was maximum with adjustable magnitude response filter.

In conclusion, adjustable magnitude response filter gave highest improvement for subjects with low frequency loss. Spectral splitting provided maximum improvement for subjects with almost flat loss, sloping high frequency loss and asymmetrical loss with less loss at mid frequencies. Temporal splitting was highly preferred by subjects with sloping high frequency loss, and mid frequency loss. Combined splitting gave maximum improvement for subjects with severe symmetrical loss and moderately severe high frequency and mid frequency loss.

While all the thirteen subjects with sensorineural loss, who participated in these experiments, were benefited by dichotic presentation, the effectiveness of the schemes varied across the subjects. An effort has been made to relate these benefits with the audiograms of the subjects. However, as mentioned earlier in the discussion of the results of Experiment II (Section 4.9, last paragraph), the effectiveness of the various dichotic presentation schemes for an individual listener, is likely to be related to the nature and extent of spread of masking. Psychoacoustic assessment of the spread of masking for these subjects may be useful in relating the suitability of the schemes. As these assessments are often tedious, particularly for children and elderly persons with hearing impairment, it may be useful to develop speech processors with the option of selecting the scheme and parameters best suited for the individual.

Another point to be noted here is that combined splitting provides periodic stimulation and relaxation for all the surviving cells. Hence from a physiological perceptive, this is a better scheme than spectral splitting which stimulates cells in alternate bands. As the improvements obtained with combined splitting are very close to those with spectral splitting, it needs to be investigated whether combined splitting can be suitable after adequate practice, for all listeners with binaural aids.

Chapter 7

SUMMARY AND CONCLUSIONS

7.1 Introduction

Sensorineural hearing impairment is characterized by frequency dependent shifts in hearing threshold, loudness recruitment, reduced frequency and temporal resolution, and increased spectral and temporal masking. Due to increase in threshold of hearing with no corresponding increase in loudness discomfort level, the dynamic range reduces. Increased spectral masking causes smearing of spectral peaks and results in degradation of speech perception. Increased temporal masking leads to increase in forward and backward masking of weak acoustic segments by adjacent strong ones. Cues like voice-onset time, formant transition, and burst duration, which are important for the identification of consonants, get masked by the following or preceding vowel segment, resulting in degraded speech perception. Hence, intelligibility of speech gets degraded for persons with sensorineural hearing impairment.

Masking takes place primarily at the peripheral level, while integration of information takes place at higher levels in the auditory system. The effect of increased masking may be reduced by splitting speech into two complementary signals in such a way that signal components, likely to mask or get masked, are presented to different ears. Speech processing schemes, using a pair of comb filters with complementary magnitude responses have helped in reducing the effect of increased spectral masking (Lunner *et al.* 1993; Chaudhari and Pandey, 1998a, b). A scheme of temporal splitting, in which speech was switched between two ears using trapezoidal fading function with an inter-aural switching period of 20 ms, has helped in reducing the effect of increased temporal masking (Jangamashetti, 2003; Jangamashetti and Pandey, 2000b).

In the scheme of spectral splitting, the sensory cells corresponding to alternate bands of the basilar membrane are always stimulated, whereas sensory cells of the other bands do not receive stimulation. In the temporal splitting scheme, all the sensory cells of the two ears get relaxed alternately for some time. A combined splitting scheme was devised to provide all the sensory cells of the basilar membrane periodic relaxation from stimulation, and thereby achieve a simultaneous reduction in the adverse effects of increased spectral and temporal masking. In an investigation in our lab, combined splitting was evaluated using a pair of time-varying comb filters with pre-calculated sets of coefficients (Jangamashetti, 2003; Jangamashetti *et al.*, 2001; Pandey *et al.*, 2001), which were selected in steps for a cyclic sweeping of magnitude responses such that the pass bands of each of these comb filter pairs are shifted in a complementary manner along the frequency axis. The scheme was implemented for 2, 4, 8, and 16 shiftings and a sweep cycle of 20 ms.

The objective of the research was to investigate the use of binaural dichotic presentation for improving the speech perception by persons with moderate bilateral sensorineural hearing impairment and by normal hearing persons under adverse listening conditions, and to find the optimal splitting scheme and associated processing parameters. For spectral splitting, the auditory critical bandwidth based comb filters, designed with linear phase response and sharp inter-band transitions (Chaudhari and Pandey 1998a, b) were found to exhibit change in perceived loudness with frequency. It was decided to design and evaluate the comb filters with perceptually balanced magnitude response, to minimize the change in perceived intensity with frequency by reducing the passband ripple, controlling the variation in inter-band crossover gain, and improving stop-band attenuation. For temporal splitting, previous investigations (Jangamashetti, 2003; Jangamashetti and Pandey, 2000b), with different fading functions and duty cycles have shown that highest improvement was obtained with trapezoidal fading functions with 3 ms transition duration and 70 % duty cycle. The effect of various inter-aural switching periods with this fading function needed to be investigated. Combined splitting has earlier been investigated (Jangamashetti, 2003; Jangamashetti *et al.*, 2001; Pandey *et al.*, 2001) with cyclic sweeping of a set of comb filters, with sweep cycle duration of 20 ms. It was decided to investigate the effect of various sweep cycle durations with different number of shiftings. Finally, an overall evaluation of the three schemes with different processing parameters was carried out, to find the optimal processing scheme and the associated parameters.

Investigations involving implementation and evaluation of the dichotic presentation schemes were conducted in two phases. In the first phase, perceptually balanced comb filters were designed with improved magnitude responses to minimize the changes in intensity perception with frequency. The design of filters involved, adjustment in magnitude response at transition crossovers, reduction in passband ripple, and increase in stop-band attenuation. Spectral splitting scheme was implemented and evaluated using the comb filters. The second phase of investigations involved implementation and evaluation of the three dichotic splitting schemes namely, spectral splitting, temporal splitting, and combined splitting with associated processing parameters, on persons with normal hearing under simulated loss and on persons with moderate bilateral sensorineural hearing impairment. Spectral splitting was evaluated for perceptually balanced comb filters, since it was found to provide better improvement compared to comb filters with sharp transitions. On the basis of the previous investigations (Jangamashetti, 2003; Jangamashetti and Pandey, 2000b), it was decided to carry out the evaluation of temporal splitting scheme for trapezoidal fading functions with 70 % duty cycle and 3 ms transition duration. Since the earlier evaluation was carried out for inter-aural switching period of only 20 ms, the present evaluation was conducted to establish the optimal value of inter-aural switching periods. Combined splitting scheme was implemented with cyclic sweeping of perceptually balanced comb filters and evaluated for various combinations of sweep cycle duration and number of shiftings. An adjustable magnitude response filter with gain variation within ± 3 dB, to partly compensate for the frequency dependent shifts in hearing thresholds, alone and cascaded with other schemes, was also evaluated.

The first phase of the investigation, involving the design of perceptually balanced comb filters, experimental evaluation, and results, is presented in Chapter 4. The second phase of the investigation is presented in Chapter 5 and Chapter 6. Combined splitting scheme with time-varying comb filter was evaluated for different sweep cycle durations and number of shiftings, to find the optimal processing condition, and is presented in Chapter 5. The overall evaluation of the three binaural dichotic presentation schemes, spectral, temporal, and combined splitting with different processing parameters, carried out by conducting listening tests, is given in Chapter 6. Listening tests were conducted on normal hearing subjects, to narrow down the processing conditions. Subsequently, listening tests were conducted on hearing impaired subjects.

In the sections to follow, a summary of the investigations, conclusions based on the test results and a few suggestions for further work are presented.

7.2 Summary of the investigation

In the first phase, magnitude response of the comb filter for spectral splitting was investigated. The comb filters based on auditory critical bandwidths, used for spectral splitting, were earlier designed with sharp transitions between bands. These filters introduce perceptual distortion because of passband ripple of up to 4 dB, stop-band attenuation as low as 10 dB, and inter-band crossover range of 0 – 10 dB. Investigations involving monaural/binaural presentation, filter design, and listening tests showed that crossovers lying in the 4 – 6 dB range resulted in perceptual balance. Comb filters were designed as 256-coefficient linear phase FIR filters, to obtain crossovers within 4 – 6 dB and the passband ripple restricted to 1 dB. These provided stop-band attenuation of 30 dB. These comb filters when tested with slowly swept sine waves showed no noticeable perceptual distortion.

To evaluate the improvements, listening tests were conducted with both the comb filters, namely the one with sharp transitions and the perceptually balanced comb filters. Evaluation was first conducted on five normal hearing subjects with simulated hearing loss (Experiment I) and subsequently on five persons with moderate bilateral hearing loss (Experiment II). The tests involving consonant identification were carried out with 12 consonants in VCV context with vowel /*a*/, in order to evaluate reduction in the effect of masking. For subjects with bilateral sensorineural loss, adjustable magnitude response filters with gain variation within ±3 dB cascaded with the perceptually balanced comb filters were also used. A computerized listening test administration set-up was used to present stimuli, obtain the subject's response, and tabulate these in the form of stimulus-response confusion matrices. The response time, recognition score, and relative information transmission for consonantal features were used for the performance evaluation.

In the first part of the second phase of investigations, the scheme of combined splitting was implemented using time-varying perceptually balanced comb filters to obtain spectral and temporal splitting together. The number of combinations of processing conditions was narrowed down by conducting listening tests, involving perceptual quality ranking by three normal hearing subjects (Experiment III). A sine wave, linearly swept from 100 Hz to 5 kHz and back over 1 minute duration, was used as the test material. Processing conditions tested were combinations of sweep cycle durations T_c = 10, 20, 50, 80, and 100 ms and number of shiftings m = 2, 4, 8, and 16. Further listening tests were carried out with a speech passage, recorded by a male and a female speaker, on three normal hearing subjects with simulated hearing loss. Processing conditions with combinations of T_c = 20, 50, 80, and 100 ms and m = 8, and 16 were used. Perceptual quality ranking was carried out through paired comparisons at different SNR conditions.

The second part of the second phase involved evaluation of the three binaural dichotic presentation schemes namely: spectral splitting to reduce the effect of increased spectral masking, temporal splitting to reduce the effect of increased temporal masking, and combined splitting to simultaneously reduce the effects of increased spectral and temporal masking, and the associated processing conditions. Listening test consisted of words presented in a randomized order from a phonetically balanced list of monosyllabic words in Marathi, Hindi, or English. This was carried out initially on seven normal hearing subjects with simulated hearing loss (Experiment IV). After narrowing down the range of processing conditions, tests were conducted on thirteen subjects with bilateral sensorineural loss (Experiment V). Adjustable magnitude response filters and their cascading with the three dichotic schemes were also evaluated. A computerized test administration set-up was used to conduct the listening tests. Subject’s response time and recognition score were used for evaluation.

For listening tests on normal hearing listeners, simulation of hearing loss was carried out, by adding broadband Gaussian noise to the signal at different levels. Noise was added to speech signal based on short-time (≈ 10 ms) signal energy keeping the SNR constant, which resulted in no background noise during silence intervals. By analyzing the results from listening tests with unprocessed speech, it was verified that addition of broadband noise at various levels resulted in degradation in response time, recognition score, and relative information transmitted for different features almost similar to that of different levels of sensorineural hearing impairment. Subjects with bilateral sensorineural loss were selected primarily on the basis of their willingness to participate in at least one set of experiments. The nature and extent of loss varied widely across these subjects.

7.3 Conclusions

The evaluation conducted in the first phase showed that spectral splitting with comb filters helps in improving the response time, recognition score, and the reception of consonantal features. In Experiment I on normal hearing subjects with simulated hearing loss, the improvement in recognition score for both types of comb filters were statistically significant for low SNR conditions. For SNR levels of -3, -6, -9, -12 and -15 dB, the range of improvements was 18.8 – 24.0 % for comb filters with sharp transitions (SpA) and 20.7 – 29.1 % for perceptually balanced comb filters (SpB). Between the two types, the perceptually balanced comb filters showed additional and statistically significant improvement, increasing from 1.6 % at -3 dB to 4.8 % at -15 dB. There was a modest reduction in subject's response time, indicating a reduction in load on perception process. The improvement in relative information transmission was significant at low SNR levels. Perceptually balanced comb filters showed higher improvement in information transmission of consonantal features compared to comb filters with sharp transitions. The relative improvement of SpB over SpA was maximum for duration feature. On the basis of 80 % recognition score, the schemes SpA and SpB provided SNR advantages of approximately 7.5 and 9 dB respectively. For relative information transmission of consonantal features, the corresponding advantages were 8 and 9 dB for overall, 7 and 8 dB for place, 7.5 and 9.5 dB for frication, and 9 and 14 dB for duration feature.

The listening tests on subjects with bilateral sensorineural hearing impairment included a scheme (SpC), with a filter with magnitude response adjusted to partly compensate for the frequency dependent shifts in hearing thresholds cascaded with spectral splitting with perceptually balanced comb filter. All the three processing schemes evaluated in this investigation provided better speech reception in listening tests for persons with mild to severe bilateral sensorineural hearing impairment. The recognition scores for unprocessed speech varied from 51.0 to 88.7 %. The percentage relative improvements for SpA, SpB and SpC, ranged between 1.3 – 27.4, 8.7 – 39.2 and 12.0 – 47.7 respectively. These improvements were statistically significant in almost all cases. Processing with perceptually balanced comb filters provided better results compared to processing with comb filters with sharp transitions. Subjects with high frequency loss were more benefited by perceptually balanced comb filters. The scheme SpC provided better intelligibility for all subjects but with varying degrees of improvement. Subject with almost flat loss (particularly below 2 kHz) showed higher improvement in recognition score and relative information transmission of features (overall and place). Also persons with asymmetrical losses were benefited from this scheme. Spectral splitting helped in reducing the response time. The relative improvements were more for perceptually balanced comb filters. Cascading of the filter with adjustable magnitude response further reduced the response time of subject with asymmetrical loss.

Thus from the results of Experiment I and Experiment II, it can be concluded that dichotic presentation with spectral splitting modestly reduced the load on perception process and significantly improved speech perception by normal hearing persons under adverse listening conditions and by persons with moderate bilateral sensorineural hearing impairment. Further spectral splitting with perceptually balanced comb filters was more effective than spectral splitting with sharp transitions.

Listening tests in Experiment III for the evaluation of combined splitting scheme using time-varying comb filters to narrow down the choice of processing parameters showed that the preferred conditions were sweep cycle duration of $T_c = 50 - 80$ ms and number of shiftings m = 8 or 16. In Experiment IV for the overall evaluation carried out with all the three splitting schemes on normal hearing subjects with simulated loss, response time increased and recognition scores decreased with decrease in SNR. The relative improvements in both were higher at low SNR levels, indicating that binaural dichotic presentation was more effective under adverse listening conditions. The improvements for the response time and recognition scores had a similar pattern across the processing conditions. The highest improvements in response time as well as in recognition score were generally provided by spectral splitting. At SNR levels of -6 and -9 dB, the spectral splitting provided improvements of 16.0 and 13.8 % in response time and 101 and 152 % in recognition score. This indicates that processing resulted in a modest reduction in load on the perception process and a large improvement in speech perception. Combined splitting with sweep cycle of 80 ms and with 8 and 16 shiftings resulted in recognition scores close to spectral splitting. The highest improvements in recognition scores for temporal splitting were for $T_c = 20 - 40$ ms, and these were lower than that for combined splitting. Another comparison between the three schemes was made on the basis of SNR advantage for the same recognition score. For 60% score, temporal splitting with T_c = 20 and 40 ms provided an SNR advantage of approximately 1.5 dB. Combined splitting gave an advantage of approximately 4 dB for T_c = 40 ms and 4.5 dB for T_c = 80 ms for both *m* = 8 and 16. Spectral splitting with perceptually balanced comb filters provided an advantage of approximately 5 dB in SNR.

In Experiment V involving listening tests on thirteen bilateral hearing impaired subjects, the optimal processing condition based on maximum improvement in recognition score varied across subjects, and was found to be related to the individual's audiogram. Further, the optimal conditions based on improvements in recognition scores and response time were generally in agreement. Under optimal processing conditions, the relative improvements ranged from 5.8 to 138 % with an average of 35 % and median of 32 %. Adjustable magnitude response filtering, spectral splitting, and temporal splitting provided improvement in recognition scores for all the hearing impaired subjects. Five subjects had highest improvement for spectral splitting. For temporal splitting scheme, inter-aural switching interval of 20 ms provided highest improvement for four subjects. Three subjects had highest improvement for combined splitting, with different combination of processing conditions (Sp_AG-CS-40/16, Sp_AG-CS-80/8, and Sp_AG-CS-80/16). For the two subjects with low frequency hearing loss, the improvement was maximum with adjustable magnitude response filter.

Thus the overall conclusions from our implementation and evaluation of dichotic presentation schemes can be summarized as follows

1) The binaural dichotic presentation schemes with spectral, temporal, and combined splitting reduced the load on speech perception process and improved speech perception by the subjects with moderate to severe bilateral loss, as well as by normal hearing subjects under adverse listening conditions.

2) In case of spectral splitting, auditory critical bandwidth based comb filters with linear phase response and perceptually balanced magnitude response with 1 dB passband ripple, inter-band crossovers adjusted within 4 – 6 dB and stop-band attenuation greater than 30 dB, were found to be superior to the comb filters with sharp inter-band transitions

3) In case of temporal splitting, employing trapezoidal fading with 70 % duty cycle and 3 ms transition duration (to suppress the inter-aural switching distortion), the optimal inter-aural switching period was $T_c = 20 - 40$ *ms.*

4) In case of combined splitting with time-varying comb filters employing cyclic sweeping of a set of perceptually balanced comb filters, the best conditions were sweep cycle duration of $T_c = 40 - 80$ *ms and number of shiftings* $m = 8$ *and 16.*

5) Evaluation of the three binaural dichotic presentation schemes, with phonetically balanced words, showed that for normal hearing subjects with simulated loss, the highest improvements in response time as well as in recognition score were generally provided by spectral splitting, closely followed by combined splitting. For 60% recognition score, the SNR advantage with best processing parameters, was 1.5 dB for temporal splitting, 4 – 4.5 dB for combined splitting, and 5 dB for spectral splitting.

6) Evaluation of the dichotic schemes showed that for subjects with sensorineural loss, the optimal conditions based on improvements in recognition scores and response time were generally in agreement. Under optimal processing conditions, the relative improvements in recognition scores ranged from 6 to 110 % for temporal splitting, -30 to 138 % for combined splitting, 6 to 121 % for spectral splitting, and 1 to 66 % for adjustable magnitude response filtering. Across the subjects, the relative improvements under optimal processing scheme and parameters ranged over 6 – 138 %, with an average of 35 %.

7) Evaluation on hearing impaired subjects showed that the effectiveness of the schemes varied across the subjects. Adjustable magnitude response filter gave highest improvement for subjects with low frequency loss. Spectral splitting provided maximum improvement for subjects with almost flat loss, sloping high frequency loss, and asymmetrical loss with less loss at mid frequencies. Temporal splitting was found to be better for subjects with sloping high frequency loss, and mid frequency loss. Combined splitting gave maximum improvement for subjects with severe symmetrical loss and moderately severe high frequency and mid frequency loss.

The effectiveness of the various dichotic presentation schemes for an individual listener, is likely to be related to the nature and extent of spread of masking. Psychoacoustic assessment of the spread of masking for these subjects may be useful in relating the suitability of the schemes. As these assessments are often tedious, particularly for children and elderly persons with hearing impairment, it may be useful to develop speech processors with the option of selecting the scheme and parameters best suited for the individual.

It is to be noted that combined splitting provides periodic stimulation and relaxation for all the surviving cells. From a physiological perceptive, this is a better scheme than spectral splitting which stimulates cells in alternate bands. Hence there is a need to investigate whether it can be suitable, with adequate practice, for all listeners with binaural aids.

7.4 Suggestions for further work

In this investigation, binaural dichotic presentation schemes involving spectral, temporal, and combined splitting, have shown improvement in perception of speech for normal hearing subjects under adverse listening conditions and for subjects with moderate bilateral loss. The scheme of adjustable magnitude response filter used to partly compensate for the frequency dependent shifts in hearing thresholds cascaded with these schemes have provided further improvement, without interfering with the improvements obtained due to reduction in the effect of masking. The research has not addressed the problem of loudness recruitment. Use of multi-band compression to reduce the effects of frequency dependent loudness recruitment along with the binaural dichotic presentation schemes needs to be investigated.

In spectral splitting, perceptually balanced comb filters based on auditory critical bandwidths provided improvement in speech intelligibility, by reducing the effect of increased spectral masking. The optimal number of bands and bandwidths for the comb filters in spectral and combined splitting for different subjects also needs to be investigated.

In the binaural dichotic schemes investigated in our research, dynamic spectral and temporal characteristics of speech signal have not been considered. Hence a perceptually salient acoustic subsegment may get split temporally or spectrally, and this may contribute to an overall speech degradation. Thus the effective advantage of reduction in the effects of masking may be getting reduced. One way of overcoming this problem may be to use temporal splitting in which the splitting is dynamically synchronized to acoustic subsegment boundaries. Similarly spectral splitting may be adjusted along spectral valleys. Use of harmonic-plus–noise model based analysis-synthesis (Stylianou, 2001; Lehana and Pandey, 2004) may be investigated.

We have employed phonetically balanced word list for an overall comparison of the three types of splitting schemes. There is a need to extend the evaluation of the schemes with the use of larger number of subjects and more types of test material.

Our investigations have not addressed the issue of possible conflicts between source localization and dichotic presentation. It has been reported (Murase *et al.*, 2004) that dichotic presentation using a low pass and a high pass filter affected sound localization for hearing impaired subjects. Preliminary investigation with perceptually balanced comb filters, in our lab, has indicated that localization gets impaired for signals with tones, but not for broadband noise and speech signal which have approximately even distribution of spectral energy between the two ears. However, this needs to be investigated further.

Issues involved in the implementation of dichotic presentation schemes in wearable hearing aids, with the possibility of selecting the schemes and the processing conditions to suit the individual user, needs to be investigated. There is a possibility that same subjects listening in different environment, and speech from different speakers (male or female), may find different processing conditions more suitable. These issues can be properly investigated with the use of wearable hearing aids. In the implementation of binaural aids, each channel will have its own microphone, amplifier, ADC-processor-DAC, power amplifier, and receiver. For spectral splitting, processors for the two aids can work independently. In case of temporal splitting, the switching of the two fading functions has to be synchronized and for combined splitting, the sweeping of the comb filter responses has to be synchronized. Hence we have to either use a single processor, or use two processors with one of them acting as the master.

Appendix A

ANALYSIS OF SPEECH SPECTRA

Spectrogram is a visual representation of a signal, as a two dimensional plot of dynamic power spectrum with time on horizontal axis, frequency on vertical axis and spectral power (in dB) indicated by the gray scale of the display (Kent and Read, 1992; Koenig *et al.,* 1946; O'Shaughnessy, 2001; Rabiner and Schafer, 1978). Wideband spectrograms of speech signal, obtained using analysis filter bandwidth of approximately 300 Hz, having good time resolution, are useful to observe the formant transitions and the pitch periods corresponding to vocal fold closing. The formant frequencies and harmonics of the fundamental frequency can be observed in narrowband spectrogram, obtained using analysis filter bandwidth of approximately 45 Hz.

A spectrogram can be digitally generated by computing the time varying log magnitude spectrum, using short time Fourier transform (Oppenheim, 1970, O'Shaughnessy, 2001; Rabiner and Schafer, 1978). The short time Fourier transform (STFT) of signal $s(n)$ is obtained by computing the discrete Fourier transform of a windowed segment

$$S(n,k) = \sum_{m=0}^{N-1} w(m)s(n-m)e^{-j2\pi mk/N} \qquad 0 \le k \le N\text{-}1 \qquad \text{(A.1)}$$

where n represents the discrete time samples, k the discrete frequency, N the DFT size and $w(m)$ is a suitably chosen window sequence. The type of window, window length L, and the sampling rate f_s determine the frequency resolution of the spectrogram. An Hamming window sequence (Harris, 1978) of L samples ($L \le N$) has a frequency resolution, given as

$$F_{rs} \approx 1.36 f_s/L \qquad \text{(A.2)}$$

Thus with f_s = 10 k Sa/s, for F_{rs} = 300 Hz and 45 Hz, we need L of 45 and 300 samples respectively.

The frequency spectrum is computed for each windowed sequence sliding across the signal, the magnitude spectrum is calculated, converted to dB scale, and displayed as a function of time and frequency. For proper display of high frequency components in the limited dynamic range, pre-emphasis is done (Oppenheim, 1970; O'Shaughnessy, 2001).

Spectrographic analysis has been carried out using a PC based spectrographic analyzer developed earlier in our lab (Ratanpal, 2000), which itself is based on earlier developments (Baragi, 1996; Chaudhari and Pandey, 1999a; Prasad, 1996, Thomas *et al*., 1994, Thomas, 1996). The signal acquisition was done through the sound card and spectrograms were displayed on the SVGA monitor. The spectral analysis and display were carried out using program "`spec2000`" written in 'C'. This program invokes "`audio`" written in Visual C++, for recording and playing back of sound. There is a provision to obtain the spectrogram of a segment of the signal, which can be selected with cursors moved using arrow keys. Along with the spectrogram, corresponding speech signal also is displayed. The speech waveform is displayed using 500 × 45 pixels and the spectrogram with 500 × 128 pixels, above the speech waveform. The intensity display with 64 gray levels for the dynamic range of the spectrogram is displayed on the side of the spectrogram. The segment selected for spectrogram display is divided into 500 overlapping time frames. An L-point Hamming window is applied on each frame, after pre-emphasis. The magnitude spectrum is obtained using a 256-point FFT. Zero padding is done to convert the L-point (selected for the desired resolution) to make the sequence length as 256. The log magnitude of the 128 frequency samples in the FFT sequence is then calculated and is linearly mapped to 64 gray levels for display. Offset and scaling factor for linear mapping are set from the minimum and maximum dB values, which should be specified to get the best contrast in the display. From the displayed spectrogram, the magnitude at a particular frequency and time can be obtained by positioning the cursor. The spectrum at some particular time instant can also be displayed. The display, which includes the spectrogram, waveform, and the gray level intensity can be stored as a Postscript file with a resolution of 256 gray levels (Adobe systems, 1988). There is a provision to output the selected segment through the left, right, or both the channels of the sound card.

Figures A.1 and A.2 show the wideband and the narrowband spectrograms of swept sine wave, white noise, and syllable /*asa*/.

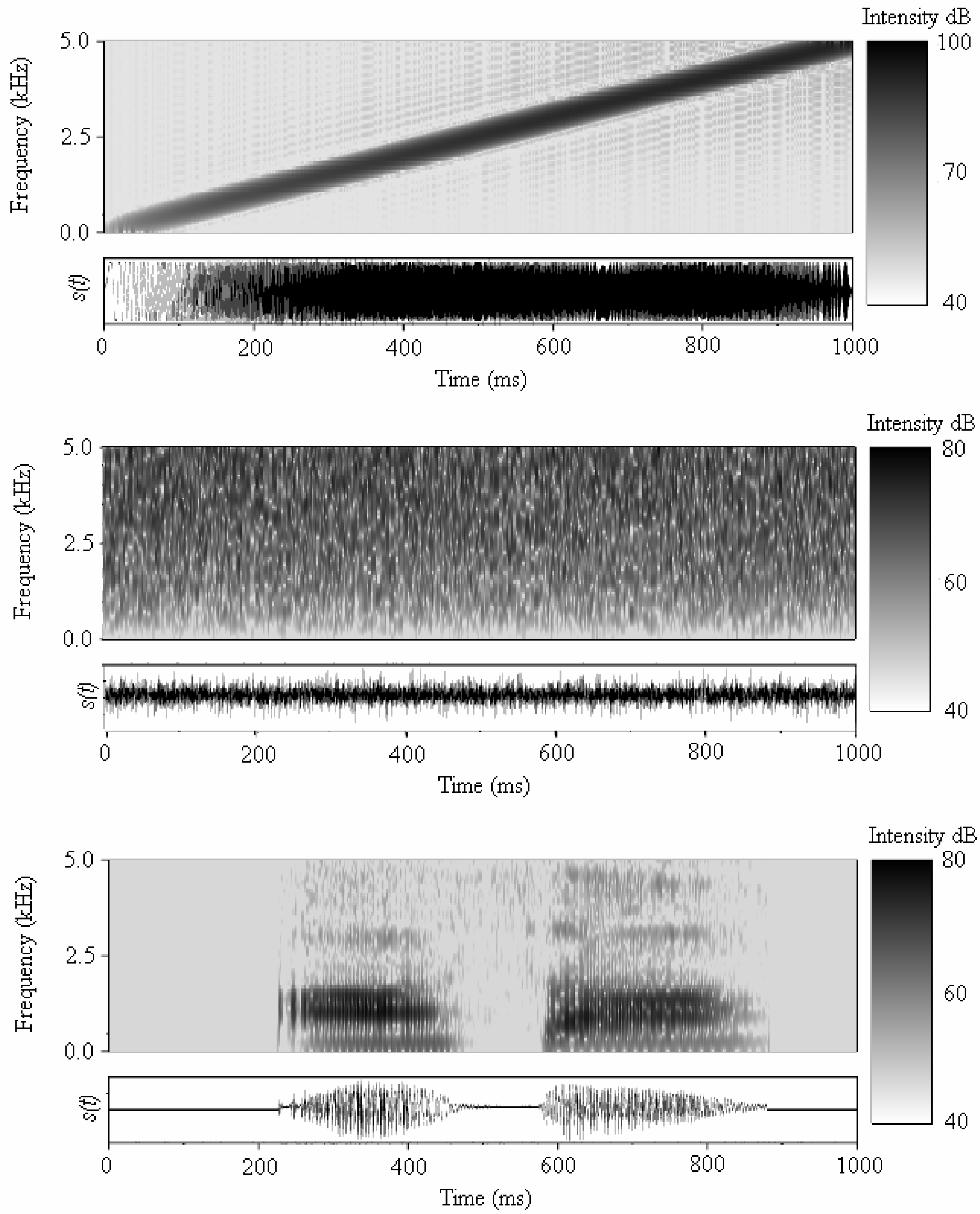


FIG. A.1. Wideband spectrograms of swept sine wave, white noise, and syllable /*asa*/.

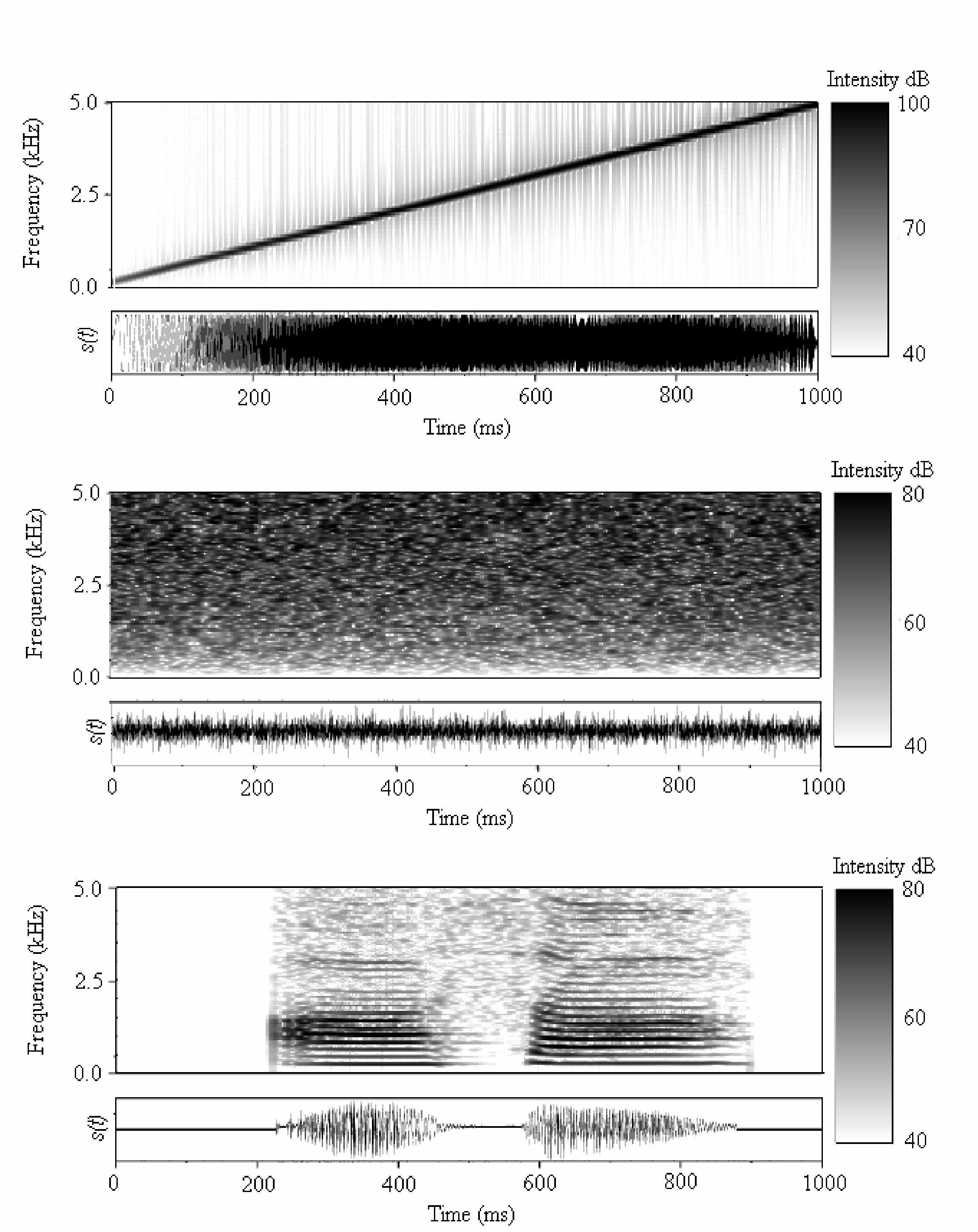


FIG. A.2. Narrowband spectrograms of swept sine wave, white noise, and syllable /*asa*/.

Appendix B

PERFORMANCE EVALUATION

B.1 Introduction

In the first phase of this research, evaluation was carried out to find effectiveness of spectral splitting to reduce the effect of increased spectral masking. Perceptually balanced comb filters were designed as described in Chapter 4 and Appendix C. The evaluation was conducted initially on normal hearing subjects with simulated hearing loss for unprocessed speech, spectral splitting with comb filters with sharp transitions, and spectral splitting with perceptually balanced comb filters. While conducting listening tests on hearing-impaired subjects, in addition to the above three schemes, a fourth scheme was considered in which an adjustable magnitude response filter was cascaded to perceptually balanced comb filters. The evaluation included closed set listening tests using twelve consonants /p, b, t, d, k, g, m, n, s, z, f, v/ in VCV context with vowel /*a*/. The response times, percentage correct recognition scores, and percentage relative improvements in response times and recognition scores were calculated. Information transmission analysis was carried out to get a measure relatively free of chance scoring or subject biases. It was also carried out to find the contribution of different consonantal features in the improvement.

In the second phase, an overall evaluation was carried out for spectral, temporal and combined splitting schemes with different processing parameters, by conducting an open set listening test using phonetically balanced monosyllabic words. The response time, recognition score, and percentage relative improvements in response times and recognition scores were calculated. This appendix describes these evaluation techniques.

B.2 Information transmission analysis

Recognition score (percent correct score) is an important indicator of effectiveness of speech processing scheme. However, it does not provide information about the distribution of incorrect responses. Stimulus-response scores represented in the form of confusion matrix, help in a detailed study of the pattern of errors (Miller and Nicely, 1955). The stimuli are shown along rows and responses along the columns and each entry in the cell represents either the frequencies or probabilities of stimulus-response pair. If a sets of n stimulus items and responses are represented as $\{x_1, x_2, \ldots x_n\}$ and $\{y_1, y_2, \ldots y_n\}$ respectively, and $N(x_i)$, $N(y_j)$, and $N(x_i, y_j)$ represent the frequencies of stimulus x_i, response y_j and the stimulus-response pair (x_i, y_j) respectively in a test of N observations, then the probabilities can be obtained as

$$p(x_i, y_j) = \frac{N(x_i, y_j)}{N} \tag{B.1}$$

$$p(x_i) = \frac{N(x_i)}{N} = \sum_{j=1}^{n} p(x_i, y_j) \tag{B.2}$$

$$p(y_j) = \frac{N(y_j)}{N} = \sum_{i=1}^{n} p(x_i, y_j) \tag{B.3}$$

In the confusion matrix, the diagonal elements indicate correct responses, whereas off-diagonal elements indicate errors. Sum of diagonal elements of a confusion matrix gives the recognition score R_s.

$$R_s = \sum_{i=1}^{n} p(x_i, y_i) \tag{B.4}$$

The recognition scores may be affected by the subject's biasing in response. It does not take into account the pattern of the distribution of errors in off-diagonal cells. These problems are taken care in the information transmission analysis (Miller and Nicely, 1955), which provides a measure 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 a stimulus x_i with response y_i, $I(x)$ and $I(y)$ respectively, are given by

$$I(x) = \text{MLP}(x) = -\sum_{i} p(x_i) \log_2 p(x_i) \quad \text{bits} \tag{B.5}$$

$$I(y) = \mathrm{MLP}(y) = -\sum_{j} p(y_j)\ \log_2\ p(y_j) \quad \text{bits} \tag{B.6}$$

MLP measure of covariance of stimulus-response is

$$I(x;y) = \mathrm{MLP}(x) + \mathrm{MLP}(y) - \mathrm{MLP}(xy)$$

$$= -\sum_{i,j} p(x_i\,, y_j)\ \log_2 \frac{p(x_i)\,p(y_j)}{p(x_i, y_j)} \quad \text{bits} \tag{B.7}$$

The relative information transmission is given by

$$I_{rel}\,(x;y) = I(x;y)/I(x) \tag{B.8}$$

Since $I(x) \geq I(x;y) \geq 0$; $1 \geq I_{rel}(x;y) \geq 0$

Relationship between the recognition score R_s and relative information transmitted I_{rel} for a special case in which the stimulus items have equal probabilities, correct responses are equally distributed among the diagonal cells, and incorrect responses are equally distributed among the off-diagonal cells is shown in Fig. B.1 (Pandey, 1987). For 'n' number of stimuli, cell entries are

$$p(x_i\,, y_j) = R_s/n\,, \qquad i = j$$

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

$$p(x_i) = p(y_j) = 1/n \tag{B.10}$$

It is observed that, chance scoring, ($R_s = 1/n$) corresponds to $I_{rel}(x;y) = 0$. Perfect scoring ($R_s = 1$) corresponds to $I_{rel}(x;y) = 1$. Zero scoring ($R_s = 0$) corresponds to a finite relative information transmission, $I_{rel}(x;y) = \log_2(n/(n\text{-}1))/\log_2(n)$. For $n \geq 12$, and recognition score in the range of 25 – 90 %, there is a near linear relationship between R_s and I_{rel} .

Generally, the stimulus items can be grouped according to features. Confusions within feature groups are more likely to occur than those among different groups. For learning the pattern of confusions, new confusion matrix can be formed by combining stimuli and responses into groups with common features. Relative information transmission of specific features can be obtained from these smaller confusion matrices derived from the original matrix.

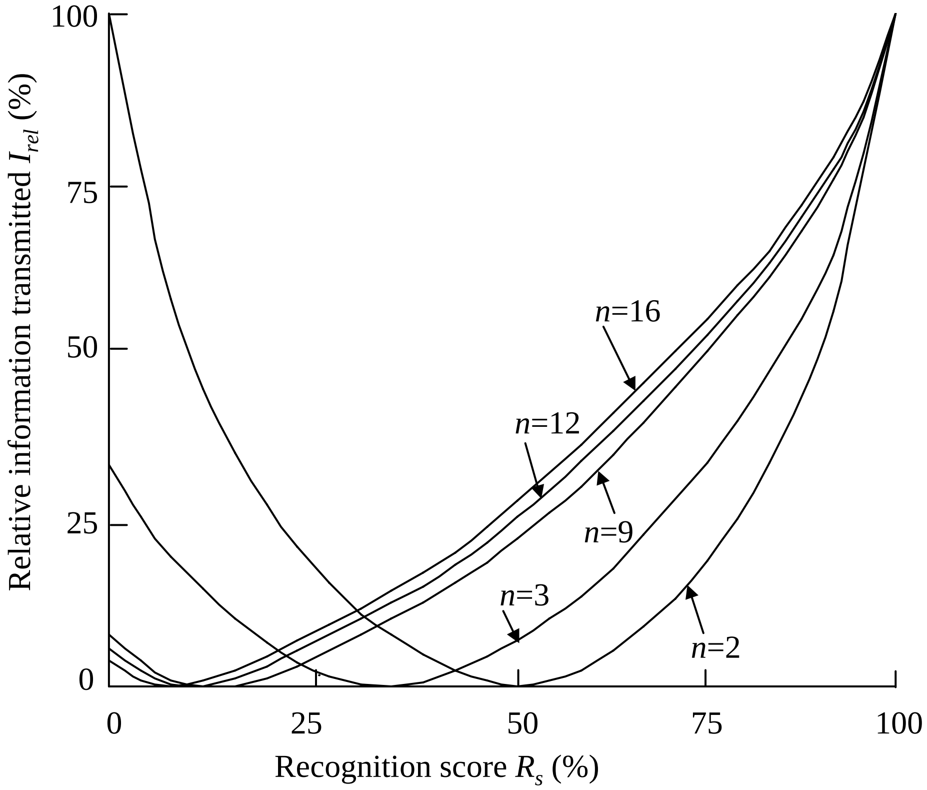


FIG. B.1. Relative information transmitted (I_{rel}) versus recognition score (R_s) for a special case when the correct responses are equally distributed among the diagonal cells and the errors are equally distributed among the off-diagonal cells in the stimulus-response confusion matrix (as given in Eqn. B.9). n = number of items. Adapted from Pandey (1987), Fig 4.1

B.3 Analysis of closed set listening tests with VCV monosyllables

Confusion matrices of several test runs carried out for an experimental condition, were combined using the program '`cummat`'. The confusion matrix was analyzed and recognition score and information transmission were calculated using program '`info`'. The program reads the necessary information, namely, number of test stimuli, total number of presentations, stimulus names, and cell entries from the input file. It obtains feature groupings from another file `infogr.dat`. The program gives overall percent scores, percent scores for different speech features, overall information transmitted and information transmission for speech features, and summary of the results.

Input file obtained by combining several confusion matrices using program '`cummat`' has the following information.

1) Number of stimuli *n*, total number of presentations.
2) "S/R", Names of test stimuli (each name can be two or three characters long separated by one or two spaces of response times in seconds, "+".
3) *n* lines of confusion matrix row data, each line having
   - two or three character stimulus name, *n* cell entries separated by one or two spaces, sum of the cell entries in the row.
4) "+", sum of the *n* row entries for each column separated by one or more spaces, total sum,
5) No. of files
   Date and time in each of the files.
6) Minimum, maximum, mean of percentage recognition score and standard deviation.
7) Minimum, maximum, mean, standard deviation of response times in seconds, total time in minutes

Feature groupings file `infogr.dat` contains the following information:

1) Number of test stimuli and number of feature classification,
2) Name of the test stimuli in the same order as in the input confusion matrix,
3) Feature classification information. Feature groups are represented by consecutive integers. The feature classification information consists of group numbers followed by feature and group labels.

The program outputs three files: `infosc.dat` (recognition scores), `infotr.dat` (information transmission analysis), and `infosu.dat` (summary of both the recognition scores and information transmission analysis).

A sample analysis and its results are given here, for listening tests conducted with processing condition SpC (speech processed with adjustable magnitude response filter cascaded with perceptually balanced comb filter) for subject SM.

output of '`cummat`'

STIMULUS-RESPONSE CONFUSION MATRIX

12 300

S/R	aPa	aBa	aTa	aDa	aKa	aGa	aMa	aNa	aSa	aZa	aFa	aVa	+
aPa	0	0	0	1	0	1	0	0	0	0	23	0	25
aBa	1	21	0	0	0	1	0	0	0	0	2	0	25
aTa	0	0	21	3	0	0	0	0	0	1	0	0	25
aDa	0	0	0	24	0	0	0	0	0	1	0	0	25
aKa	1	0	0	1	21	1	0	1	0	0	0	1	25
aGa	0	2	1	2	0	9	2	2	3	1	0	3	25
aMa	0	0	1	1	0	0	22	0	0	1	0	0	25
aNa	1	0	2	0	0	0	0	22	0	0	0	0	25
aSa	0	0	0	0	0	0	0	0	25	0	0	0	25
aZa	0	0	0	0	0	1	0	1	0	23	0	0	25
aFa	1	0	1	0	1	0	0	0	0	0	22	0	25
aVa	0	2	0	0	0	0	0	0	0	3	0	23	25

```
  +    3   25   26   32   22   13   24   26   28   27   47   27   300
No. of files: 5
--- 13-08-2002    11:07:41
--- 13-08-2002    11:17:52
--- 13-08-2002    11:28:16
--- 13-08-2002    11:47:24
--- 13-08-2002    12:07:30
71.7  81.7  77.7  4.8
5.44   6.38  6.03  0.41  3.90  9.0
```

1) File: `infogr.dat` (feature groupings for information transmission analysis).

```
12  6
 aPa  aBa  aTa  aDa  aKa  aGa  aMa  aNa  aSa  aZa  aFa  aVa
  0    0    0    0    0    0    0    0    1    1    0    0    Duration    SH  LO
  0    0    0    0    0    0    0    0    1    1    1    1    Fricative   ST  FR
  0    0    0    0    0    0    1    1    0    0    0    0    Nasality    OR  NA
  1    1    1    1    1    1    3    3    2    2    2    2    Manner      OS  FR  NA
  1    1    2    2    3    3    1    2    2    2    1    1    Place       FN  MD  BK
  0    1    0    1    0    1    1    1    0    1    0    1    Voicing     UV  VO
```

2) Analysis results: Percentage scores, file `infosc.dat`

```
NO. OF STIMULI:12
** PERCENTAGE SCORES **
* (12): OVERALL

   S/R  APa  aBa  aTa  aDa  aKa  aGa  aMa  aNa  aSa  aZa  aFa  aVa
   aPa   0    0    0    1    0    1    0    0    0    0   23    0
   aBa   1   21    0    0    0    1    0    0    0    0    2    0
   aTa   0    0   21    3    0    0    0    0    0    1    0    0
   aDa   0    0    0   24    0    0    0    0    0    1    0    0
   aKa   1    0    0    1   21    1    0    1    0    0    0    1
   aGa   0    2    1    2    0    9    2    2    3    1    0    3
   aMa   0    0    1    1    0    0   22    0    0    1    0    0
   aNa   1    0    2    0    0    0    0   22    0    0    0    0
   aSa   0    0    0    0    0    0    0    0   25    0    0    0
   aZa   0    0    0    0    0    1    0    1    0   23    0    0
   aFa   1    0    1    0    1    0    0    0    0    0   22    0
   aVa   0    2    0    0    0    0    0    0    0    3    0   23
         3   25   26   32   22   13   24   26   28   27   47   27

  Correct: 77.7

* (2): DURATION
```

```
S/R  SH   LO
SH   98    3
LO    5   96

Correct:  97.0

* (2): FRICATION
S/R  ST   FR
ST   82   18
FR    8   94

Correct:  85.7

* (2): NASALITY
S/R  OR   NA
OR   98    3
NA   12   88

Correct:  96.0

* (3): MANNER
S/R  OS   FR   NA
OS   74   24    4
FR    6   94    2
NA   10    3   88

Correct:  82.3

* (3): PLACE
S/R  FN   MD   BK
FN   94    5    3
MD    1   99    1
BK   16   22   63

Correct:  90.3

* (2): VOICING
S/R  UV   VO
UV   92    9
VO    7   94

Correct:  93.0
```

3) Analysis results: information transmission, file `infotr.dat`

```
** INFORMATION TRANSMISSION **

* (12): OVERALL
Stimulus info =  3.5832
Response info =  3.4472
```

```
Transn info =  2.6909
Perc transn =  75.1

* (2): DURATION
Stimulus info =  0.6497
Response info =  0.6870
Transn info =  0.4931
Perc transn =  75.9

* (2): FRICATION
Stimulus info =  0.9180
Response info =  0.9855
Transn info =  0.4102
Perc transn =  44.7

* (2): NASALITY
Stimulus info =  0.6497
Response info =  0.6497
Transn info =  0.4254
Perc transn =  65.5

* (3): MANNER
Stimulus info =  1.4587
Response info =  1.4823
Transn info =  0.7546
Perc transn =  51.7

* (3): PLACE
Stimulus info =  1.4829
Response info =  1.4011
Transn info =  0.9549
Perc transn =  64.4

* (2): VOICING
Stimulus info =  0.9796
Response info =  0.9812
Transn info =  0.6161
Perc transn =  62.9
```

4) Analysis of results: summary of information transmission, file `infosu.dat`

```
NO. OF STIMULI: 12

  COR  ERR     IS     IR     IT  RTR      FEATURE      N
   78  100   3.58   3.45   2.69   75      OVERALL     12
   97   13   0.65   0.69   0.49   76     DURATION      2
   86   64   0.92   0.99   0.41   45    FRICATION      2
   96   18   0.65   0.65   0.43   65     NASALITY      2
   82   79   1.46   1.48   0.75   52       MANNER      3
   90   43   1.48   1.40   0.95   64        PLACE      3
   93   31   0.98   0.98   0.62   63      VOICING      2
```

B.4 Analysis of open set listening tests with phonetically balanced monosyllables

For the second set of listening tests, sets of phonetically balanced monosyllabic words in three different languages (English, Hindi, Marathi) were used. In each language there were 50 to 60 words. The listening test procedure is given in Appendix C. A program ”stest”, written in ‘C’ was used for the automated listening test evaluation and analysis of the results. The program uses signal information files named as "ma_list", "hi_list" and "en_list", which contained the monosyllables for Marathi, Hindi, English languages respectively. The words listed in the chosen signal information file were appended by the processing identification and the corresponding speech stimuli were presented to the subject

A single listening test consisted of presentation of each listed word 3 times. The presentation order was randomized (controlled by the seed value). The maximum number of consecutive presentations for a monosyllable was kept two. The result of the test was stored in a text file with name “ssstt”, where ‘sss’ and ‘tt’ represent subject identification and test number respectively. The file contained the words presented in the order of presentation with “correct” and “wrong” written in the same line for correctly and wrongly responded words respectively along with the response time. The number of correctly responded words, the percentage recognition scores, and response times (total, average and standard deviation separately for correctly and wrongly responded words) were stored in this file.

A sample output file is given here, for listening tests conducted with phonetically balanced monosyllables in Marathi, for processing condition Sp_AG-TS-40 (adjustable magnitude response filter cascaded with the scheme temporal splitting with inter-aural switching period of 40 ms with trapezoidal fading function with 70 % duty cycle and 3 ms transition duration) for subject SM.

```
The response file name :sm05.txt
The processing identification :_xs
The seed number :7

THE WORDS PRESENTED IN SEQUENCES    RESPONSE    RES. TIME

            swad                    correct     2.409912
            hath                    correct     1.760010
```

```
phal                   correct     1.810059
      •
      •
      •

dhol                   correct     2.199951
mal                          correct     2.140137
kush                   correct     1.810059

Total words : 141
Number correctly recognized : 118
Percentage recognition score : 83.68

Total time : 334.22
Average time : 2.37

correct resp. time :   total 271.21       mean 2.29
                             std. dev. 0.745
                             maximum 9.83      minimum 1.70

wrong resp. time :           total 63.01  mean 2.74
                             std. dev. 0.424
                             maximum 4.17      minimum 2.26
```

Appendix C

COMB FILTER DESIGN

C.1 Filter design method

The comb filters for binaural splitting of speech were designed as linear phase FIR filters. This design was carried out using frequency sampling techniques (Ifeachor and Jervis, 1997; Oppenheim *et al.*, 1999; Proakis and Manolakis, 1997; Rabiner and Gold, 1998), since it is a convenient technique for designing filters with frequency response not conforming to a prototype.

In frequency sampling technique, the desired frequency response $H_d(e^{j\omega})$ is sampled at a set of uniformly spaced frequencies,

$$\omega = 2\pi k/N$$

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

$$0, 1,.2, \bullet\bullet\bullet, \frac{N-1}{2} \qquad \text{for } N \text{ even} \tag{C.1}$$

The filter coefficients are the IDFT of this set of samples.

$$h(n) = \frac{1}{N}\sum_{k=0}^{N-1} H(k) e^{j2\pi nk/N} \tag{C.2}$$

The filter response $H(e^{j\omega})$ will coincide with $H_d(e^{j\omega})$ at $\omega = 2\pi k/N$, but may have large variation at other frequencies. The design process basically involves iteratively modifying

$H_d(e^{j\omega})$ specifications, such that $H(e^{j\omega})$ matches with the original $H_d(e^{j\omega})$ under certain design constraints.

In our design, the filter has a linear phase response, and we specify only the magnitude response. Thus the specified magnitude response is combined with a linear phase response corresponding to a delay of *N*/*2* samples.

$$H(k) = |H(k)| e^{-j2\pi k(N-1)/2N} \tag{C.3}$$

This is used to calculate *N*–point FIR impulse response $h(n)$ by (C.2). From the impulse response, frequency response $H(e^{j\omega})$ is calculated at *L* uniformly spaced samples as

$$H(e^{j2\pi k/L}) = \sum_{n=0}^{L-1} h(n) e^{-j2\pi kn/L}, \qquad k = 0, 1, 2, \text{ - - - - - } L\text{-}1. \tag{C.4}$$

A program "`modfilt`" developed by Kasthuri (1997) and modified by Ratanpal (2000) has been used for designing the filter. For *N*-coefficient filter, the *N*/*2* uniformly spaced frequency samples of the magnitude response are entered graphically. The magnitude of the samples can be varied in steps of 0.006 on a linear scale. From the magnitude response, *N*-point sampled frequency response is obtained by (C.3). *N*-point impulse response samples are calculated using (C.2) and these are stored as filter coefficients in a text file. The interpolated frequency response is obtained by equation (C.4) for *L* uniformly spaced samples. The value of *L* can be selected between *N* and 17*N*. The interpolated magnitude response in linear scale can be observed graphically along with the specified response. There is a provision to store the interpolated frequency response in linear scale as well as in dB scale separately in "`.ps`" format. A previously calculated set of filter coefficients can also be used as input for starting the filter design.

C.2 Optimizing comb filters

Using frequency sampling technique, comb filters were designed with 256 coefficients with sharp transitions between pass and stop bands initially. The resulting magnitude responses exhibit passband and stop-band ripples. Further there was no control over the gain cross-over (transition) point between two corresponding comb filters.

Rabiner *et al.*, (1970) described a method for improving the filter response with a reduction in passband ripple and an increase in minimum attenuation in stop bands, by sacrificing the sharp transition from pass band to stop band. The samples in the pass band were considered as constrained samples taking a magnitude of 1, some of samples lying close to the pass band were taken as unconstrained taking values between 0 and 1, and the remaining samples of the stop band were constrained with value 0. The magnitude of the unconstrained samples lying in the transition region was varied to obtain maximum stop band-attenuation. They applied this method to several prototype filters and obtained reduction in approximation error. They used linear programming based optimization technique and determined the magnitude of the transition band sample for different low pass and band pass filters with different number of coefficients and with one, two, and three transition samples.

For optimizing the comb filters, the sharp transitions were sacrificed, and stop band samples lying close to pass bands were considered as transition (unconstrained) samples and their magnitudes were varied. Since each comb filter has 18 transitions, it was difficult to find the optimal magnitude for each of the unconstrained transition samples. For deciding the initial value of the unconstrained sample in the iteration process, each of 18 pass bands was considered separately, i.e a band pass filter was designed with each of the bands, with one or two transition samples separately on either side of the pass band. The transition samples were adjusted to obtain minimum ripple in the pass band and maximum attenuation in the stop band.

After conducting listening tests for perceptual balance in monaural/binaural presentations, it was been found that perceptual balance was obtained with the binaural level 4 to 6 dB less than the monaural level. So the crossover points were constrained to lie between 4 to 6 dB. The passband ripple was constrained to 1 dB to avoid perceptual changes and the stop-band attenuation was maximized. The steps involved in the design of perceptually balanced comb filters were the following.

(i) Feed in the magnitude at each pass band and stop band edges in the graphical interface with uniformly sampled frequency points. The samples in the pass band and stop band were given a magnitude of 1 and 0 respectively. For the unconstrained samples in the transition region, initial values obtained considering each auditory band pass filters separately was taken. For low frequency bands, one transition sample only could be accommodated.

(ii) Obtain the filter coefficients.

(iii) Observe the passband ripple and stop-band attenuation in the interpolated frequency response.

(iv) Adjust the transition samples and repeat the steps (ii) and (iii) until passband ripple is less than 1 dB, with maximum stop-band attenuation.

(v) Repeat steps (i) to (iv) for the complementary magnitude response.

(vi) The magnitude response of the two comb filters were overlapped and all the crossover points were checked to find whether it is lying between 4 to 6 dB below the pass band response. If not repeat steps (i) to (v).

Automating the process of the design of the comb filters with crossover adjusted to lie in the specified range was very difficult. So the above mentioned iterative method was used for designing the comb filters. Further the possibility of applying Taguchi techniques (Phadke, 1989) for optimizing the magnitude response of the comb filters was explored. Hence the filter design program was modified. The new program “`amodf11c`”, used an input file in “`.txt`” format, which contained the information about the position of the samples of the magnitude response. These details were stored in the file, as starting sample and ending sample for each pass and stop bands. The program calculates the maximum pass band ripple of the interpolated magnitude response, for various possible values of transition samples varied in steps of 0.2 on the linear scale. The program finds the maximum, average, and variance for the pass band ripple and the minimum, average, and variance for the stop-band attenuation. For finding the magnitude at crossovers, “`ccomb2`” was used. For optimizing the filter design for different combination of transition sample magnitude, the maximum, average, and variance of the passband ripple, the minimum, average and variance of the stop-band attenuation and the different crossovers at the different transitions were considered. Applying Taguchi techniques to this data, the optimal magnitude for transition samples for separate band pass filters could be found. For a pair of comb filters with approximately 60 transition samples, with due consideration to ripple at all pass bands, attenuation at all stop bands and gain adjustment at seventeen crossover points, the technique seemed to be too tedious. The method needs to be explored further.

C.3 Listening tests for perceptual balance in monaural/binaural presentation

To find the level difference in monaural-binaural presentation for perceptual balance, various stimuli (tones, noise, and vowel /*a*/) were presented monaurally and binaurally one after the other with a presentation gap. The procedure is described in Chapter 4 and a flow chart of the steps involved is shown in Fig C.1. The program "`mono_bin`" written in C was used to present the signals to the subject through headphones (Telephonics TDH-39P). The test signals stored in the PC were passed through two 12-bit DAC channels of PC based PCL-208 data acquisition card (Dynalog Micro Systems, 1989), pair of smoothening low pass filter, and two audio amplifiers, with separate gain adjustment for each ear. The stimuli were of 1 s duration, with 1 s interval between monaural and binaural presentations. Presentations were repeated with a time gap of 2 s.

C.4 Combined splitting using time-varying comb filters

Combined splitting was carried using time-varying comb filters with 2, 4, 8, and 16 shiftings. The center frequencies and bandwidths for the eighteen critical bands used for 16 shiftings are given in Table C.1. For combined splitting with 8 shiftings the bands corresponding to the alternate center frequencies are used. Similarly bandwidths are selected for 4 and 2 shiftings. The eighteen critical bands used for spectral splitting are shown in bold.

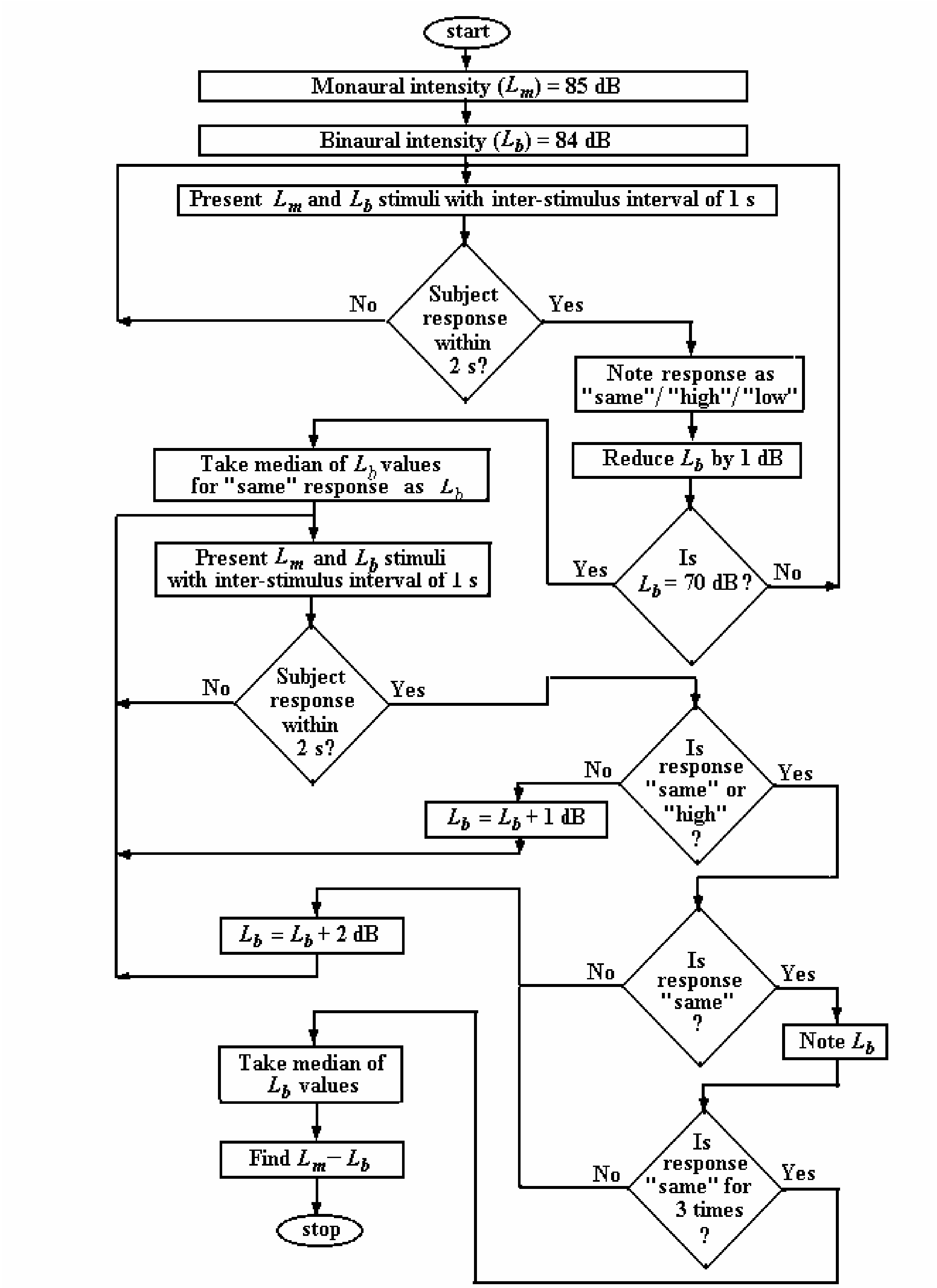


FIG. C.1. Flow chart showing the steps involved in finding the level of monoaural/ binaural presentation for perceptual balance

TABLE C.1. Bandwidths for 18 critical band filters with 16 shiftings between odd-even-odd bands

Filter for left ear			Filter for right ear		
Band	Center frequency (kHz)	Pass band (kHz)	Band	Center frequency (kHz)	Pass band (kHz)
1	0.150	0.100-0.200	2	0.250	0.200-0.300
	0.163	0.113-0.213		0.263	0.213-0.313
	0.175	0.125-0.225		0.275	0.225-0.325
	0.188	0.138-0.238		0.288	0.238-0.338
	0.200	0.150-0.250		0.300	0.250-0.350
	0.213	0.163-0.263		0.313	0.263-0.363
	0.225	0.175-0.275		0.325	0.275-0.375
	0.238	0.188-0.288		0.338	0.288-0.388
	0.250	0.200-0.300		0.150	0.100-0.200
	0.263	0.213-0.313		0.163	0.113-0.213
	0.275	0.225-0.325		0.175	0.125-0.225
	0.288	0.238-0.338		0.188	0.138-0.238
	0.300	0.250-0.350		0.200	0.150-0.250
	0.313	0.263-0.363		0.213	0.163-0.263
	0.325	0.275-0.375		0.225	0.175-0.275
	0.338	0.288-0.388		0.238	0.188-0.288
3	0.350	0.300-0.400	4	0.450	0.395-0.505
	0.363	0.313-0.413		0.465	0.410-0.520
	0.375	0.325-0.425		0.480	0.425-0.536
	0.387	0.338-0.438		0.495	0.439-0.551
	0.400	0.350-0.450		0.510	0.454-0.566
	0.413	0.362-0.464		0.525	0.468-0.582
	0.425	0.375-0.480		0.540	0.482-0.598
	0.438	0.385-0.492		0.555	0.496-0.614
	0.450	0.395-0.505		0.350	0.300-0.400
	0.465	0.410-0.520		0.363	0.313-0.413
	0.480	0.425-0.536		0.375	0.325-0.425
	0.495	0.439-0.551		0.387	0.338-0.438
	0.510	0.454-0.566		0.400	0.350-0.450
	0.525	0.468-0.582		0.413	0.362-0.464
	0.540	0.482-0.598		0.425	0.375-0.480
	0.555	0.496-0.614		0.438	0.385-0.492
5	0.570	0.510-0.630	6	0.704	0.632-0.773
	0.586	0.526-0.646		0.721	0.648-0.790
	0.602	0.542-0.662		0.738	0.666-0.808
	0.618	0.558-0.677		0.756	0.682-0.825
	0.634	0.574-0.693		0.773	0.699-0.842
	0.651	0.589-0.713		0.790	0.717-0.862
	0.669	0.603-0.733		0.806	0.735-0.881
	0.687	0.618-0.753		0.823	0.752-0.901
	0.704	0.632-0.773		0.570	0.510-0.630
	0.721	0.648-0.790		0.586	0.526-0.646
	0.738	0.666-0.808		0.602	0.542-0.662

TABLE C.1. (Contd.)

Filter for left ear		
Band	Center frequency (kHz)	Pass band (kHz)
	0.756	0.682-0.825
	0.773	0.699-0.842
	0.790	0.717-0.862
	0.806	0.735-0.881
	0.823	0.752-0.901
7	0.840	0.770-0.920
	0.858	0.787-0.938
	0.877	0.805-0.956
	0.895	0.822-0.974
	0.913	0.839-0.992
	0.936	0.859-1.015
	0.959	0.880-1.037
	0.981	0.900-1.060
	1.004	0.920-1.082
	1.024	0.940-1.105
	1.044	0.959-1.127
	1.064	0.979-1.150
	1.084	0.998-1.172
	1.106	1.019-1.197
	1.127	1.039-1.221
	1.149	1.060-1.246
9	1.170	1.080-1.270
	1.195	1.105-1.295
	1.221	1.130-1.321
	1.246	1.154-1.346
	1.271	1.179-1.371
	1.296	1.202-1.398
	1.322	1.225-1.426
	1.347	1.248-1.453
	1.372	1.271-1.480
	1.400	1.296-1.509
	1.428	1.322-1.537
	1.455	1.347-1.566
	1.483	1.372-1.594
	1.513	1.399-1.626
	1.542	1.426-1.658
	1.572	1.453-1.690
11	1.601	1.480-1.722
	1.631	1.510-1.754
	1.662	1.540-1.787
	1.692	1.570-1.819
	1.722	1.599-1.851
	1.757	1.629-1.889
	1.792	1.660-1.926

Filter for right ear		
Band	Center frequency (kHz)	Pass band (kHz)
	0.618	0.558-0.677
	0.634	0.574-0.693
	0.651	0.589-0.713
	0.669	0.603-0.733
	0.687	0.618-0.753
8	1.004	0.920-1.082
	1.024	0.940-1.105
	1.044	0.959-1.127
	1.064	0.979-1.150
	1.084	0.998-1.172
	1.106	1.019-1.197
	1.127	1.039-1.221
	1.149	1.060-1.246
	0.840	0.770-0.920
	0.858	0.787-0.938
	0.877	0.805-0.956
	0.895	0.822-0.974
	0.913	0.839-0.992
	0.936	0.859-1.015
	0.959	0.880-1.037
	0.981	0.900-1.060
10	1.372	1.271-1.480
	1.400	1.296-1.509
	1.428	1.322-1.537
	1.455	1.347-1.566
	1.483-	1.372-1.594
	1.513	1.399-1.626
	1.542	1.426-1.658
	1.572	1.453-1.690
	1.170	1.080-1.270
	1.195	1.105-1.295
	1.221	1.130-1.321
	1.246	1.154-1.346
	1.271	1.179-1.371
	1.296	1.202-1.398
	1.322	1.225-1.426
	1.347	1.248-1.453
12	1.861	1.720-2.000
	1.896	1.753-2.038
	1.931	1.786-2.075
	1.966	1.818-2.113
	2.001	1.851-2.150
	2.041	1.888-2.193
	2.081	1.926-2.236

TABLE C.1. (Contd.)

Filter for left ear		
Band	Center frequency (kHz)	Pass band (kHz)
	1.826	1.690-1.963
	1.861	1.720-2.000
	1.896	1.753-2.038
	1.931	1.786-2.075
	1.966	1.818-2.113
	2.001	1.851-2.150
	2.041	1.888-2.193
	2.081	1.926-2.236
	2.120	1.963-2.278
13	2.160	2.000-2.321
	2.203	2.038-2.369
	2.246	2.076-2.417
	2.288	2.113-2.464
	2.331	2.151-2.512
	2.376	2.193-2.560
	2.421	2.235-2.607
	2.465	2.277-2.654
	2.510	2.319-2.701
	2.560	2.365-2.756
	2.611	2.410-2.811
	2.661	2.456-2.866
	2.711	2.501-2.921
	2.764	2.551-2.979
	2.817	2.601-3.037
	2.870	2.651-3.094
15	2.923	2.701-3.152
	2.982	2.753-3.219
	3.042	2.806-3.286
	3.101	2.858-3.353
	3.160	2.910-3.420
	3.225	2.970-3.490
	3.290	3.030-3.560
	3.355	3.090-3.630
	3.420	3.150-3.700
	3.495	3.215-3.783
	3.570	3.280-3.865
	3.645	3.345-3.948
	3.720	3.410-4.030
	3.801	3.483-4.123
	3.885	3.555-4.215
	3.968	3.628-4.308
17	4.050	3.700-4.400
	4.135	3.778-4.493

Filter for right ear		
Band	Center frequency (kHz)	Pass band (kHz)
	2.120	1.963-2.278
	1.601	1.480-1.722
	1.631	1.510-1.754
	1.662	1.540-1.787
	1.692	1.570-1.819
	1.722	1.599-1.851
	1.757	1.629-1.889
	1.792	1.660-1.926
	1.826	1.690-1.963
14	2.510	2.319-2.701
	2.560	2.365-2.756
	2.611	2.410-2.811
	2.661	2.456-2.866
	2.711	2.501-2.921
	2.764	2.551-2.979
	2.817	2.601-3.037
	2.870	2.651-3.094
	2.160	2.000-2.321
	2.203	2.038-2.369
	2.246	2.076-2.417
	2.288	2.113-2.464
	2.331	2.151-2.512
	2.376	2.193-2.560
	2.421	2.235-2.607
	2.465	2.277-2.654
16	3.420	3.150-3.700
	3.495	3.215-3.783
	3.570	3.280-3.865
	3.645	3.345-3.948
	3.720	3.410-4.030
	3.801	3.483-4.123
	3.885	3.555-4.215
	3.968	3.628-4.308
	2.923	2.701-3.152
	2.982	2.753-3.219
	3.042	2.806-3.286
	3.101	2.858-3.353
	3.160	2.910-3.420
	3.225	2.970-3.490
	3.290	3.030-3.560
	3.355	3.090-3.630
18	4.700	4.400-5.000
	4.792	4.455-5.130

TABLE C.1. . (Contd.)

Filter for left ear		
Band	Center frequency (kHz)	Pass band (kHz)
	4.220	3.855-4.585
	4.305	3.932-4.678
	4.390	4.010-4.770
	4.468	4.108-4.823
	4.545	4.205-4.885
	4.623	4.303-4.943
	4.700	4.400-5.000
	4.792	4.455-5.130
	4.885	4.510-5.560
	4.978	4.565-5.390
	5.070	4.620-5.520
	5.162	4.707-5.617
	5.255	4.765-5.715
	5.346	4.881-5.811

Filter for left ear		
Band	Center frequency (kHz)	Pass band (kHz)
	4.885	4.565-5.390
	4.978	4.620-5.520
	5.070	4.510-5.560
	5.162	4.707-5.617
	5.255	4.765-5.715
	5.346	4.881-5.811
	4.050	3.700-4.400
	4.135	3.778-4.493
	4.220	3.855-4.585
	4.305	3.932-4.678
	4.390	4.010-4.770
	4.468	4.108-4.823
	4.545	4.205-4.885
	4.623	4.303-4.943

Appendix D

HARDWARE SET-UP

D.1 Introduction

In the first set of experiments, i.e. to evaluate the perceptually balanced comb filters, closed set listening tests were carried using twelve consonants /p, b, t, d, k, g, m, n, s, z, f, v/ in VCV context with vowel /*a*/. The set of syllables recorded earlier in the lab by Jangamashetti (2003) was used. Further experimental evaluation for finding the optimal value of number of shiftings and sweep cycle duration with time varying comb filters was carried out using phonetically balanced monosyllabic words. The set-up for speech acquisition and listening test are presented in the next two sections. The last section describes calibration of headphones.

D.2 Set-up for signal acquisition

The signal acquisition set-up for recording nonsense syllables for the first set of experiments is shown in Fig D.1. The microphone (B&K 4176) attached to the sound level meter (B&K 2235) was used for acquiring the speech signal. The sound level meter had built-in preamplifier, an input attenuator, a frequency weighting filter, and a buffer amplifier, and provided ac output. The recording was done with A-weighting filter. The range switch was kept in 20-90 dB range, which gave the ac signal output of 1 V rms for 1 kHz acoustic tone of 90 dB SPL. The ac signal from the sound level meter was passed through an amplifier, an anti-aliasing low pass filter and analog-to-digital converter (ADC) of the PC based DSP card TI/TMS320C25. The signal from the sound level meter was amplified to fully use the input range of the ADC. An anti-aliasing low pass filter was used to limit the signal within 5 kHz. The filter was an active elliptic filter of the 7th order with cut-off frequency of 4.6 kHz, passband ripple of 0.3 dB, stop-band attenuation of 40 dB and transition band of 4.6-5 kHz. The speaker uttered each syllable three times, and the clearest one was selected as part of the test material.

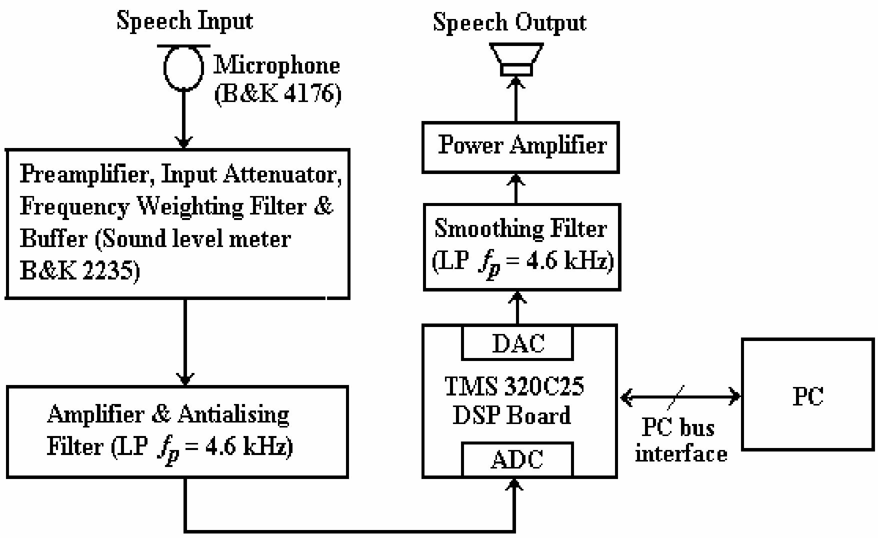


FIG. D.1. Experimental set-up for acquisition and analysis of speech using DSP card TMS 320C25

For the second set of listening tests, sets of phonetically balanced monosyllabic words in different languages (English, Hindi, Marathi) were used. In each language, there were 50 to 60 words. The word lists were obtained from Ali Yaver Jung National Institute for Hearing Handicapped, Mumbai, where these lists were used to evaluate the speech discriminating capacity of hearing impaired persons. For recording these words, multimedia sound card attached to the PC was used instead of ADC attached to DSP boards. The set-up is shown in Fig D.2. Software "`goldwave`" was used for this purpose. The line-in volume control was kept at 40% while recording. The 16-bit ADC of the sound card has comparatively linear characteristics for input voltages up to 1.2 V pp at 40% line-in volume. A pure tone of 1 kHz and 1.2 V p-p resulted in integers over ± 26627 range. The sound level meter was set to C-weighting filter, while recording. The range switch was kept in 40-110 dB range. At 90 dB SPL, the ac signal output was 0.1 V rms for 1 kHz tone. The ac signal from the sound level meter was passed through an amplifier, an anti-aliasing low pass filter and to the line-in terminal of the sound card attached to the PC. The speaker was instructed to speak at normal conversational level, and the distance of the microphone was adjusted such that the sound level meter indicated 70-75 dB SPL. After amplification and filtering, the maximum level of the signal was less than 0.9 V pp. Out of the three utterance of the same word, the cleanest one was included in the test material.

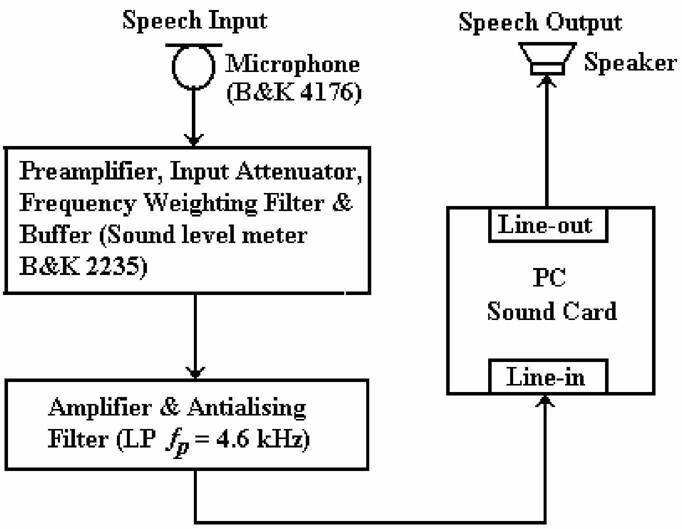


FIG. D.2. Experimental set-up for acquisition and analysis of speech using PC sound card.

D.3 Listening test set-up

The set-up used for conducting listening tests is shown in Fig. D.3. The subject seated inside the acoustically isolated room had a terminal to respond and headphones (TDH-39P) to listen the speech presented. The PC kept outside the acoustically isolated room was interfaced to the inside terminal through RS-232 asynchronous serial port. The speech stored in the PC was presented to the headphones at 10 k samples/s through the two independent 12-bit digital-to analog (DAC) channels of the PCL-208 data acquisition card, followed by a pair of smoothing low pass filters and audio amplifiers. The entire automated listening test was controlled by the PC. It included presenting the speech signals to headphones, displaying the stimuli information and obtaining responses along with response times for each presentation, etc. The subject terminal was used for displaying the response choices and obtaining responses from its keyboard. The smoothing filters were 7th order elliptical filters. The audio amplifiers were class B push-pull amplifiers, with logarithmic volume control.

The listening test set-up of Fig D.3 using PCL-208 data acquisition card, external audio amplifier and smoothing filter, was not a portable set-up. Since there was difficulty in getting sufficient number of hearing impaired subjects to come to our laboratory for participating in listening tests, the set-up was modified to use sound card of a Notebook PC to present speech to the subject. The line-out of the sound card was connected to the headphones via an audio amplifier or power amplifier in a dual channel audiometer. Provision was made to conduct the listening test with and without a subject terminal. The test procedure is described in Section E.10 as part of the software description. The listening set-up is shown in Fig D.4.

D.4 Calibration of headphones

In the listening tests, TDH-39P headphones were used. It was necessary to present the speech stimuli at a calibrated sound level. The set-up used to obtain the electro-acoustic characteristics and calibration of headphones is shown in Fig D.5. Function/arbitrary waveform generator (HP33120A) with the audio amplifier provided the input to the artificial ear (B & K 4153) fitted with microphone (B & K 4176), followed by the sound level meter (B & K 2235), headphone (TDH-39P) to be calibrated, and digital multimeter (HP34401A). The headphone was placed on the artificial ear and a force of 600 mg was applied to hold the headphone in place, using the spring tension mechanism of the artificial ear. Microphone housed inside the artificial ear sensed the signal, which was read by the sound level meter.

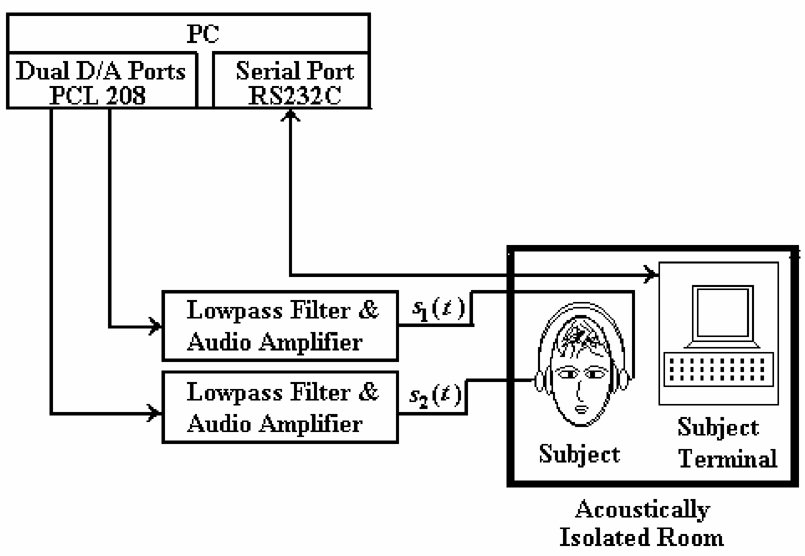


FIG. D.3. Automated computed listening test set-up for closed set evaluation of VCV syllables.

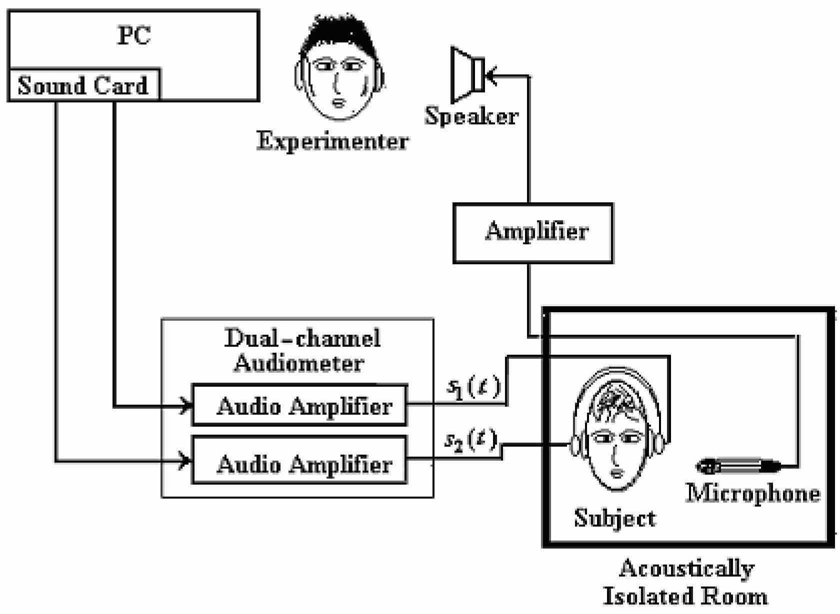


FIG. D.4. Computerized listening test set-up for open set evaluation of phonetically balanced mono syllables.

The input voltage in dBm measured by the multimeter (with reference to 1 mW power in a 50 Ω resistance, i.e. Vref = 0.224 V) required to produce sound level of 100 dB SPL in the artificial ear was measured at different frequencies, to obtain the electro-acoustic characteristics of the headphone. The characteristic obtained for the two headphones, shown in Fig D.6 are very similar and the variation is within 10 dB over the frequency range of 125 Hz to 5 kHz.

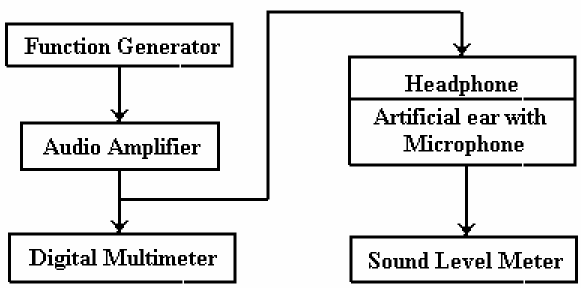


FIG. D.5. Experimental set-up for the measuring the electroacoustic response of headphones

While conducting listening tests, the speech stimuli was presented at the most comfortable level of the subject, which was maintained over the various test sessions. The electric input voltage in dBm, was used for monitoring the acoustic output level. A 1 kHz pure tone with acoustic output of 75 dB SPL corresponded to electrical input of –40 dBm (i.e. 2.24 mV rms).

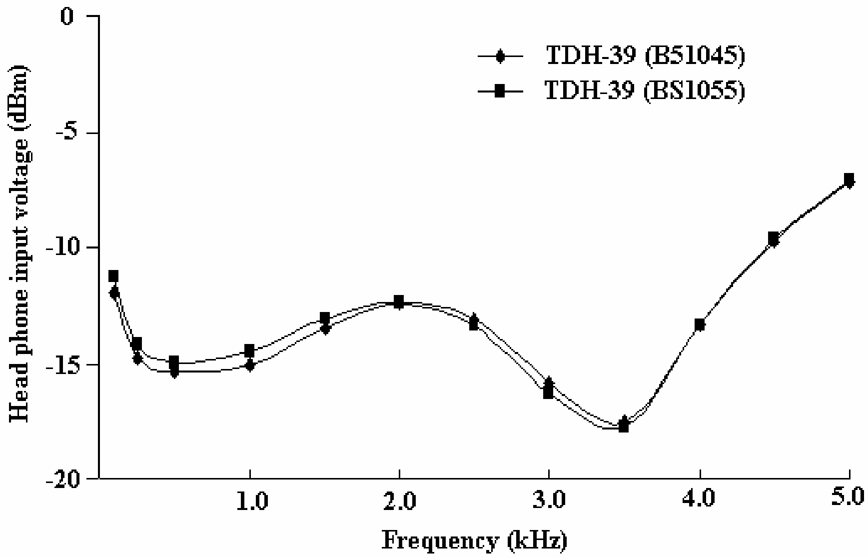


FIG. D.6. Electroacoustic characteristics of the headphones. Input voltage (dBm) versus frequency for 100 dB SPL output.

Appendix E

SOFTWARE FOR SPEECH PROCESSING AND LISTENING TESTS

E.1 Introduction

The objective of speech processing was to reduce the effect of frequency dependent shifts in hearing thresholds, increased spectral masking, and increased temporal masking, for persons with moderate bilateral sensorineural type of hearing loss. Adjustable magnitude response filtering was carried out to reduce the effect of frequency dependent shifts in hearing thresholds. Spectral splitting of speech was carried out using a pair of comb filters with bands corresponding to auditory critical bands, to reduce the effect of spectral masking. To reduce the effect of temporal masking, speech was split using trapezoidal fading functions. Cyclic sweeping of bands of the auditory critical bands was carried out in time varying comb filter, to simultaneously reduce the effects of spectral and temporal masking. Further listening tests were carried for evaluating the different speech processing schemes. The programs used for speech processing and listening test evaluation are listed in Table E.1, and briefly described in the following sections.

E.2 Filter design

Filters were designed as FIR, using frequency sampling techniques. A program "`modfilt.c`" developed by Kasthuri (1997) and modified by Ratanpal (2000) was used. For designing *N*-coefficient linear phase FIR filter, the magnitude of the samples are entered graphically and the filter coefficient are calculated and stored as a text file (details are given in C.1). The input parameters are

```
Filter specifications from data file (y/n):
Length of filter ( ≤ 400 ):
Choose response (M: magnitude\B: magnitude and phase):
```

The format of the ".txt" file containing N filter coefficients is

No. of filter coefficients <cr> b_0 <cr> b_1 <cr> - <cr> b_{m-1}.

E.3 Adjustable magnitude response filtering

To partially compensate for the frequency dependent shifts, speech was processed with adjustable magnitude response filters. The frequency response of these filters with adjustable gain in the range of –3 to +3 dB was obtained by interpolating the pure tone audiogram of the subject for the two ears. The filters were 256-coefficient linear phase FIR, designed using the program mentioned in E2. In the first phase of investigation a scheme of adjustable magnitude response filter cascaded with perceptually balanced comb filters was evaluated. The program agdico.c was used to simultaneous processing with adjustable magnitude response. The input parameters are

```
Enter the name of the file to be filtered:
Enter the name of the first coeff. file:
Enter the name of the second coeff. file:
Enter the name of the ag_coefficient file (right):
Enter the name of the ag_coefficient file (left):
Enter the name of the processed file (right):
Enter the name of the processed file (left):
Enter the name of the processed file:
```

In the second phase of investigations, a program "inag_b.c" was used for processing set of phonetically balanced monosyllables, using the adjustable magnitude response filters. The list of words was stored in one ".txt" file. The speech signal files were named with the monosyllable itself, followed by '_' symbol and a suffix, denoting the processing. For example, the speech file containing the monosyllable "ace" will have a file name "ace_pp", where "pp" represents the type of processing. For the processing of speech signal files, only the suffix needs to be inputted. The program prompts to enter the following input parameters.

```
Name of the ag_coefficient file right
Name of the ag_coefficient file left
Input speech file name suffix
Processed speech file name suffix
Output file name
```

Output file stored each of the input speech file names along with the processed file names for left and right ear separately and combined.

E.4 Spectral splitting

To reduce the effect of increased spectral masking, the speech was split using a pair of comb filters with complementary magnitude response, each filter having 9 pass bands corresponding to auditory critical bands. The design of perceptually balanced comb filters is given in Appendix C. The linear phase FIR filters for right and left ear are realized with 256 coefficients by frequency sampling techniques, for off-line processing of the speech signals. The program "`dico`" written in C, is a modified version of "`dicho`" developed earlier by Ratanpal (2000). The program prompts the user to input the name of the input file (in "`.wav`" format), the two coefficients files (in "`.txt` format), and the left and right output files (in "`.wav`" format).

The program "`dico`" could be used for ".wav" file with 10 kSa/s for 1 s duration. The program was modified to "`indico`" for processing long duration (≈ 30 min) speech files for quality assessment listening tests. Second phase of listening tests were carried out with phonetically balanced monosyllables in different languages. Such sets contained 50-60 monosyllables, each stored as a separate file to facilitate randomized presentations. To process the monosyllables of a set together, the program was further modified, to "`indico_b2`". The list of words was stored in one "`.txt`" file. As mentioned in the previous section, for the processing of speech signal files, only the suffix needs to be inputted. The input parameters for processing are

```
Input speech file name suffix
Processed speech file name suffix
Output file name
```

Output file stored each of the input speech file names along with the processed file names for left and right ear separately and combined. The processing was done with coefficients of perceptually balanced comb filters.

E.5 Temporal splitting

For reducing the effect of increased temporal masking, the speech was split into various segments and adjacent segments are presented to two different ears. Program "split" written in

C, by Jangamashetti (2003) was used for this purpose initially to split 12 nonsense syllables in VCV context. For quality assessment of speech text of long duration in ".wav" format, the program was modified, "`magtem`". For processing the speech signal files of the list of phonetically balanced monosyllables (numbering 50-60 each in different languages), program "`sptem_b3`" was used. The input parameters for processing are

```
Duration of one cycle in ms
Input speech file name suffix
Processed speech file name suffix
Output file name
```

The duty cycle and transition duration were fixed at 70 % and 3 ms respectively.

E.6 Optimizing the parameters of time varying comb filters

To simultaneously reduce the effect of increased spectral and temporal masking, a scheme of combined splitting was implemented using time varying comb filters. Each time varying comb filter consists of a number of comb filters (denoted as number of shiftings, *m*), which are used in a cyclic manner to process the speech for left and right ear. The pass bands of these comb filters are shifted along the frequency axis such that a sweeping of bands takes place. The frequency range between two alternate critical bands of the comb filter for spectral splitting, was divided into *m* equal parts, to obtain center frequencies of the pass bands of successive comb filters in a time varying comb filter with *m* shiftings. The filter coefficients for the *m* comb filters were obtained by frequency sampling technique as detailed in Appendix C. The pre-calculated set of coefficients were cyclically swept with a cycle duration of T_p ms.

A program "`shdi`" written in C, prompts the user to input the number of shiftings, names of the input file to be processed and the output processed files for the left and right ear separately. The duration of sweep cycle was fixed at 20 ms. The program loads the required sets of coefficients. The number of shiftings used were 16, 8, 4, and 2. Totally there were 16 comb filters. For 16 shiftings, all 16 sets of coefficients were used. The complementary pairs were formed as 1-9, 2-10, 3-11, 4-12, … , 8-16, 9-1, 10-2, … , 16-8. For 8 shifts the odd pairs only were used. For 4 shifts, only the odd pairs from pairs used for 8 shifts were required. For 2 shiftings, the pairs needed were 1-9 and 9-1. Each pair is designed for perceptual balance as discussed in Appendix C. The filter coefficients for all the 16 comb filters were calculated and stored in "`.txt`" format.

Further the program was modified to process speech files of longer duration in "`.wav`" format for quality evaluation. The program "`inshdi`" could process for different sweep cycle duration. Listening tests for finding the optimal parameters of the time varying comb filters like sweep cycle duration and number of shiftings, were conducted with phonetically balanced monosyllabic words in different languages. Such sets contained 50-60 words. Hence, provision for processing all the words in a set together, was provided in the program "`insh_b2`". The words in one language are stored in a "`.txt`" file. For processing of speech files, only the file name suffix is inputted. The input parameters for processing are

```
Duration of one cycle in ms
Number of shiftings (16, 8, 4, or 2)
Input speech file name suffix
Processed speech file name suffix
Output file name
```

E.7 Monaural/binaural hearing level difference

Listening tests were conducted with various types of signals, to find the level difference in monaural/binaural presentations for the same perceptions. The presentation procedure is detailed in Appendix C. A program was written, "`mono_bin`" in C, to present the signals for monaural and binaural presentation through two 12-bit DAC channels of PC based PCL-208 data acquisition card (Dynalog Micro Systems, Bombay, 1989). The user prompts for the program are

```
file names for monaural and binaural presentation
time between the monaural/binaural presentations
time between inter-presentations
```

E.8 Simulation of sensorineural hearing loss

For all the speech processing schemes, initial evaluation was done by conducting listening tests, on normal hearing subjects with simulated hearing loss. Sensorineural hearing loss was simulated by adding broadband noise with constant short-time signal-to-noise ratio (SNR). A program "`snr`" developed earlier by Prasad (1996), and modified by Chaudhari (2000) was used to add noise to the speech signal at various SNR conditions. The energy in the signal was calculated for a short time of 10 ms and noise was added to the signal in such a way that SNR remains constant. The simulation was done for each speech stimulus separately. The program prompts the user to enter the input and output speech file names and the SNR (dB) value *k*. The program accepts the noise file "`noise.bin`" from the working directory. A block of 100 samples (duration of 10 ms with sampling rate of 10 k Sa/s) of the output signal $x(n)$ were calculated from 100 samples of the input signal $s(n)$, and 100 samples of noise file $e(n)$. The output $x(n)$ was obtained by adding the input signal $s(n)$ with $e(n)$ scaled by a factor β such that the output signal energy has an SNR = k dB in every block of 100 samples.

$$x(n) = s(n) + \beta e(n)$$

where $$\beta^2 = \frac{{\sigma_s}^2}{{\sigma_e}^2 \times 10^{0.1k}}$$

$${\sigma_s}^2 = \sum s^2(n)\,, \quad {\sigma_e}^2 = \sum e^2(n)$$

The program "`snr`" was modified to "`snr21a`" for adding broad-bandwith constant short-time SNR to speech stimuli in "`.wav`" format. The program was further modified to "`snr_b_lr`" for adding noise to a set of phonetically balanced words for left and right ears and to combine them to be presented through the sound card. The speech files for right and left ear were given "R" and "L" suffixes, following the processing identification.

```
Input speech file name suffix
Processed speech file name suffix
Output file name
Required SNR in dB
```

E.9 Listening tests with VCV syllables

For the evaluation of perceptually balanced comb filters, listening test was conducted using 12 nonsense syllables in VCV context. The automated listening test used a program "`test`". One listening test consists of 60 presentations, with each syllable presented 5 times. For randomizing the presentation and the display at the subject terminal (for giving the response), different sets of randomized list were made using a program "`tlist`". For information transmission analysis, a program "`info`" was used. These three programs were initially developed in Fortran by Pandey (1987), and were written in Pascal and 'C' by Thomas (1996) and Chaudhari (2000) respectively. The program "`test`" used two DAC channels of PCL-208 data acquisition card (Dynalog Micro Systems, Bombay, 1989) for outputting the signal at 10 k samples/s. The communications with the subject terminal, i.e. display of test instructions and the reception of subject response, were done through the COM port. The program records the subject response and response time after each presentation. It also records the scores in the form of confusion matrix. The experimenter was prompted to input the following parameters.

```
Subject identification (max 5 ch)
Test listening of specified sounds (y/n)
Test number (1-200)
Speech list file number (1-25)
Test using (DAC channel 1/ channel 1&2)(1/0)
Response feedback while hearing (y/n)
Signal info file
```

To minimize the biasing effect, the presentations and the order of display at the subject terminal were randomized. Certain uniformity constraints were followed in randomizing the presentation order (Pandey, 1987). They were as follows:

Overall uniformity : In an experiment with N presentations and n stimuli, each stimulus need to be presented N/n times.
Mid-range uniformity : In M consecutive presentations, any stimulus should not occur more than $M/n+2$ times.
Short-range uniformity : More than three consecutive presentation of an item should not be there.

Using "`tlist`" 25 files (`slist1, slist2, - - - slist25`) of randomized lists were generated to provide different randomization in each test. The information about the stimuli for the test was provided in a file in "`.dat`" format, which has to be entered at "`Signal info file`" prompt. The data in this file are in the following format:

```
Sampling frequency in Hz
Number of items in the test set
Names of test items (to be displayed as response choices)
Names of data files containing signal samples
```

The subjects were given instructions before the beginning of the test session, as given in Appendix F. Subjects could have listening practice at the start of each test run to become familiar with the words. At the subject terminal, all the 12 speech stimuli were displayed along with the key to be pressed to listen to each of them. While undergoing the test, after listening the speech stimuli through the headphones, the subject responded by pressing the corresponding key as fast as possible. Test progressed only after the response from the subject. The program compared the test stimuli and the response, and the score was updated. The response time for each presentation was stored for analysis. For each processing condition first test run was with feedback. Five test runs without feedback with consistent results were considered for analysis. Stimulus response was stored in the form of confusion matrix. Each entry in the matrix corresponded to the frequency of occurrence of a stimulus-response pair, in which the diagonal elements represented correct recognition. Percentage recognition score and response time statistics were also displayed along with the confusion matrix. These results were stored in a "`.res`" file with name having a combination of subject initials and test number.

E.10 Listening tests with phonetically balanced monosyllables

For the second set of listening tests, sets of phonetically balanced monosyllabic words in three different languages (English, Hindi, Marathi) were used. In each language there were 50 to 60 words. The word lists were obtained from Ali Yaver Jung National Institute for Hearing Handicapped, Mumbai, where these lists are used to evaluate the discriminating capacity of hearing impaired persons. These words were recorded with a sampling rate of 10 k samples/s, from native speakers with "clear" voice quality. The listening test set-up is discussed in Appendix D. Since it was difficult to get sufficient number of hearing impaired subjects in our laboratory for participating in the listening tests, the set-up was modified to use sound card of a Notebook PC. A new program "`stest`" was written in 'C', for conducting the listening test with the modified set-up. For presenting the monosyllabic words in "`.wav`" format to the subject, this program invokes "`simple`" written by Ratanpal (2000) in Visual C++.

The program had the facility to present the words in the order given in a list for listening practice, any number of times as required by the subject. The subject seated in an acoustically isolated chamber, was given a copy of the list while listening. The subject could listen to specific test items for more number of times. Once the subject is familiar with the words, the tests started. Provision was provided to conduct the listening test with and without a subject terminal. There were four options to the method of conducting listening tests. The first two options did not use a subject terminal. In the first option, speech syllables were presented one after the other at a specified inter-presentation interval The interval fixed for a particular test, had to be sufficient for the subject to write down the response before the presentation of the next word. The second option was used where a microphone kept in the acoustically isolated room could be used to send the verbal response of the subject to the experimenter sitting outside. The subject repeated the words as listened and the experimenter sitting outside compared the response with the presented word and keyed in 'y' or 'n'. The response time also was noted. Percentage correct recognition score and average response times were calculated. The third option was similar to the first, but there was no fixed inter-presentation interval. The subject wrote down what he/she heard and pressed the key of the subject terminal for the presentation to continue. The response time in addition to recognition score were recorded. The fourth option was very much similar to the second one and was used when there was no provision for hearing the subject's verbal response outside the acoustically isolated room. Here the subject repeated the words he/she has heard through the headphone and the experimenter sitting near the terminal inside the acoustically isolated room typed in 'y' or 'n' on the subject terminal keyboard. The speech presented was displayed on the terminal, which was seen only by the experimenter typing the response. The response time also will be recorded. The program prompts the user to input the following parameters, before the listening practice.

```
Signal info file
Processing identification
Listening to become familiar with words (y or n)
No of times words need to be presented
Inter-presentation delay in sec
Press 'e' to exit in between if required
```

The "`signal info files`" were three "`.txt`" files named as "`marathi`", "`hindi`" and "`english`", which contained the mono-syllables in the respective languages, to be presented to the subjects. The words listed in the chosen "`signal info file`" were appended by the processing identification and the corresponding speech stimuli were presented to the subject in the order listed (printed copy given to the subject), with the inter-presentation delay inputted. The speech file would be displayed on the monitor of the experimenter. The subject had to go through the list of words given to him/her while listening to the words to obtain the listening practice, any number of times. If presentation of any of the words, had to repeated then the subject had to mention the numbers corresponding to the words to be repeated from the list. The prompt follows,

```
From
To
No of times
```

The presentation of words corresponding to numbers 'from' and 'to' would be done as many number of times as mentioned. The same prompt appears,

```
Listening to be repeated (y or n)
```

If 'y' was pressed, the previous prompt appears again. Else the next prompt,

```
Ready to start the listening test (y or n)
Subject identification (max 5 letters)
Test number
Presentation method
(0-delay, 1-console key press, 2-terminal key press,
3- terminal yes/no )
```

Presentation method gives the four different options for conducting the listening tests. If ‘0’ was pressed, an additional prompt appears for the inter-presentation delay to be inputted

```
Inter-presentation delay in sec
```

Next prompt for all the options

```
Number of words presented
Number of times each word presented
Number of consecutive presentations allowed
Seed (for randomized presentation)
```

New screen appears

```
Total presentations in this test :
*** START***
Press 'e' to exit in between if required
     1.   speech file ...
```

One listening test consisted of presentation of each word in the list 3 times. The presentation order was randomized (controlled by the seed value inputted). The maximum number of consecutive presentation for a monosyllable was limited to two. For maintaining mid-range uniformity, in M consecutive presentations, any stimulus should not occur more than $M/n+2$ times, where n is the number of stimuli.

The present set of listening tests were conducted with presentation method option ‘2’. The result of the test was stored in a “`.txt`” file with name in “`ssstt`” formant, where ‘`sss`’ and ‘`tt`’ represent subject identification and test number respectively. The file contained the words presented in the order of presentation with “correct” and “wrong” written in the same line for correctly and wrongly responded words respectively along with response time. The number of correctly responded words, the percentage recognition scores, and response times (total, average and standard deviation separately for correctly and wrongly responded words) were stored in this file.

TABLE E.1. List of programs for speech processing and listening tests

Program	Purpose	Remarks
`modfilt.c`	Filter design using frequency sampling techniques	Developed by Kasthuri (1997) and modified by Ratanpal (2000)
`agdico`	Adjustable magnitude response filtering cascaded with spectral splitting	
`inag_b.c`	Adjustable magnitude response filtering for batch processing of 50-60 words	
`dicho.c`	Spectral splitting using comb filters with 128 coefficients	Developed by Ratanpal (2000)
`dico.c`	Spectral splitting using comb filters with 256 coefficients	Modified from "`dicho.c`"
`indico.c`	Spectral splitting using comb filters with 256 coefficients for processing speech of longer duration (30 m) in "`.wav`" format for quality assessment.	Modified from "`dico.c`"
`indico_b2.c`	Spectral splitting using comb filters with 256 coefficients for batch processing of 50-60 words	Modified from "`indico.c`"
`split.c`	Temporal splitting using trapezoidal fading functions with different duty cycles with 20 ms transition durations for processing of speech in ".bin"	Developed by Jangamashetti (2003)
`mgtem.c`	Temporal splitting using trapezoidal fading functions with different duty cycles and transition durations for processing of speech of longer duration (30 m) in "`.wav`" format for quality assessment.	Modified from "`split.c`"
`sptem_b3.c`	Temporal splitting using trapezoidal fading functions with different duty cycles and transition durations for batch processing of 50-60 words.	Modified from "`mgtem.c`"
`shdi.c`	Combined splitting with time-varying comb filters with fixed duty cycle 20 ms and with 2, 4, 8, and 16 shiftings	Modified from "`shdi.c`"
`inshdi.c`	Combined splitting with different sweep cycles and with 2, 4, 8, and 16 shiftings for processing speech files of longer duration (30 m) in "`.wav`" format for quality assessment.	

TABLE E.1. (Contd.)

Program	Purpose	Remarks
`insh_b2.c`	Combined splitting with different sweep cycles and with 2, 4, 8, and 16 shiftings for batch processing of 50-60 words.	Modified from "`inshdi.c`"
`mono_bin.c`	Monaural and binaural presentation of signals	
`snr.c`	Simulation of sensorineural hearing loss, by adding broadband noise with constant short-time SNR	Developed by Prasad (1996) and modified by Chaudhari (2000)
`snr21a.c`	Simulation of sensorineural hearing loss, by adding broadband noise with constant short-time SNR for speech of longer duration in "`.wav`" format	Modified from "`snr.c`"
`snr_b_lr.c`	Simulation of sensorineural hearing loss, by adding broadband noise with constant short-time SNR for batch processing of 50-60 words for left and right ear and to combine them to be used in sound card.	Modified from "`snr21a.c`"
`test.c`	Automated listening test consisting of 60 randomized presentation for the 12 CVC syllables, with each syllable presented 5 times.	Developed in Fortran by Pandey (1987), in Pascal by Thomas (1996) and in 'C' by Chaudhari (2000).
`tlist.c`	To generate the randomized list for presentation and display	do
`info.c`	Information transmission analysis	do
`stest.c`	Automated listening test for randomized presentation of a set of phonetically balanced monosyllabic words, in different languages	
`simple.cpp`	This was invoked by "stest.c" for sending the syllables through the soundcard of the Notebook PC.	Developed by Ratanpal(2000)

Appendix F

INSTRUCTIONS TO SUBJECTS

F.1 Introduction

The first set of listening tests was conducted to evaluate the perceptually balanced comb filters. A fully automated computerized listening test set up was used for the closed set evaluation of twelve consonants /p, b, t, d, k, g, m, n, s, z, f, v/ in VCV context with vowel /*a*/. The instructions to the subjects participating in these listening tests are given in F.2. Further an overall experimental evaluation was carried out for the three splitting schemes with different processing parameters, using phonetically balanced monosyllabic words. The instructions given to the subjects before participating in this listening test are given in F.3. Forms used for collecting background information are given in subsequent sections.

F.2 Instruction to subjects participating in listening tests with VCV syllables

The purpose of the listening tests is to evaluate the performance of speech processing schemes developed to increase the intelligibility of speech for normal persons with normal hearing under adverse listening conditions and for persons with bilateral sensorineural type of hearing loss.

You will be seated in an acoustically isolated room, in front of a computer terminal (a keyboard to give your response and a monitor to observe the randomized list of presentations with the corresponding key to be pressed). Your task will be to listen the test sounds (stimuli) presented through the headphones to both the ears, identify the sound and respond by pressing appropriate key. The intensity of sound will be adjusted to your most comfortable level of hearing.

The test material consists of 12 English nonsense syllables / aPa, aBa, aTa, aDa, aKa, aGa, aMa, aNa, aSa, aZa, aFa, aVa /. Each of the syllables will be presented 5 times in random order. Thus in a test run there will be a total of 60 (12 × 5) presentations. The instruction to be followed while undergoing a test run will be displayed on the monitor before the start of each test run. Provision is given to listen to the test stimuli any number of times in any sequence to become familiar with the sounds. You have to select the sound you need to hear and press the corresponding key from the list of choices displayed on the monitor. Before each presentation, presentation number, and response choices (randomized) with the corresponding key to be pressed will be displayed. A message "Listen" will be displayed followed by the presentation of speech sound through the headphones. You have to identify the sound heard and press the appropriate response key on the keyboard as quickly as possible. Presentations will not be repeated. Test will not proceed further unless you press a key. If by chance, you miss a presentation, press any key other than the valid response keys. The next presentation will be after a brief pause (2-5 s). Time taken by one test run consisting of 60 presentations will be 5 to 8 minutes for normal subjects. Hearing-impaired subjects may take more time (15 to 20 minutes). The complete evaluation for an experimental condition may include number of test runs. The first of these will be with feedback: during which your response and an "Ok" message will be displayed in case of correct response else the right choice will be indicated. Further test runs will be without feedback: only your response will be displayed. A number of such test runs will be carried out.

Instructions displayed on the subject terminal screen at the beginning of the test

```
Welcome to this computerized listening session.
Instructions follow...

Test listening of specified stimuli.
You can listen to the stimuli in any order you want by hitting
the appropriate key as listed below.

Sounds-----> aPa aBa aTa aDa aKa aGa aMa aNa aSa aZa aFa aVa
Press Key-->  1   2   3   4   5   6   7   8   9   0   -   =

Press ‘e’ or ‘E’ when you have finished test listening.
```

After this the instructions for the actual listening test are displayed

***** Consonant Identification Test *****

Your task is to identify the presented sound from among the following:

aPa, aBa, aTa, aDa, aKa, aGa, aMa, aNa, aSa, aZa, aFa, aVa

After listening to the sound, please hit the corresponding key as quickly as possible. A presentation will not be repeated. If you are not sure, you can guess. The test will not proceed if you do not respond. If you missed a presentation, and cannot even guess, you may hit a key other than the valid choices.

Please hit any key when ready for the test.

It was verbally explained to the subject, that the association between the keys and syllables would be a different one in each presentation.

F.3 Instruction to subjects participating in the listening tests with phonetically balanced monosyllables

The listening tests are being conducted to evaluate speech processing schemes to find the most suitable one for improving speech perception for persons having sensorineural type of hearing loss.

The test material consists of phonetically balanced monosyllabic words in three languages namely, English, Hindi and Marathi. The lists of monosyllables are shown in Tables F.1, F.2 and F.3 for English, Hindi and Marathi respectively. You can choose one of these languages (most commonly used by you), for carrying out the listening tests. The tests will be conducted in an acoustically isolated room. The words will be presented to the two ears through headphones. The intensity of sound in each ear will be adjusted to your most comfortable level of hearing.

A printed copy of the list of monosyllables in your choice of language will be given to you. For listening practice before each test, for each of the syllables will be presented in the order given in the list. You can listen to them several times until you become familiar. There is a provision to listen to certain syllables as per your choice, for more times.

Each test run consists of presentation of each word three times. The presentation is randomized and the same word may appear twice consecutively. You have to listen to each word carefully and repeat it through the microphone as fast as you can. We are interested in your response as well as the response time. Presentation will not be repeated. Presentation will continue only after you repeat what you have heard. If the word presented is not heard clearly, you have to choose the most likely one from the list, or just say "pass". Time taken for each test run will be 10-15 minutes.

F.4 Form for recording background information of the normal and hearing impaired subjects

SUBJECT BACKGROUND INFORMATION

Date __/__/____

Name ________________________________ Code ____________

Address ____

____

Phone ( ) ___ _______Extension _____

____

Sex _______ Age _____________

Occupation: ____

Place of birth: ____

First language: ____

Other languages: ____

Handedness: Left / Right

History of noise exposure: _____________________________

History of hearing problems: _____________________________

____

Other remarks: ____

F.5 Form for subject's willingness to participate

CONSENT FORM

I have carefully read the test instructions provided by Ms. Alice N. Cheeran (Ph.D. Scholar, IIT Bombay) for participation in listening experiments for evaluation of speech processing schemes. I am willing to participate in tests conducted by her.

Signature: ______________________________

Name: ______________________________

Address: ______________________________

____

Date: ______________________________

TABLE F.1 Phonetically balanced Marathi monosyllables

1	आज	17	बस	33	तीन
2	ऊन	18	माल	34	तोल
3	काम	19	मोर	35	दूध
4	स्वाद	20	मूल	36	दोष
5	गाल	21	मन	37	दोन
6	घाव	22	मोल	38	धूळ
7	घात	23	रस	39	नोट
8	छत	24	लाज	40	पाप
9	भेट	25	लूट	41	फूल
10	ठीक	26	शाप	42	बोध
11	डाळ	27	जात	43	रोज
12	तेल	28	खूष	44	लाल
13	नळ	29	गीत	45	साफ
14	नाक	30	धाप	46	हात
15	आग	31	चीर	47	पाय
16	फळ	32	मठ		

TABLE F.2 Phonetically balanced Hindi monosyllables

1	अब	19	घात	37	कब
2	आज	20	दिन	38	पल
3	रस	21	भीड	39	जान
4	खाट	22	पाप	40	भेंट
5	घाव	23	नोट	41	साफ
6	खुद	24	नाक	42	हाथ
7	मोल	25	दोष	43	दूध
8	लाज	26	लाल	44	छेद
9	देर	27	वोट	45	फूल
10	आप	28	शोर	46	तीन
11	गीत	29	सच	47	हार
12	माल	30	धूल	48	खुश
13	मोड	31	दाम	49	जीत
14	रात	32	रोज	50	शेर
15	रेल	33	सब	51	शोक
16	ठीक	32	दाग	52	याद
17	तेल	33	और	53	मार
18	दस	34	बस	54	भूल

TABLE F. 3 Phonetically balanced English monosyllables

1	ace	21	ham	41	ice
2	chew	22	deaf	42	them
3	toe	23	odd	43	knew
4	bathe	24	bell	44	ill
5	wait	25	one	45	cap
6	sea	26	tip	46	low
7	raw	27	skin	47	give
8	kit	28	dad	48	thing
9	air	29	him	49	pet
10	knees	30	earn	50	hit
11	yard	31	flap	51	chest
12	more	32	die	52	law
13	own	33	could	53	you
14	him	34	owl	54	bin
15	day	35	hurt	55	key
16	ache	36	up	56	live
17	she	37	jawl	57	room
18	not	38	will	58	and
19	shoe	39	felt	59	none
20	call	40	jam	60	else

Appendix G

AUDIOGRAM AND

ADJUSTABLE MAGNITUDE RESPONSE FILTERS

To partially compensate for the frequency dependent shifts in hearing thresholds, we have used adjustable magnitude response filters with gain corresponding to the audiogram for the different ears of the subjects, with maximum gain variation of ±3 dB. Listening tests were conducted for speech processed with such filters. The adjustable magnitude gain filters were cascaded with the schemes of spectral, temporal, and combined splitting for the various processing parameters. This appendix contains the pure tone audiogram and the corresponding adjustable magnitude filter response used for all the subjects who participated in the listening tests. The figures are arranged in the alphabetical order of the subject codes.

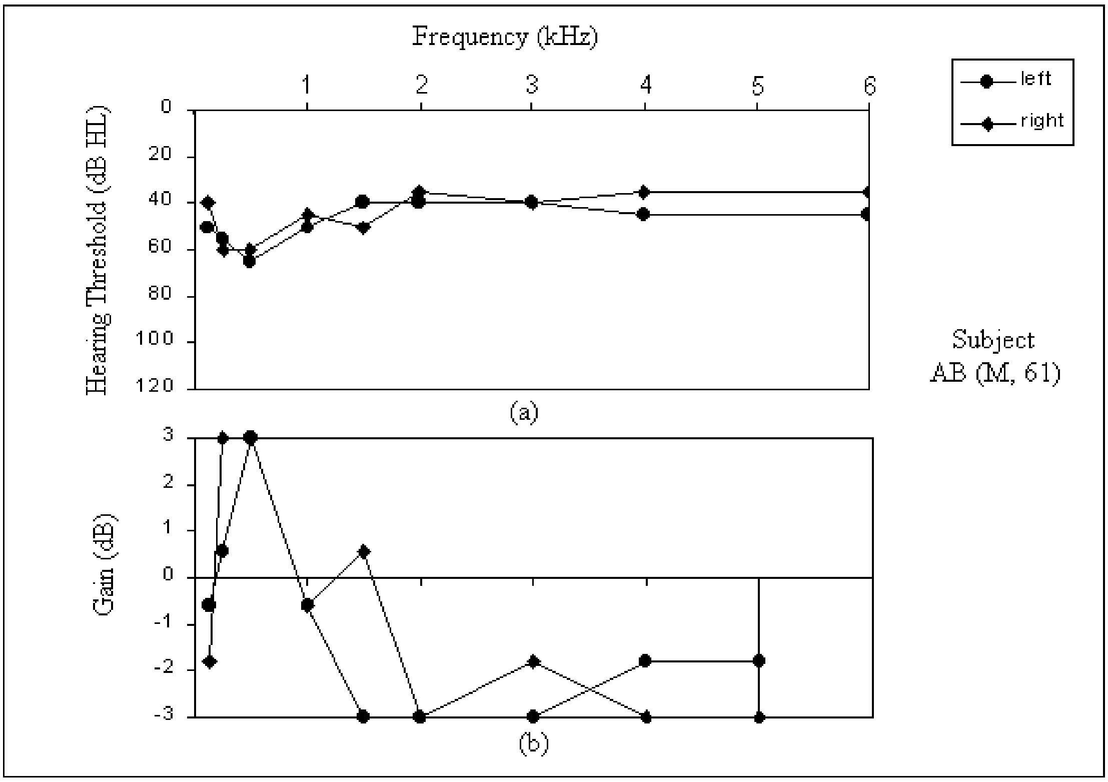


FIG. G.1. Pure tone audiogram and adjustable magnitude response filter for subject AB

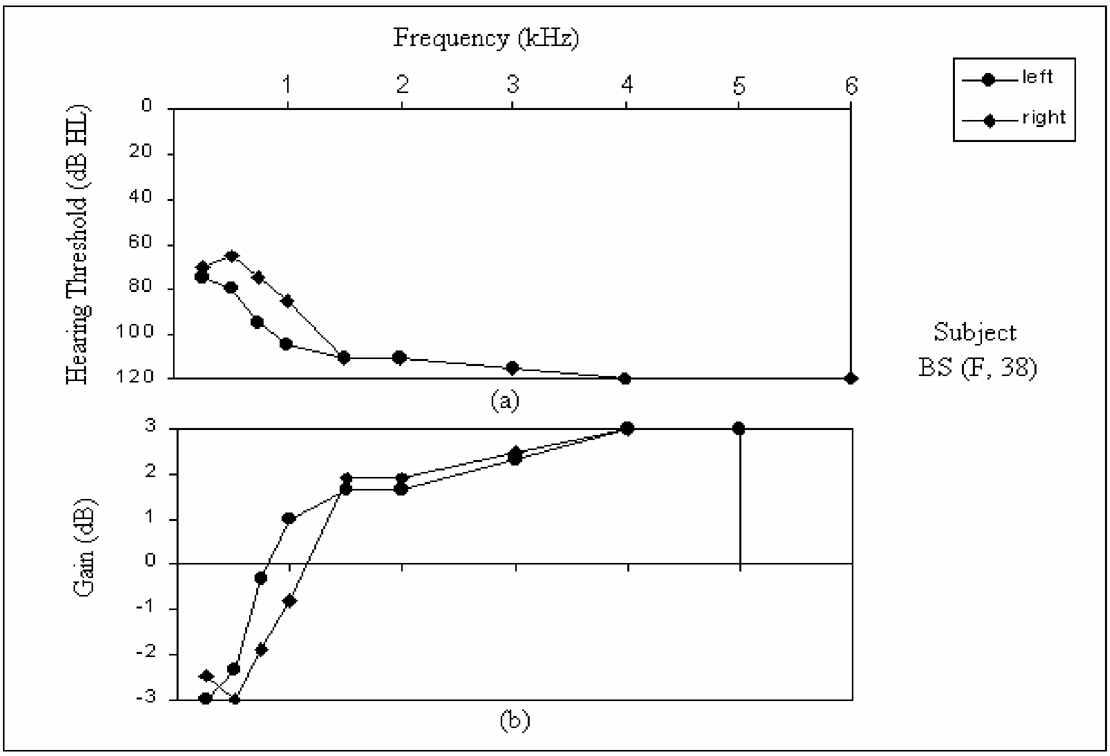


FIG. G.2. Pure tone audiogram and adjustable magnitude response filter for subject BS

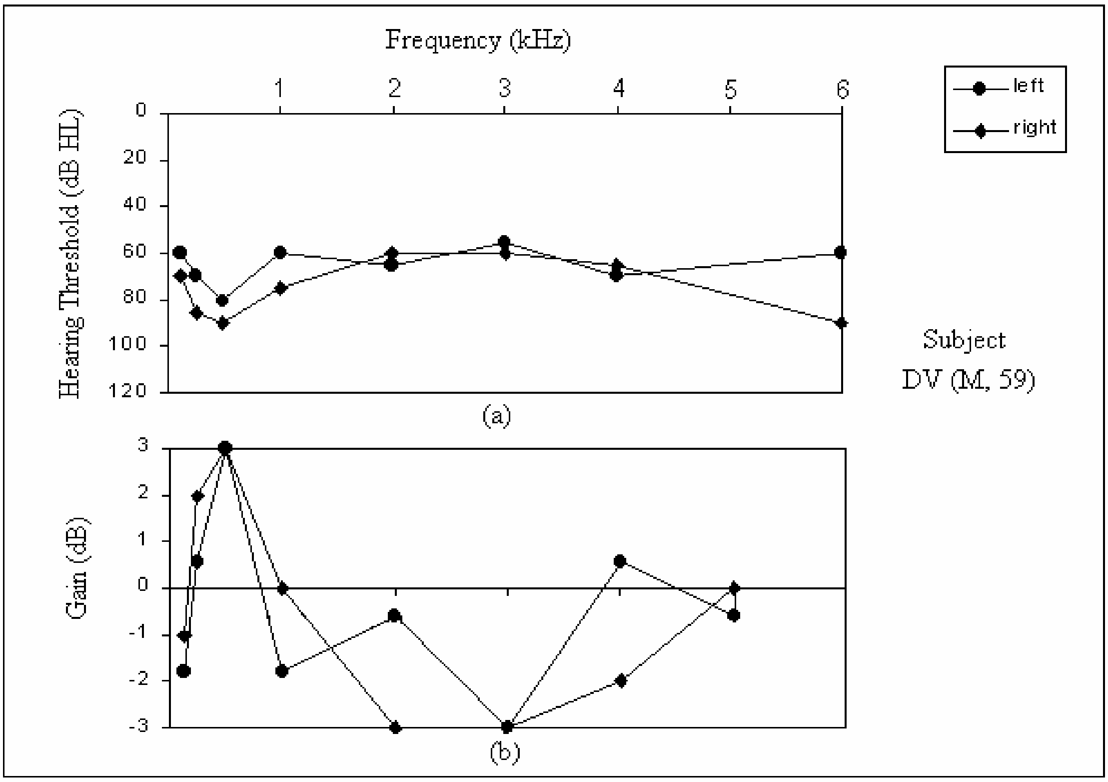


FIG. G.3. Pure tone audiogram and adjustable magnitude response filter for subject DV

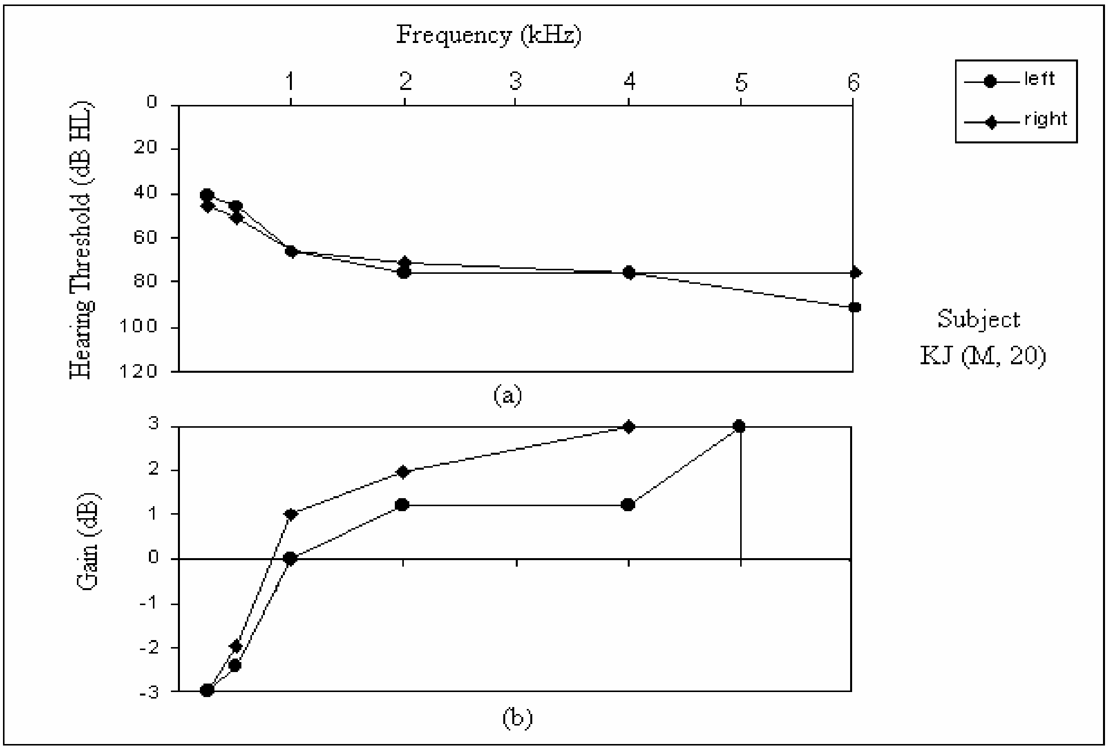


FIG. G.4. Pure tone audiogram and adjustable magnitude response filter for subject KJ

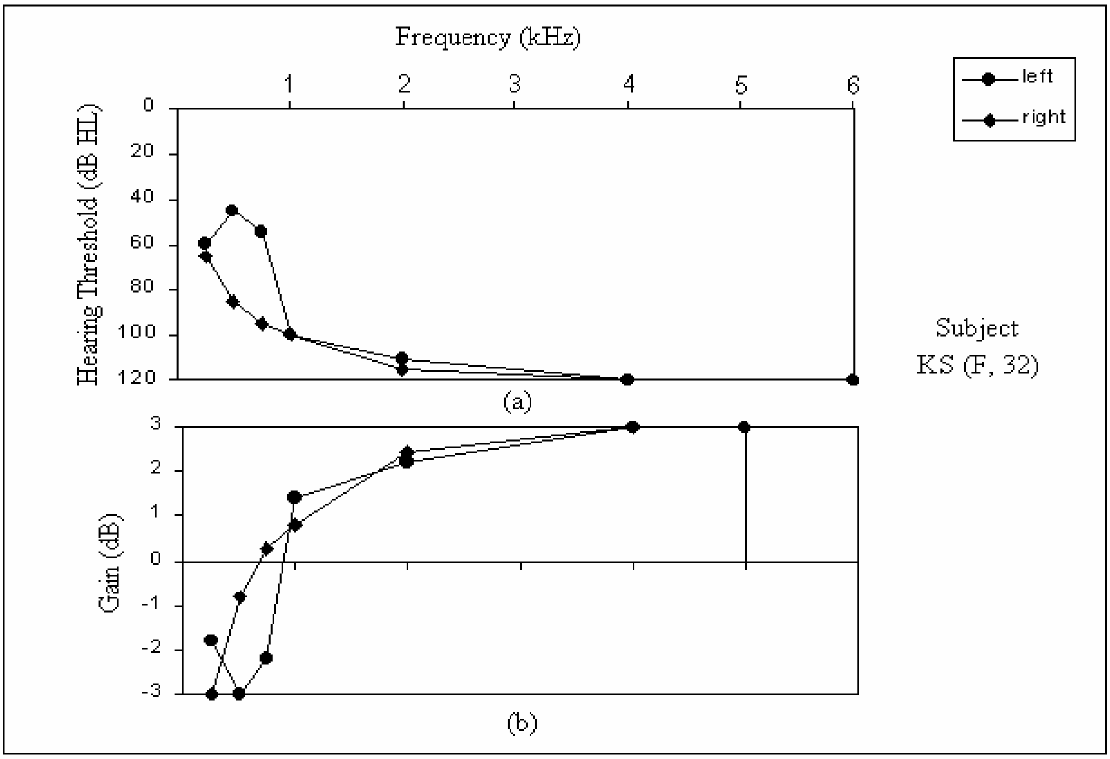


FIG. G.5. Pure tone audiogram and adjustable magnitude response filter for subject KS

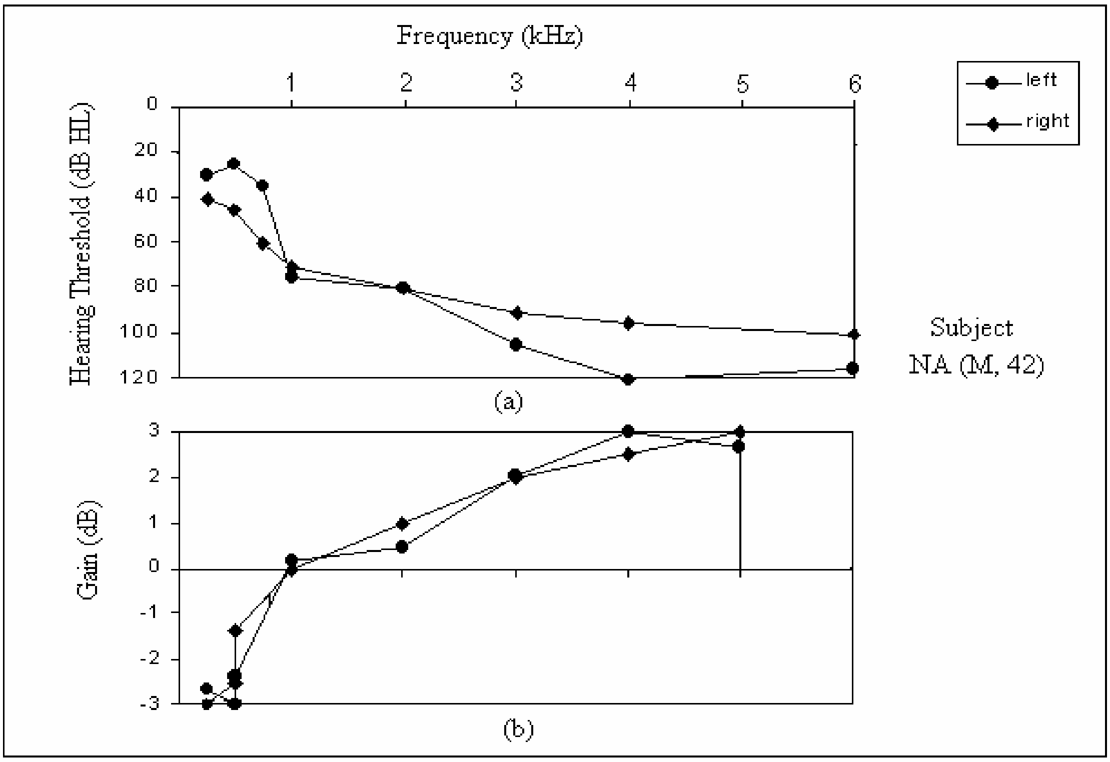


FIG. G.6. Pure tone audiogram and adjustable magnitude response filter for subject NA

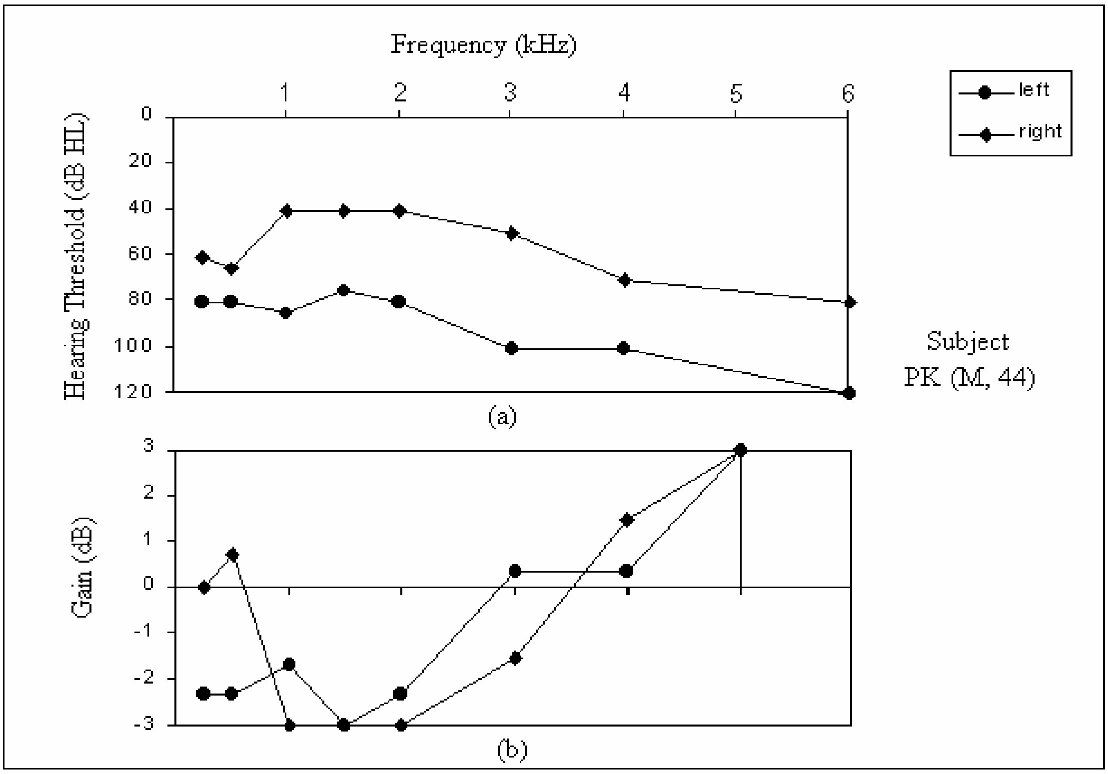


FIG. G.7. Pure tone audiogram and adjustable magnitude response filter for subject PK

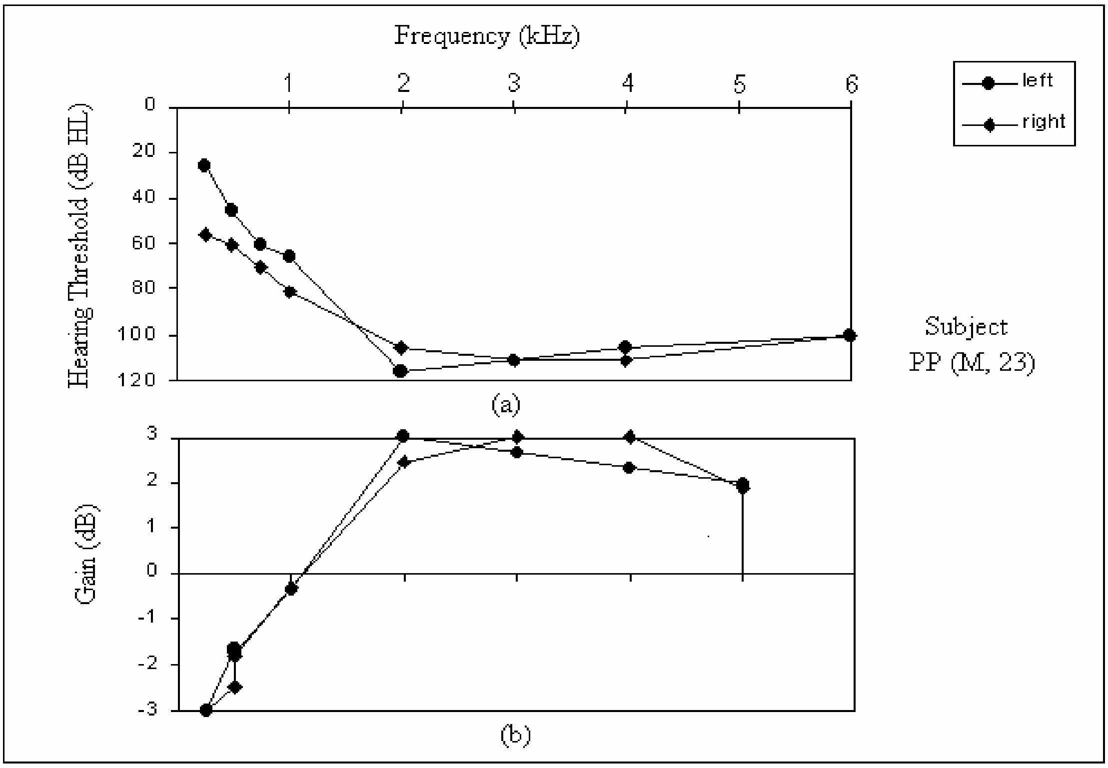


FIG. G.8. Pure tone audiogram and adjustable magnitude response filter for subject PP

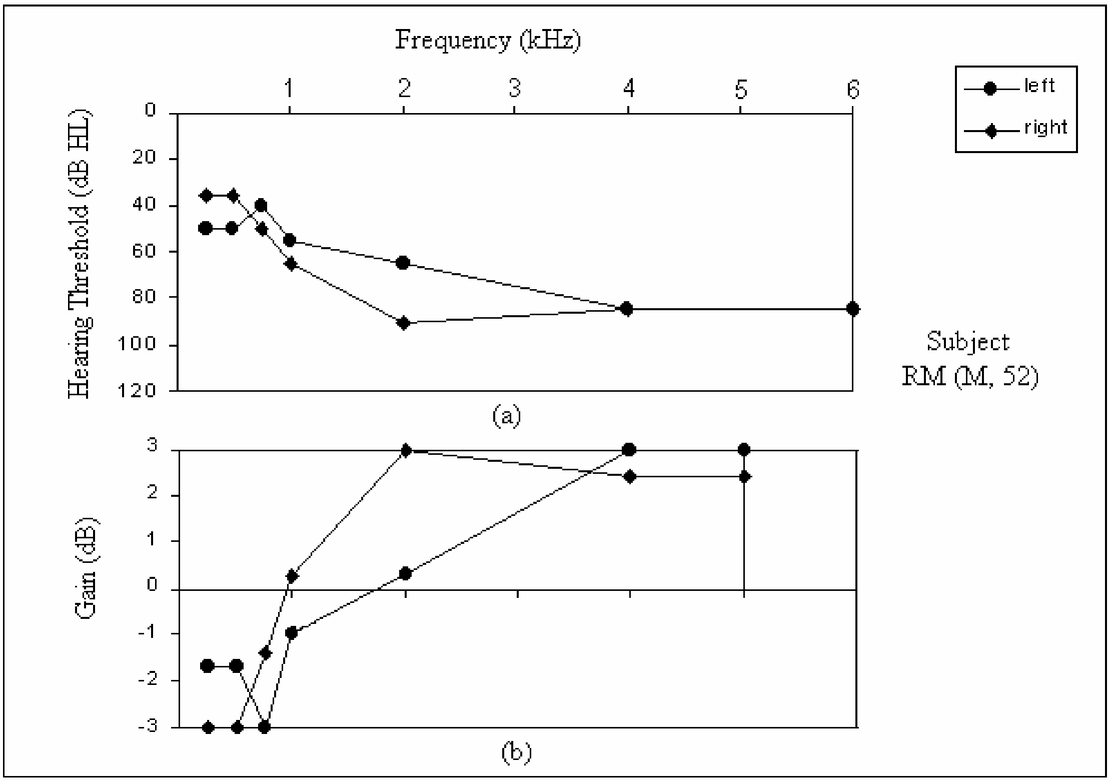


FIG. G.9. Pure tone audiogram and adjustable magnitude response filter for subject RM

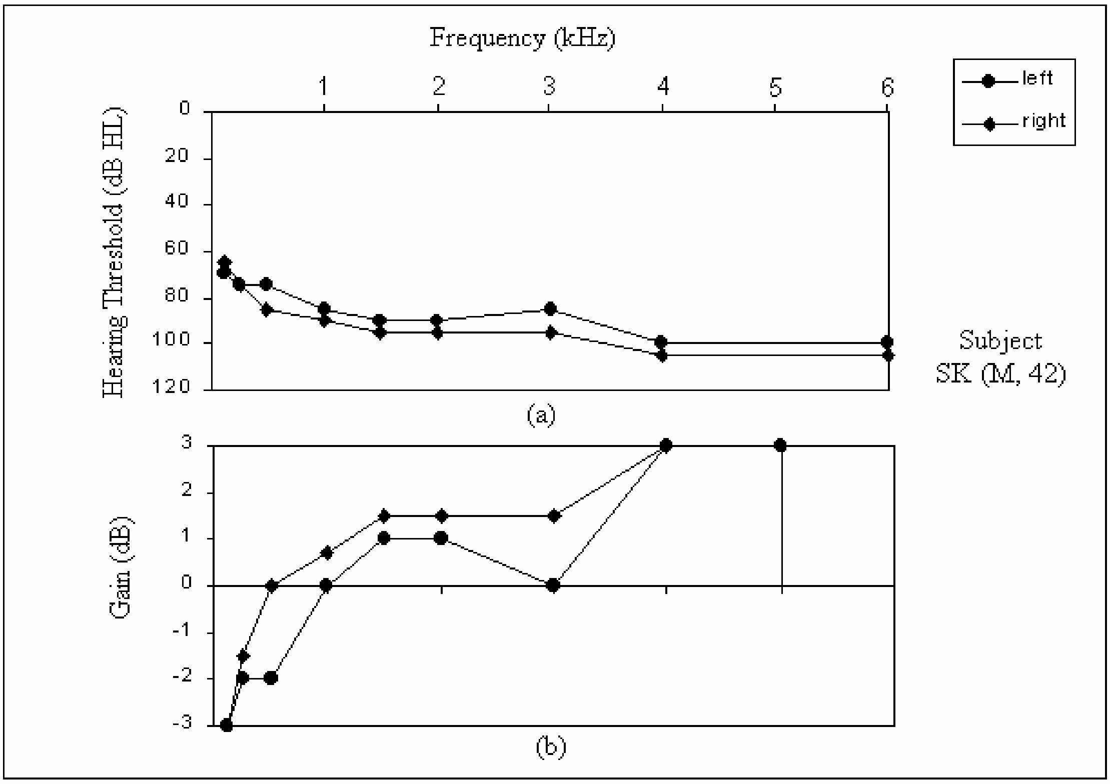


FIG. G.10. Pure tone audiogram and adjustable magnitude response filter for subject SK

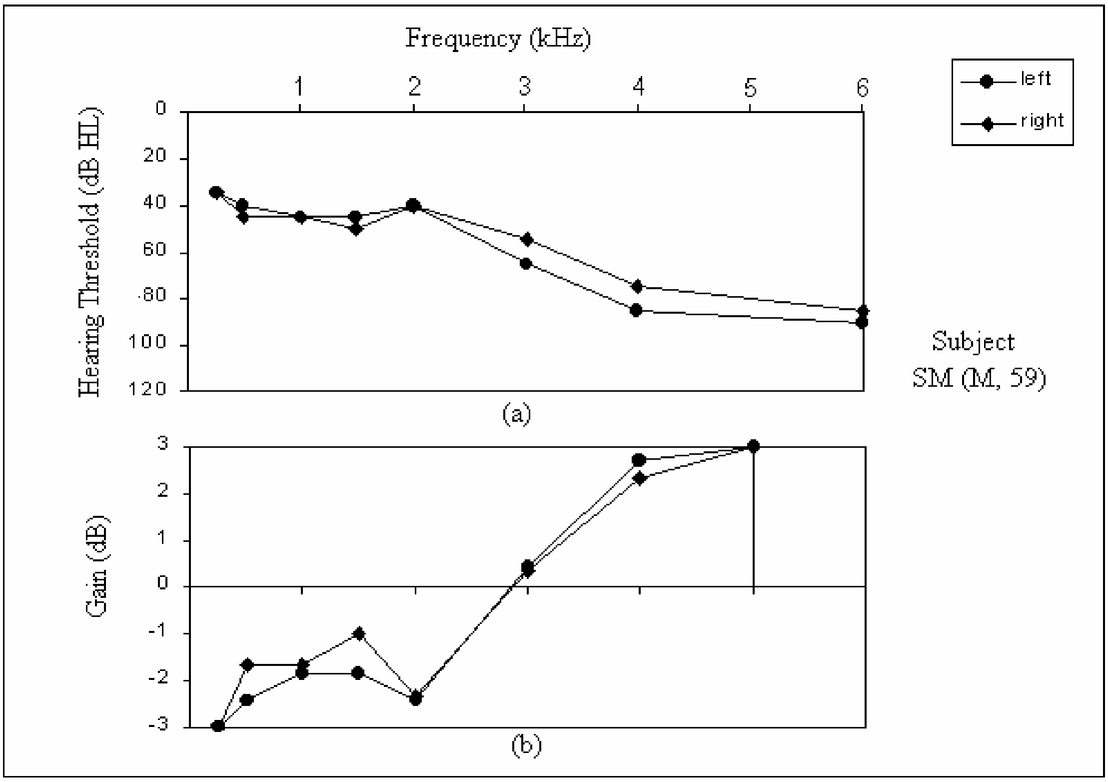


FIG. G.11. Pure tone audiogram and adjustable magnitude response filter for subject SM

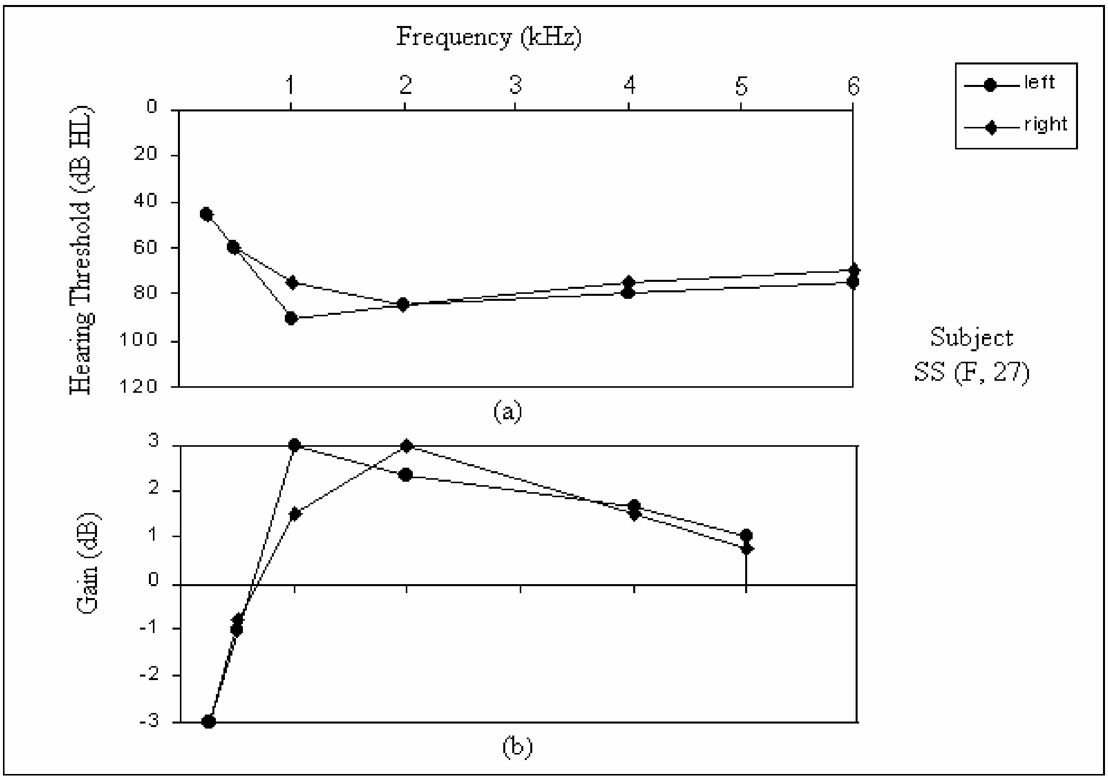


FIG. G.12. Pure tone audiogram and adjustable magnitude response filter for subject SS

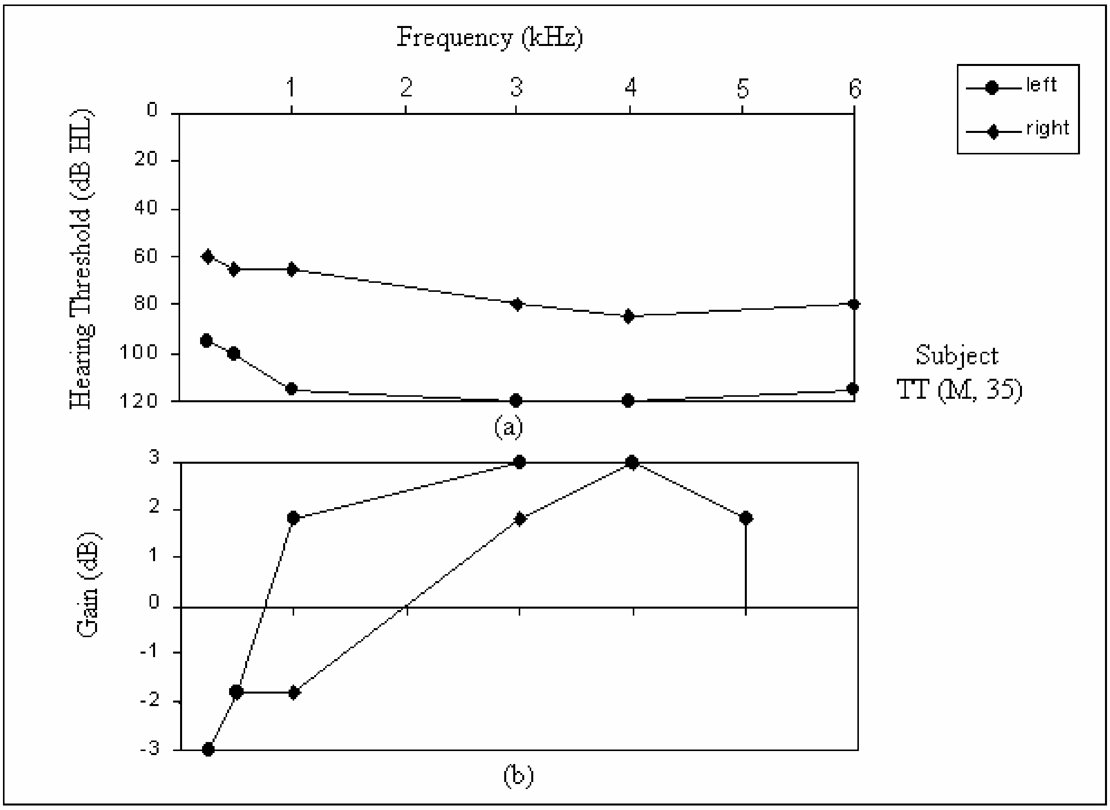


FIG. G.13. Pure tone audiogram and adjustable magnitude response filter for subject TT

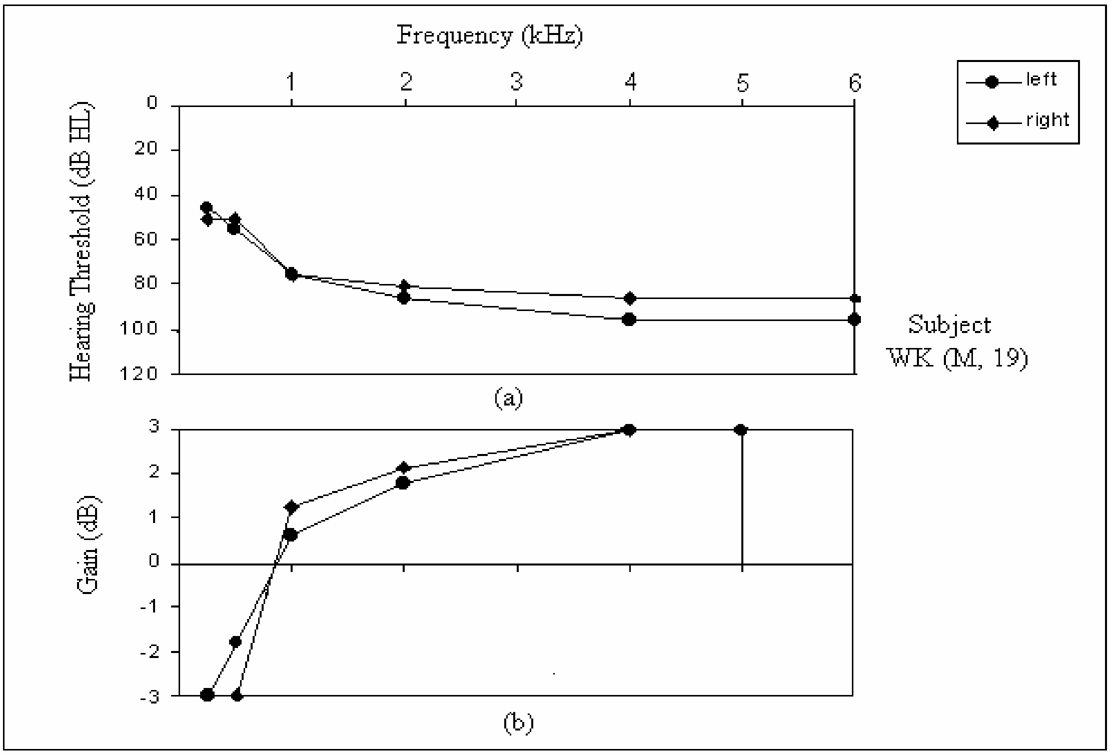


FIG. G.14. Pure tone audiogram and adjustable magnitude response filter for subject WK

Appendix H

STATISTICAL ANALYSIS OF TEST RESULTS

TABLE H.1. Experiment I. Averaged percentage relative improvements and paired t-test significance levels *p* for different pairing of processing schemes and conditions at SNR = –6, –9, –12 and –15 dB, for spectral splitting with comb filters with sharp transitions (SpA) and perceptually balanced comb filters (SpB). In each cell, the value within parentheses is the *p* value.

(A) SNR = -6 dB

Proc. cond.	SpA	SpB
Su	19.1 (0.01)	21.5 (0.008)
SpA		2.1 (0.1)

(B) SNR = -9 dB

Proc. cond.	SpA	SpB
Su	18.3 (0.02)	21.4 (0.003)
SpA		3.0 (0.1)

(C) SNR = -12 dB

Proc. cond.	SpA	SpB
Su	24.0 (0.01)	29.7 (0.01)
SpA		4.2 (0.02)

(D) SNR = -15 dB

Proc. cond.	SpA	SpB
Su	21.7 (0.009)	27.5 (0.003)
SpA		4.8 (0.05)

TABLE H.2. Experiment II. Averaged percentage relative improvements and paired t-test significance levels *p* for different pairing of processing schemes and conditions for hearing impaired subjects, for spectral splitting with comb filters with sharp transitions (SpA), perceptually balanced comb filters (SpB), and adjustable magnitude response filter cascaded with perceptually balanced comb filters (SpC). In each cell, the value within parentheses is the *p* value.

Proc. cond.	SpA	SpB	SpC
Su	12.4 (0.01)	19.3 (0.003)	25.2 (0.002)
SpA		6.1 (0.003)	11.2 (0.005)
SpB			4.82 (0.02)

TABLE H.3. Experiment IV. Two-way analysis of variance (ANOVA) on percentage relative improvements (with reference to unprocessed speech) averaged across subjects, for SNR and processing schemes and conditions (a) all processing schemes and conditions (Sp_SS, Sp_TS-20, Sp_TS-40, Sp_TS-80, Sp_CS-20/4, Sp_CS-20/8, Sp_CS-20/16, Sp_CS-40/4, Sp_CS-40/8, Sp_CS-40/16, Sp_CS-80/4, Sp_CS-80/8, Sp_CS-80/16, Sp_CS-120/8, Sp_CS-120/16, Sp_CS-160/8, Sp_CS-160/16), (b) temporal splitting with the 3 conditions, (c) combined splitting with the 13 conditions, and (d) the 3 schemes with corresponding "optimal" conditions.

.

Processing	Source of variation	df	Mean sum of squares	*F*	*p*
All processing schemes and conditions	SNR	5	22,792	77.94	0.01
	Processing	16	1,347	4.61	0.01
	Residual	80	292		
Temporal splitting	SNR	5	93	0.28	NS
	Processing	2	363	1.09	NS
	Residual	10	333		
Combined splitting	SNR	5	23,509	969.81	0.01
	Processing	12	298	12.32	0.01
	Residual	60	24		
Sp_SS, Sp_TS-20, Sp_CS-80/16	SNR	5	4,864	8.71	0.01
	Processing	2	2,340	4.19	0.05
	Residual	10	559		

TABLE H.4. Experiment IV. Repeated measures analysis of variance (ANOVA) on percentage relative improvements (with reference to unprocessed speech) at different SNR conditions, for subjects and processing schemes and conditions (Sp_SS, Sp_TS-20, Sp_TS-40, Sp_TS-80, Sp_CS-20/4, Sp_CS-20/8, Sp_CS-20/16, Sp_CS-40/4, Sp_CS-40/8, Sp_CS-40/16, Sp_CS-80/4, Sp_CS-80/8, Sp_CS-80/16, Sp_CS-120/8, Sp_CS-120/16, Sp_CS-160/8, Sp_CS-160/16).

.

SNR	Source of variation	df	Mean sum of squares	*F*	*p*
∞ dB	Subjects	6	54	7.62	0.01
	Processing	16	197	28.10	0.01
	Residual	96	7		
3 dB	Subjects	6	222	10.58	0.01
	Processing	16	234	11.15	0.01
	Residual	96	21		
0 dB	Subjects	6	398	16.12	0.01
	Processing	16	239	9.68	0.01
	Residual	96	25		
-3 dB	Subjects	6	1,685	34.75	0.01
	Processing	16	797	16.44	0.01
	Residual	96	49		
-6 dB	Subjects	6	8,373	46.23	0.01
	Processing	16	5,753	31.77	0.01
	Residual	96	181		
-9 dB	Subjects	6	11,222	28.05	0.01
	Processing	16	13,736	34.33	0.01
	Residual	96	400		

TABLE H.5. Experiment IV. Repeated measures analysis of variance (ANOVA) on percentage relative improvements (with reference to unprocessed speech) at SNR = -6 dB and -9 dB, for subjects and processing conditions of (a) temporal splitting, (b) combined splitting, and (c) the 3 schemes with corresponding "optimal" conditions.

(A) SNR = -6 dB

Processing scheme	Source of variation	df	Mean sum of squares	*F*	*p*
Temporal splitting	Subjects	6	472	3.26	0.05
	Processing	2	5,486	37.80	0.01
	Residual	12	145		
Combined splitting	Subjects	6	7,965	67.80	0.01
	Processing	12	526	4.48	0.01
	Residual	72	118		
Sp_SS, Sp_TS-20, Sp_CS-80/16	Subjects	6	1,613	5.98	0.01
	Processing	2	10,189	37.76	0.01
	Residual	12	270		

(B) SNR = -9 dB

Processing scheme	Source of variation	df	Mean sum of squares	*F*	*p*
Temporal splitting	Subjects	6	821	9.71	0.01
	Processing	2	7,956	94.81	0.01
	Residual	12	84		
Combined splitting	Subjects	6	9,646	27.01	0.01
	Processing	12	1,612	4.51	0.01
	Residual	72	357		
Sp_SS, Sp_TS-20, Sp_CS-80/16	Subjects	6	2,809	3.13	0.05
	Processing	2	29,296	32.62	0.01
	Residual	12	898		

TABLE H.6. Experiment IV. Averaged percentage relative improvements and paired t-test significance levels *p* for different pairing of processing schemes and conditions at SNR = -6 dB and -9 dB, for the processing conditions of (a) temporal splitting, (b) combined splitting, and (c) the 3 schemes with corresponding “optimal” conditions. In each cell, the value within parentheses is the *p* value.

(A) SNR = -6 dB

(a) Temporal splitting

Proc. cond.	Sp_TS-20	Sp_TS-40	Sp_TS-80
Su	28.9 (0.04)	27.5 (0.07)	-20.3 (0.06)
Sp_TS-20		-1.3 (0.5)	-37.9 (0.000)
Sp_TS-40			-36.4 (0.003)

(b) Combined splitting

Proc. cond.	Sp_CS-20/4	Sp_CS-20/8	Sp_CS-20/16	Sp_CS-40/4	Sp_CS-40/8	Sp_CS-40/16	Sp_CS-80/4	Sp_CS-80/8	Sp_CS-80/16	Sp_CS-120/8	Sp_CS-120/16	Sp_CS-160/8	Sp-CS-160/16
Su	58.6 (0.001)	66.3 (0.000)	59.5 (0.000)	75.7 (0.000)	69.7 (0.000)	74.5 (0.000)	77.0 (0.000)	80.2 (0.000)	87.1 (0.000)	75.2 (0.000)	70.6 (0.000)	84.7 (0.000)	79.3 (0.000)
Sp_CS-20/4		5.4 (0.3)	0.8 (0.5)	10.9 (0.1)	7.6 (0.2)	10.9 (0.2)	12.4 (0.1)	14.8 (0.08)	18.5 (0.04)	11.2 (0.1)	7.8 (0.02)	16.8 (0.05)	13.3 (0.08)
Sp_CS-20/8			-4.3 (0.3)	5.3 (0.3)	-1.9 (0.4)	4.8 (0.3)	6.5 (0.2)	8.8 (0.2)	12.4 (0.07)	5.4 (0.2)	2.5 (0.03)	11.1 (0.09)	7.8 (0.2)
Sp_CS-20/16				10.1 (0.1)	6.7 (0.2)	9.7 (0.1)	11.8 (0.09)	14.3 (0.06)	18.1 (0.02)	10.7 (0.1)	7.2 (0.2)	16.3 (0.03)	12.7 (0.05)
Sp_CS-40/4					-3.0 (0.3)	-0.3 (0.5)	1.5 (0.4)	3.9 (0.3)	7.3 (0.2)	0.6 (0.5)	-2.6 (0.1)	5.7 (0.2)	2.6 (0.4)
Sp_CS-40/8						2.6 (0.4)	4.7 (0.3)	7.0 (0.2)	10.9 (0.1)	3.8 (0.3)	0.7 (0.1)	9.3 (0.1)	5.9 (0.2)
Sp_CS-40/16							2.1 (0.4)	4.4 (0.3)	8.5 (0.2)	1.4 (0.4)	-1.6 (0.3)	6.9 (0.2)	3.6 (0.3)
Sp_CS-80/4								2.3 (0.4)	6.1 (0.2)	-0.7 (0.4)	-3.5 (0.4)	4.7 (0.3)	1.6 (0.5)
Sp_CS-80/8									3.8 (0.3)	-2.8 (0.3)	-5.4 (0.4)	2.6 (0.4)	-0.7 (0.4)
Sp_CS-80/16										-5.9 (0.2)	-8.3 (0.4)	-0.9 (0.4)	-4.0 (0.3)
Sp_CS-120/8											-2.6 (0.3)	5.6 (0.2)	2.2 (0.4)
Sp_CS-120/16												8.6 (0.1)	5.4 (0.2)
Sp_CS-160/8													-2.8 (0.3)

(c) Three schemes with corresponding "optimal" conditions.

Proc. cond.	Sp_SS	Sp_TS-20	Sp_CS-80/16
Su	100.8 (0.000)	28.9 (0.04)	87.1 (0.000)
Sp_SS		-34.7 (0.000)	-6.4 (0.1)
Sp_TS-20			45.1 (0.000)

(B) SNR = -9 dB

(a) Temporal splitting

Proc. cond.	Sp_TS-20	Sp_TS-40	Sp_TS-80
Su	33.3 (0.02)	26.2 (0.06)	-30.4 (0.007)
Sp_TS-20		-5.9 (0.3)	-46.7 (0.000)
Sp_TS-40			-43.2 (0.000)

(b) Combined splitting

Proc. cond.	Sp_ CS- 20/4	Sp_ CS- 20/8	Sp_ CS- 20/16	Sp_ CS- 40/4	Sp_ CS- 40/8	Sp_ CS- 40/16	Sp_ CS- 80/4	Sp_ CS- 80/8	Sp_ CS- 80/16	Sp_ CS- 120/8	Sp_ CS- 120/16	Sp_ CS- 160/8	Sp- CS- 160/16
Su	83.4 (0.000)	87.3 (0.000)	95.3 (0.000)	103.1 (0.000)	92.0 (0.000)	116.4 (0.000)	110.7 (0.000)	137.6 (0.000)	126.2 (0.000)	105.5 (0.000)	111.9 (0.000)	108.1 (0.000)	109.7 (0.000)
Sp_CS-20/4		2.5 (0.5)	7.7 (0.3)	11.9 (0.2)	4.2 (0.4)	19.1 (0.1)	15.8 (0.1)	31.6 (0.06)	24.8 (0.04)	12.4 (0.2)	16.0 (0.2)	13.5 (0.2)	15.4 (0.2)
Sp_CS-20/8			5.3 (0.4)	10.4 (0.2)	2.1 (0.4)	16.5 (0.08)	13.8 (0.1)	29.0 (0.01)	22.2 (0.03)	10.2 (0.2)	14.1 (0.4)	11.7 (0.2)	13.7 (0.1)
Sp_CS-20/16				4.4 (0.3)	-1.4 (0.5)	10.9 (0.1)	8.5 (0.2)	21.6 (0.01)	15.7 (0.03)	5.5 (0.3)	9.2 (0.2)	7.0 (0.2)	7.5 (0.2)
Sp_CS-40/4					-4.3 (0.3)	7.1 (0.2)	4.6 (0.3)	16.9 (0.03)	11.3 (0.07)	1.9 (0.4)	5.3 (0.4)	3.3 (0.4)	3.2 (0.3)
Sp_CS-40/8						15.1 (0.1)	12.3 (0.2)	28.3 (0.03)	21.1 (0.06)	8.5 (0.3)	12.3 (0.5)	9.8 (0.2)	12.7 (0.2)
Sp_CS-40/16							-2.2 (0.4)	9.9 (0.1)	4.3 (0.3)	-5.0 (0.3)	-1.5 (0.4)	-3.5 (0.4)	-3.0 (0.3)
Sp_CS-80/4								12.8 (0.1)	7.0 (0.2)	-2.7 (0.4)	0.8 (0.3)	-1.3 (0.5)	-6.0 (0.4)
Sp_CS-80/8									-4.4 (0.2)	-12.4 (0.06)	-9.3 (0.2)	-10.7 (0.1)	-11.1 (0.06)
Sp_CS-80/16										-8.3 (0.1)	-4.9 (0.1)	-6.7 (0.2)	-6.9 (0.1)
Sp_CS-120/8											3.6 (0.3)	1.4 (0.4)	2.6 (0.4)
Sp_CS-120/16												-2.0 (0.4)	-0.8 (0.4)
Sp_CS-160/8													1.6 (0.5)

(c) Three schemes with corresponding "optimal" conditions.

Proc. cond.	Sp_SS	Sp_TS-20	Sp_CS-80/8
Su	151.7 (0.000)	33.3 (0.02)	137.6 (0.000)
Sp_SS		-45.8 (0.000)	-4.2 (0.2)
Sp_TS-20			78.8 (0.000)

TABLE H.7. Experiment V. Repeated measures analysis of variance (ANOVA) on percentage relative improvements (with reference to unprocessed speech) for subjects and processing schemes and conditions (a) different processing schemes and conditions (Sp_AG-SS, Sp_ AG-TS-20, Sp_ AG-TS-40, Sp_ AG-CS-20/8, Sp_ AG-CS-20/16, Sp_ AG-CS-40/8, Sp_ AG-CS-40/16, Sp_ AG-CS-80/8, Sp_ AG-CS-80/16), (b) temporal splitting with 2 conditions, (c) combined splitting with 6 conditions, and (d) schemes with corresponding "optimal" conditions.
.

Processing	Source of variation	df	Mean sum of squares	*F*	*p*
All processing schemes and conditions	Subjects	12	9,889	38.68	0.01
	Processing	9	3,658	14.31	0.01
	Residual	108	256		
Temporal splitting	Subjects	12	1,209	24.57	0.01
	Processing	1	480	9.75	0.01
	Residual	12	49		
Combined splitting	Subjects	12	7220	16.24	0.01
	Processing	5	2475	5.57	0.01
	Residual	60	444		
Sp_AG, Sp_SS, Sp_TS-20, Sp_CS-40/16	Subjects	12	3,730	10.92	0.01
	Processing	3	283	0.83	NS
	Residual	36	342		

TABLE H.8. Experiment V. Repeated measures analysis of variance (ANOVA) on percentage recognition scores for subjects and different processing schemes and conditions (Su, Sp_AG, Sp_AG-SS, Sp_ AG-TS-20, Sp_ AG-TS-40, Sp_ AG-CS-20/8, Sp_ AG-CS-20/16, Sp_ AG-CS-40/8, Sp_ AG-CS-40/16, Sp_ AG-CS-80/8, Sp_ AG-CS-80/16).
.

Processing	Source of variation	df	Mean sum of squares	*F*	*p*
All processing schemes and conditions	Subjects	12	2,624	21.18	0.01
	Processing	10	1389	11.21	0.01
	Residual	120	124		

TABLE H.9. Experiment V. Difference between averaged recognition scores and Tukey's HSD (honestly significant difference) *p* value for different pairing of processing schemes and conditions. In each cell, the value within parenthesis is the *p* value. Based on the values in Table H. 7b, the *q* values and HSD for significance levels are: p = 0.01, q = 5.38, HSD = 16.61; p = 0.05, q = 4.6, HSD = 14.3; p = 0.1, q = 4.28, HSD = 13.2.

Proc. cond.	Sp_AG	Sp_AG-SS	Sp_ AG-TS-20	Sp_ AG-TS-40	Sp_ AG-CS-20/8	Sp_ AG-CS-20/16	Sp_ AG-CS-40/8	Sp_ AG-CS-40/16	Sp_ AG-CS-80/8	Sp_ AG-CS-80/16
Su	11.1 (NS)	15.1 (0.05)	14.5 (0.05)	10.6 (NS)	-15.2 (0.05)	-14.4 (0.05)	0.2 (NS)	4.5 (NS)	3.9 (NS)	6.7 (NS)
Sp_AG-		4.0 (NS)	3.4 (NS)	-0.5 (NS)	-26.3 (0.01)	-25.5 (0.01)	-10.9 (NS)	-6.6 (NS)	-7.2 (NS)	-4.4 (NS)
Sp_AG-SS			-0.6 (NS)	-4.5 (NS)	-30.3 (0.01)	-29.5 (0.01)	-14.9 (0.05)	-10.5 (NS)	-11.1 (NS)	-8.4 (NS)
Sp_AG-TS-20				-3.9 (NS)	-29.7 (0.01)	-28.9 (0.01)	-14.3 (0.05)	-9.9 (NS)	-10.5 (NS)	-7.8 (NS)
Sp_AG-TS-40					-25.8 (0.01)	-25.0 (0.01)	-10.4 (NS)	-6.1 (NS)	-6.7 (NS)	-3.9 (NS)
Sp_AG-CS-20/8						0.8 (NS)	15.4 (0.05)	19.8 (0.01)	19.2 (0.01)	21.9 (0.01)
Sp_AG-CS-20/16							14.6 (0.05)	19.0 (0.01)	18.4 (0.01)	21.1 (0.01)
Sp_AG-CS-40/8								4.4 (NS)	3.8 (NS)	6.5 (NS)
Sp_AG-CS-40/16									-0.6 (NS)	2.1 (NS)
Sp_AG-CS-80/8										2.7 (NS)

TABLE H.10. Experiment V. Averaged percentage relative improvements and paired t-test significance levels *p* for different pairing of processing schemes and conditions for the processing conditions of (a) temporal splitting, (b) combined splitting, and (c) Three schemes with corresponding "optimal" conditions. In each cell, the value within parentheses is the *p* value

(a) Temporal splitting

Proc. cond.	Sp_AG-TS-20	Sp_AG-TS-40
Su	30.3 (0.03)	21.7 (0.09)
Sp_AG-TS-20		-5.9 (0.3)

(b) Combined splitting

Proc. cond.	Sp_AG-CS-20/8	Sp_AG-CS-20/16	Sp_AG-CS-40/8	Sp_AG-CS-40/16	Sp_AG-CS-80/8	Sp_AG-CS-80/16
Su	-16.7 (0.04)	-13.4 (0.04)	9.2 (0.5)	24.1 (0.3)	14.4 (0.3)	19.4 (0.2)
Sp_AG-CS-20/8		4.4 (0.5)	60.4 (0.03)	85.0 (0.007)	75.0 (0.01)	85.0 (0.005)
Sp_AG-CS-20/16			51.7 (0.03)	79.3 (0.006)	65.7 (0.01)	74.9 (0.005)
Sp_AG-CS-40/8				11.5 (0.3)	6.4 (0.3)	12.1 (0.2)
Sp_AG-CS-40/16					-1.6 (0.5)	3.4 (0.4)
Sp_AG-CS-80/8						5.4 (0.4)

(c) Three schemes with corresponding "optimal" conditions.

Proc. cond.	Sp_AG	Sp_AG-SS	Sp_AG-TS-20	Sp_AG-CS-40/16
Su	21.8 (0.09)	31.4 (0.03)	30.3 (0.03)	16.4 (0.3)
Sp_AG		7.1 (0.3)	6.4 (0.3)	-6.1 (0.2)
Sp_AG-SS			-0.5 (0.5)	-13.1 (0.06)
Sp_AG-TS-20				-18.5 (0.06)

REFERENCES

Adobe Systems Inc. (**1988**). *Postscript Language Reference Manual* (Addison Wesley, Boston, Massachusetts).

Apoux, F., Crouzet, O., and Lorenzi, C. (**2000a**). "Consonant recognition in noise with temporal cues I:. Effects of temporal enhancement on identification performance," J. Acoust. Soc. Am. **107**, 2913.

Apoux, F., Crouzet, O., and Lorenzi, C. (**2000b**). "Consonant recognition in noise with temporal cues II:. Effects of temporal enhancement on response times," J. Acoust. Soc. Am. **107**, 2913.

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

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

Baer, T., and Moore, B. C. J. (**1993**). "Effects of spectral smearing on the intelligibility of sentences in noise," J. Acoust. Soc. Am. **94**, 1229-1241.

Baer, T., and Moore, B. C. J. (**1994**). "Effects of spectral smearing on the intelligibility of sentences in the presence of interfering speech," J. Acoust. Soc. Am. **95**, 2277-2280.

Bailey, T. J. (**1981**). *Statistical Methods in Biology* (John Wiley, New York).

Baker, R. J., and Rosen, S. (**2002**). "Auditory filter nonlinearity in mild/moderate hearing impairment," J. Acoust. Soc. Am. **111**, 1330-1339.

Baragi, A. B. N. (**1996**). "*A speech training aid for the deaf*," M. Tech. Dissertation, Dept. of Elect. Engg., IIT Bombay, India.

Bess, F. H., and Humes, L. E. (**1995**). *Audiology: The Fundamentals*, 2nd Ed. (Williams and Wilkins, Maryland, Pennsylvania).

Biswas, A. (**1995**). *Clinical Audio-vestibulometry for Otologists and Neurologists* (Bhalani Publishing, Mumbai, India).

Boothroyd, A. (**1984**). "Auditory perception of speech contrasts by subjects with sensorineural hearing loss," J. Speech Hear. Res. **27**, 134-144.

Boothroyd, A., Mulheam, B., Gong, J., and Ostroff, J. (**1999**). "Effects of spectral smearing on phoneme and word recognition," J. Acoust. Soc. Am. **100**, 1807-1818.

Borden, G. J., Harris, K. S., and Raphael, L. J. (**1994**). *Speech Science Primer* (Williams and Wilkins, Maryland, Pennsylvania).

Carney, A., and Nelson, D. (**1983**). "An analysis of psychophysical tuning curves in normal and pathological ears," J. Acoust. Soc. Am. **73**, 268-278.

CHABA Working Group on Communication Aids for Hearing-Impaired (**1991**). "Speech-perception aids for hearing-impaired people: Current status and needed research," J. Acoust. Soc. Am. **90**, 637-683.

Chaudhari, D. S. (**2000**). "*Dichotic presentation for improving speech perception by persons with bilateral sensorineural hearing impairment,*" Ph.D. Thesis, School of Biomedical Engg., IIT Bombay, India.

Chaudhari, D. S., and Pandey, P. C. (**1998a**). "Dichotic presentation of speech signal using critical filter band for bilateral sensorineural hearing impairment," *Proc. 16th Int. Congress Acoust.* (*ICA*), Seattle, Washington.

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

Chaudhari, D. S., and Pandey, P. C. (**1998c**). "Real time speech processing for dichotic presentation for binaural hearing aids," *Proc. Nat. Conf. Biomedical Engg.* (*NCBME'98*), Manipal, Karnataka, India, II-45.

Chaudhari, D. S., and Pandey, P. C. (**1999a**). "A DSP board based spectrographic analyzer," *Nat. Conf. Electronic Materials, Devices, and Systems*, Gulbarga, Karnataka, India.

Chaudhari, D. S., and Pandey, P. C. (**1999b**). "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.

Crain, T. R., and Yund, W. (**1995**). "The effect of multichannel compression on vowel and stop consonant discrimination in normal-hearing and hearing-impaired subjects," Ear Hear. **16**, 529-543.

Crandall, I. B. (**1917**). "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).

Crystal, H., and House, A. (**1982**). "Segmental durations in connected speech signals: Preliminary results" J. Acoust. Soc. Am. **72**, 705-716.

Danaher, E. M., Wilson, M., and Pickett, J. M. (**1978**). "Backward and forward masking in listeners with severe sensorineural hearing loss," Audiol. **17**, 324-338.

Daniel, W. W. (**1999**). *Biostatistics: A Foundation for Analysis in the Health Science*, 7th Ed. (John Wiley, New York).

Derleth, R. P., Dau, T., and Kollmeier, B. (**2001**). "Modeling temporal and compressive properties of the normal and impaired auditory system," Hear. Res. **159**, 132-149.

Deutsch, l. J., and Richards, A. M. (**1979**). *Elementary Hearing Science* (Allyn and Bacon, Boston, Massachusetts).

Diehl, R., and Walsh, M. A. (**1989**). "An auditory basis for the stimulus length effect in the perception of stops and glides" J. Acoust. Soc. Am. **85**, 2154-2164.

Dorman, M. F., and Dougherty, K. (**1981**). "Shifts in phonetic identification with changes in signal presentation level," J. Acoust. Soc. Am. **69**, 1439-1440.

Dorman, M. F., and Hannley, M. T. (**1985**). "Identification of speech and speech like signals by hearing impaired listeners" in *Speech Science*, edited by Daniloff, R.G. (Taylor and Francis, London), pp. 111-153.

Dubno, J. R., and Dirks, D. D. (**1989**). “Auditory filter characteristics and consonant recognition for hearing-impaired listeners,” J. Acoust. Soc. Am. **85**, 1666-1675.

Dubno, J. R., and Levitt, H. (**1981**). “Predicting consonant confusions from acoustic analysis,” J. Acoust. Soc. Am. **69**, 249-261.

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

Dubno, J. R., Dirks, D. D., and Morgan, D. E. (**1984**). “Effects of age and mild hearing loss on speech recognition in noise,” J. Acoust. Soc. Am. **76**, 87-96.

Dubno, J. R., Dirks, D. D., and Schaefer, A. B. (**1987**). “Effects of hearing loss on utilization of short-duration spectral cues in stop consonant recognition,” J. Acoust. Soc. Am. **81**, 1940-1977.

Dynalog Microsystems (**1989**). *User’s Manual, PCL-208* (*PCL-718*) *Data Acquisition Card* (Dynalog Microsystems, Mumbai, India).

Elliott, L. L. (**1962**). “Backward masking: Monotic and dichotic conditions,” J. Acoust. Soc. Am. **34**, 1108-1115.

Fitzgibbons, P. J. (**1983**). “Temporal gap detection in noise as a function of frequency, bandwidth, and level” J. Acoust. Soc. Am. **74**, 67-72.

Fitzgibbons, P. J., and Gordon-Salent, S. (**1987**). “Temporal gap resolution in listeners with high-frequency sensorineural hearing loss” J. Acoust. Soc. Am. **81**, 133-137.

Fitzgibbons, P. J., and Wightman, F. L. (**1982**). “Gap detection in normal and hearing-impaired listeners” J. Acoust. Soc. Am. **72**, 761-765.

Flanagan, J. L. (**1972**). *Speech Analysis Synthesis and Perception* (Springer-Verlag, New York).

Fletcher, H. (**1953**). *Speech and Hearing in Communication* (Van Nostrand, New York).

Florentine, M., and Buus, S. (**1984**). “Temporal gap detection in sensorineural and simulated hearing impairments ,” J. Speech Hear. Res. **27**, 449-455.

Florentine, M, Fastl, H., and Buus, S. (**1988**). “Temporal integration in normal, cochlear impairment, and impairment simulated by masking,” J. Acoust. Soc. Am. **84**, 195-203.

French, N. R., and Steinberg, J. C. (**1947**). “Factors governing the intelligibility of speech sounds,” J. Acoust. Soc. Am. **19**, 90-119.

Gabrielsson, A., Schenkman, B. N., and Hagerman, B. (**1988**). “The effects of different frequency response on sound quality judgements and speech intelligibility,” J. Speech Hear. Res. **31**, 166-177.

Ganong W. F. (**1993**). *Review of Medical Physiology*, 16th Ed. (Prentice Hall, Englewood Cliffs, New Jersey).

Gatehouse, S., and Gordon, J. (**1990**). “Response times to speech stimuli as measures of benefit from amplification ,” Br. J. of Audiol. **24**, 63-68.

Gelfand, S. A. (**1998**). *Hearing: An Introduction to Psychological and Physiological Acoustics*, 3rd Ed. (Marcel Dekker, New York).

Glasberg, B. R., and Moore, B. C. J. (**1984**). “Dynamic range and asymmetry of the auditory filter,” J. Acoust. Soc. Am. **76**, 419-427.

Glasberg, B. R., and Moore, B. C. J. (**1986**). "Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments," J. Acoust. Soc. Am. **79**, 1020-1033.

Glasberg, B. R., and Moore, B. C. J. (**1987**). "Gap detection and masking in hearing-impaired and normal-hearing subjects," J. Acoust. Soc. Am. **81**, 1546-1556.

Glasberg B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched noise data," Hear. Res. **47**, 103-138.

Glasberg, B. R., Moore, B. C. J., and Bacon, S. P. (**1987**). "Gap detection and masking in hearing-impaired and normal-hearing subjects," J. Acoust. Soc. Am. **81**, 1546-1565.

Gordon-Salent, S. (**1984**). "Effects of reducing low-frequency amplification on consonant perception in quiet and noise ," J. Speech Hear. Res. **27**, 483-493.

Gordon-Salent, S. (**1986**). "Recognition of natural and time/intensity altered CVs by young and elderly subjects with normal hearing," J. Acoust. Soc. Am. **80**, 1599-1607.

Gordon-Salent, S. (**1987**). "Effect of acoustic modification on consonant recognition by elderly hearing-impaired subjects," J. Acoust. Soc. Am. **81**, 1199-1202.

Gutnick, H. N, (**1982**). "Consonant-feature transmission as a function of presentation level in hearing-impaired listeners," J. Acoust. Soc. Am. **72**, 1124-1130.

Hall, J. W., and Harvey, A. D. G. (**1985**). "Diotic loudness summation in normal and impaired hearing," J. Speech Hear. Res. **28**, 445-448.

Hall, J. W., and Wood, E. J. (**1984**). "Stimulus duration and frequency discrimination for normal-hearing and hearing-impaired subjects," J. Speech and Hear. Res. **27**, 252-256.

Hall, J. W., Tyler, R. S., and Fernandes, M. A. (**1984**). "Factors influencing the masking level difference in cochlear hearing-impaired and normal-hearing listeners," J. Speech Hear. Res. **27**, 145-154.

Hannley, M., and Dorman, M. F. (**1983**). "Susceptibility to intraspeech spread of masking in listeners with sensorineural hearing loss," J. Acoust. Soc. Am. 74, 40-51.

Harris, F. J. (**1978**). "On the use of windows for harmonic analysis with the discrete fourier transform," Proc. of the IEEE **66**, 51-83.

Hartmann, W. M (**1998**). *Signals, Sounds, and Sensation* (Springer-Verlag, New York).

Helfer, K. S. (**1994**). "Binaural cues and con sonant perception in reverberation and noise," J. Speech Hear. Res. **37**, 429-438.

Hellman, R. P., Zwislocki, J. (**1963**). "Monaural loudness function at 1000 cps and interaural summation," J. Acoust. Soc. Am. **35**, 856-865.

Hicks, M. L., and Bacon, S. P. (**1999**). "Psychophysical measures of auditory nonlinearities as a function of frequency in individuals with normal hearing," J. Acoust. Soc. Am. **105**, 326-338.

Hickson, L., Dobb, B., and Byrne, D. (**1994**), "Constant perception with linear and compression amplification," Scand. Audiol. **24**, 175-184.

Hidaka, H., Sasaki, N., Kawase, T., Takasaka, T., Ozawa, K., Suzuki, Y., and Sone, T. (**1998**). "Apparent effects of the use of digital hearing aid "CLAIDHA" on several hearing functions of impaired listeners," *Proc. 16th Int. Congress Acoust.* (*ICA*), Seattle, Washington, 2625-2626.

Horwitz, A. R., Turner, C. W., and Fabry, D. A. (**1991**). "Effects of different frequency response strategies upon recognition and preference for audible speech stimuli," J. Speech Hear. Res. **34**, 1185-1196.

Hou, Z., and Pavlovic, C. V. (**1994**). "Effects of temporal smearing on temporal resolution, frequency selectivity, and speech intelligibility," J. Acoust. Soc. Am. **96**, 1325-1340.

Humes, L. E., Dirks, D D, Bell, T. S., and Kincaid, G. E. (**1987**). "Recognition of nonsense syllables by hearing-impaired listeners and by noise-masked normal hearers," J. Acoust. Soc. Am. **81**, 765-773.

Humes, L. E., Espinoza-Varas, B., and Watson, C. S. (**1988**). "Modeling sensorineural hearing loss. I. Model and retrospective evaluation," J. Acoust. Soc. Am. **83**, 188-202.

Ifeachor, E. C., and Jervis, B. W. (**1997**). *Digital Signal Processing, A Practical Approach* (Addison Wesley, Boston, Massachusetts).

Irwin. R. J., and McAuley, S. F. (**1987**). "Relations among temporal acuity, hearing loss, and the perception of speech distorted by noise and reverberation," J. Acoust. Soc. Am. **81**, 1557-1565.

Jangamashetti, D. S (**2003**). "*Binaural dichotic presentation to reduce the effects of temporal and spectral masking due to sensorineural hearing loss*," Ph.D. Thesis, Dept. of Elect. Engg., IIT Bombay, India.

Jangamashetti, D. S., and Pandey, P. C. (**2000a**). "Dichotic presentation with inter-aural switching for reducing the effects of temporal masking due to sensorineural hearing loss," *Proc. National Conf. Biomedical Engg.*, Roorkee, India, 346-353.

Jangamashetti, D. S., and Pandey, P. C. (**2000b**). "Inter-aural switching with different fading functions for dichotic presentation to reduce the effect of temporal masking in sensorineural hearing loss," *Proc. 4th World Multi Conf. Systemics, Cybernetics, and informatics* (*SCI'2000*), Orlando, Florida, VI-434-439.

Jangamashetti, D. S., Pandey, P. C., and Cheeran, A. N. (**2001**). "Time varying comb filters to reduce effect of spectral and temporal masking in sensorineural hearing impairment," *Proc. Int. Conf. on Biomedical Engg.* (*ICMBE*), Bangalore, India, 258-263.

Jenison, R. L., Greenberg, S., Kluender, K. R., and Rhode, W. S. (**1991**). "A composite model of the auditory periphery for the processing of speech based on the filter response functions of single auditory-nerve fibers," J. Acoust. Soc. Am. **90**, 773-786.

Jesteadt. W. (**1997**). Eds. *Modeling Sensorineural Hearing Loss* (Lawrence Erlbaum, Mahwah, New Jersey).

Jesteadt. W., and Wier, C. C. (**1977**). "Comparison of monaural and binaural discrimination of intensity and frequency," J. Acoust. Soc. Am. **61**, 1599-1603.

Johnson, D., Whaley, P., and Dorman, M. F. (**1984**). "Processing of cues for stop consonant voicing by young hearing-impaired listeners," J. Speech Hear. Res. **27**, 112-118.

Kamm, C. A., Dirks, R., and Carterette, E. C. (**1982**). "Some effects of spectral shaping on recognition of speech by hearing-impaired listeners," J. Acoust. Soc. Am. **71**, 1211-1224.

Kasthuri, P. (**1997**). "*Real-time filter design with tunable frequency response using TMS320C50 processor*," Project Report, Signal Processing and Instrumentation Lab, Dept. of Elect. Engg., IIT Bombay.

Kates, J. M. (**1994**). “Speech enhancement based on a sinusoidal model,” J. Speech Hear. Res. **37**, 449-464.

Kelly, J. P. (**1991**). “Hearing,” in *Principles of Neural Sciences*, edited by Kandel, E. R., Schwartz, J. H., and Jessel, T. M (Elsevier, New York).

Kennedy, E., Levitt, H., Neuman, A. C., and Weiss, M. (**1998**). “Consonant-vowel intensity ratios for maximizing consonant recognition by hearing-impaired listeners,” J. Acoust. Soc. Am. **103**, 1098-1114.

Kent, R.D., and Read, C. (**1992**). *The Acoustic Analysis of Speech* (Singular Publishing, San Diego, California).

Koenig, W., Dunn, H. K., and Lacy, L. Y. (**1946**). “The sound spectrograph,” J. Acoust. Soc. Am. **18**, 19-49

Ladefoged, P. (**1982**). *A Course in Phonetics*, 2nd Ed. (Harcourt Brace Jovanovich, New York).

Leek, M. R., and Summers, V. (**1996**). “Reduced frequency selectivity and the preservation of speech contrast in noise,” J. Acoust. Soc. Am. **100**, 1796-1806.

Leek, M. R., Dorman, M. F., and Summerfield, Q. (**1987**). “Minimum spectral contrast for vowel identification by normal-hearing and hearing-impaired listeners,” J. Acoust. Soc. Am. **81**, 148-154.

Lehana, P. K., and Pandey, P. C. (**2004**) “Harmonic plus noise model based speech synthesis in Hindi and pitch modification”, *Proc. 18th Int. Congress Acoust.* (*ICA*), Kyoto, Japan, 3333-3336.

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

Lippmann, R., Braida, L., and Durlach, N. (**1981**). “Study of multichannel amplitude compression and linear amplification for persons with sensorineural hearing loss,” J. Acoust. Soc. Am. **69**, 524-534.

Lim, D. J. (**1986**). “Functional structure of the organ of Corti: A review,” Hear. Res. **22**, 117-146.

Lim, D. J. (**1980**). “Cochlear anatomy related to cochlear micromechanics: A review,” J. Acoust. Soc. Am. **67**, 1686-1695.

Loizou, P. C. (**1999**). “Signal processing techniques for cochlear implants,” IEEE Engg. Med. and Biol. **18**, 34-46.

Lunner, T. (**1997**). “*A digital filter bank hearing aid*,” Ph.D. Thesis, Dept. of Neuroscience and Locomotion, Linkoping Univ., Linkoping, Sweden.

Lunner, T., Arlinger, S., and Hellgren, J. (**1993**) “8-channel digital filter bank for hearing aid use: Preliminary results in monaural, diotic and dichotic modes,” Scand. Audiol., Suppl. **38**, 75-81.

Lyregaard, P. E. (**1982**). “Frequency selectivity and speech intelligibility in noise,” Scand. Audiol. Suppl. **15**, 113-122.

Mackenna, B. R., and Callander, R. (**1997**). *Illustrated Physiology*, 6th Ed. (Churchill Livingstone, New York).

Marks, L. E. (**1987**). “Binaural versus monaural loudness: Supersummation of tone partially masked by noise,” J. Acoust. Soc. Am. **81**, 122.

Marks, L. E. (**1978**). “Binaural summation of the loudness of pure tones,” J. Acoust. Soc. Am. **64**, 107-113.

Massaro, D. W., and Oden, G. C. (**1980**). “Evaluation and integration of acoustic features in speech perception,” J. Acoust. Soc. Am. **67**, 996-1013.

Meftah, M., and Boudelaa, S. (**1996**). “How facilitatory can lexical information be during word recognition? Evidence from moroccon Arabic,” *Proc. Int. Conf. Spoken Language Processing* (*ICSLP*), Philadelphia, Pennsylvania, vol. 1, 74-77.

Miller, G. A., and Nicely, P. E. (**1955**). “An analysis of perceptual confusions among some English consonants,” J. Acoust. Soc. Am. **27**, 338-352.

Minifie, F. D. (**1994**). Eds. *Introduction to Communication Science and Disorders* (Singular Publishing, San Diego, California).

Montgomery, A. A., and Edge, R. A. (**1988**). “Evaluation of two speech enhancement techniques to improve intelligibility for hearing-impaired adults,” J. Speech Hear. Res. **31**, 386-393.

Moore, B. C. J. (**1978**). “Psychophysical tuning curves measured in simultaneous and forward masking,” J. Acoust. Soc. Am. **63**, 524-532.

Moore, B. C. J. (**1986**). *Frequency Selectivity in Hearing* (Academic, London).

Moore, B. C. J. (**1997**). *An Introduction to Psychology of Hearing*, 4th Ed. (Academic, London).

Moore, B. C. J. (**1998**). “Psychoacoustics of cochlear impairment and design of hearing aids,” *Proc. 16th Int. Congress Acoust.* (*ICA*), Seattle, Washington, 2105-2108.

Moore, B. C. J., and Glasberg, B. R. (**1983**). “Suggested formulae for calculating auditory filter bandwidths and excitation patterns,” J. Acoust. Soc. Am. **74**, 750-753.

Moore, B. C. J., and Glasberg, B. R. (**1993**). “Simulation of the effects of loudness recruitment and threshold elevation on the intelligibility of speech in quiet and in a background noise,” J. Acoust. Soc. Am. **94**, 2050-2062.

Moore, B. C. J., Alcantara, J. I., and Glasberg, B. R. (**1998**). “Development and evaluation of a procedure for fitting multi-channel compression hearing aids,” Br. J. Audiol. **32**,177-195.

Moore, B. C. J., Glasberg, B. R., and Simpson, A. (**1992**). “Evaluation of a method of simulating reduced frequency selectivity,” J. Acoust. Soc. Am. **91**, 3402-3423.

Moore, B. C. J., Glasberg, B. R., and Stone, M. A. (**1993**), “Effect on the speech reception threshold in noise of the recovery time of the compressor in the high-frequency channel of a two-channel aid,” Scand. Audiol. **38**, 82-91.

Moore, B. C. J, Peters, R. W., and Glasberg, B. R. (**1990**). “Auditory filter shapes at low center frequencies,” J. Acoust. Soc. Am. **88**, 132-140.

Murase, A., Nakajima, F., Sakmoto, S., Suzuki, Y., and Kawase, T. (**2004**). “Effect of sound localization with dichotic-listening hearing aids,” *Proc. 18th Int. Congress Acoust.* (*ICA*), Kyoto, Japan, II-1519-1522.

Myers, J. L., and Well, W. A. (**2003**). *Research Design and Statistical Analysis* (Lawrence Erlbaum, Mahwah, New Jersey).

Nabelek, I. V. (**1983**). "Performance of hearing-impaired listeners under various types of amplitude compression," J. Acoust. Soc. Am. **74**, 776-791.

Nejime, Y., and Moore, B. C. J. (**1997**). "Simulation of the effect of threshold elevation and loudness recruitment combined with reduced frequency selectivity on the intelligibility of speech in noise ," J. Acoust. Soc. Am. **102**, 603-615.

Nejime, Y., and Moore, B. C. J. (**1998**). "Evaluation of the effects of speech-rate slowing on speech intelligibility in noise using a simulation of cochlear hearing loss," J. Acoust. Soc. Am. **103**, 572-576.

Nelson, D. A. (**1991**). "High-level psychophysical tuning curves: Forward masking in normal_hearing and hearing-impaired listeners," J. Speech Hear. Res. **34**, 1233-1249.

Nelson, D. A., and Fortune, T. W. (**1991**). "High-level psychophysical tuning curves: Simultaneous masking different noise bandwidths," J. Speech Hear. Res. **34**, 374-378.

Newby, H. A. (**1979**). *Audiology, 4th Ed.* (Prentice Hall, Englewood Cliffs, New Jersey).

Nilsson, M., Soli, S. D., and Sullivan, J. A. (**1994**). "Development of the hearing in noise test for the measurement of speech reception thresholds in quite and in noise," J. Acoust. Soc. Am. **95**, 1085-1099.

Oppenheim, A. V. (**1970**). "Speech spectrograms using the fast Fourier transform," IEEE Spectrum, 7, 57-62

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

O'Shaughnessy, D. (**1987**). *Speech Communication: Human and Machine* (Addison Wesley, Boston, Massachusetts).

O'Shaughnessy, D. (**2001**). *Speech Communication: Human and Machine*, 2nd Ed. (Universities Press, Hyderabad, India).

Oxenham, A. J., and Plack, C. J. (**1997**). "A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing," J. Acoust. Soc. Am. **101**, 3666-3675.

Oxenham, A. J., Moore, B. C. J., and Vickers, D. A. (**1997**). "Short-term temporal integration: Evidence for the influence of peripheral compression," J. Acoust. Soc. Am. **101**, 3676-3687.

Pandey, P. C. (**1987**). "*Speech processing for cochlear prostheses*," Ph.D. Thesis, Dept. of Elect. Engg., Univ. of Toronto, Canada.

Pandey, P. C., Jangamashetti, D. S., and Cheeran, A. N. (**2001**), "Binaural dichotic presentation to reduce the effect of increased temporal and spectral masking in sensorineural hearing impairment," J. Acoust. Soc. Am. **110**, 2705.

Pandey, P. C. Kunov, H., and Abel, S. M. (**1987**). "A speech processor providing fricative and low-frequency periodicity information for single channel cochlear prostheses," *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing* (*ICASSP '87),* 34.3.

Patterson, R. D. (**1976**). "Auditory filter shapes derived with noise stimuli," J. Acoust. Soc. Am. **59**, 640-654.

Payton, K. L, Uchanski, R. M., and Braida, L. D. (**1994**). "Intelligibility of conversational and clear speech in noise and reverbation for listeners with normal and impaired hearing," J. Acoust. Soc. Am. **95**, 1581-1592.

Phadke, M. S. (**1989**). *Quality Engineering Using Robust Design* (Prentice Hall, Englewood Cliffs, New Jersey).

Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1985**). "Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech," J. Speech Hear. Res. **28**, 96-103.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1986**). "Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech," J. Speech Hear. Res. **29**, 434-446.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1989**). "Speaking clearly for the hard of hearing III: An attempt to determine the contribution of speaking rate to differences in intelligibility between clear and conversational speech," J. Speech Hear. Res. **32**, 600-603.

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

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

Plomp, R. (**1988**). "The negative effect of amplitude compression in multichannel hearing aids in the light of modulation-transfer function," J. Acoust. Soc. Am. **83**, 2322-2327.

Plomp, R. (**1994**). "Noise, amplification, and compression: Consideration of three main issues in hearing aid design," Ear Hear. **15**, 2-12.

Prasad, V. V. R. (**1996**). "*Speech processing for single channel aid*," M. Tech. Dissertation, Dept. of Elect. Engg., IIT Bombay, India.

Proakis, J. G., and Manolakis, D. G. (**1997**). *Digital Signal Processing Principles, Algoritms, and Applications* (Prentice Hall, New Delhi, India).

Rabiner, L. R., and Gold, B. (**1998**). *Theory and Application of Digital Signal Processing* (Prentice Hall, New Delhi, India).

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

Rabiner, L.R., Gold, B., and McGonegal, C. A. (**1970**). "An approach to the approximation problem for nonrecursive digital filters," IEEE Trans. Audio and Electroacoustics, AU-**16**, 83-105.

Rand, T. C. (**1974**). "Dichotic release from masking for speech," J. Acoust. Soc. Am. **55**, 678-680.

Ratanpal, M. S. (**2000**). "*Speech processing for binaural presentation*," M. Tech. Dissertation, Dept. of Elect. Engg. IIT Bombay.

Reed, C. M, Hicks, B. L., Braida, L. D., and Durlach, N. I. (**1983**). "Discrimination of speech processed by low-pass filtering and pitch-invariant frequency lowering," J. Acoust. Soc. Am. **74**, 409-419.

Revoile, S. G., Holden-Pitt L., and Pickett, J. M. (**1985**). "Perceptual cues to the voiced-voiceless distinction of final fricatives for listeners with-impaired and normal-hearing," J. Acoust. Soc. Am. **77**, 1263-1265.

Revoile, S. G., Holden-Pitt, L., Pickett, J. M., and Brandt, F. (**1986**). “Speech cue enhancement for the hearing impaired I. Altered vowel durations for perception of final fricative voicing,” J. Speech Hear. Res. **29**, 240-255.

Revoile, S. G., Kozma-Spytek, L., Holden-Pitt L., Pickett, J. M., and Droge, J. (**1991a**). “VCV vs CVC for stop/fricative distinctions by hearing-impaired and normal-hearing listeners,” J. Acoust. Soc. Am. **89**, 457-460.

Revoile, S. G., Pickett, J. M., and Holden, L. D. (**1982**). “Acoustic cues to final stop voicing for impaired- and normal hearing listeners,” J. Acoust. Soc. Am. **72**, 1145-1154.

Revoile, S, G, Pickett, J. M., and Kozma-Spytek, L. (**1991b**). “Spectral cues to perception of / d,n,l / by normal- and impaired- hearing listeners,” J. Acoust. Soc. Am. **90**, 787-798.

Revoile, S. G., Pickett, J. M., Holden-Pitt, L., Talkin, D., and Brandt, F. D. (**1987**). “Burst and transition cues to voicing perception for spoken initial stops by impaired- and normal-hearing listeners,” J. Speech Hear. Res. **30**, 3-12.

Rintelmann, W.E. (**1991**). Eds. *Hearing Assessment*, 2ndEd. (Allyn and Bacon, Boston, Massachusetts).

Rosen, S., Baker, R.J., and Darling, A. (**1998**). “Auditory filter nonlinearity at 2 kHz in normal hearing listeners,” J. Acoust. Soc. Am. **103**, 2539-2550.

Sandlin, R. E. (**1988**). Eds. *Handbook of Hearing Aid Amplification, Vol. 1, Theoretical and Technical Considerations* (College-Hill, Boston, Massachusetts).

Sandlin, R. E. (**1994**). Eds. *Understanding Digitally Programmable Hearing Aids* (Allyn and Bacon, Boston, Massachusetts).

Sandlin, R. E. (**2000**). Eds. *Text Book of Hearing Aid Amplification Technical and Clinical Considerations*, 2nd Ed. (Singular Publishing, San Diego, California).

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

Scharf, B. (**1969**). “Dichotic summation of loudness,” J. Acoust. Soc. Am. **45**, 1193-1205.

Shailer, M. J., and Moore, B. C. J. (**1983**). “Gap detection as a function of frequency, bandwidth, and level,” J. Acoust. Soc. Am. **74**, 467-473.

Shailer, M. J., and Moore, B. C. J. (**1985**). “Detection of temporal gaps in bandlimited noise: Effects of variation in bandwidth and signal-to-masker ratio,” J. Acoust. Soc. Am. **77**, 635-639.

Shailer, M. J. Moore, B. C. J., Glasberg, B. R., Watson, N., and Harris, S. (**1990**). “Auditory filter shapes at 8 and 10 kHz,” J. Acoust. Soc. Am. **88**, 141-148.

Simon, C. (**1978**). “On the use of comfortable listening levels in speech experiments,” J. Acoust. Soc. Am. **64**, 744-750.

Smiarowski, R. A., and Carhart, R. (**1975**). “Relation among temporal resolution, forward masking, and simultaneous masking,” J. Acoust. Soc. Am. **57**, 1169-1173.

Snedecor, G. W., and Cochran, W. G. (**1980**). *Statistical Methods* (The Iowa State Uni. Press, Ames, Iowa).

Spelman, F. A. (**1999**). “The past, present, and future of cochlear prostheses,” IEEE Engg. Med. and Biol. **18**, 27-33.

Stevens, K. N. (**1980**). “Acoustic correlates of some phonetic categories,” J. Acoust. Soc. Am. **68**, 836-842.

Stevens, K. N., and Blumstein, S. E, (**1978**). “Invaient cues for place of articulation in stop consonants,” J. Acoust. Soc. Am. **64**, 1358-1368.

Stevens, S. S. (**1955**). “The measurement of loudness,” J. Acoust. Soc. Am. **27**, 815-829.

Stone, M. A., and Moore, B. C. J. (**1992**). “Syllabic compression: Effective compression ratios for signals modulated at different rates,” Br. J. of Audiol. **26**,351-361.

Studebaker, G. A., and Hockberg, I. (**1993**). *Acoustical Factors Affecting Hearing Aid Performance*, 2nd Ed. (Allyn and Bacon, Boston, Massachusetts).

Stylianou, Y. (**2001**). “Applying the harmonic plus noise model in concatenative speech synthesis,” IEEE Trans. Speech Audio Processing vol. **9**, 21-29.

Summers, V., and Leek, M. R. (**1992**). “The role of spectral and temporal cues in vowel identification by listeners with impaired hearing,” J. Speech Hear. Res. **35**, 1189-1199.

Summers, V., and Leek, M. R. (**1994**). “The internal representation of spectral contrast in hearing-impaired listeners,” J. Acoust. Soc. Am. **95**, 3518-3528.

Summers, V., and Leek, M. R. (**1997**). “Intraspeech spread of masking in normal-hearing and hearing-impaired listeners,” J. Acoust. Soc. Am. **101**, 2866-2876.

ter Keurs, M., Festen, J. M., and Plomp, R. (**1992**). “Effect of spectral envelope smearing on speech reception. I,” J. Acoust. Soc. Am. **91**, 2872-2880.

ter Keurs, M., Festen, J. M., and Plomp, R. (**1993**). “Effect of spectral envelope smearing on speech reception. II,” J. Acoust. Soc. Am. **93**, 1547-1552.

Thomas, T.G. (**1996**). “*Experimental evaluation of improvement in speech perception with consonantal intensity and duration modification*,” Ph.D. Thesis, Dept. of Elect. Engg., IIT, Bombay, India.

Thomas, T. G. Pandey, P. C., and Agashe, S. D. (**1994**). “A PC-Based multi-resolution spectrograph,” J. IETE (India) **40**, 105-108.

Thomas, T. G. Pandey, P. C., and Agashe, S. D. (**1996**). “Effect of consonantral intensity and duration modification on speech perception by listeners with simulated hearing impairment,” J. Acoust. Soc. India **24**, VI-4.1-4.5.

Turek, S. V. G., Dorman, M. F., Franks, J. R., and Summerfield, Q. (**1980**). “Identification of synthetic / bdg / by hearing-impaired listeners under momotic and dichotic formant presentation,” J. Acoust. Soc. Am. **67**, 1031-1040.

Turner, C. W., and Robb, M. P. (**1987**). “Audibility and recognition of stop consonants in normal and hearing-impaired subjects,” J. Acoust. Soc. Am. **81**, 1566-1573.

Turner, C. W., Fabry D. A., Barrett, S., and Horwitz, A. R. (**1992**). “Detection and recognition by normal-hearing and hearing-impaired listeners,” J. Speech Hear. Res. **35**, 942-949.

Turner, C. W., Smith S. J., Aldridge P. L., and Stewart, S. L. (**1997**). “Formant transition duration and speech recognition in normal and hearing-impaired listeners,” J. Acoust. Soc. Am. **101**, 2822-2825.

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

Tyler, R. S., Summerfield, Q., Wood, E. J., and Fernandes, M. A. (**1982**). “Psychoacoustic and phonetic temporal processing in normal and hearing-impaired listeners,” J. Acoust. Soc. Am. **72**, 740-752.

Tyler, R. S., Wood, E. J., and Fernandes, M. (**1983**). "Frequency resolution and discrimination of constant and dynamic tones in normal and hearing-impaired listeners," J. Acoust. Soc. Am. **74**, 1190-1199.

Uchanski, R. M., Choi. S. S., Braida, L. D , Reed C. M., and Durlach, N. I. (**1996**). "Speaking clearly for the hard of hearing IV: Further studies of the role of speaking rate," J. Speech Hear. Res. **39**, 494-509.

Van Dijkhuizen, J. N., Festen, J. M., and Plomp, R. (**1991**). "The effect of frequency-selective attenuation on the speech-reception threshold of sentences in conditions of low-frequency noise," J. Acoust. Soc. Am. **90**, 885-894.

Van Tasell, D. J., Hagen, L. T, Koblas, L. L., and Penner, S. G. (**1982**). "Perception of short-term spectral cues for stop consonant place by normal and hearing-impaired subjects," J. Acoust. Soc. Am. **72**, 1771-1780.

Veerarajan, T. (**2003**). *Probability, Statistics and Random Processes* (Tata McGraw-Hill, New Delhi, India).

Villchur, E. (**1977**). "Electronic models to simulate the effect of sensory distortions on speech perception by the deaf," J. Acoust. Soc. Am. **62**, 665-674.

Villchur, E. (**1974**). "Simulation of the effect of recruitment on loudness relationships in speech," J. Acoust. Soc. Am. **56**, 1601-1611.

Villchur, E. (**1973**). "Signal processing to improve speech intelligibility in perceptive deafness," J. Acoust. Soc. Am. **53**, 1646-1657.

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

Wang, M.D., and Bilger, R.C., (**1973**). "Consonant confusions in noise: A study of perceptual features," J. Acoust. Soc. Am. **54**, 1248-1266 .

Williams, P.L., Warwide, R., Dyson, M., and Bannister, L.H. (**1989**). *Gray's Anatomy.* (Churchill Livingstone, London, UK).

Wilson, R.H., Arcos J.T., and Jones, H.C. (**1984**). "Word recognition with segmented-alternated CVC words: A preliminary report on listeners with normal hearing," J. Speech Hear. Res. **27**, 378-386.

Yost, W.A. (**1994**). *Fundamentals of Hearing: An Introduction* (Academic, San Diego, California).

Yund, E.W., and Buckles, K.M. (**1995a**). "Discrimination of multichannel-compressed speech in noise: Long-term learning in hearing-impaired subjects," Ear Hear. **16**, 417-427.

Yund, E. W., and Buckles, K. M. (**1995b**). "Multichannel compression hearing aids: Effect of number of channels on speech discrimination in noise," J. Acoust. Soc. Am. **97**, 1206-1223.

Zeng, F., and Turner, C. W. (**1990**). "Recognition of voiceless fricatives by normal and hearing-impaired- subjects," J. Speech Hear. Res. **33**, 440-449.

Zhou, B. (**1995**). "Auditory filter shapes at high frequencies," J. Acoust. Soc. Am. **98**, 1935-1942.

Zwicker, E. (**1961**). "Subdivision of the audible range into critical bands (Frequenzgruppen)," J. Acoust. Soc. Am. **33**, 248.

Zwicker, E., Flottorp, G., and Stevens, S.S. (**1957**). "Critical band width in loudness summation," J. Acoust. Soc. Am. **29**, 548-557.

LIST OF PUBLICATIONS

1. A. N. Cheeran and P. C. Pandey, "Speech processing for hearing aids for moderate bilateral sensorineural hearing loss," in *Proc. of IEEE Int. Conf. on Acoust., Speech and Signal Processing* (Montreal, Quebec), 2004, pp. IV-17–20.
2. A. N. Cheeran and P. C. Pandey, "Evaluation of speech processing schemes using binaural dichotic presentation to reduce the effect of masking in hearing-impaired listeners," in *Proc. of 18th Int. Congress on Acoust.*. (Kyoto, Japan), 2004, pp. II-1523 –1526.
3. A. N. Cheeran, and P. C. Pandey, "Binaural dichotic presentation for moderate bilateral sensorineural hearing-impaired," in *Proc of Int. Conf. on Systemics, Cybernatics and Informatics* (Hyderabad, India), 2004, pp. 377–382.
4. A. N. Cheeran, and P. C. Pandey, "dichotic presentation for binaural hearing aids using perceptually balanced critical bandwidth based comb filters," in *Proc. 7th World Multi Conf. on Systemics, Cybernatics, and Informatics* (Orlando, Florida), 2003, pp. IV-360 –365.
5. A. N. Cheeran, and P. C. Pandey, "Dichotic presentation to overcome the effect of increased spectral masking and frequency dependant hearing threshold shifts in persons with bilateral sensorineural hearing impairment," *J. Acoust. Soc. Am.*, vol. 112, pp. 2359, 2003.
6. A. N. Cheeran, and P. C. Pandey, "Optimizing the sweep cycle of time-varying comb filters for binaural dichotic presentation in sensorineural hearing impairment," in *Proc. 14th Int. Conf. Digital Signal Processing* (Santorini, Greece), 2002, pp. 1145 –1148.
7. A. N. Cheeran, P. C. Pandey, and D.S. Jangamashetti, "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), 2002, pp. 971–974.
8. A. N. Cheeran, P. C. Pandey, and D.S. Jangamashetti, "Optimal sweep cycle for time-varying comb filters for binaural dichotic presentation to improve speech perception in sensorineural hearing impairment," *J. Acoust. Soc. Am.*, vol. 111, pp. 2426, 2002.
9. A. N. Cheeran, P. C. Pandey, and D.S. Jangamashetti, "Comb filters for binaural dichotic presentation to improve speech perception by persons with bilateral sensorineural hearing impairment," *J. Acoust. Soc. Am.*, vol. 110, pp. 2705, 2001.
10. P. C. Pandey, D. S. Jangamashetti, and A. N. Cheeran, "Binaural dichotic presentation to reduce the effect of increased temporal and spectral masking in sensorineural hearing impairment," *J. Acoust. Soc. Am.*, vol. 110, pp. 2705, 2001.
11. D. S. Jangamashetti, P. C. Pandey, and A. N. Cheeran, "Time varying comb filters to reduce effect of spectral and temporal masking in sensorineural hearing impairment," in *Proc. Int. Conf. Biomedical Engineering* (Bangalore, India), 2001, pp. 258–263.

Acknowledgements

It is my pleasure to acknowledge the support of numerous people who have helped me physically and morally during the tenure of my research. I give all glory to God the Almighty who has bestowed me with adequate strength all through this period of many ups and downs.

Special thanks and devotion are due unto my thesis supervisor Dr. P. C. Pandey, who despite his busy schedule extended guidance without which this endeavor would not have been completed. I am indebted to him for helping me to complete this thesis.

I express my sincere gratitude to all the IIT professors particularly Dr. S. D. Agashe, Dr. V. M. Gadre, and Dr. Preeti Rao for their helpful suggestions and comments.

Heartfelt thanks also goes to my colleagues in SPI lab whom I have interacted with during the course of my research. Some of them are Praveen, Milind, Vinod, Vidyadhar, and Dinesh. Outside the lab, Asha and Sincy have given me much encouragement during the coffee break occasions.

My sincere gratitude goes towards Prof. R. Rangasayee, Director, Dr. V. Basavaraj, Dy. Director, and Prof. Rajiv Jelvi, Head of Audiology Dept., of Ali Yavar Jung National Institute for the Hearing Handicapped (AYJNIHH), Mumbai, for sparing their time and supporting me in this endeavor. I appreciate the benefit of access to their library.

I specially thank all the normal hearing and hearing impaired subjects who have participated in the listening tests. This thesis would not have seen light, if they had not “lent their ears” of support.

Thanks to the authorities of VJTI, who have sponsored me for the Ph.D programme under QIP scheme, especially to the former principal Dr. S. D. Varwandkar who has always been an encouragement. Also I should thank Dr. M. Dixit, the Director, and Dr. B.K. Lande, Head of Elect. Engg. Dept., for their support. I appreciate the timely help and encouragements of many of my colleagues, especially Mrs. J.O. Chandle and Mrs. A. Jeyakumar.

Words are not sufficient to convey my feelings towards my husband Noble and children Joe, Ann, and Merin who have patiently endured long hours of my staying away from home. I have deprived them of my love, presence, and support many a times during the tenure of research. They remain among those who motivated me continually.

I also remember my mother and other members of my family and church, who have prayed for me and provided the needed courage.

The chain of gratitude is never-ending. There are several others who have played significant role behind my success. Though I am not able to mention their names, deep in my heart there is an overwhelming feeling of warmth and gratitude.

Alice N. Cheeran

SYNOPSIS

Introduction

Sensorineural hearing impairment is characterized by frequency dependent shifts in hearing threshold, loudness recruitment, reduced frequency and temporal resolution, and increased spectral and temporal masking [1, 2]. Due to increase in threshold of hearing with no corresponding increase in loudness discomfort level, the dynamic range reduces. Increased spectral masking causes smearing of spectral peaks and results in degradation of speech perception [3]. Increased temporal masking leads to increase in forward and backward masking of weak acoustic segments by adjacent strong ones. Cues, like voice-onset-time, formant transition, and burst duration, which are important for the identification of consonants, get masked by the following or preceding vowel segment, resulting in degraded speech perception.

Masking takes place primarily at the peripheral level, while integration of information takes place at higher levels in the auditory system [4]. The effect of increased masking may be reduced by splitting speech into two complementary signals in such a way that signal components, likely to mask or get masked, are presented to different ears. Speech processing schemes, using a pair of comb filters with complementary magnitude responses, have helped in reducing the effect of increased spectral masking [5, 6]. A scheme of temporal splitting, in which speech was switched between two ears using trapezoidal fading function with an inter-aural switching period of 20 ms, has helped in reducing the effect of increased temporal masking [7].

In the scheme of spectral splitting, the sensory cells corresponding to alternate bands of the basilar membrane are always stimulated, whereas sensory cells of the other bands do not receive stimulation. In the temporal splitting scheme, all the sensory cells of the two ears get relaxed alternately for some time. A combined splitting scheme was devised to provide all the sensory cells of the basilar membrane periodic relaxation from stimulation, and thereby achieve a simultaneous reduction in the adverse effects of increased spectral and temporal masking. In an earlier investigation [8], combined splitting was evaluated using a pair of time-varying comb filters with pre-calculated sets of coefficients, which were selected in steps for a cyclic sweeping of magnitude responses such that the pass bands of each of these comb filter pairs are shifted in a complementary manner along the frequency axis. The scheme was implemented for 2, 4, 8, and 16 shiftings, for a sweep cycle of 20 ms.

The objective of the research is to investigate the use of binaural dichotic presentation for improving the speech perception by persons with moderate bilateral sensorineural hearing impairment and by normal hearing persons under adverse listening conditions, and to find the optimal splitting scheme and associated processing parameters. For spectral splitting, the auditory critical bandwidth based comb filters designed with linear phase response and sharp inter-band transitions [6], were found to exhibit change in perceived loudness with frequency. It was decided to design and evaluate the comb filters with perceptually balanced magnitude response, to minimize the change in perceived intensity with frequency by reducing the passband ripple, controlling the variation in inter-band crossover gain, and improving stop-band attenuation. For temporal splitting, previous investigations [7], with different fading functions and duty cycles have shown that highest improvement was obtained with trapezoidal fading functions with 3 ms transition duration and 70 % duty cycle. The effect of various inter-aural switching periods with this fading function needed to be investigated. Combined splitting has earlier been investigated [8] with cyclic sweeping of a set of comb filters, with sweep cycle duration of 20 ms. It was decided to investigate the effect of various sweep cycle durations with different number of shiftings. Finally, an overall evaluation of the three schemes with different processing parameters was carried out, to find the optimal processing scheme and the associated parameters.

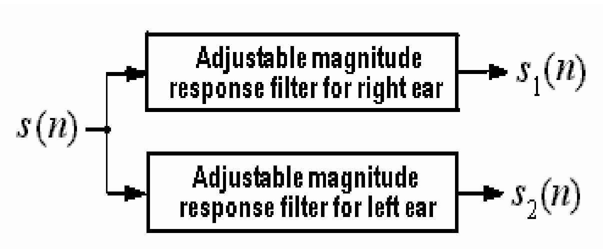
Investigations

Investigations involving implementation and evaluation of the dichotic presentation schemes were conducted in two phases. First phase of this research involved design, implementation, and evaluation of perceptually balanced comb filters. These were designed, based on 18 auditory critical bandwidths over 5 kHz range, as described by Zwicker [9], with improved magnitude responses to minimize the changes in intensity perception with frequency. The design of filters involved adjustment in magnitude response at transition crossovers, reduction in passband ripple, and increase in stop-band attenuation. Spectral splitting scheme was implemented and evaluated using these comb filters on normal hearing subjects with simulated loss and hearing impaired subjects. The second phase of investigations involved implementation and evaluation of speech processing schemes namely, spectral splitting, temporal splitting, and combined splitting for different processing conditions on normal hearing persons with simulated loss and on persons with moderate bilateral sensorineural hearing impairment. Spectral splitting was evaluated for perceptually balanced comb filters, since it was found to provide better improvement compared to comb filters with sharp transitions. On the basis of the results of the previous investigations [6], it was decided to carry out the evaluation of temporal splitting scheme for trapezoidal fading functions with 70 % duty cycle and 3 ms transition duration. Since the earlier evaluation was carried out for inter-aural switching period of only 20 ms, the present evaluation was conducted to establish the optimal value of inter-aural switching period. Combined splitting scheme was implemented with cyclic sweeping of perceptually balanced comb filters and evaluated for various combinations of sweep cycle duration and number of shiftings. An adjustable magnitude response filter with gain variation within ± 3 dB, to partly compensate for the frequency dependent shifts in hearing thresholds, alone and cascaded with other schemes, was also evaluated.

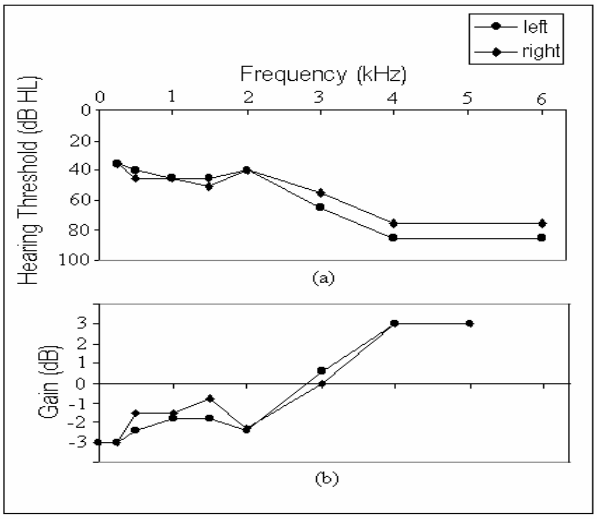
The schematic representations of the schemes are shown Figs 1 – 4. Figure 1 shows a schematic representation of the adjustable magnitude response filter and the pure tone audiogram and the desired magnitude response of the two compensating filters for one subject. A schematic representation of the adjustable magnitude response filter cascaded with spectral splitting scheme and the perceptually balanced comb filters used for spectral splitting are shown in Fig. 2. Similar representations for temporal and combined splitting schemes are shown in Fig. 3 and Fig. 4 respectively.

For listening tests on normal hearing listeners, simulation of hearing loss was carried out, by adding broadband Gaussian noise to the signal at different levels. Noise was added to speech signal based on short-time (≈ 10 ms) signal energy keeping the SNR constant, which resulted in no background noise during silence intervals. By analyzing the results from listening tests with unprocessed speech, it was verified that addition of broadband noise at various levels resulted in degradation in response time, recognition score, and relative information transmitted for different features, almost similar to that of different levels of sensorineural hearing impairment. Subjects with bilateral sensorineural loss were selected primarily on the basis of their willingness to participate in at least one set of experiments. The nature and extent of loss varied widely across these subjects.

In the first phase, magnitude response of the comb filter for spectral splitting was investigated. The comb filters based on 18 auditory critical bandwidths over 5 kHz range, used for spectral splitting, were earlier designed with sharp transitions between bands [6]. These filters introduced perceptual distortion because of passband ripple of up to 4 dB, stop-band attenuation as low as 10 dB, and inter-band crossover range of 0 – 10 dB. Investigations involving monaural/binaural presentation, filter design, and listening tests showed that crossovers lying in the 4 – 6 dB range resulted in perceptual balance. Comb filters were designed as 256-coefficient linear phase FIR filters, to obtain crossovers within 4 – 6 dB and the passband ripple restricted to 1 dB. These provided stop-band attenuation of 30 dB. These comb filters when tested with slowly swept sine waves showed no noticeable perceptual distortion.

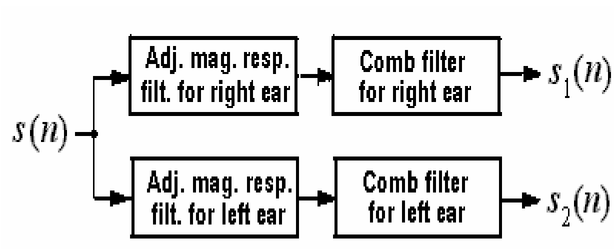


(a)

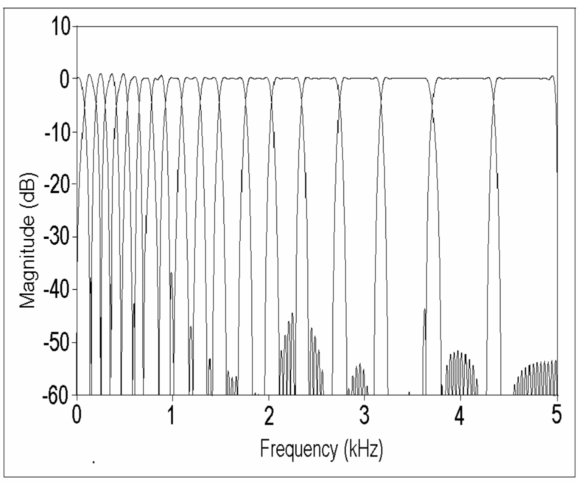


(b)

Fig. 1. Adjustable magnitude response filter (a) schematic, (b) pure tone audiogram of a subject and the corresponding magnitude responses.

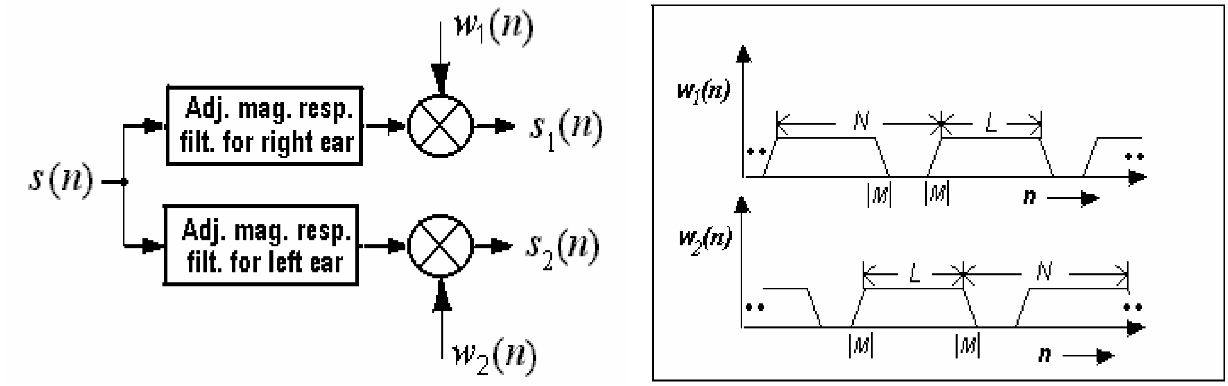


(a)



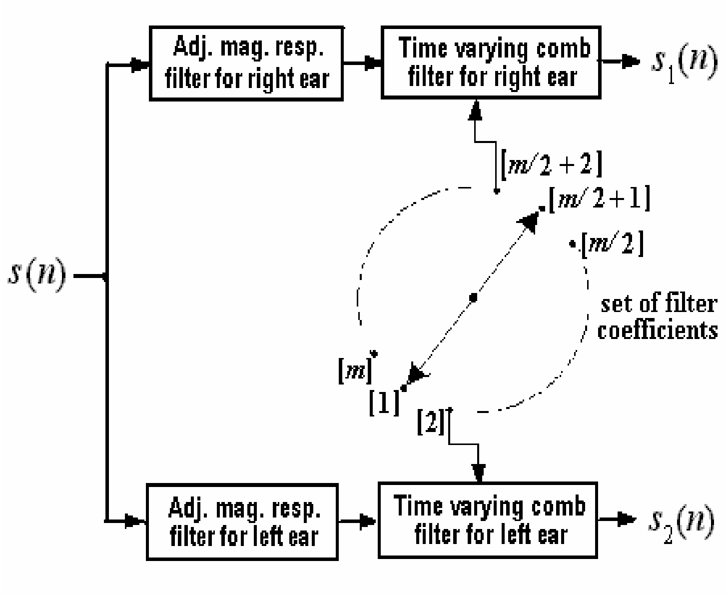
(b)

Fig. 2. Adjustable magnitude response filter cascaded with spectral splitting (a) Schematic, (b) magnitude response of the comb filter pair.



(a) (b)

Fig. 3. Adjustable magnitude response filter cascaded with temporal splitting (a) schematic (b) trapezoidal fading functions with inter-aural switching period of *N samples,* duty cycle *L*/*N,* transition duration of *M samples.*



(a)

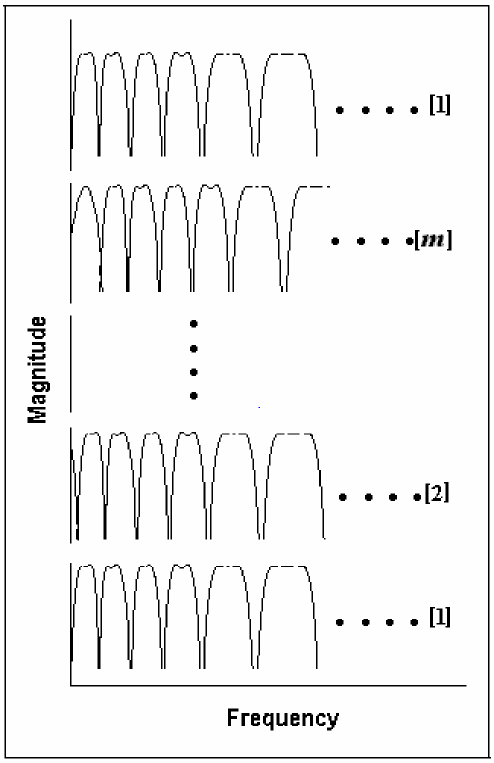


Fig. 4. Adjustable magnitude response filter cascaded with combined splitting scheme (a) schematic and (b) magnitude response of a time-varying comb filter

(b)

To evaluate the improvements, listening tests were conducted with both the comb filters, namely the one with sharp transitions and the perceptually balanced comb filters. Evaluation was first conducted on five normal hearing subjects with simulated loss and subsequently on five persons with moderate bilateral hearing loss. The tests involving consonant identification were carried out with 12 consonants in VCV context with vowel /*a*/, in order to evaluate reduction in the effect of masking. For subjects with bilateral sensorineural loss, adjustable magnitude response filters with gain variation within ±3 dB cascaded with the perceptually balanced comb filters were also used. A computerized listening test administration set-up was used to present stimuli, obtain the subject's response, and tabulate these in the form of stimulus-response confusion matrices. The response time, recognition scores, and relative information transmitted for consonantal features [10] were used for the performance evaluation.

The second phase involved evaluation of the three binaural dichotic presentation schemes, namely, spectral splitting to reduce the effect of increased spectral masking, temporal splitting to reduce the effect of increased temporal masking, and combined splitting to simultaneously reduce the effects of increased spectral and temporal masking, and the associated processing conditions. Listening test consisted of words presented in a randomized order from a phonetically balanced list of monosyllabic words in Marathi, Hindi, or English, was carried out initially on seven normal hearing subjects with simulated loss. Based on the results on normal hearing subjects the range of processing conditions was narrowed down, and tests were conducted on thirteen subjects with bilateral sensorineural loss. Adjustable magnitude response filters and their cascading with the three dichotic schemes were also evaluated. A computerized test administration set-up was used to conduct the listening tests. Subject’s response time and recognition score were used for evaluation and were analyzed.

Conclusions

The overall conclusions from our implementation and evaluation of dichotic presentation schemes can be summarized as follows

1) The binaural dichotic presentation schemes with spectral, temporal, and combined splitting reduced the load on speech perception process and improved speech perception by the subjects with moderate to severe bilateral loss, as well as by normal hearing subjects under adverse listening conditions.

2) In case of spectral splitting, auditory critical bandwidth based comb filters with linear phase response and perceptually balanced magnitude response with 1 dB passband ripple, inter-band crossovers adjusted within 4 – 6 dB and stop-band attenuation greater than 30 dB, were found to be superior to the comb filters with sharp inter-band transitions

3) In case of temporal splitting, employing trapezoidal fading with 70 % duty cycle and 3 ms transition duration (to suppress the inter-aural switching distortion), the optimal inter-aural switching period was T_c = 20 – 40 ms.

4) In case of combined splitting with time-varying comb filters employing cyclic sweeping of a set of perceptually balanced comb filters, the best conditions were sweep cycle duration of T_c = 40 – 80 ms and number of shiftings m = 8 and 16.

5) Evaluation of the three binaural dichotic presentation schemes, with phonetically balanced words, showed that for normal hearing subjects with simulated loss, the highest improvements in response time as well as in recognition score were generally provided by spectral splitting, closely followed by combined splitting. For 60% recognition score, the SNR advantage with best processing parameters, was 1.5 dB for temporal splitting, 4 – 4.5 dB for combined splitting, and 5 dB for spectral splitting.

6) Evaluation of the dichotic schemes showed that for subjects with sensorineural loss, the optimal conditions based on improvements in recognition scores and response time were generally in agreement. Under optimal set of processing parameters, the relative improvements in recognition scores ranged from 6 to 110 % for temporal splitting, -30 to 138 % for combined splitting, 6 to 121 % for spectral splitting, and 1 to 66 % for adjustable magnitude response filtering. Across the subjects, the relative improvements under optimal processing scheme and parameters ranged over 6 – 138 %, with an average of 35 %.

7) Evaluation on hearing impaired subjects showed that the effectiveness of the schemes varied across the subjects. Adjustable magnitude response filter gave highest improvement for subjects with low frequency loss. Spectral splitting provided maximum improvement for subjects with almost flat loss, sloping high frequency loss, and asymmetrical loss with less loss at mid frequencies. Temporal splitting was found to be better for subjects with sloping high frequency loss, and mid frequency loss. Combined splitting gave maximum improvement for subjects with severe symmetrical loss and moderately severe high frequency and mid frequency loss.

The effectiveness of the various dichotic presentation schemes for an individual listener is likely to be related to the nature and extent of spread of masking. Psychoacoustic assessment of the spread of masking for these subjects may be useful in relating the suitability of the schemes. As these assessments are often tedious, particularly for children and elderly persons with hearing impairment, it may be useful to develop speech processors with the option of selecting the scheme and parameters best suited for the individual.

It is to be noted that combined splitting provides periodic stimulation and relaxation for all the surviving cells. Hence from a physiological perceptive, this is a better scheme than spectral splitting which stimulates cells in alternate bands.

Some of the suggestions for further work are: (a) use of multi-band compression to reduce the effects of frequency dependent loudness recruitment along with the binaural dichotic presentation schemes, (b) finding optimal number of bands and bandwidths for the comb filters in spectral and combined splitting for different subjects, (c) development and evaluation of techniques to avoid splitting of perceptually salient acoustic subsegments, (d) listening tests involving larger number of subjects and more types of test material, (e) studying the effect of dichotic presentation on sound localization, and (f) implementation of dichotic presentation schemes in wearable hearing aids and testing.

References

[1] T. Baer and B. C. J. Moore, "Effects of spectral smearing on the intelligibility of sentences in noise," *J. Acoust. Soc. Am.*, vol. 94, pp. 1229-1241, 1993.

[2] B. C. J. Moore, "Psychoacoustics of cochlear hearing impairment and design of hearing aids," in *Proc. of Int. Congress on Acoust.* (Seattle, Washington), 1998, pp. 2105-2108.

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

[4] B. C. J. Moore, *An Introduction to Psychology of Hearing.* London: Academic, 1997.

[5] T. Lunner, S. Arlinger, and J. Hellgren, "8-channel digital filter bank for hearing aid use: Preliminary results in monaural, diotic and dichotic modes," *Scand. Audiol.*, Suppl. 38, pp. 75-81, 1993.

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

[7] D. S. Jangamashetti and P. C. Pandey, "Inter-aural switching with different fading functions for dichotic presentation to reduce the effect of temporal masking in sensorineural hearing loss," in *Proc. 4th World Multi Conf. on Systemics, Cybernatics, and Informatics* (Orlando, Florida), 2000, pp.434-439.

[8] P. C. Pandey, D. S. Jangamashetti, and A. N. Cheeran, "Binaural dichotic presentation to reduce the effect of increased temporal and spectral masking in sensorineural hearing impairment," *J. Acoust. Soc. Am.*, vol. 110, pp. 2705, 2001.

[9] E. Zwicker, “Subdivision of the audible range into critical bands (Frequenzgruppen),” *J. Acoust. Soc. Am.*, vol. 33, pp. 248, 1961.

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