BINAURAL DICHOTIC PRESENTATION TO REDUCE THE EFFECTS OF TEMPORAL AND SPECTRAL MASKING DUE TO SENSORINEURAL HEARING LOSS

DAKSHAYANI. S. JANGAMASHETTI

DEPARTMENT OF ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY, BOMBAY
POWAI, MUMBAI 400 076, INDIA

2003

BINAURAL DICHOTIC PRESENTATION TO REDUCE THE EFFECTS OF TEMPORAL AND SPECTRAL MASKING DUE TO SENSORINEURAL HEARING LOSS

submitted in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy

by

Dakshayani S. Jangamashetti
(Roll No. 97407302)

Supervisor: Prof. P. C. Pandey

Department of Electrical Engineering
Indian Institute of Technology, Bombay

2003

Dedicated to

My

teachers, parents, in-laws,

husband Suresh, and daughter Akhila

Thesis Approval

Thesis entitled: ***Binaural dichotic presentation to reduce the effects of temporal and spectral masking due to sensorineural hearing loss***

by ***D. S. Jangamashetti***

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

Internal Examiner : ______________________ (Prof. S. D. Agashe)

External Examiner : ______________________ (Prof. S. Sanyal)

Supervisor : ______________________ (Prof. P. C. Pandey)

Chairman : ______________________ (Prof. G.G.Ray)

Date: March 10, 2003

INDIAN INSTITUTE OF TECHNOLOGY, BOMBAY, INDIA

CERTIFICATE OF COURSE WORK

This is to certify that Mrs. D. S. Jangamashetti was admitted to the candidacy of the Ph.D. Degree on 25-07-1997, after successfully completing all the courses required for the Ph.D. Degree programme. The details of course work done are given below.

Sr. No.	Course No.	Course Name	Credits
1	EE603	Digital signal processing and its applications	6
2	BM603	Physiology for engineers	6
3	EE698	Special topics in electrical engineering: Speech processing	6
4	EES801	Seminar	4
		Total Credits	22

I.I.T. Bombay

Dy. Registrar (Academic)

Dated:

D. S. Jangamashetti. “Binaural dichotic presentation to reduce the effects of temporal and spectral masking due to sensorineural hearing loss”, *Ph.D. Thesis*, Department of Electrical Engineering, IIT Bombay, 2002, Supervisor: Prof. P. C. Pandey.

Abstract

Sensorineural hearing loss is characterized by reduced frequency and temporal resolution, and increased spectral and temporal masking. Consonantal place and duration features, cued by temporal and spectral properties, are not well perceived by persons with sensorineural loss. Earlier investigations have shown that splitting the signal into frequency bands and presenting alternate bands to the two ears helps in improving the place perception by reducing the effect of increased spectral masking. Splitting the speech signal temporally into segments and presenting the adjacent segments to the two ears is likely to reduce the effect of increased temporal masking and improve the duration perception. Combining the scheme of temporal splitting with spectral splitting for binaural dichotic presentation may help in improving the perception of various consonantal features by decreasing the effects of temporal and spectral masking.

The present research deals with implementation and evaluation of the schemes of temporal splitting and combined splitting for binaural dichotic presentation for improving speech reception by persons using binaural hearing aids. The scheme of temporal splitting uses step and trapezoidal fading functions along with overlap. The scheme of combined splitting uses a pair of time-varying comb filters with pass bands corresponding to auditory critical bands. Experimental evaluation was carried out through listening tests involving identification of twelve English consonants in vowel-consonant-vowel context with vowel /*a*/. Evaluation was conducted in two phases: first for different processing conditions on normal hearing subjects with simulated loss and later for selected processing conditions on subjects having bilateral sensorineural loss.

Temporal splitting with inter-aural switching period of 20 ms and step transitions showed an improvement in speech quality, response time, recognition score and relative information transmission of consonantal features, particularly for duration and place, with highest improvements for duty cycle of 70%. Trapezoidal fading function was employed for reducing the spectral distortions and thereby further improving speech reception. Best results were obtained for 2 ms transition. For combined splitting, each time-varying comb filter is realized as a set of cyclically selected comb filters such that a cyclic sweeping of magnitude responses occurs in 20 ms. Comb filters are 256-coefficient linear phase FIR filter, with magnitude responses optimized for minimizing perceived spectral distortion by having low pass band ripple, high stop band attenuation, and perceptual balance at inter-band crossovers. Listening tests showed improvement in response time, recognition scores, and transmission of features particularly duration and place, indicating reduction in the effects of temporal and spectral masking. Improvements were highest for 4 and 8 filter sets.

An overall evaluation of the schemes of temporal, spectral, and combined splitting was done by conducting listening tests on five persons with moderate to severe bilateral sensorineural hearing loss by selecting the processing parameters that resulted in maximum improvements in earlier tests. The processing schemes have shown improvement in response time, recognition scores, and information transmission for consonantal features particularly for duration and place features, the extent of improvements being related to the hearing loss configuration.

Contents

Appendices

List of figures

List of tables

List of symbols and abbreviations

Symbols

d	duty cycle
$e(n)$	broad band noise
F1	first formant
F2	second formant
$I(x)$	information measure of stimulus x, in bits
$I(y)$	information measure of response y, in bits
$I(x;y)$	information transmitted from stimulus x to response y in bits
$I_{rel}\,(x;y)$	information transmitted relative to stimulus information
L	"on" period
m	number of shiftings
M	transition duration
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 stimulus response pair (x_i,y_j)
R_s	recognition or articulation score
$s(t)$	analog signal
$s(n)$	digital signal
T_p	inter-aural switching period in ms
T_{on}	"on" interval in ms
T_s	switching interval in ms
$w(n)$	fading function
β	scaling factor
σ_s	short-time energy of the signal
σ_e	short-time energy of broad band noise

Abbreviations

ADC	analog-to-digital converter
AGC	automatic gain control
AI	articulation Index
AVC	automatic volume control
Avg.	average
B&K	Brüel and Kjær
BTE	behind the ear
CB	critical band
CF	characteristic frequency
CM	cochlear microphonics
CS	combined spectral and temporal splitting
C/V	consonant-to-vowel
CV	consonant-vowel
CVR	consonant-to-vowel intensity ratio
DAC	digital-to-analog converter
dB	decibel
df	degree of freedom
DL	difference limen
DSP	digital signal processing
ERB	equivalent rectangular bandwidth
FFT	fast Fourier transform
FIR	finite impulse response
FTD	Formant transition duration
HL	hearing level
HP	Hewlett-Packard
HTL	hearing threshold level
Hz	Hertz
ITC	in the canal
ITE	in the ear
MLP	mean logarithmic probability
PC	personal computer
PCL	PC-LabCard
PTA	pure tone threshold average
PTC	psychophysical tuning curves
R.D.	relative decrease
R.I.	relative improvement

RS	recognition score
RT	response time
S	subject
s.d.	standard deviation
SINFA	sequential information transmission analysis
SL	sensational level
SNR	signal-to-noise ratio
S_p	processed speech
SPL	sound pressure level
SR	sampling rate
SRT	speech reception threshold
sSRT	sentence speech reception threshold
SS	spectral splitting
S_u	unprocessed speech
TI	Texas Instruments
TS	temporal splitting
TS_ST	temporal splitting with step transition
TS_TR	temporal splitting with trapezoidal transition
ULL	uncomfortable listening level
VCV	vowel-consonant-vowel
VOT	voice-onset-time
VC	vowel-consonant

Chapter 1

Introduction

1.1 Overview

Sensorineural hearing loss, which occurs due to damage to hair cells in the cochlea and degeneration of auditory nerve fibers, is characterized by elevated thresholds, loudness recruitment, reduced frequency and temporal resolution, and increased spectral and temporal masking. To address the problem of elevated threshold and reduced dynamic range due to loudness recruitment, hearing aids employ frequency selective amplification and compression techniques respectively.

Masking is a phenomenon where presence of one signal component results in in-audibility of the neighboring signal component. Increased spectral masking associated with broad auditory filters results in smearing of spectral peaks and valleys, and leads to difficulty in perception of consonantal place feature. Increased forward and backward temporal masking of weak acoustic segments by strong ones causes reduction in discrimination of voice-onset-time, formant transition, and burst duration that are required for consonant identification. Thus the overall effect of two types of masking is a difficulty in discrimination of consonants, resulting in relatively degraded speech perception by persons with sensorineural loss.

Presentation of different signals to the two ears is known as dichotic binaural presentation. Masking takes place primarily at the peripheral auditory system. In speech perception, the information received from both the ears gets integrated. Hence splitting of speech signal into two complementary signals such that signal components likely to mask each other get presented to the different ears can be used for reducing the effect of increased masking. This technique can be used for improving speech reception by persons with moderate bilateral sensorineural loss, i.e. residual hearing in both the ears.

For reducing the effect of increased spectral masking, the speech processing schemes that are investigated are based on spectral splitting of speech signal into bands for dichotic binaural presentation (Chaudhari and Pandey, 1998a,b; Lunner *at al*., 1993). The scheme developed by Chaudhari and Pandey (1998a,b) employs, two comb filters having complementary magnitude responses, each with 9 pass bands corresponding to critical bandwidths of auditory filters. The scheme has shown improvement in place feature perception, in persons with bilateral sensorineural hearing impairment. In this scheme, sensory cells corresponding to alternate filter bands of the basilar membrane are always stimulated, whereas sensory cells in the intervening bands are not stimulated. Provision of relaxation time between the segments of speech, thereby providing relaxation for the sensory cells, may help in reducing the effect of increased temporal masking. In this research, a method of temporal splitting is investigated to split the speech signal into two for binaural dichotic presentation, with adjacent segments presented to the two ears, in such a way that all the sensory cells of a ear go through cycles of stimulation and relaxation. Combining the scheme of temporal splitting with spectral splitting for binaural dichotic presentation may help in improving the perception of various consonantal features by decreasing the effects of temporal and spectral masking. A method for splitting of speech spectrally and temporally for binaural dichotic presentation with different frequency bands swept between the two ears, such that all the sensory cells of the basilar membrane get periodic relaxation from stimulation is proposed and investigated.

1.2 Research objective

Objective of the research was to investigate speech processing schemes of temporal splitting and combined spectral and temporal splitting for binaural dichotic presentation in order to improve speech reception by persons with moderate bilateral sensorineural hearing loss, and to study the effect of processing parameters.

Speech processing schemes have been developed and implemented. Experimental investigations are carried out for studying their effectiveness in improving speech reception.

1.3 Thesis outline

Chapter 2 describes the auditory system, types of hearing impairment, effects of sensorineural hearing loss on speech perception, hearing aids, and speech processing schemes for improving the speech reception. In Chapter 3, some of speech processing schemes that use binaural dichotic presentation are reviewed, followed by description of the scheme proposed and evaluation methods. In Chapter 4, implementation and evaluation of the schemes of temporal splitting with two fading functions are described. In Chapter 5, implementation and evaluation of the scheme of combined spectral and temporal splitting are given, followed by discussion of results. Chapter 6 discuses the experimental evaluation of the schemes of temporal and combined splitting along with the scheme of spectral splitting with persons having sensorineural hearing impairment, using optimal parameters for processing that gave maximum improvements in experiments with subjects with simulated sensorineural hearing loss. Summary of the work done, conclusions drawn from the present research, and suggestions for further work are given in Chapter 7.

Appendices deal with supplementary information related to experimental set-up and procedures. Spectrographic analysis set-up is described in Appendix A. Performance evaluation method is given in Appendix B. Appendix C deals with design of comb filters for spectral splitting. Programs for listening tests and the hardware used are described in Appendices D and E respectively. Test instructions are given in Appendix F. Simulation of sensorineural hearing loss is explained in Appendix G.

Chapter 2

Auditory system and hearing impairment

2.1 Introduction

This chapter describes the physiology and function of the auditory system, hearing impairment, and perceptual consequences of sensorineural hearing loss on speech perception. Speech processing schemes for the hearing aids are also reviewed.

2.2 Peripheral auditory system

Outer ear, middle ear, and the inner ear are the three main parts of the peripheral auditory system as shown in Fig 2.1.

2.2.1 The outer ear and middle ear

The pinna (auricle) and the auditory meatus (auditory canal) together form the outer ear. The pinna apart from protecting the ear from foreign bodies, modifies the incoming signal at high frequencies and thus helps in sound localization. The meatus introduces an increase in sound pressure of about 10–20 dB over the frequency range from 2 kHz to 7 kHz (Flanagan, 1972; Pickles, 1982; Wall, 1995). The tympanic membrane separates the outer ear from the middle ear.

The middle ear, an air filled cavity, is composed of three small bones known as ossicles. These are the malleus (hammer), the incus (anvil) and the stapes (stirrup). Malleus is attached to the tympanic membrane at one end and to the incus at the other end. Incus makes contact with the stapes, the footplates of which are attached to the membrane of oval window (starting point of the inner ear). The middle ear provides impedance matching between the air in the meatus and the cochlear fluid. This is contributed mainly, due to the large difference in area of the tympanic membrane and the oval window, and partly by the lever action of the ossicles. The impedance matching provides a pressure gain of about 30 dB in the middle ear (Pickles, 1982; Wall, 1995). The pressure gain versus frequency exhibits bandpass characteristic, with a nearly flat top in the frequency range of 500 Hz to 4 kHz, with maximum being at about 1 kHz (Moore, 1997; Pickles, 1982). For intense sounds, the muscles connecting the ossicles contract and prevent the transformer action to take place. Hence most of the sound is reflected back. This is known as acoustic reflex, and it plays an important role in protecting the inner ear from excessive intense sounds. However, due to latency in the acoustic reflex, transient signals are likely to reach the inner ear and cause damage.

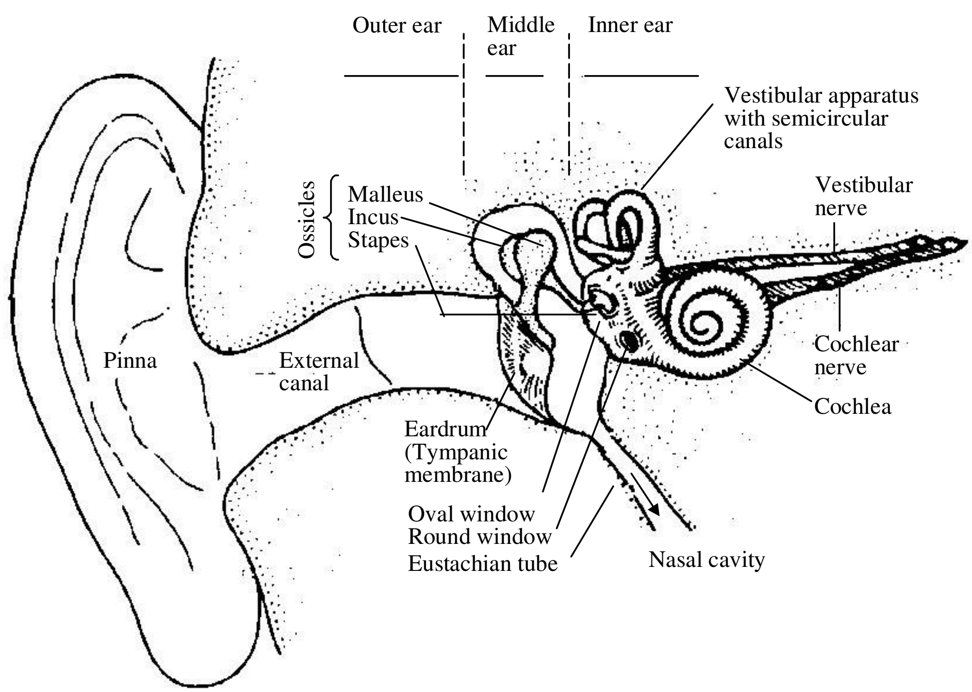


FIG. 2.1. Peripheral part of the auditory system. Adapted from: Flanagan (1972), Fig. 4.1.

2.2.2 The inner ear

The cochlea in the inner ear is a fluid filled bony structure of length about 3–3.5 cm, which is coiled like a spiral into about two and one-half turns (Babsky, 1970; Ganong, 1991; Gelfand, 1990; Levitt *et al.*, 1980; Møller, 2000; Moore, 1997; Previte, 1983). Figure 2.2 shows the longitudinal cross-section of the cochlea. The end at which the cochlea begins is known as the basal end and its end point is known as the apical end. The cochlea is divided along its length into three chambers namely: scala vestibuli, scala media, and scala tympani. Scala vestibuli and scala media are separated by a thin membrane called Reissner's membrane. Scala media is separated from scala tympani by basilar membrane. The two outer chambers, the scala vestibuli and scala tympani, contain perilymph (similar in composition to extracellular fluid) and open at the apex through a small opening called the helicotrema. The center chamber, scala media, is filled with endolymph (similar in composition to intracellular fluid). The basilar membrane is stiff and narrow at the basal end, and less stiff and broad at the apex. Pressure variations at the tympanic membrane are transmitted to the oval window, resulting in cochlear fluid movement which in turn cause upward and downward movement of the basilar membrane. The point at which maximum vibration occurs depends on the mechanical properties of the basilar membrane and the frequency of the input signal. The high frequency signals produce maximum deflection at the base while low frequency signals produce maximum deflection at the apex (Guyton, 1986; Kandel *et al.*, 1991; Moore, 1997; Pickles, 1982). Figure 2.3 shows the vibration of the basilar membrane for low, medium and high frequencies.

The organ of Corti, which contains sensory hair cells, is situated on the basilar membrane. Figures 2.4 and 2.5 show the cross section of cochlea and the cross section of organ of Corti respectively. There are about 3,500 inner hair cells arranged in a single row near the modiolus, and there are 20,000 to 25,000 outer hair cells arranged in 3 to 5 rows (Babsky, 1970; Gulick, 1971; Moore, 1997; Pickles, 1982; Wall, 1995;). Each hair cell has hairs called stereocilia on their top. On each outer hair cell there are nearly 100–200 hairs arranged in 3 to 4 rows forming V or W shape and are embedded in the tectorial membrane, which lies above the Organ of Corti. Each inner hair cell has about 40–60 hairs arranged in three rows (Brobeck, 1973; Moore, 1997; Pickles, 1982; Wall, 1995). The hairs on the inner hair cells are not embedded in tectorial membrane, but are loosely attached (Moore, 1997;

Pickles, 1982). The hair cells contact at their base, directly to the axons of neurons whose cell bodies lie in spiral ganglion (Kandel *et al.*, 1991; Pickles, 1982). The upward and downward movement of the basilar membrane cause inward-upward and backward-downward motion of the tectorial membrane. This relative motion results in displacement of the stereocilia, which in turn produces an electrical potential (depolarizing-hyperpolarizing) called the cochlear microphonics (CM) across the membrane of the hair cell at its apical end (Pickles, 1982). The depolarization potentials are accompanied by release of neurotransmitter, which excite the neuronal axons that synapse the base of the hair cells, which in turn initiate neural spikes (action potentials) and these are carried to the higher auditory levels (Kandel *et al.*, 1991; Pickles, 1982;).

There are about 30,000 sensory neurons in man that convey auditory information from cochlea to the central nervous system (brain stem) through the auditory nerve fibers (Brobeck, 1973; Flanagan, 1972; Kandel *et al.*, 1991; Pickles, 1982; Previte, 1983). About 90–95% of the afferent nerve fibers are innervated to the inner hair cells, each inner hair cell being connected to about 20 nerve fibers. Remaining 5–10% of the afferent nerve fibers make contact with the outer hair cells (Kandel *et al.*, 1991; Pickles, 1982). Inner hair cells play a role in detection of the signal and excitation of the afferent fibers. Outer hair cells enhance the basilar membrane responses to low-level sounds thus increasing the sensitivity, and they sharpen the tuning (frequency selectivity) of the basilar membrane (Moore, 1998; Pickles, 1982; Wall, 1995). Role of hair cells in the transduction mechanism is discussed later in subsection 2.4.4

In the auditory nerve each neuron shows a random and spontaneous firing in the absence of input signal, and this rate increases with input signal. Considering the sound intensity needed to raise the firing rate just above the spontaneous firing rate as threshold, the threshold versus frequency relationship for each neuron is a V-shaped tuning curve. The frequency for which the threshold is lowest is known as characteristic frequency (CF) of the neuron. The tuning of the nerve fiber corresponds to the tuning of the hair cell that innervates the nerve fiber. The characteristic frequency of the neurons varies with the site of hair cell innervation: from high at the basal end to low at the apical end. Further, tuning curves at the high frequency side are steeper than at the low frequency side (Moore, 1997; Kandel *et al.*, 1991).

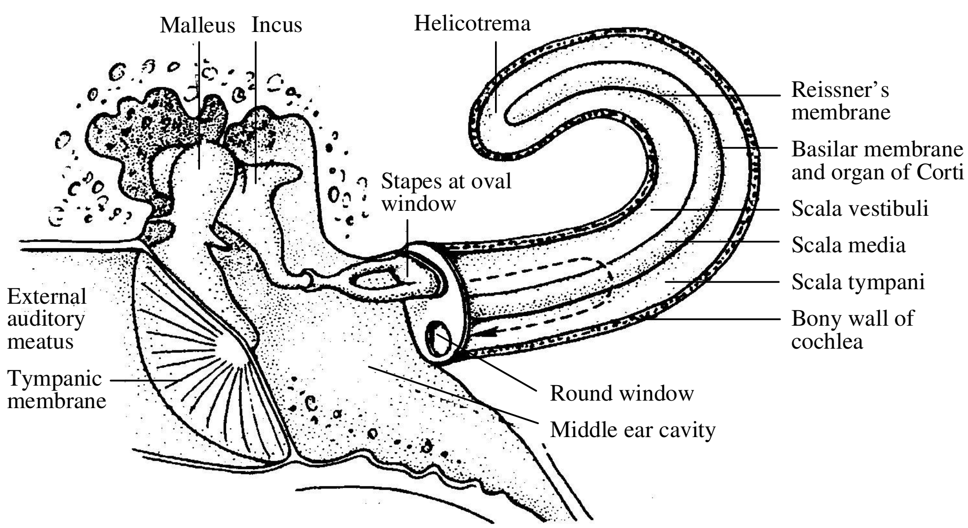


FIG. 2.2. Longitudinal cross-section of the cochlea. Adapted from Kandel (1991), Fig.32-2.

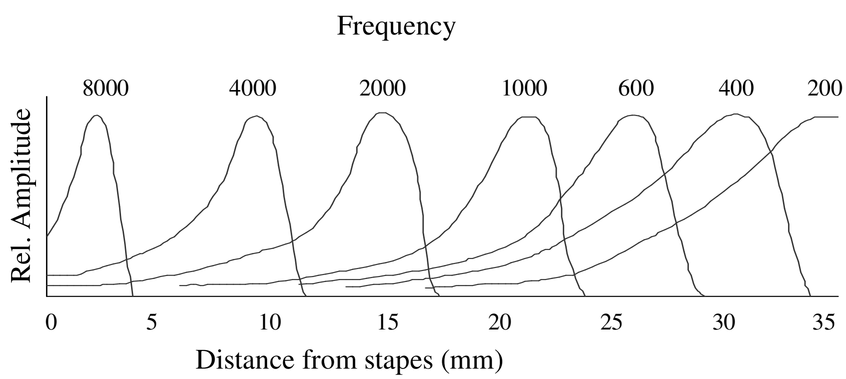


FIG. 2.3. Amplitude patterns of vibration of basilar membrane for different frequencies. Adapted from: Guyton (1986), Fig. 61-6.

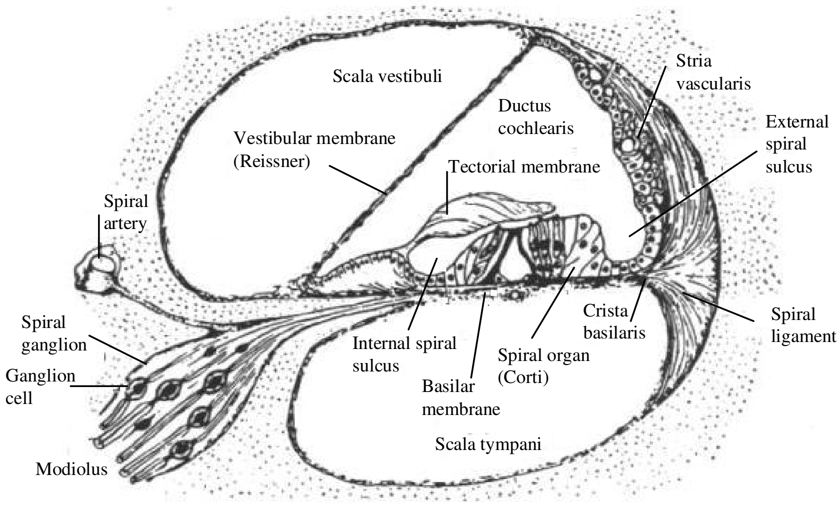


FIG. 2.4. Cross section of cochlea. Adapted from Zemlin (1998), Fig. 6-81.

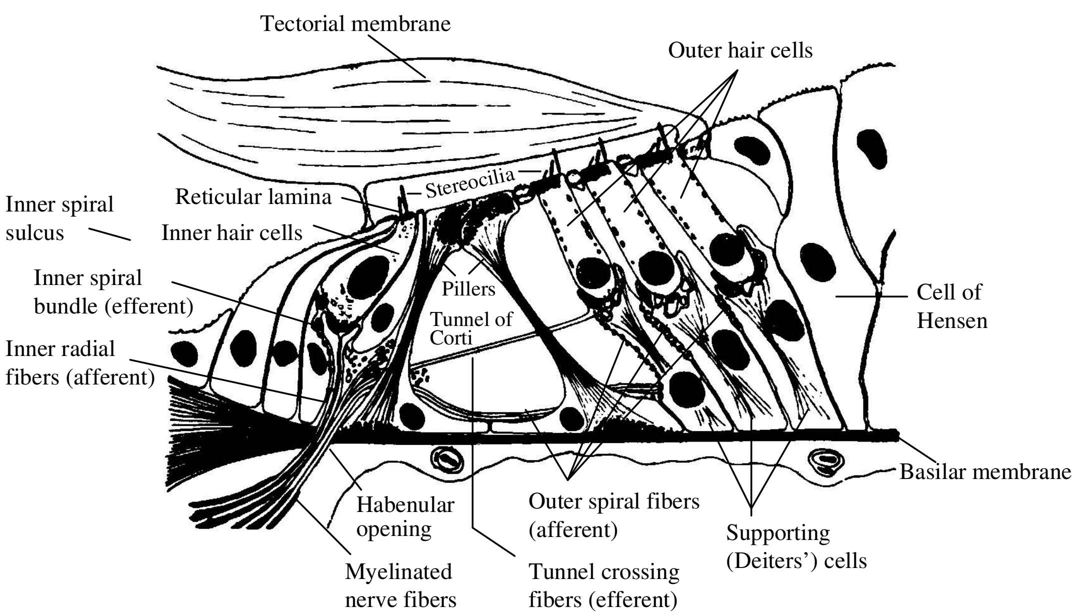


FIG. 2.5. Cross-section of organ of Corti. Adapted from Wall (1995), Fig. 1.15.

Auditory nerve fibers carry the information from the hair cells to the cochlear nucleus in the brain stem. Each auditory nerve fiber on entering the nucleus, branches to ventral and dorsal cochlear nucleus. The fibers are arranged in an orderly manner, reflecting an orderly arrangement of characteristic frequencies of the neuron they innervate, leading to tonotopic frequency mapping. The tonotopic frequency mapping is maintained at higher auditory processes to some extent. The output axons of the dorsal nucleus bypass the superior olivary complex (next nucleus in the auditory pathways), and end in the nuclei of the lateral lemniscus and the inferior colliculus mainly on the opposite side. The neurons of the ventral nucleus end in the superior olivary complex of both sides and to a lesser extent in the nuclei of the lateral lemniscus. Inferior colliculus is the main receiving part for the ascending pathways from the superior olivary complex. Inferior colliculus receives afferents bilaterally from the superior olivary complex and contralaterally from the dorsal cochlear nucleus. The inferior colliculus sends the information to the auditory cortex through the medial geniculate. Auditory paths cross at or near the level of medulla and at inferior colliculus (Moore, 1997; Pickles, 1982).

2.3 Hearing impairment

Hearing impairment is broadly categorized as conductive loss, sensorineural loss, central impairment, and functional impairment depending upon the location of damage in the auditory system (Levitt *et al*., 1980; Moore, 1997; Pickles, 1982).

Conductive loss occurs due to damage in the outer ear, ear drum or middle ear, obstructing the transmission of the sound to the inner ear. Conductive loss produces attenuation of the stimulus and results in increase in both the hearing threshold levels and uncomfortable loudness levels. The causes include: fixation of stapes (due to disease otosclerosis) due to growth of bone on the oval window, perforation of the tympanic membrane, and fluid, wax, or pus formation in the ear canal and middle ear (otitis media). Conductive loss can be treated medically or surgically in most cases.

Sensorineural hearing loss occurs due to damage to the transduction mechanism of the inner ear. The loss due to damage in the cochlea is known as cochlear (sensory) loss and that occurring due to abnormality in the auditory nerve is retrocochlear (neural) hearing loss. The cochlear hearing loss is caused by exposure to intense sound, drugs, Meniere's syndrome (endolymphatic hydrops which results in low frequency hearing loss), infections, or congenital defects leading to loss of cochlear hair cells. Exposure to noise cause alterations in the stereocilia, destruction of the supporting structure, and loss of hair cells and thereby degeneration of the neurons leading to loss of sensitivity at high frequencies. Toxic materials cause loss of outer hair cells at the basal end leading to loss of sensitivity at high frequency. The diseases like mumps, measles, meningitis, and influenza also result in loss of hair cells and neurons. Retrocochlear hearing loss is associated with damage to auditory nerve fibers or neurons, caused by tumors in the eighth cranial nerve and internal auditory canal, multiple sclerosis, or hemorrhage.

Central hearing impairment is associated with reduced ability for speech comprehension and is caused by damage to the brain cortex due to cerebral hemorrhage, meningitis, skull trauma, or congenital defects. Subjects with central hearing impairment alone, show normal peripheral hearing (Wall, 1995). The fourth type of hearing impairment, the functional impairment is related to the psychological factors and cognitive processes.

Age related auditory dysfunction is commonly termed as presbycusis, which include degenerative changes in the auditory system: stiffening of the basilar membrane, devascularization (destruction of blood vessels) of the cochlea leading to loss of hair cells in the organ of Corti and degeneration of neurons in the entire auditory system, weakening of the stria vascularis (blood vessels connecting the nerve fibers) and spiral ganglion, and abnormalities in the brain stem. Loss is bilaterally symmetrical, with greater hearing loss at high frequencies, and increased difficulty in understanding speech. The condition worsens with increasing age. Aging is also associated with decreased cognitive processing which leads to decreased ability for speech comprehension (Levitt *et al.*, 1980; CHABA, 1988).

Other important hearing disorders associated with sensorineural hearing loss are tinnitus and diplacusis. Tinnitus often referred to as "ringing" in the ear, is the perception of ringing, hissing, buzzing, or other sounds in the ears in the absence of any external stimulus (Janssen *et al.*, 2000; Pickles, 1982; Wall, 1995). It may be caused by an amplifying and mechanically active physiological process. Activation of the hair cells lead to feed back of some mechanical energy into the basilar membrane. The feed back is believed to result from fluid flows across cell membranes associated with ion flows, or from active motility of the stereocilia due to actin-myosin interactions. There is back and forth movement of the mechanical energy along the cochlea, further stimulating the hair cells, leading to self-sustained oscillation (Pickles, 1982). However Tinnitus can arise with a large variety of causes, and also in cases with total deafness where the cell population is completely absent. Tinnitus is often associated with the disease Meniere's syndrome (Wall, 1995). Other causes include, exposure to loud sound and effect of toxic drugs (Janssen *et al.*, 2000), leading to spontaneous discharge of hair cells and neurons. Tinnitus is also caused by otosclerosis, ear infection, tumors, meningitis, and head and neck injury. Partial obstruction of blood vessels and sudden contraction of inter-aural muscles may also result in tinnitus (Gulick, 1971). Tinnitus, if severe, affects speech perception significantly.

Diplacusis is a type of hearing disorder in which a pure tone of one pitch presented to both ears is heard with two different pitches. This disorder is caused by exposure to intense noise, fatigue, or mild injury to the organ of Corti, Meniere's disease (Pickles, 1982; Van den Brink, 1970; Ward, 1963). According to Van den Brink (1970), origin of diplacusis can be in higher centers (auditory nerve, cochlear nucleus, and inferior collicilus). Diplacusis results in reduced speech intelligibility.

In acoustic measurements, sound level is often given in dB, taking sound pressure of 20 μ Pa as the reference level and the sound level measured with respect to this reference level is known as sound pressure level (SPL). Sound level in dB with reference to the threshold level of the listener is referred as sensational level (SL). In clinical practice, the sound level of pure tones is given in dB by taking the average hearing threshold of normal hearing young adults as the reference, and this is known as hearing level (HL). The average hearing threshold is frequency dependent, and hence SPL corresponding to a given HL varies with frequency. Hearing impairment is clinically assessed using audiometry, which involves measurement of hearing threshold level (HTL) and uncomfortable listening level (ULL) for pure tone stimuli, speech reception threshold, speech discrimination score at the most comfortable listening level, etc. Thresholds as a function of tone frequency are plotted as audiogram. These and other test results are used for assessing the level of hearing impairment as well as for diagnosing the cause of hearing impairment (Biswas, 1995; Coleridge Smith and Scurr, 1988; Moore, 1997; Severns, 1985).

2.4 Sensorineural hearing loss

Sensorineural hearing loss is caused by exposure to intense sound, toxic drugs, Meniere's syndrome, infections, congenital defects leading to loss of cochlear hair cells, damage to auditory neurons caused by tumors, multiple sclerosis, or hemorrhage. The audiograms of the sensorineural loss show typical shapes depending on the pathology. A flat audiogram represents weakening and damaging of the nerve fibers due to salicylate poisoning (Guyton, 1986; Wall, 1995). Stiffening of the basilar membrane and damage to the organ of Corti (loss of hair cells and supporting structures due to toxic drugs) result in high frequency hearing loss. Noise trauma is associated with a notch around 4 kHz in the audiogram (Pickles, 1982; Wall, 1995). Meniere's disease results in elevated thresholds at low frequencies (Wall, 1995). Congenital sensorineural damage, produce a trough shaped audiogram. Diseases like typhoid, meningitis, mumps, etc. cause damage to the inner ear and auditory nerves giving rise to typical audiogram patterns. Typhoid shows a moderate to severe bilateral sensorineural hearing loss, meningitis is associated with bilateral and profound deafness, and mumps shows usually a unilateral very severe or profound deafness (Biswas, 1995). Apart from the elevated thresholds, sensorineural hearing loss is characterized by loudness recruitment, reduced frequency selectivity and temporal resolution, and increased spectral and temporal masking. These characteristics are explained in the following subsections.

2.4.1 Loudness recruitment

Loudness recruitment is an unusually rapid growth of perceived loudness as the sound intensity is increased. Loudness recruitment results due to damage to the outer hair cells. Outer hair cells increase the sensitivity of the transduction mechanism for low-input sound levels, leaving the response to high-level sounds unaffected. This is known as compressive nonlinearity feature of the basilar membrane. Loss of this active process of the cochlea leads to elevated thresholds, without affecting the response to high-level sounds, leading to loudness recruitment (Moore, 1997; Moore and Glasberg, 1993; Oxenham and Plack, 1997; Pickles, 1982). Loudness recruitment results in reduced dynamic hearing range (range between hearing threshold level and uncomfortable loudness level) and hence loss of speech intelligibility. For normal hearing, dynamic range is about 85–95 dB. For moderate sensorineural loss, the dynamic range is between 50–60 dB. In extreme cases this range may be as small as 10–20 dB (Stone and Moore, 1992).

2.4.2 Frequency selectivity and spectral masking

Frequency selectivity of the auditory system is described by two properties: frequency discrimination and frequency resolution. Ability of the auditory system to discriminate, successively presented tones that differ in their frequency content is known as frequency discrimination. The frequency difference limen (smallest difference) is about 0.2% or 0.3% of the tone (Flanagan, 1972; Pickles, 1982). Frequency resolution refers to the ability of the auditory system to resolve among simultaneously presented different frequency components in a complex signal. Masking is a phenomenon in which threshold of audibility of one signal component is raised by the presence of another component (Dolan and Small, 1984; Moore, 1997; O' Shaughnessy, 2001; Tanner, 1958;). A signal is most likely to get masked by another signal with frequency components close to, or the same as, that of the signal. The masking effect is used to explain and quantify frequency selectivity. It is found that the frequency selectivity is reduced in persons with sensorineural hearing loss (Carney and Nelson, 1983; Moore, 1997; Tyler *et al.*, 1983; Zwicker and Schorn, 1978).

Peripheral auditory system can be modeled as a bank of band pass filters, called auditory filters with overlapping pass bands. Each location on the basilar membrane behaves like a filter with different center frequency (Moore, 1997). Fletcher (1953) determined the shape of the auditory filter by measuring the threshold of a tone as function of the bandwidth of a band pass noise masker centered at the tone frequency. The threshold of the signal increases at first with an increase in the noise bandwidth, but then remains constant irrespective of increase in noise bandwidth. Fletcher (1953) called this bandwidth at which the signal threshold ceased to increase, as critical band (Moore, 1997). Critical bandwidths (CB) are 15–20% of the center frequency above 1 kHz and nearly constant below 500 Hz (Moore, 1982; Pickles, 1982; Zwicker, 1961). The critical band concept is useful to explain the frequency selectivity of the auditory system.

To derive the shape of the auditory filter, Patterson (1976) used a sinusoidal tone as signal and noise notched at the signal frequency as masker. Threshold of the signal as a function of notch width is measured. This is plotted to give the relative response of the auditory filter centered at the tone frequency. The 3 dB bandwidths of the auditory filter are found to be between 10% and 15% of the center frequency. Bandwidth of a rectangular filter, which has a transmission in its pass band equal to the maximum transmission of the specified filter, is called as equivalent rectangular bandwidth (ERB) of the filter. The ERB's of the auditory filter obtained by Patterson (1976) are between 11% and 17% of the center frequency. For frequencies above 1 kHz, these values are very close to the values of the CBs reported earlier by Zwicker (1961). However, at low frequencies, the values are lower. Figure 2.6 shows the CBs of the auditory filters reported by Zwicker (1961) and ERBs of the auditory filters reported by Glasberg and Moore (1990). The latter show a continuous decrease as the center frequency decreases below 500 Hz.

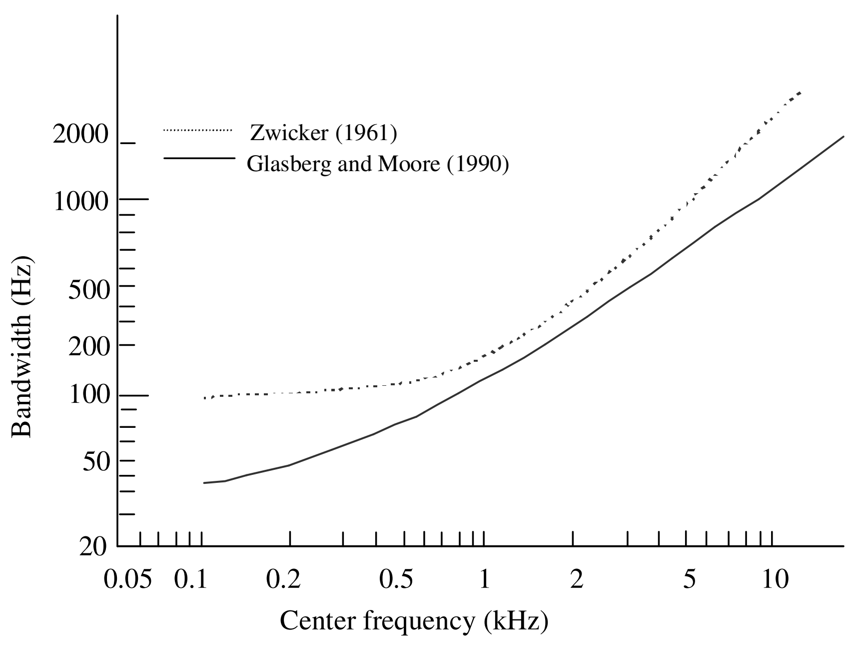


FIG. 2.6. Bandwidth as a function of center frequency of the auditory filters. Curve with dotted line shows critical bandwidth estimates obtained by Zwicker (1961), solid curve shows the estimates of equivalent rectangular bandwidth obtained by Glasberg and Moore (1990). Source: Moore (1997), Fig. 3.10.

Wider auditory filters are observed in listeners with sensorineural hearing loss (Carney and Nelson, 1983; Glasberg and Moore, 1986) and the slope of the skirts decrease with increasing age (Glasberg *et al*., 1984). Broad auditory filters result in increased upward spread of masking (masking of low energy-high frequency components by high energy-low frequency components) in sensorineural impairment (Florentine *et al.*, 1980). Dubno and Schaefer (1992), Hanley and Dorman (1983), and Leek and Summers (1996) have reported that, frequency selectivity for listeners with cochlear hearing loss is poorer than normal leading to reduced speech perception.

2.4.3 Temporal resolution and temporal masking

Ability to follow variations in time pattern of sounds is known as temporal resolution. It is given by the minimum duration of a silence interval between two stimuli that can be detected (Fitzgibbons and Wightman, 1982; Moore, 1997; O'Shaughnessy, 1987; Strouse *et al.*, 1998; Tyler *et al.*, 1982). Changes in time pattern of a signal are usually associated with changes in magnitude spectra, which pose difficulty in measuring true value of temporal resolution. Broadband white noise has the property that, its long-term spectrum does not change with the introduction of a gap, hence it has been often used for measuring temporal resolution. Various researchers have measured the temporal resolution (gap threshold) using different types of test stimuli at various presentation levels and using different methods for normal and sensorineural hearing impaired subjects. They have reported gap threshold in the range of 2.3–7.4 ms for normal hearing persons. The gap thresholds for impaired persons are 50–100% higher than those for normal hearing subjects (Abel, 1971; Fitzgibbons and Wightman, 1982; Florentine and Buus; 1984; Plomp, 1964; Shailer and Moore, 1985; Smiarowski and Carhart, 1975; Snell, 1997; Strouse *et al.*, 1998; Tyler *et al.*, 1982)

Florentine and Buus (1984), and Shailer and Moore (1983) have shown that, persons with poor high-frequency hearing have larger gap thresholds than persons with good high-frequency hearing, indicating that loss at high frequency is associated with reduced temporal resolution. Temporal resolution improves systematically with an increase in octave band center frequency (Fitzgibbons and Wightman, 1982). Larsby and Arlinger (1998) have reported larger gap thresholds (poor temporal resolution) in elderly normal hearing subjects. Moore *et al.* (1992), Schneider *et al.* (1994), Snell (1997), and Strouse *et al.* (1998) have also reported degraded temporal resolution in elderly persons.

Other temporal characteristics, which are associated with hearing are temporal integration, temporal difference limen, and gap difference limen. Temporal integration is an estimation of the ability to integrate energy over time. Temporal difference limen refers to an increment in duration necessary to detect a difference in the duration of a noise burst. Gap difference limen is an increment in duration necessary to detect a difference in the duration of a silent interval between two noise bursts (Tyler *et al.*, 1982). In persons with sensorineural loss, these characteristics are poorer than normal (Dorman *et al.*, 1985; Fitzgibbons and Wightman, 1982; Glasberg *et al.*, 1987; Tyler *et al.*, 1982). These degraded characteristics and reduced temporal resolution are associated with increased forward and backward masking of weak signal components by adjacent strong ones. In forward masking, signal follows the masker whereas, in backward masking signal precedes masker. Both forward and backward masking are caused by temporal overlap of cochlear responses (O' Shaughnessy, 1987).

Forward masking occurs due to reduction in sensitivity of recently stimulated cells or persistence of neural activity produced by the masker (Fitzgibbons and Wightman, 1982; Jesteadt *et al.*, 1982; Plack and Oxenham, 1998; Smiarowski and Carhart, 1975). It is believed that forward masking is related to the adaptation and fatigueness of the auditory system (Dolan and Small 1984; Duan and Canlon, 1996; Moore, 1997; O'Shaughnessy, 1987). Backward masking occurs due to a blocking phenomenon, in which the auditory processing of a signal is interrupted by a following intense noise (masker) if the noise arrives before the perception of the signal is completed (O' Shaughnessy, 1987). 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. Backward masking is effective up to 20 ms (O'Shaughnessy, 1987; Smiarowski and Carhart, 1975; Wilson and Carhart, 1971). Kidd *et al.* (1984) have reported that forward masking increases as a function of masker level, more rapidly for hearing impaired persons than normal. Danaher *et al*. (1978) and Danaher and Picket (1975) have shown that, persons with sensorineural impairment exhibit backward and forward masking effects for a period of about 100–200 ms. Gehr and Sommers (1999) studied backward masking in young and elderly normal hearing listeners, to see the effect of age. Thresholds were measured for signal-masker delays of 1, 2, 4, 6, 8, 10 and 20 ms. Signal was a 10 ms, 500 Hz sinusoid and masker was a 50 ms segment of white noise. It was found that, masked thresholds were higher at all signal-masker delays for older subjects than younger. Younger and older adults exhibited masking of 19 dB and 40 dB respectively at shortest delay (1 ms). Across the set of delays tested, differences in masked thresholds for the two groups ranged from 22 to 25 dB. At 20 ms delay, older adults exhibited 25 dB of masking, while younger adults exhibited 5 dB of masking for signal-masker delays greater than 6 ms. Thus, older subjects exhibit backward masking for larger delays between signal and masker.

Danaher *et al.* (1978) studied forward and backward masking effect of F1 on the discrimination of F2 transitions and forward and backward masking of pure tones, in normal hearing persons and persons with moderate to severe sensorineural hearing loss. It is reported that discrimination thresholds for F2 transitions are always higher for hearing impaired subjects and both backward and forward masking effects extend up to about 100–200 ms. They measured the thresholds of pure-tone probes of 800 Hz and 1100 Hz each of length 20 ms preceded or followed by a masker. Masker was a noise with a formant-shaped spectrum having a single resonance centered at 600 Hz. Silent intervals of 10, 30, 100, or 300 ms were used between the probe and masker. The mean threshold shifts at -10 ms and -30 ms were 24–26 dB and 10–18 dB respectively. Threshold shifts at +10 and +30 ms were less than those at -10 and -30 ms indicating more backward masking effect than forward masking. Threshold shifts for probe of 1100 Hz were slightly lower than those for 800 Hz probe. It is reported that masking contribution of sensitivity shifts is limited to a range of only 50 ms before and after the first formant masker.

Elliot (1975) studied the effect of backward masking in persons with sensorineural hearing impairment. Threshold of a pure tone of 4000 Hz of length 20ms was measured in the presence of a 90 dB, 50 ms white noise. It is reported that effect of backward masking extends to about 100 ms compared to 10 ms for normal listener and amounts of threshold elevation in the 30–50 ms were substantial.

For normal hearing listeners forward and backward masking effects occur for delays up to 200–250 ms and 10–20 ms respectively (Dolan and Small, 1984; Elliot, 1962; Elliot, 1975; O'Shaughnessy, 1987; Raab, 1961; Smiarowski and Carhart, 1975; Wilson and Carhart, 1971). Danaher *et al.* (1978), Danaher and Picket (1975), and Elliot (1975) reported a backward masking effect for interval about 100–200 ms for persons with sensorineural hearing loss. Listeners with sensorineural loss show a forward masking effect for delay up to 200 ms (Danaher *et al.*, 1978). On the basis of the literature reviewed it can be inferred that for hearing impaired persons, the masking extends up to about 200 ms in both forward and backward directions.

2.4.4 Role of hair cells in transduction mechanism

Researchers have studied the role of hair cells in the transduction mechanism and role of auditory nerve fibers in conveying the information to the higher levels of the brain. Outer hair cells enhance the basilar membrane responses to low-level sounds thus increasing the sensitivity and they also sharpen the tuning (frequency selectivity) of the basilar membrane (Moore, 1997; Moore, 1998; Pickles, 1982; Wall, 1995). Inner hair cells play role in detection of the signal and excitation of the nerve fibers. Outer hair cells control the sensitivity of the inner hair cells, in such a way that it is high at low levels of vibration, and progressively decreases for higher levels of vibrations (Moore, 1997).

Damage to the inner hair cells (e.g. noise trauma producing changes in the stereocilia) reduces their transduction sensitivity for basilar membrane vibrations, leading to increased hearing thresholds at all the frequencies, without affecting the tuning curves (frequency selectivity). There is no loudness recruitment and hence the dynamic range does not get affected. Outer hair cells are more susceptible to damage than the inner hair cells over the greater length of the cochlea especially at the basal and middle turns. Ototoxic drugs damage mainly the outer hair cells. Damage to outer hair cells cause reduction of sensitivity at low sound levels and loss of frequency selectivity. Reduction in sensitivity at low sound levels results in loudness recruitment and thereby reduced dynamic range. Severe spectral masking occurs due to broadening of the tuning curves (Moore, 1998). As the place representation of frequency along the basilar membrane is preserved in the auditory nerve, damage to the auditory nerves generally impairs the “place” coding of the frequencies, resulting in reduced frequency selectivity and also in loudness recruitment and reduced dynamic hearing range (Pickles, 1982).

2.5 Perceptual consequence of sensorineural hearing loss

Temporal variations in level in normal conversational speech cover a range of about 30 dB. This is in excess compared to the available dynamic range for most hearing impaired listeners

(Moore, 1997; Stone and Moore, 1992). Acoustic elements associated with vowels are more intense than those associated with consonants. Thus, a person with recruitment may be able to hear vowels, but weaker consonants may be inaudible.

Poor frequency selectivity reduces the difference in the amplitude of the spectral peaks and troughs in the internal auditory representation of the acoustic signal. The minimum difference in amplitude between spectral peaks and troughs sufficient for vowel identification required by normal hearing persons is 1–2 dB, whereas hearing impaired persons require 6–7 dB. Persons with moderate hearing loss can still identify the vowels, because in natural speech vowels have peak-to-trough differences of at least 8 to 10 dB. Persons with severe to profound hearing loss, find difficulty in identifying the vowels (Leek *et al.*, 1987).

Flanagan (1972) has reported smallest detectable changes or difference limens (DLs) for frequency and intensity for synthetic sounds. DL for overall intensity of a synthetic vowel is about 1.5 dB. For voiced and unvoiced sounds (fricative consonants) DLs for overall intensity are 1.5 dB and 0.4 dB respectively. Amplitude DLs for first and second formants are 1.5 dB and 3 dB respectively. Spectral peaks and notches are differentially perceptible if their Q's (ratio of center frequency to bandwidth) are greater than about 5 and 8 respectively. DL for fundamental frequency is 0.3 to 0.5% of the fundamental frequency. Threshold for detecting a difference in the frequencies of two successively presented pure tones is about 0.1%. DLs for first and second formant frequencies are 3–5% of the formant frequencies. DL for formant bandwidth is 20–40%.

Reduced frequency selectivity due to broad auditory filters is associated with reduced spectral contrast (smearing of spectral peaks and valleys) and masking of certain frequency components. Sources of masking can be from environmental noise or from the primary signal itself. In speech signal, energy content of the first formant of the vowel is usually higher than that of the second and higher formants and hence second and higher formants are masked by the first formant. Also, first formant interferes with second formant transitions, which are important cues for consonant identification.

Hearing impaired persons show deficit in using the information contained in the rapid formant transition which are important for place identification (Dorman *et al.*, 1985). This occurs due to upward spread of masking leading to poor place identification (Danaher and Picket, 1975; Hanley and Dorman, 1983; Turek *et al*., 1980). Fricatives have more energy in the high frequencies. Persons with hearing loss at high frequency will have difficulty in perceiving these.

In speech signals, the consonants which are major carriers of information have low energy compared to vowels (Crandall, 1917). Therefore, there is possibility of masking of consonantal segment by vowels. Typically, stop bursts are 30 dB weaker than following vowels, and they occur within the temporal range of backward masking, hence they get confused in persons with sensorineural hearing loss. Poor temporal analysis results in significant reduced speech intelligibility. Voice-onset-time (VOT) is one of the major parameters, distinguishing voiced and unvoiced stop consonants. Tyler *et al.* (1982) conducted experiments on hearing impaired subjects to see the effect of degraded temporal parameters on speech perception. Subjects were required to identify and discriminate synthetic voiced and unvoiced stop consonants, which are distinguished by differences in VOT. Reduced ability to discriminate VOT cues, is reported in their study. Snell (1997) measured the gap thresholds for noise-burst stimuli in young and elderly listeners with pure-tone thresholds less than 20 dB HL over 250–4000 Hz range. It is reported that, gap thresholds in elderly listeners were significantly larger. It is found that, poor temporal resolution is unrelated to degree of hearing loss. Strouse *et al*. (1998) studied the effect of age on temporal processing of speech signals. Elderly persons with normal hearing thresholds (pure-tone thresholds less than 20 dB HL from 250 to 6000 Hz) and young normal hearing persons were tested with synthesized continuum of consonant-vowel (CV) syllables ranging from /b*a*/ – /p*a*/. It is reported that, elderly listeners have shown reduced sensitivity to difference in VOT as compared to younger listeners. Syllable /b*a*/ synthesized with VOT of 0, 10, 20 ms was correctly recognized with percentage scores of 98, 95, 82 by young listeners and 93, 91, 72 by elderly listeners respectively. For syllable /p*a*/ synthesized with VOT of 30, 40, 50, and 60 ms the corresponding percentage scores are 75, 88, 92, 98, and 50, 70, 80, 92 respectively. The mean phonetic boundary of VOT for /b*a*/ – /p*a*/ discrimination was 27 ms for young adults and was 32 ms for the elderly.

It has been observed that hearing impaired persons identify consonants in initial position better than consonants in final position (Dubno *et al*., 1982; Picheny, 1986;

Resnick *et al*., 1975; Yund and Buckles, 1995a). However some studies have reported better consonant recognition in syllable final position (Pols and Schouten, 1978; Wang and Bilger, 1973). Voiced consonants are identified better than unvoiced consonants (Dubno and Levitt, 1981; Resnick *et al*., 1975). However Gordon-Salant (1986) has reported higher recognition scores for unvoiced consonants than voiced for both young and elderly normal hearing persons. Dubno *et al*. (1982) and Walker *et al*. (1984) have reported higher scores for manner feature than place feature for hearing impaired subjects.

In case of unvoiced stops /p, t, k/ with vowel /a/, stop /t/ has high frequency noise burst, and it is distinguished easily from /p/ and /k/ which have low frequency noise bursts. Stop /k/ has slightly higher frequency noise as compared to /p/. Spectral masking may lead to difficulty in discrimination of /p/ and /k/. Among the voiced stops /b, d, g/ followed by vowel /a/, formant F2 rises for /b/ and falls in case of /d/ and /g/. As the vowel formant is relatively audible, /b/ is easily distinguished from /d/ and /g/, but there is confusion between /d/ and /g/. Compared to stop consonants, nasals /m, n/ are more intense and slightly longer, where as the fricatives /s, z/ are intense and longest. The fricatives /s, z/ have energy concentration at high frequency as compared to fricatives /f, v/ (Flanagan, 1972; Miller and Nicely, 1955). The VOTs of consonants /d/ and /g/ are very close and lead to confusion between them (Kewley-Port, 1982). Perception of various consonants according to place of articulation depends on the variations in second and third formant transitions, while perception of manner requires comparable variations in first formant (Liberman, 1957).

Above discussion infers that the consonantal segments in which formant transitions, VOTs, and noise bands do not differ widely, will have perceptual confusions due to masking.

2.6 Speech processing for sensorineural loss

Conventional electroacoustic hearing aids are commonly used for addressing the problem of elevated thresholds. The components used in these hearing aids are: a microphone, electronic filter, controls for adjusting the amplification and overall shape of the frequency response, circuits for limiting the amplified signals to a comfortable level, an earphone, a battery for providing the power source, and flexible tubing and an earmold, for coupling the output of the earphone to the external ear canal. Conventional hearing aids are classified according to their size and the way of wearing, as body-worn hearing aid, eyeglass hearing aid, behind-the-ear (BTE) hearing aid, in-the-ear (ITE) hearing aid, and in-the-canal (ITC) aid (CHABA, 1991; Wall, 1995). BTE hearing aid is relatively small and very popular, the electronic components of which are contained in a small elliptical case that fits behind the ear. The receiver delivers the signals to the ear canal through a flexible tube terminating in an earmold. These are the hearing aids most commonly used by persons suffering from severe to profound hearing loss. ITE hearing aid is even a smaller one. This is made up of a small plastic case containing all the components. It is molded to fit into outer portion of the external ear and the concha. Persons having mild to moderate and moderately severe hearing loss use this type of aid. ITC aid is the smallest among all hearing aids that fits entirely in the ear canal. Because of small size, available power output is low. Hence patients with mild or moderate, flat, and gradual sloping hearing losses generally use these aids (CHABA, 1991; Wall, 1995).

Acoustic amplification for making weak signals audible makes intense signals uncomfortably loud due to loudness recruitment in sensorineural hearing loss. To overcome this problem, compression techniques are used in hearing aids. In these techniques the amplification decreases with intensity, such that the wide dynamic range of input signal gets compressed in a smaller dynamic range at the output.

Frequency transposition technique is a method in which energy in the high frequency region is transposed to low frequency region. Various speech-enhancing schemes have attempted to reduce the effect of increased spectral and temporal masking, in cases of sensorineural hearing loss. To reduce the effect of spectral masking, schemes are based on the alteration of energy of certain frequency components. Schemes for enhancing the speech perception for the persons with reduced temporal resolution are based on the use of clear speech.

Some of the studies that have used compression technique, frequency transposition technique, and the schemes that have attempted to reduce the effect of increased spectral and temporal masking are reviewed in the following subsections.

2.6.1 Compression techniques

In compression techniques, weaker signals are amplified more than stronger ones so that wide dynamic range of the input signal gets compressed to smaller dynamic range. Automatic volume control, compression limiting, syllabic compression, and multiband compression systems are different compression techniques (CHABA, 1991; Lunner, 1997; Moore, 1997; Moore, 1998). Automatic volume control systems have long time constants (> 150 ms) and change their gain slowly with changes in signal level. Compression limiting systems have short time constants (attack time < 5 ms and release time ≈ 20–100 ms). Syllabic compression systems have short time constants (20 ms to 100 ms), same as the duration of individual syllables and raise the level of weak consonantal segments avoiding the more intense sounds becoming uncomfortably loud (Lunner, 1997; Moore, 1998). Since the hearing loss is frequency dependent, provision for different amount of compression in two or more frequency bands, i.e. multiband compression are found advantageous. This method will make relatively weak high-frequency components in speech, audible (Villchur, 1973). Some of the studies that use compression technique are reviewed in the following paragraphs.

Van Dijkhuizen *et al.* (1991) studied the effect of multiband compression on the masked speech reception threshold (SRT) in conditions of low-frequency noise for both normal hearing and hearing impaired listeners. Three compression factors (0%, 50%, and 100%) were used. Increase in level with addition of noise in the octave bands was compressed with the above mentioned compression factors. Four octave bands of 0.25–0.5, 0.5–1, 1–2, 2–4 kHz were employed. An improvement of 4 dB in masked speech reception threshold for a compression factor of 100% is reported. Authors have stated that there is reduction in upward spread of masking due to multiband compression. Lunner (1997) has pointed out that, this scheme may not help in improving speech intelligibility in noise, since compression causes reduction in internal variations (differences in the level of various phonemes of the speech signal itself) and negative effects may increase with increased compression ratio and the number of channels. In multi band compression method, 2–3 bands are sufficient to achieve adequate compensation for recruitment. Since the compression reduces the spectral contrasts in complex signals, use of large number of bands may be disadvantageous in hearing impaired persons with reduced frequency selectivity (Moore, 1997).

Kiessling and Steffens (1991) studied the effect of a 3-channel compression system on subjects with flat and sloping hearing loss. Speech recognition scores were 7 to 20% higher compared to single channel. To determine the effect of number of bands on speech discrimination of mild to moderately severe hearing-impaired subjects, Yund and Buckles (1995a,b) used multichannel compression hearing aids with 4, 6, 8, 12, and 16 frequency channels. Test material consisted of a closed set of nonsense syllables. Sixteen hearing impaired subjects showed increase in speech discrimination when the number of channels increased from 4 to 8, but did not show significant change when the number of channels varied from 8 to 16. In a study by Walker *et al.* (1984), compression using 6 channels found beneficial only for some subjects (subjects had moderate sensorineural hearing loss) under some listening conditions.

Asano *et al.* (1991) reported a digital hearing aid which employed processing to compensate for narrow dynamic range, with the objective of avoiding the spectral flattening associated with multi band compression. Input signal was segmented into 8 ms window, and each block was processed with an FIR filter, with frequency-gain characteristic determined by the short-time magnitude spectrum of the segment and loudness compensation functions (relation between the loudness for normal listeners and that for the impaired listener). Spectral values in 6 octave bands were used for determining the frequency response, and filter coefficients were calculated using frequency sampling technique. The scheme was evaluated by conducting tests on 13 sensorineural impaired subjects with monosyllabic speech as test material. Higher scores compared to linear amplification were reported.

2.6.2 Frequency transposition techniques

The common characteristic of sensorineural hearing impairment is greater loss in the high frequency range. Johannson *et al.* (1966) investigated a scheme of frequency lowering. In this method, transpose of the speech energy in the high frequency region into low-frequency region made the high frequency components audible and resolvable. In the experiment conducted, children having profound hearing loss showed improvements in the discrimination of fricatives and other phonemes. Foust and Gengel (1973) tested the frequency transposing technique on subjects with moderate to profound hearing loss using monosyllabic words with unvoiced stops, and voiced and unvoiced fricatives. It is reported that, the subjects who had undergone training showed 6–10% improvement in discrimination ability, also subjects with normal hearing up to 4 kHz showed better performance with the scheme.

Reed *et al*. (1983) studied the discrimination of speech processed by frequency lowering (preserving the fundamental frequency) and low pass filtering. Low pass filtering was used to simulate the high frequency hearing loss. Frequency lowering was achieved with nonuniform compression of short-term spectral envelope. There were four steps of signal processing: segmentation, warping, expanding and time aliasing, and resynthesis. The speech waveform was first segmented into temporally consecutive intervals: pitch periods for voiced sounds and intervals of random length for unvoiced sounds. Each segment was processed by the linear, time-variant “warping” operation to achieve merely a nonuniform monotonic transformation of the frequency axis of the spectrum of the segment. The resulting segments were expanded in time to accomplish a lowering of the warped spectrum and time aliased to compensate for the prolongation that accompanies warping and expansion. Frequency lowering scheme was superior to low-pass filtering for contrasts for fricative sounds but inferior for contrasts for nasals and semivowels. Place contrasts of fricatives and manner contrasts of affricates with plosives or with fricatives were more discriminable with frequency lowering scheme than with the low pass filtering.

In speech processing technique for single channel cochlear prosthesis (Pandey, 1987; Pandey *et al*., 1987), provision of information about the high frequency noise burst along with the information about low frequency periodicity improved the recognition scores for normal hearing subjects and for patients with extracochlear stimulation. For normal hearing subjects, high frequency information made the sounds perceptually distinct without adversely affecting the response time. Provision of high frequency noise burst information, did not adversely affect the reception of voicing (which was provided by low frequency periodicity) and place (through lip reading) for cochlear implant patients.

2.6.3 Processing to reduce the effect of degraded frequency selectivity

Some of the schemes that attempted to reduce the effect of increased spectral masking are reviewed in this section.

Summers and Leek (1997) investigated the effect of attenuating the energy of first formant of vowel, in reducing the upward spread of masking. Six synthetic CV stimuli /b*a*/, /d*a*/, /g*a*/, /be/, /de/, and /ge/ were used as test materials. It is reported that attenuation in the first formant energy of the vowel up to 18 dB improves the recognition score of consonants by 6–7%. In an earlier investigation by Skinner (1980), attenuation of low frequency components of monosyllables relative to high frequency components resulted in improved intelligibility for hearing impaired listeners.

Jamieson *et al.* (1985) have reported that, narrowing the formant bandwidths of vowel, improved the vowel identification significantly. Summerfield *et al.* (1985, as reported by Leek and Summers, 1996) studied the effect of modifying formant bandwidths on identification of synthetic vowel-consonant-vowel (VCV) syllables. Narrowing the bandwidths helps to increase spectral contrast, whereas widening the bandwidths tends to result in flatter input spectra. Authors have found poorer performance with increasing bandwidth and improved consonant identification with narrowing the bandwidths in both normal hearing and hearing impaired subjects. However, improvement was not statistically significant for hearing impaired listeners.

Simpson *et al.* (1990, as reported by Leek and Summers, 1996) investigated a scheme of pre-processing the signal to compensate for reduced frequency resolution. Aim of the processing technique was to improve the spectral contrast in speech, which is contaminated by background noise. Magnitude spectrum was obtained by using a bank of filters with bandwidths corresponding to auditory filters (normal) and was enhanced by spectral convolution with difference-of-Gaussians filter. Speech was resynthesized by sampling the enhanced magnitude spectrum and associating it with the original phase spectrum. They have reported that there was consistent increase in intelligibility for hearing impaired listeners tested in noise.

2.6.4 Processing to reduce the effect of temporal masking

The speech used in normal everyday situations is known as “conversational speech”. On the other hand the speech that occurs when one is trying to improve communication in difficult situations, as speaking in a noisy environment or speaking to a hearing impaired person, is known as “clear speech” (Picheny *et al.*, 1985, 1986). Changes in the conversational context, sentence structure, vocabulary, speaking rate and stress, pronunciation of individual words and vocal efforts play a significant role in enhancing the clarity of the speech. Acoustic features between clear and conversational speech can be classified into three levels: global, phonological, and phonetic changes. Global changes include changes in speaking rates, long term spectra, fundamental frequency, and number of pauses. Phonological changes consist of insertion, deletion, and feature changes of phonemes. Changes in acoustic properties of individual words like spectra, amplitude, and duration are the phonetic changes. Consonant-to-vowel (C/V) intensity ratio and consonant duration are the two important features that contribute to clear speech. In clear speech, there is increase in C/V intensity ratio and decrease in average speaking rate. Burst, closure, formant transition, and VOT are various acoustic segments associated with articulation and perception of stop consonants that contribute to the increased duration in clear speech. In conversational speech, the vowels are modified or reduced and stop bursts are often not released, whereas in clear speech vowels are less modified and stop bursts are always released. Picheny *et al.* (1986) have shown that intelligibility with clear speech is about 17% higher than conversational speech, for persons with sensorineural loss.

Ono *et al*. (1982) have shown that for unvoiced plosives, an increase of consonant amplitude by 10–20 dB and insertion of a silent pause between consonant and vowel to prevent masking by a following vowel, improved recognition score by 46.7%. Gordon-Salant (1986) conducted experiments to investigate the effect of C/V intensity ratio and consonant duration on consonant recognition in noise. Nonsense syllables in CV context were used at SNR of 6 dB. The C/V intensity ratio was increased by 10 dB and consonant duration was increased by 100%. Consonant recognition was measured for normal hearing listeners at presentation levels of 75 dB SPL and 90 dB SPL respectively. Improvements in the recognition scores of 13% at 75 dB SPL and 9% at 90 dB SPL are reported for C/V intensity modification. In consonant duration modification, replicating each period of the waveform throughout the consonant increased the duration of the voiced consonants. Whereas for the unvoiced consonants the duration was increased by partitioning the frication or burst noise into segments of approximately 20 ms and replicating each segment following the original. This increase in consonant duration has shown slight improvement in recognition score. In a later study, Gordon-Salant (1987) tested the scheme on persons with sensorineural impairment and observed that 10 dB increase in C/V intensity ratio resulted in 16% and 11% improvement at presentation levels of 75 dB SPL and 90 dB SPL respectively. Reasons for lack of significant improvement with duration cue enhancement, as brought out by the author, are due to: (i) the specific nature of the acoustic characteristics implemented with the modification, and (ii) the acoustic segments associated with consonant phonemes increased in a nonuniform manner in clear speech.

In a study by Montgomery and Edge (1988), increase in C/V intensity ratio resulted in 10–12% improvement in recognition scores at 65 dB SPL presentation level. Modest benefit of 5% has been reported for consonant duration enhancement, in which consonant duration was lengthened by replicating several short subsegments. Reason given by the author for lack of significant improvement with duration increase is that, increasing the consonant duration may not duplicate the articulatory efforts present in the clear speech. In a study by Freyman and Nerbonne (1989), increasing C/V intensity ratio resulted in improvement in recognition scores for normal hearing persons for CV syllables presented in noise.

Thomas (1996) and Thomas *et al*. (1996) have shown that, increasing the C/V intensity ratio improves the recognition scores. Six synthetic stop consonants /p, t, k, b, d, g/ with vowel /*a*, i, u/ were used in vowel-consonant (VC) and CV contexts. C/V intensity ratios were modified by +3, +6, +9, and +12 dB. To simulate hearing loss in normal hearing persons, stimuli were presented at 12 dB and 6 dB SNR. In CV context, 12 dB C/V intensity ratio modification has shown 26% and 28% improvement in score for 12 dB SNR and 6 dB SNR respectively. Higher scores are reported for /*a*/ context than for /i, u/ context. In VC context, 12 dB C/V intensity ratio modification resulted in 8% and 20% increase in score for 12 dB SNR and 6 dB SNR respectively. Scores were higher for VC position than CV position, indicating that increase in C/V intensity ratio is more effective in reducing forward masking than backward masking. In the VC context, high C/V intensity ratio resulted in some vowel confusions.

Kennedy *et al*. (1998) used VC nonsense syllables with three vowel contexts /*a*, u, i/. C/V intensity ratios were adjusted according to consonant type, vowel environment and audiogram configurations. Eighteen subjects with sensorineural hearing impairment participated in the test. For voiced consonants, with C/V ratio adjusted, percentage recognition scores averaged across the subjects increased from 65.8 to 73.4, 46.8 to 62.6, and 45.8 to 63.8 for vowel contexts /a/, /u/, and /i/ respectively. It is reported that individualized adjustment of the C/V ratio for each subject and consonant-vowel combination can produce substantial improvements in consonant recognition.

In a study by Revoile *et al.* (1985), severely hearing-impaired listeners showed significant reduced voicing perception for final stops, especially for final fricatives upon duration reductions of the vowels preceding voiced consonants and corresponding duration increases for the vowels preceding voiceless consonants in spoken syllables. Revoile *et al.* (1986) investigated the effect of modifying the vowel duration for consonant recognition in severely/profoundly hearing impaired persons. Vowel duration cue was enhanced as a means of improving voicing perception for final fricative consonants. Spoken words /bæf, bæs, bæv, bæz/ were used for testing voicing perception with vowels unmodified and with enhancement of the vowel duration cue. Vowel durations were modified by deletion or replication of pitch periods. Vowels were lengthened by 100 ms to 150 ms from their normal duration before voiced fricatives and shortened by the same amount before the unvoiced fricatives. A subgroup of 14 listeners showed an improvement of 20–40% in recognition score.

In the study by Thomas (1996) and Thomas *et al.* (1996), the acoustic segments (burst, closure, formant transition duration, and VOT) were lengthened during consonant duration modification tests. Speech stimuli were used in CV context with vowel /a/. Burst duration, formant transition duration (FTD), and VOT were increased by 100%, 50%, and 100% respectively. Overall duration of each syllable was kept at 300 ms by adjusting the vowel duration. At 12 dB SNR, one of the four subjects has shown improvement of 12% when burst duration was doubled. At 6 dB SNR, 3 subjects have shown an average of 6% increase in recognition. In formant transition duration test, one subject has shown an increase of 13% for doubling of formant transition. At 6 dB SNR, three subjects have shown an average of 12% increase in score for 50% increase in FTD. There was no improvement with 100% increase in FTD. Average scores decreased by 4% at 12 dB SNR when FTD was doubled. For 6 dB SNR, scores increase by 4% for 50% FTD increase and then reduced by 4% for 100% FTD increase. In VOT modification, one subject has shown improvement, while average scores decrease with increase in VOT. It is reported that, formant transition duration of up to 50%

may be combined with burst duration modification for obtaining better performance at higher noise levels.

Turner *et al.* (1997) investigated the effect of increasing the formant transition duration on speech intelligibility in hearing impaired persons. Synthetic CV syllables /b*a*, d*a*, g*a*/ were used in the experiment. Formant transition durations were varied from 5 ms to 160 ms. There were no improvements in recognition scores for longer (slower) formant transition durations.

Nejime and Moore (1998) studied the effect of speech-rate slowing on speech intelligibility in persons with simulated hearing loss. Simulation was done for elevated thresholds, loudness recruitment, and reduced frequency selectivity. Two expansion rates of 1.25 and 1.5 were used for slowing the speech rate. The processing slows the speed of speech without changing its pitch. The algorithm expands the duration of speech segments whose power exceeds a certain threshold, but does not manipulate segments whose power is below the threshold. Thus a nonuniform time expansion is used. They found no improvement in recognition score. Artificially slowing of speech rate may not be suitable in improving speech intelligibility in hearing impaired with elevated thresholds, loudness recruitment and reduced frequency selectivity.

Listening with two ears is known as binaural hearing. Presenting same signal to both the ears is known as diotic presentation and presenting two different signals to the two ears is referred to as dichotic presentation. Researches have reported the benefit of binaural dichotic mode of hearing in improving speech perception by persons with sensorineural hearing loss. In Chapter 3, some of the schemes that have employed binaural dichotic presentation to reduce the effect of spectral and temporal masking are reviewed and the proposed scheme is explained.

Chapter 3

Binaural dichotic presentation

3.1 Introduction

Sensorineural hearing loss is characterized by increased threshold of hearing, reduced dynamic hearing range, degraded frequency selectivity and temporal resolution, and increased spectral and temporal masking. Some of the studies to solve the problem of elevated threshold and reduced dynamic hearing range are discussed in the previous chapter. Masking is a phenomenon, in which presence of one signal component elevates the threshold of neighboring signal component. To reduce the effect of increased spectral and temporal masking, studies that are based on consonant enhancement technique are also discussed in the previous chapter. Listening with both the ears is known as binaural hearing. The ability to perceptually combine binaurally received signals from two ears improves speech perception under adverse listening conditions. Binaural listening offers better overall sound quality and intelligibility, more relaxed listening, and it also helps in source localization (Moore, 1997; Pickles, 1982). Many studies have shown the benefit of binaural hearing over the monaural hearing (Brooks, 1984; Hawkins and Yacullo, 1984; Hirsh, 1950; Jerger *et al*., 1961). Presenting the same signal to both the ears is known as diotic presentation and presenting two different signals to the two ears is referred to as dichotic presentation. Here we discuss the use of binaural dichotic presentation as a means of reducing the effects of masking for persons with moderate bilateral sensorineural loss.

Masking takes place primarily, at the level of peripheral auditory system. In speech perception, the information received from both the ears gets integrated. Hence splitting of information in speech signal for presenting signals to the two ears, in some sort of a complementary fashion, to provide relaxation for sensory cells of the basilar membrane, may help in reducing the effect of increased masking and thereby improve the speech reception in cases of bilateral sensorineural hearing impairment with some residual hearing. To reduce the effect of spectral masking, studies that are investigated are based on splitting of speech signal into odd and even bands for binaural dichotic presentation (Chaudhari and Pandey, 1998a,b; Lunner *et al*., 1993; Lyregaard, 1982). These studies are briefly reviewed in this chapter. Next processing schemes for reducing the effect of both spectral and temporal masking are proposed. This chapter also presents experimental methods used for evaluation of the schemes.

3.2 Review of dichotic presentation

Reduced frequency selectivity due to broad auditory filters is associated with increased spectral masking. Sources of masking can be from environmental noise or from the primary signal itself. In speech signal, energy content of the first formant of the vowel is usually higher than that of the second and higher formants and hence second and higher formants are masked by the first formant. Also, first formant interferes with second formant transitions, which are important cues for consonant identification (Danaher and Pickett, 1975; Summers and Leek, 1997). Increased temporal masking associated with reduced temporal resolution results in, increased forward and backward masking of weak acoustic segments by strong ones. Important cues namely voice-onset-time, formant transition, and bursts that are required for consonant identification get masked by the preceding and following vowel segments, resulting in phonemic confusion (O'Shaughnessy, 1987).

To reduce the spectral masking, schemes based on separating various frequency bands for binaural dichotic presentation have been reported (Chaudhari and Pandey, 1998a,b; Lunner *et al*., 1993; Lyregaard, 1982).

Danaher and Pickett (1975) used dichotic presentation to study the discrimination of second formant transitions in synthetic vowels (stimuli approximating VCV syllables), in subjects with sensorineural hearing loss. It is reported that, dichotic presentation of F1 and F2 (produced by two-formant synthesizer) improved the discrimination performance by reducing the masking of F2 due to F1. To reduce the upward spread of masking in hearing-impaired listeners, Turek *et al*. (1980) used synthetic CV syllables /b*a*, d*a*, g*a*/ and presented F1 to one ear and F2/F3 to the other. Ten subjects having moderate flat to sloping high frequency sensorineural hearing loss were tested for identification performance, at most comfortable listening level. It is reported that, only few subjects showed benefit of dichotic hearing over the monotic hearing.

Based on the assumption that degraded frequency selectivity is peripheral, Lyregaard (1982) developed a scheme of separating the spectrum bands for presenting the alternate bands to the two ears, with the objective of bypassing the poor frequency selectivity at the peripheral level. He used a comb filter of constant bandwidths to separate the spectral components into two parts for binaural dichotic presentation. Filters were realized using sum and difference of the original signal with a delayed version (analog delays). The study used three bandwidths of 200, 500, and 800 Hz. Sentences consisting of two-syllable words were used as test material. The hearing impaired subjects had 50 dB HL and degraded frequency selectivity. No significant improvement with dichotic over the diotic was reported. Lack of significant improvement may be due to inappropriate filtering, inadequate listening experience, and non-feasibility of binaural fusion for two dissimilar signals.

Lunner *et al.* (1993) investigated an 8-channel digital filter bank for hearing aid use. Filters used were complementary interpolated linear phase FIR filters of 700 Hz constant bandwidth. Frequency response of the filter bank was adjusted according to the subject's hearing thresholds. Odd numbered band signals were fed to one ear and even numbered band signals were presented to the other ear. Three subjects aged between 38 to 69 years, having bilateral, symmetrical, moderate sensorineural hearing loss participated in the test. The scheme was implemented using digital signal processor TI/TMS320C25. Speech materials consisting of lists of five-word sentences were presented in the presence of background noise. For 50% correct recognition score, the test results showed an improvement of 2 dB in signal-to-noise ratio due to dichotic presentation.

Another scheme, for reducing the effect of increased spectral masking, investigated at IIT Bombay (Chaudhari, 2000; Chaudhari and Pandey, 1998a,b) used critical bands corresponding to auditory filters. These bandwidths as described by Zwicker (1961), are about 15–17% of the center frequency for 1–5 kHz range and are constant at about 100 Hz below

500 Hz. In this scheme, 18 bands have been used to cover the frequency range of 5 kHz. The scheme was implemented in real time using two TI/TMS320C50 processors, initially for constant gain in all bands (Chaudhari and Pandey, 1998b). In their later study, as a way of partial matching of the filter response to the frequency characteristics of individual subjects, gains for the bands were adjusted in the range -3 dB to +3 dB. Twelve consonants /p, t, k, b, d, g, m, n, s, z, f, v/ in VCV and CV context with vowel /*a*/ as in father, were used as test material. With constant gain in all bands percentage relative improvement in recognition score for hearing subjects with bilateral sensorineural impairment, ranged from 9.2 to 23.6 in VCV context and 14.4 to 19.2 in CV context. Information transmission analysis indicated that, all the six features (duration, frication, nasality, manner, place, and voicing) contributed to the improvement, with maximum improvement in relative information with the place feature. With adjustable gain, percentage relative improvement in recognition score ranged from 2.2 to 6.4 in VCV and from 1.6 to 7.8 in CV contexts respectively, with respect to processing with constant gain. In VCV context, almost all the subjects showed modestly higher transmission for place and manner feature(s). In CV context, two subjects showed highest improvement for place feature. The place information is related to frequency resolving capacity of the auditory processing. The maximum contribution by place feature in improving the recognition score in both schemes indicated that, the scheme of spectral splitting helps in reducing frequency masking.

In an attempt to reduce the temporal masking effect, Lunner *et al*. (1993) combined temporal splitting along with spectral splitting for dichotic presentation. In his scheme, a symmetrical inter-aural switching of odd and even filter bands (obtained with spectral splitting) was used with a period of 20 ms. During first 10 ms odd bands were presented to the left ear and even bands to the right ear, while in the next 10 ms odd bands were presented to the right ear and even bands to left ear. In this scheme, the peripheral auditory system of each ear is alternately relaxed in specific frequency bands due to inter-aural switching. The relaxation period should help in reducing the problem associated with reduced temporal resolution. However, no improvement in recognition score was reported. Also a poor sound quality was reported related to inter-aural switching.

3.3 Proposed scheme

The previous section has described earlier reported schemes making use of binaural dichotic presentation for reducing the effect of increased spectral masking. Our objective is to develop and investigate schemes making use of dichotic presentation to reduce the effects of increased temporal masking and to reduce the effects of temporal as well as spectral masking.

Reduced temporal resolution which is associated with forward and backward masking of low-energy components by adjacent high-energy components, degrades the speech perception in persons with sensorineural loss (Strouse *et al.*, 1998; Tyler *et al.*, 1982). Masking takes place primarily at the level of peripheral auditory system. Splitting the speech temporally into a number of segments and presenting adjacent segments to the different ears may reduce the overlap of cochlear responses of adjacent segments and help in reducing the effect of increased temporal masking by providing relaxation period for sensory cells of the basilar membrane. In the real-time processing scheme reported by Lunner *et al.* (1993), briefly described in the previous section, for combining the spectral and temporal splitting for dichotic presentation, odd and even bands were alternately switched between the two ears. There was no improvement in the recognition score. Also a poor sound quality was reported related to inter-aural switching. Deterioration of speech quality and lack of improvement in recognition scores may have been caused by abrupt transitions during switching of odd and even bands between the two ears.

Provision of overlap during inter-aural switching and use of smooth inter-aural fading might improve the quality of the perceived signal, by avoiding the perception of temporal gaps and spectral distortion during switching. Hence the objective of the research is to investigate the scheme of temporal splitting of speech and to study the effect of inter-aural overlap and various fading functions in reducing the effect of temporal masking.

In the scheme of temporal splitting for dichotic presentation, speech is split into subsegments for presenting adjacent segments to the two ears. This may reduce the backward and forward masking that takes place at the peripheral level, between the adjacent segments. The two fading functions used for temporal splitting are: symmetrical inter-aural switching with overlap and step transition, and symmetrical inter-aural switching with overlap and trapezoidal transition. Provision of overlap (duty cycles greater than 50%) may help in improving the speech intelligibility by avoiding the perception of temporal gaps during switching. Hence it is proposed to carry out a study to determine the effect of different duty cycles. Use of step transition may result in spectral distortion deteriorating the speech quality. Employment of trapezoidal variation during inter-aural switching may help in decreasing these distortions and thereby may improve the speech quality. Towards this end, inter-aural switching with different overlaps is studied. Next the scheme is investigated with a smooth fading function involving linear and logarithmic amplitude variation. It was found that, the intelligibility of the signals processed with linear variation was better than those processed with logarithmic variation. Hence linear variation has been used in this investigation and this is referred to as inter-aural switching with trapezoidal fading.

In the scheme of spectral splitting, sensory cells corresponding to alternate bands of the basilar membrane are always stimulated, whereas sensory cells of other bands are always relaxing in both the ears. In temporal splitting, all the sensory cells of the ears get relaxed alternately for some time. In the second phase of this research, to reduce the effect of increased spectral and temporal masking simultaneously, a scheme of combined spectral and temporal splitting for binaural dichotic presentation has been investigated using time-varying comb filters. With the combined splitting, all the sensory cells of the basilar membrane get periodic relaxation from stimulation.

In the scheme of spectral splitting for binaural dichotic presentation investigated by Chaudhari and Pandey (1998a,b) reviewed earlier, comb filters were designed for sharp transitions. In ideal splitting, any spectral component would be presented to one ear. However, with filters with finite crossover in the magnitude response, the components lying in the transition regions are presented to both the ears. With same intensity, binaurally presented components will be louder than monaurally presented components. If the magnitude response is not properly shaped at the transitions, the components lying in the overlapped region will be perceived with different loudness and will reduce the speech quality. Hence it was decided to conduct loudness perception tests to determine the difference in intensity for same perception in monaural and binaural presentations using different types of stimuli. Comb filters were designed with different levels at crossovers between adjacent bands and were tested by sweeping long duration sine tones of narrow frequency ranges (to cover a pair of adjacent bands). Comb filters with minimum pass band ripple, maximum stop band attenuation, and inter-band crossover level that produced minimum spectral distortion were considered as optimized comb filters and were used in spectral splitting.

In the scheme of combined spectral and temporal splitting, two time-varying comb filters with complementary magnitude responses are cyclically swept between left and right ears to obtain the combination of spectral and temporal spectral splitting (spectral splitting due to separating the odd and even bands for presenting to the two ears and temporal splitting due to sweeping of the bands between the two ears). In this scheme, it is proposed to study the effect of different number of filter sets that form a time-varying comb filter and to obtain optimum filter set which gives maximum improvement in speech perception.

In temporal splitting scheme, the inter-aural switching period has to be appropriately selected. Towards this end, various psycho-acoustic studies on temporal processing in speech were reviewed. Maximum rise time for a stimulus onset yielding an overshoot in auditory neural firings is about 20 ms which is often used as an integration time of auditory processing (O'Shaughnessy, 1987). Sound detectability is generally constant over the range of durations from about 15 ms to 200 ms, but falls off for durations smaller or greater than this (Buus *et al*., 1999; Green *et al*., 1957; Moore, 1997; O'Shaughnessy, 1987; Plomp and Bouman, 1959; Sheely and Bilger, 1964). Speech subsegments with important acoustic cues are of the order of about 20 ms. Hence inter-aural switching period of 20 ms has been used for temporal splitting of speech signal. The same duration has been used for cyclic sweeping of time-varying comb filters for combined spectral and temporal splitting.

3.4 Evaluation method

Verification of implementation of speech processing schemes is carried out by spectrographic and other analyses of processed output signal. However, evaluation of the effectiveness of the scheme for improving speech reception is carried out by listening tests. For this, methods and test material, measures used for evaluation, subjects, and presentation methods and level should be appropriately selected.

3.4.1 Methods and test material

Speech processing schemes for hearing aids may be evaluated on the basis of perceived sound quality and intelligibility test (Gabrielsson *et al*., 1988; Punch *et al*., 1980). In perceived sound quality judgement test, clarity and loudness of the sound are considered.

In speech intelligibility test, test materials such as words, nonsense syllables, and sentences are presented over the headphones to the subjects and subject's correct responses are noted. Gordon-Salant (1993), Humes *et al*. (1987), and Picheny *et al*. (1986) have measured consonant recognition scores (articulation score) in their study. Arlinger and Dryselius (1990), Dubno *et al*. (1984), Lunner *et al*. (1993), Lyregaard (1982), and Van Dijhuizen *et al*. (1991) have evaluated their speech processing schemes by a method in which processed speech was mixed with noise and signal-to-noise ratios for 50% correct recognition score (speech reception thresholds) were measured.

Recognition score though useful may be affected by subject's response bias. Information transmission analysis (Miller and Nicely, 1955) is another way of evaluation, which provides a measure of covariance between stimuli and responses. Chance scoring and pattern of errors are taken into consideration in the information transmission analysis. Pandey (1987) has worked out a relationship between relative measure of information transmitted and recognition score for a special case when the correct responses are equally distributed among the diagonal entries and the errors equally distributed among the off-diagonal entries in the confusion matrix. This measure can also be applied for different phonemic features, to evaluate the relative importance of the features. Information transmission analysis has been used in many studies (Apoux *et al*., 2000a,b; Apoux *et al*., 2001; Chaudhari and Pandey, 1999; Donald and Nelson, 2000; Dorman *et al*., 1990; Faulkner and Rosen, 1999; Faulkner *et al*., 2000; Fu *et al*., 1998; Fu and Shannon, 1999; Geurts and Wouters. 1999; Hou and Pavlovic, 1994; Kiefer *et al*., 2000; Loizou *et al*., 2000; Lorenzi *et al*., 1999; Miller and Nicely, 1955; Pandey, 1987; Tye-Murray *et al*., 1995; Tyler and Moore, 1992; Van der Horst *et al*., 1999; Wang and Bilger, 1973). The consonantal features may be classified as voicing, place, manner, nasality, frication, and duration (Gordon-Salant, 1986; Ladefoged, 1982; Miller and Nicely, 1955; Tye-Murray *et al*., 1995; Tyler and Moore, 1992; Wang and Bilger, 1973). Nasality feature is cued by nasal formant, first formant transitions, and intensity changes. Voicing and duration features are recognized by fundamental frequency and timing cues. High-frequency turbulences signal frication and spectral changes signal the place feature (Tye-Murray *et al*., 1995). Oral stop, fricative, or nasal stop categorizes manner feature. Tyler and Moore (1992) measured the consonant recognition in cochlear-implant subjects by using nonsense syllables. They performed information transmission analysis to determine the perception of different features of speech. Tye-Murray *et al*. (1995) also used information transmission analysis, to study the abilities of cochlear-implant children in utilizing and producing the speech features. Earlier Wang and Bilger (1973) have used sequential information transmission analysis (SINFA) which provides an indication of the contribution of independent feature.

Response time statistics is another way of assessing the speech processing scheme. When the recognition scores and information transmission are same, the scheme that requires less response time can be considered as superior. Decrease in the subject's response time, indicates lessening of load on the perception process. This method has been used by Apoux *et al*. (2000a,b), Apoux *et al*. (2001), Chaudhari and Pandey (1998), Gatehouse and Gordon (1990), Meftah and Boudelaa (1996), Pandey *et al*. (1987), and Thomas *et al*. (1996).

.

Test materials can be nonsense monosyllables, disyllables, words, and nonsense sentences. With nonsense syllables as test material, subject's errors reflect acoustic confusions without the influence of semantic constraints. Monosyllabic words are the components of meaningful speech, which do not require syntactical cues (CHABA, 1988). Dubno *et al*. (1984) and Nilson *et al*. (1994) have used spondees as test material for measuring speech reception thresholds. Spondees spoken as isolated utterances or in carrier phrases may not represent the normal spectral weighting, level fluctuations, intonations, pauses, etc., associated with conversational speech, hence they are less representative of natural language communication than sentences (Nilson *et al*., 1994). The recognition of words in sentences is typically better than the recognition of isolated words, due to the addition of grammatical and contextual cues.

To study the perceptual confusion, Miller and Nicely (1955) used 16 consonants /p, t, k, f, θ, s, ʃ, b, d, g, v, δ, z, □, m, and n/ in CV contexts with vowel /*a*/. These 16 consonants cover almost three quarters of the consonants that are used in normal speech. Consonants were chosen because they are considered to be important for intelligibility and information contained in them is less understood than that in vowels. Similar consonants have been used by Wang and Bilger (1973) in VC and CV contexts with vowels /a, i, u/ to study the perceptual features. Gordon-Salant (1986) used /b, d, g, p, t, k, m, n, s, z, ʃ, θ, δ, v, f, w, j, l, r/ paired with three vowels /a, i, u/ in CV context, to evaluate the efficiency of three acoustic modifications derived from clear speech for improving consonant recognition by young and elderly normal hearing subjects. Same consonant set has been employed by Dubno and Levitt (1981) in both CV and VC contexts. Dubno and Schaefer (1992) determined the frequency selectivity and consonant recognition in normal hearing and hearing impaired listeners by using stimuli in both VC and CV contexts. Humes *et al.* (1987) and Skinner *et al.* (1997) have used nonsense syllables in CV, VC, and VCV context in their study.

Picheny *et al.* (1986) studied the variations in the intelligibility of speech produced for hearing-impaired listeners using 50 nonsense sentences containing key words. Sentences were used because it was required to manipulate the prosodic and phonological variables. Nonsense sentences were used to avoid the semantic effect. Drawback with the sentences as test material is that there are chances of omission of words, leading to more errors.

To increase the contribution of acoustic properties and to minimize the linguistic effects, in our study we use a closed set of twelve nonsense syllables in VCV context, using consonants /p, b, t, d, k, g, m, n, s, z, f, v/ with vowel /*a*/ as in f*a*ther. These are used for studying the reception of consonantal features of voicing, place, manner, nasality, frication, and duration (Miller and Nicely, 1955; Ladefoged, 1982), with groupings as given in Table. 3.1. In our work, listening tests are carried out for consonantal identification. Response times, percentage correct recognition scores and relative information transmitted for consonantal features (details are given in Appendix B) are used for evaluating the effectiveness of the speech processing schemes.

3.4.2 Presentation and subjects

Conducting listening tests to evaluate the schemes for speech processing at various processing conditions is very time consuming. Hence it will be difficult to test different processing conditions directly on the hearing impaired subjects. Also, there are difficulties in having a

TABLE 3.1. Grouping of consonants by features.

Feature	Groups
Voicing (2)	Unvoiced : / *p t k s f* /, Voiced : / *b d g m n z v* /
Place (3)	Front : / *p b m f v* /, Middle : / *t d n s z* /, Back : / *k g* /
Manner (3)	Oral stop : / *p b t d k g* /, Fricative : / *s z f v* /, Nasal : / *m n* /
Nasality (2)	Oral : / *p b t d k g s z f v* /, Nasal : / *m n* /
Frication (2)	Stop : / *p b t d k g m n* /, Fricative : / *s z f v* /
Duration (2)	Short : / *p b t d k g m n f v* /, Long : / *s z* /

large number of hearing impaired subjects with bilateral hearing loss willing to participate in these experiments. Therefore, in our study, it was decided to evaluate the schemes initially with normal hearing persons with simulated sensorineural hearing loss and later with persons having sensorineural hearing loss. Towards this end, various studies that make use of simulating the sensorineural hearing loss were reviewed. Some of these are discussed here.

Leek *et al*. (1987) determined the minimum spectral contrast for vowel identification, by using broadband noise to simulate the elevated thresholds of range 72–75 dB in normal hearing subjects. Humes *et al*. (1987) used spectrally shaped noise to produce masked thresholds identical to the quite thresholds of hearing-impaired subjects. They measured speech recognition performance and reported that masking noise can simulate the sensorineural hearing loss. Moore and Glasberg (1993), and Nelson (1990) have also used broadband noise to simulate the elevated thresholds.

Dubno and Schaefer (1992) compared the frequency selectivity and consonant recognition among hearing impaired and masked normal hearing listeners. Spectrally shaped broadband noise was used to match the thresholds of normal hearing subjects to those of hearing impaired. The performance of masked normal hearing subjects was identical to the performance of the hearing impaired subjects.

Florentine and Buus (1984) determined the temporal gap thresholds in sensorineural and simulated hearing impairments. Spectrally shaped noise was used to simulate masked thresholds in normal listeners, similar to the quite thresholds of hearing impaired listeners. Some researchers have reported that broadband noise produces reduced dynamic hearing range and loudness recruitment (Dubno and Schaefer, 1992; Lochner and Burger, 1961; Stevens, 1965). Based on these studies, we can say that broad band noise can simulate most of the characteristics of the sensorineural hearing loss.

In our study, we have decided to use broadband Gaussian noise band limited to the band of speech, for simulating sensorineural hearing loss. Different levels of SNRs have been used to simulate hearing loss of varying degrees. Noise addition to the speech signal was done maintaining the SNR constant on the basis of short-time (≈ 10 ms) energy of the speech signal. Thus during silence segments, there would not be any background noise. Effect of SNR on recognition score, response time, relative information transmitted for different features for VCV syllables was tested on five normal hearing subjects and is presented in Appendix G.

Researchers have studied the effect of presentation level on consonant identification performance. Speech discrimination performance deteriorates at low (< 30 dB) and high (> 90 dB) presentation levels (Dorman and Dougherty, 1981; Gordon-Salant, 1987; Montgomery and Edge, 1988; Simon, 1978). Dorman and Dougherty (1981) studied the effect of presentation levels of 55, 70, and 90 dB SPL in identifying the stimuli /b*a*/, /d*a*/, and /g*a*/ by normal hearing subjects. They have found that the performance decreases for 90 dB SPL presentation level. Simon (1978) has reported that, for normal hearing subjects, consonant labelling performance was poor when the presentation level falls below 35 dB SPL. Gordon-Salant (1986) studied the effects of acoustic modification on consonant recognition by elderly hearing-impaired subjects. Test stimuli consisted of 19 consonants paired with three vowels. Two presentation levels of 75 and 90 dB SPL were used. Recognition scores were lower for 90 dB SPL presentation level compared to 75 dB SPL. Montogomery and Edge (1988) evaluated the speech processing schemes of amplitude enhancement of consonants and increased consonant duration. Two presentation levels of 65 and 95 dB SPL were used. In amplitude enhancement scheme, significant improvement in recognition scores is reported for presentation level of 65 dB. Dorman *et al.* (1990), Freyman and Nelson (1986) have used a presentation level of 65 dB and 75 dB respectively in their study.

In our work, it was decided to use presentation level at most comfortable listening level for the individual subjects. For each listening condition, presentation level was selected by the subject for each ear and was maintained constant for all the tests of a particular listening condition.

3.5 Summary

In this chapter, schemes that use binaural dichotic presentation are reviewed followed by the brief description of the proposed schemes and evaluation methods to be used. Implementation and experimental evaluation of the proposed schemes are given the following chapters.

Chapter 4

Temporal splitting using step and trapezoidal fading

4.1 Introduction

Implementation and evaluation of the schemes of temporal splitting using inter-aural fading with step and trapezoidal transition for binaural dichotic presentation to reduce the effect of increased temporal masking associated with sensorineural hearing loss are presented in this chapter. Evaluation was performed by conducting listening tests on normal hearing subjects with simulated hearing loss. Response times, recognition scores, and information transmitted for unprocessed and processed speech stimuli were used for performance evaluation.

4.2 Temporal splitting using step and trapezoidal fading

A symmetrical inter-aural switching period of 20 ms was used. Initially the stimuli were processed with 50% duty cycle and the processed stimuli were tested by listening. It was found that, the use of 50% duty cycle resulted in deterioration of intelligibility of the signal due to perception of temporal gaps. The perception of temporal gaps can be avoided by providing overlap between segments presented to the two ears. Thus the effect of different duty cycles (overlap periods) in improving the perception by normal hearing persons under noise conditions to simulate hearing loss has been studied.

Temporal splitting of speech for dichotic presentation was done by using the scheme shown in Fig. 4.1. The outputs $s_1(n)$ and $s_2(n)$ of temporal processing were obtained by multiplying the input $s(n)$ with fading functions $w_1(n)$ and $w_2(n)$ respectively. The two fading functions with step transition as shown in Fig. 4.2, provide a symmetrical overlap, each having a duty cycle $d = L/N$, where N = inter-aural switching duration and L = "on" period.

$$
\begin{aligned}
w_1(n) &= 1, && 0 \le (n)_N \le L - 1 \\
&\ 0, && \text{otherwise} \\
w_2(n) &= 0, && L - N/2 + 1 \le (n)_N \le N/2 \\
&\ 1, && \text{otherwise}
\end{aligned}
\tag{4.1}
$$

For an inter-aural switching period of N samples with 50% duty cycle, alternate segments of $N/2$ samples are presented to the same ear and adjacent segments get presented to different ears, i.e. while first $N/2$ samples of the signal are presented to left ear, right ear is relaxed and during next $N/2$ samples of signal presentation to the right ear, left ear is relaxed. This scheme may help in reducing the temporal masking. However, it is likely to introduce perceptual gaps around the transitions related to inter-aural switching. Splitting the signal with duty cycles greater than 50% will provide certain overlap between the two presentations and will help in reducing the perception of inter-aural switching. Duty cycles greater than 50% will leave less time available for relaxation of the auditory sensors and nerve fibers. This scheme is referred as TS_ST (temporal splitting with step transition). In this scheme, effect of different duty cycles was studied.

While providing relaxation interval for improving the speech perception degraded due to increased temporal masking, use of step transition during switching may result in spectral distortion that deteriorate speech quality. Inter-aural switching with trapezoidal fading function may help in improving the speech perception during switching. Keeping the duty cycle constant at optimal value (as obtained in the scheme of step transition along with overlap, TS_ST), effect of different transition durations in improving the perception for normal hearing persons under adverse listening condition has been studied. During transition period, the amplitude was varied (a) as a linear (b) logarithmic function of time. Comparison of the perceptual quality and clarity indicated that linear variation was better. This scheme is referred as TS_TR (temporal splitting with trapezoidal transition).

The fading function for trapezoidal transition along with overlap is shown in Fig. 4.3. The outputs $s_1(n)$ and $s_2(n)$ of temporal processing were obtained by multiplying the input $s(n)$

with fading functions $w_1(n)$ and $w_2(n)$ respectively. The two windows have a symmetrical overlap, each having a duty cycle $d = L/N$, where N = inter-aural switching duration, L = "on" period and M is the transition duration for inter-aural switching. For fading functions with trapezoidal transitions as shown in Fig. 4.3,

$$
\begin{aligned}
w_1(n) = \;& 1, && 0 < (n)_N \le L - 1 \\
& 1 - ((n)_N - L - 1) / M, && L \le (n)_N \le L + M - 1 \\
& 0, && L + M \le (n)_N \le N - M - 1 \\
& ((n)_N - N + M - 1) / M && N - M \le (n)_N \le N - 1 \\
\\
w_2(n) = \;& 1 - ((n)_N - L + N/2 - 1) / M, && L - N/2 \le (n)_N \le L - N/2 + M - 1 \\
& 0, && L - N/2 + M \le (n)_N \le N/2 - M - 1 \\
& ((n)_N - N/2 + M - 1) / M, && N/2 - M \le (n)_N \le N/2 \\
& 1 && \text{otherwise}
\end{aligned}
\tag{4.2}
$$

In both the schemes an inter-aural switching period of 20 ms was used. In the scheme of temporal splitting with step transition, effect of duty cycles of 100 (unprocessed), 75, 70, and 60%, (i.e. "on" period of 20, 15, 14, and 12 ms respectively) was studied. In the scheme of temporal splitting with trapezoidal transition, duty cycle of 70% (optimal value in the scheme of temporal splitting with step transition) and three transition durations 0, 1, 2, and 3 ms were used. Transition duration of 0 ms refers to step transition and 3 ms refers to maximum transition duration. Signal processing was done offline, using a program written in C. This program provides facility to select inter-aural switching period, duty cycle, and logarithmic or linear amplitude variation, and transition duration.

The process of splitting was verified by obtaining the spectrograms of the processed outputs using a spectrographic analysis set-up (as described in Appendix A). Figures 4.4, 4.5 and 4.6 show the wideband (Δf = 300 Hz) spectrograms of a swept sine wave, random white noise, and speech syllable /*asa*/ respectively for the scheme of temporal splitting with step fading. Each figure shows spectrograms of unprocessed signal and the two processed signals having complementary symmetry (processed with 70% duty cycle and inter-aural switching duration of 20 ms).

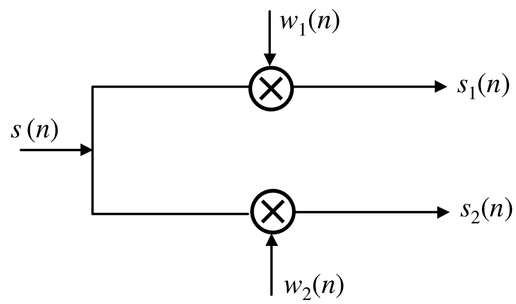


FIG.4.1. Temporal splitting of the signal for dichotic presentation.

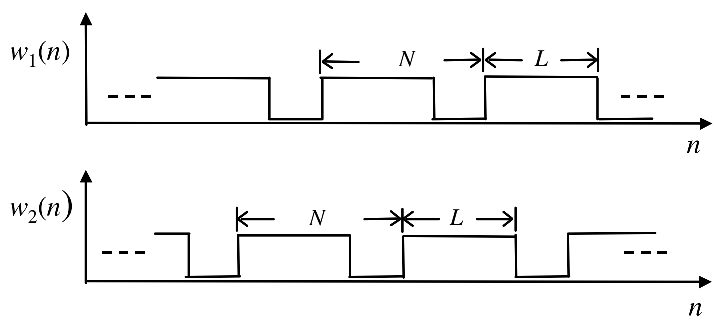


FIG. 4.2. Inter-aural fading with step transition and inter-aural overlap.

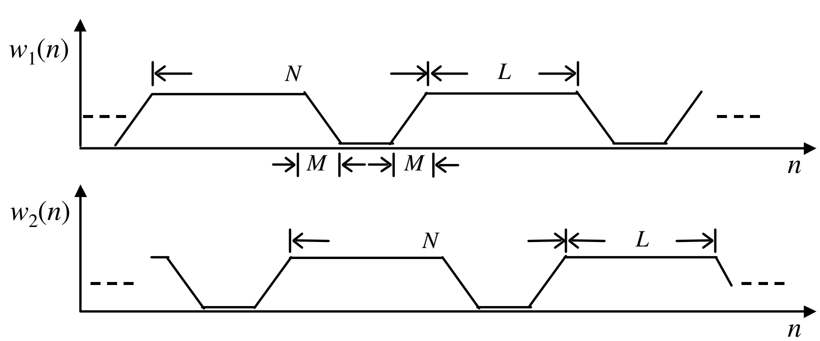


FIG. 4.3. Inter-aural fading with trapezoidal transition and inter-aural overlap.

Figures 4.7, 4.8 and 4.9 show the wideband (Δf = 300 Hz) spectrograms for a swept sine wave, random white noise, and speech syllable /*asa*/ respectively for the scheme of temporal splitting with trapezoidal transition. Each figure shows spectrograms of unprocessed signal and two processed signals having complementary symmetry (processed with inter-aural switching period of 20 ms, 70% duty cycle, and trapezoidal transition duration of 2 ms).

4.3 Listening tests

Listening tests were conducted to evaluate the effectiveness of the processing schemes. Tests were conducted on normal hearing subjects with simulated sensorineural hearing loss. In these tests, subjects were asked to identify 12 English consonants in VCV context presented over the headphones. Listening tests involved (a) diotic presentation of unprocessed speech and (b) dichotic presentation of processed speech. In the following subsections, description about the test stimuli, subjects participated in the test, and experimental procedure are given.

4.3.1 Test material and subjects

Twelve English consonants /p, b, t, d, k, g, m, n, s, z, f, v/ were used in the VCV context with vowel /*a*/ as in "father". Nonsense syllables were used to minimize the contribution of linguistic factors. These stimuli were acquired at a sampling rate of 10 k samples/s, using a microphone, a preamplifier, an input attenuater, a weighting filter, buffer amplifier (microphone B&K 4176 along with amplifier/attenuater of sound level meter B&K 2235), an amplifier, anti-aliasing filter with lowpass cut-off frequency of 4.6 kHz, and a 16-bit ADC of TI/TMS320C25 DSP board interfaced to a PC (details of hardware are given in Appendix E). The acquired signal waveform can be outputted through the DAC of the DSP board for hearing. The set-up can also be used for spectrographic analysis of the waveforms (as described in Appendix A). Figure 4.10 shows the block diagram of the signal acquisition system. During signal acquisition, each test stimulus (VCV syllable) was spoken a number of times by a female speaker (author). These stimuli were analyzed with spectrographic analyzer and were heard for testing the clarity. Properly recorded stimuli were selected and used in the processing.

In the scheme of temporal splitting with step transition (TS_ST), three normal hearing subjects (AB: M 28, BT: M 20, SR: F 32) participated in the listening tests. In the scheme of temporal splitting with trapezoidal transition (TS_TR), five normal hearing subjects (AB: M 28, BT: M 20, AC: F 37, JK: M 25, LT: F 24) participated in the listening tests. All the subjects had pure-tone hearing thresholds less than 20 dB HL in the frequency range of 125 Hz to 6 kHz.

Simulation of sensorineural loss in normal hearing persons was done by adding broadband noise to the speech stimuli. The noise used was broadband Gaussian noise (from function generator HP 33120A). Noise addition to the speech signal was done maintaining the SNR constant on the basis of short-time (≈ 10 ms) energy of the speech signal (Appendix D). Thus during silence segments, there will be no background noise.

In the scheme of temporal splitting with step transition, test material was added with broad band noise at SNRs ∞, 6, 3, 0, and -3 dB to simulate sensorineural hearing loss of varying degrees in normal hearing subjects. In the scheme of temporal splitting with trapezoidal transition, SNRs of ∞, 6, 3, 0, -3 and -6 dB were used.

4.3.2 Experimental set-up and procedure

Figure 4.11 shows the experimental set-up used for conducting the listening tests. The computerized test administration system consists of a PC interfaced, through RS232C serial port, to the subject terminal (VT-220) which is placed in an acoustically isolated chamber. Processed and unprocessed stimuli were outputted at a sampling rate of 10 k samples/s through the two 12-bit DAC channels of PC based data acquisition card (PCL-208 from Dynalog Micro Systems, Mumbai). The signals, after passing through a pair of smoothing low pass filters with cut-off frequency 4.6 kHz and a pair of audio amplifiers, were given to the two calibrated headphones Telephonics TDH-39P (calibration of headphones is described in Appendix E). Stimuli were presented at most comfortable listening level for the individual subjects and were in the range 70–80 dB SPL. The presentation level was kept constant for a particular subject throughout the test, as the loudness affects the clarity of the speech. PC was used for controlling the entire test and the subject terminal was used for displaying the

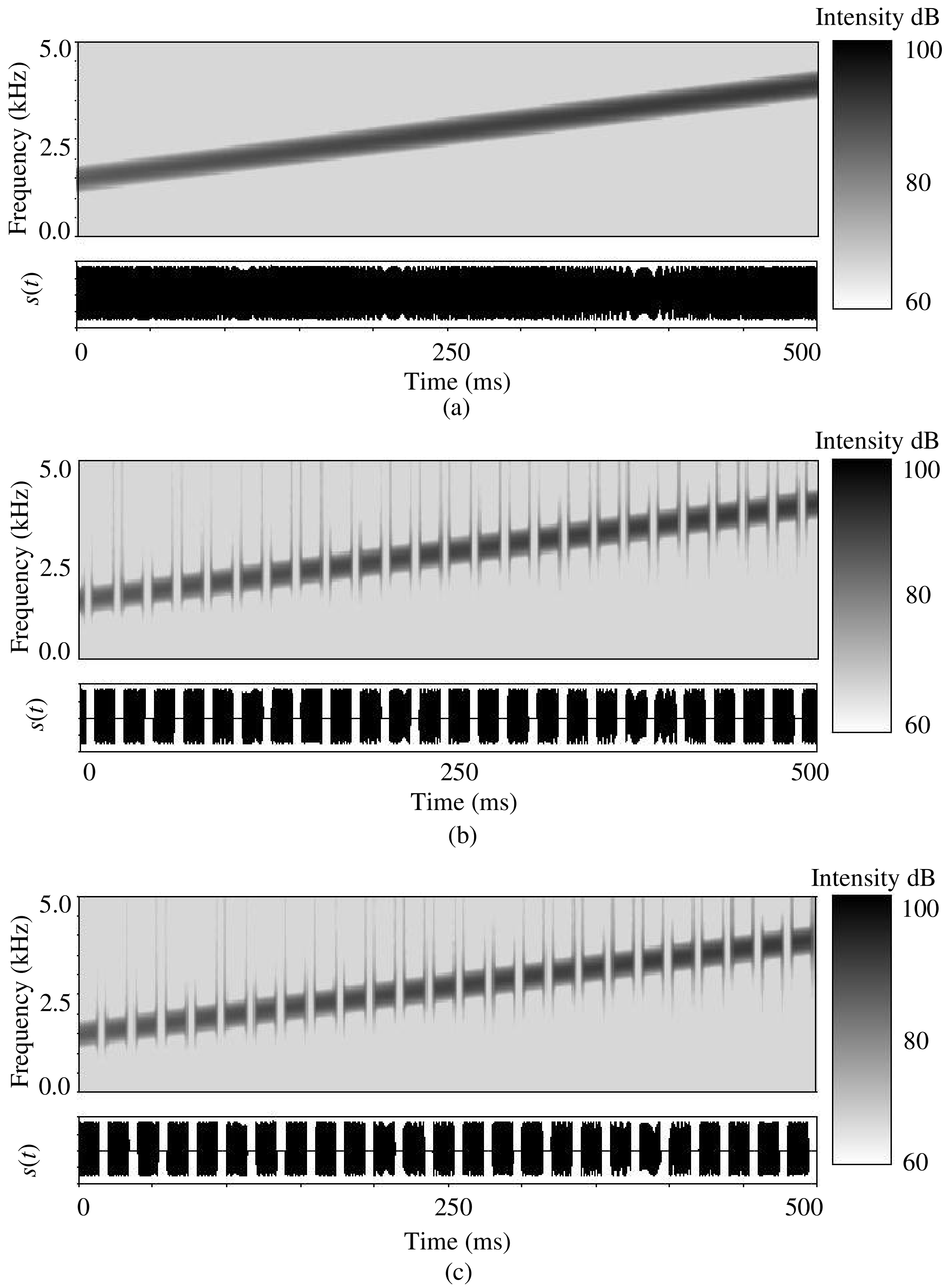


FIG. 4.4. Inter-aural switching with step transition (TS_ST): Wide-band spectrograms ($\Delta f \approx 300$ Hz) of swept sine wave. Inter-aural switching interval=20 ms and duty cycle=70%. (a) unprocessed (b) processed, left ear (c) processed, right ear.

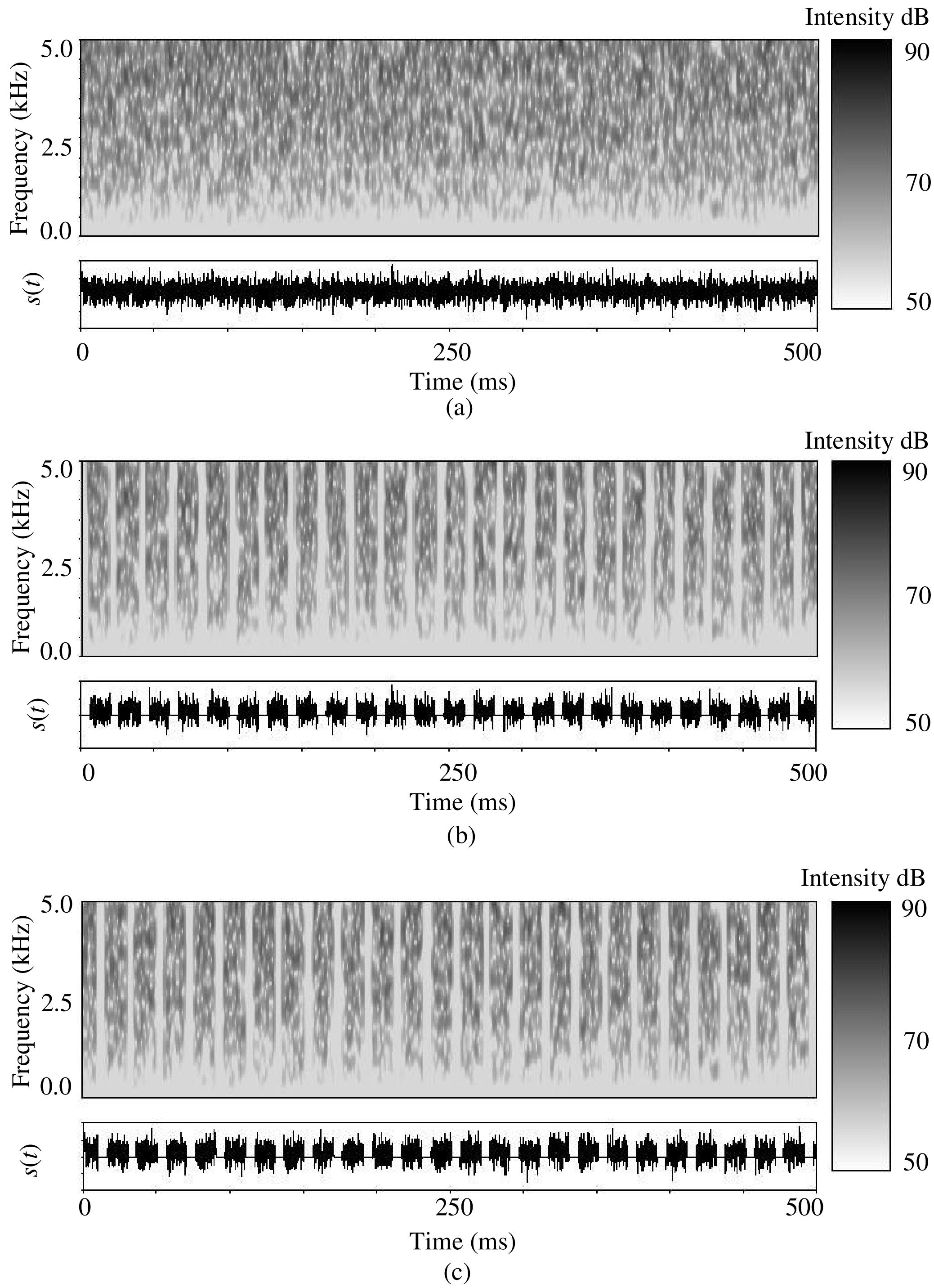


FIG. 4.5. Inter-aural switching with step transition (TS_ST): Wide-band spectrograms ($\Delta f \approx 300$ Hz) of random noise. Inter-aural switching interval=20 ms and duty cycle=70%. (a) unprocessed (b) processed, left ear (c) processed, right ear.

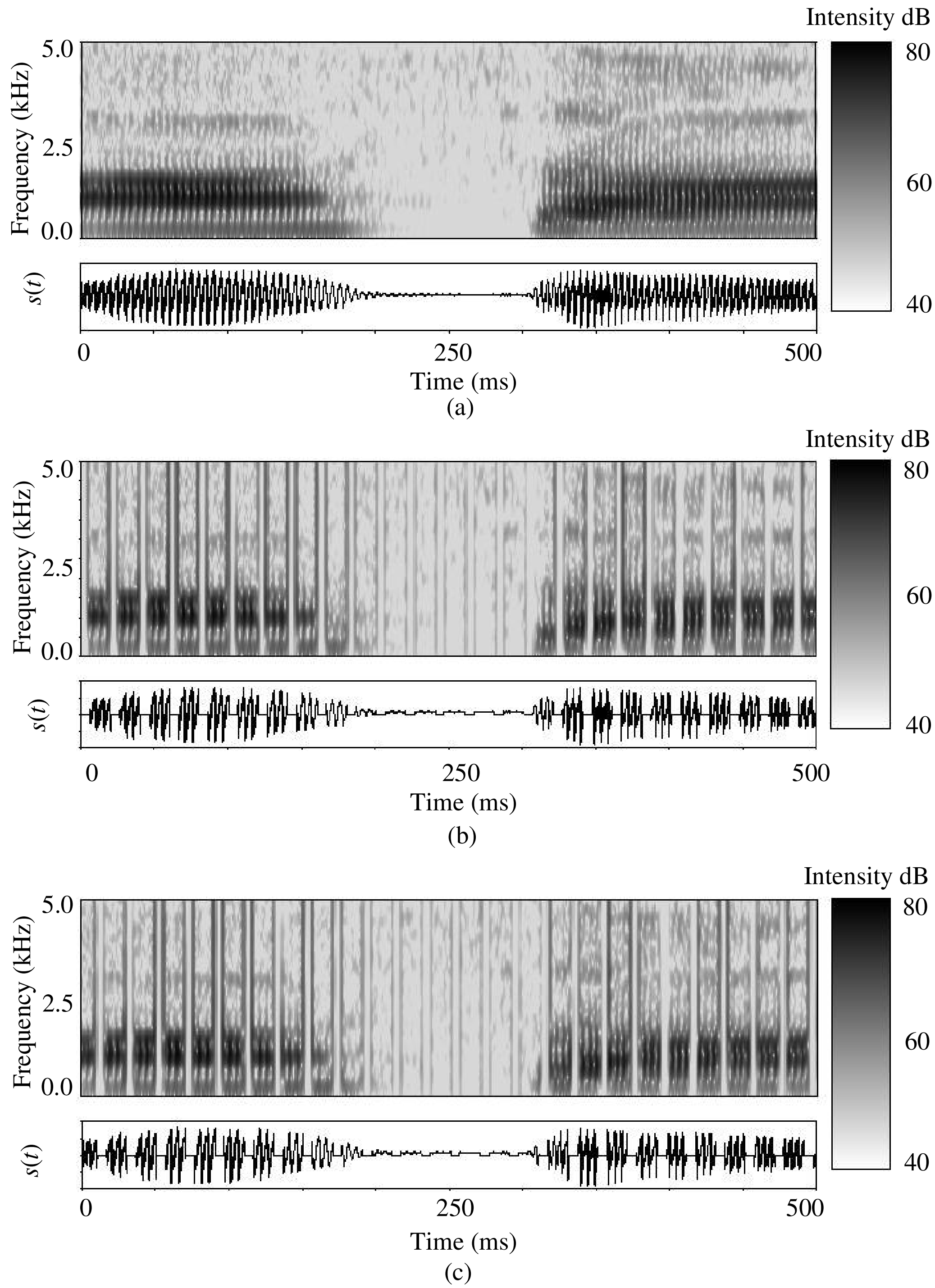


FIG. 4.6. Inter-aural switching with step transition (TS_ST): Wide-band spectrograms ($\Delta f \approx 300$ Hz) of speech syllable /*asa*/. Inter-aural switching interval=20 ms and duty cycle=70%. (a) unprocessed (b) processed, left ear (c) processed, right ear.

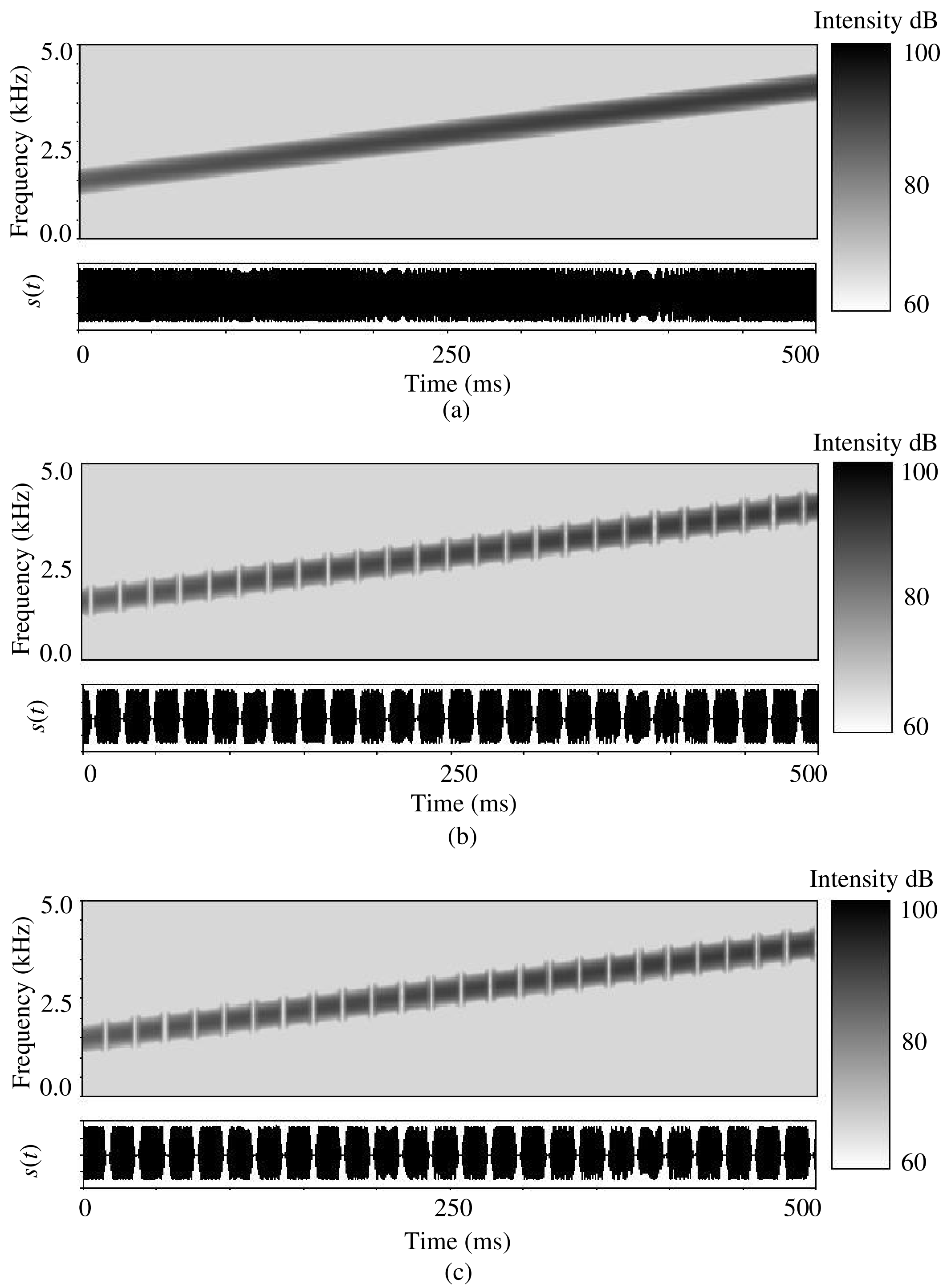


FIG. 4.7. Inter-aural switching with trapezoidal transition (TS_TR): Wide-band spectrograms ($\Delta f \approx 300$ Hz) of swept sine wave. Inter-aural switching interval=20 ms, duty cycle=70% and transition duration=2 ms. (a) unprocessed (b) processed, left ear (c) processed, right ear.

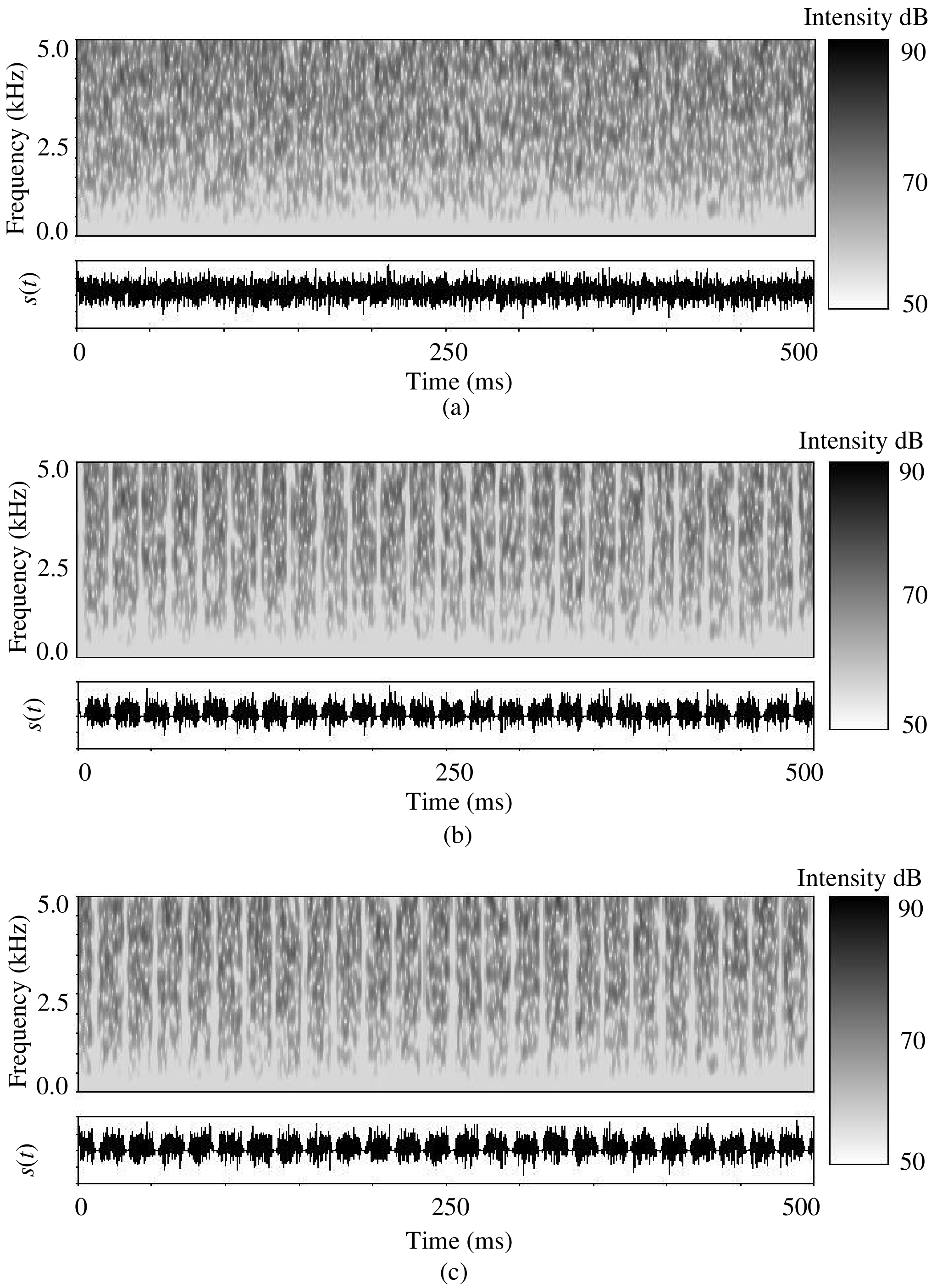


FIG. 4.8. Inter-aural switching with trapezoidal transition (TS_TR): Wide-band spectrograms ($\Delta f \approx 300$ Hz) of random noise. Inter-aural switching interval=20 ms, duty cycle=70% and transition duration=2 ms. (a) unprocessed (b) processed, left ear (c) processed, right ear.

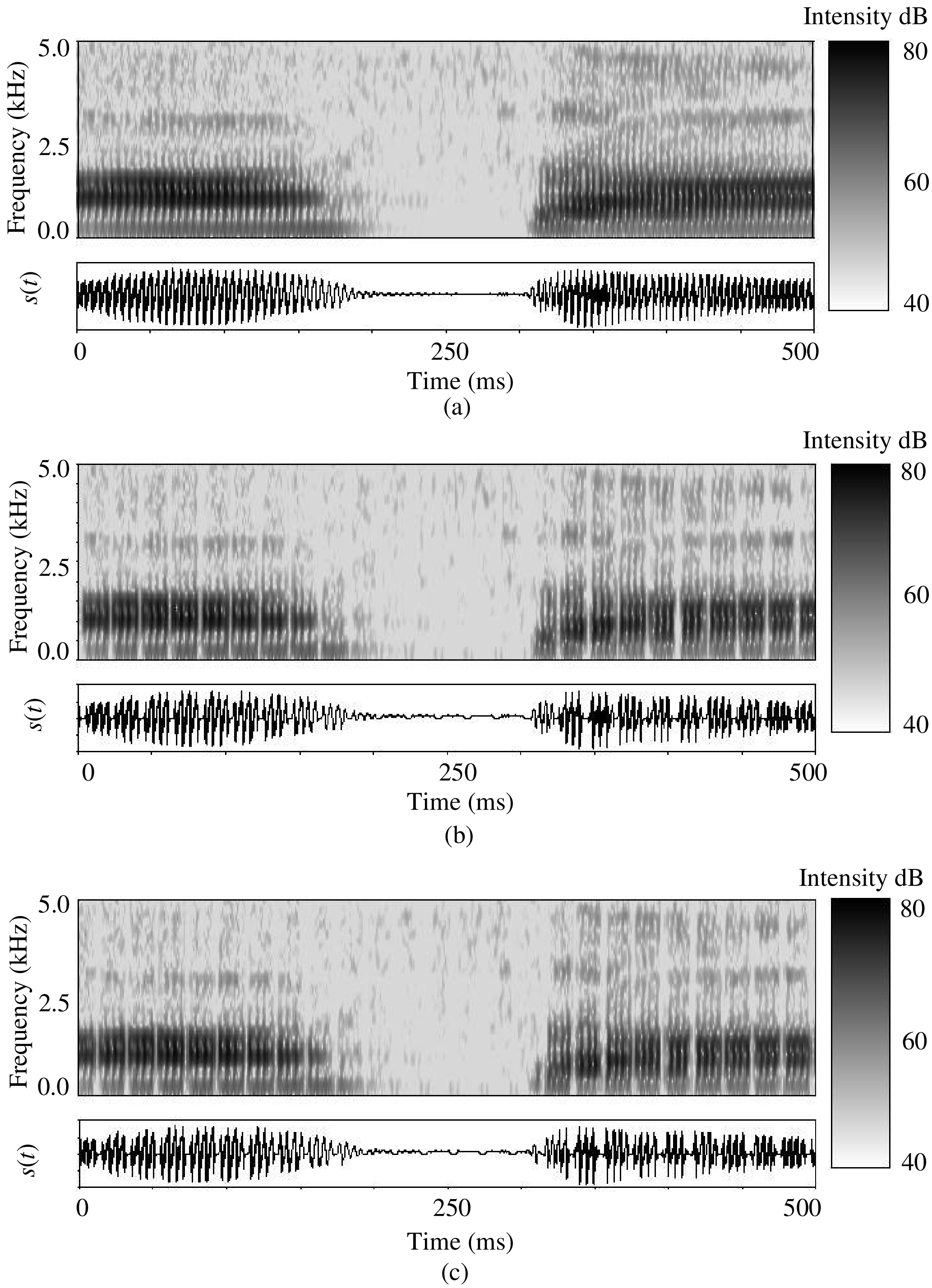


FIG. 4.9. Inter-aural switching with trapezoidal transition (TS_TR): Wide-band spectrograms ($\Delta f \approx 300$ Hz) of speech syllable /*asa*/. Inter-aural switching interval=20 ms, duty cycle=70% and transition duration=2 ms. (a) unprocessed (b) processed, left ear (c) processed, right ear.

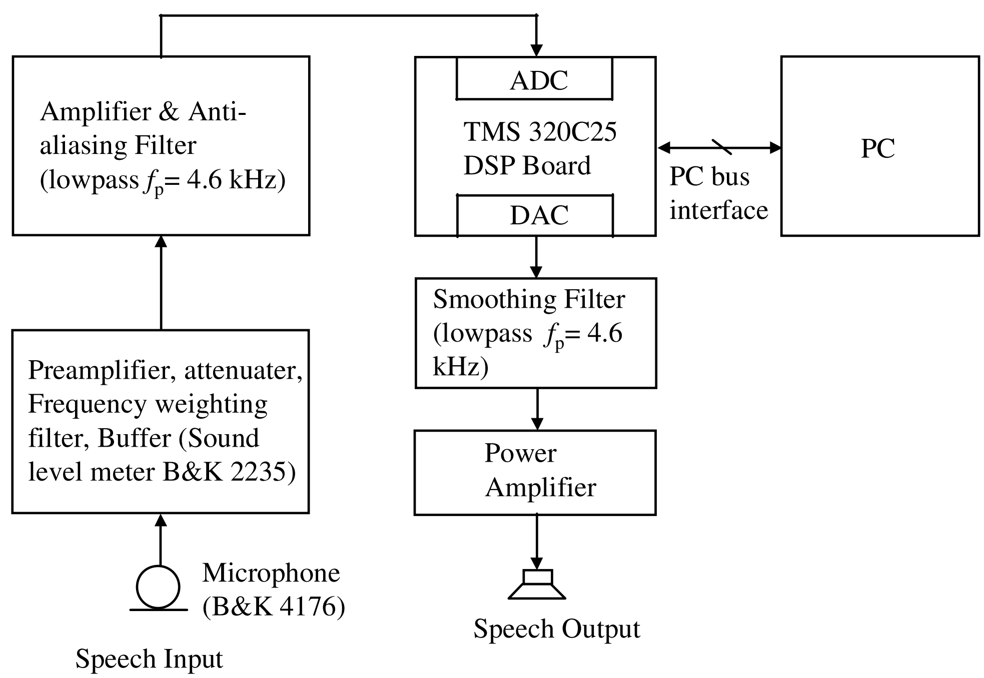


FIG. 4.10. Experimental set-up for speech signal acquisition and analysis.

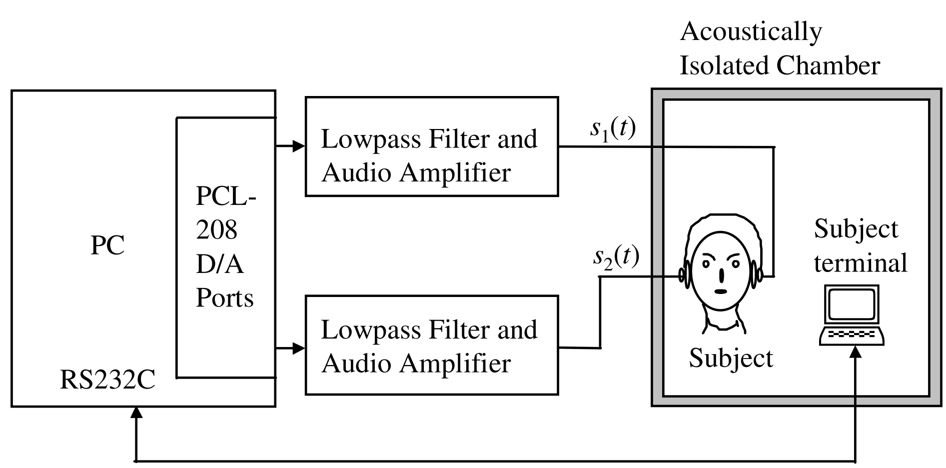


FIG. 4.11. Listening test set-up.

response choices and obtaining the subject's responses from its keyboard. Software and hardware for experimental control are described in Appendix D and E respectively.

Tests were conducted in an acoustically isolated chamber. Procedure of listening test (described in Appendix D) was explained to the subject. At the beginning of each test, subject could listen to the test material as many number of times as he/she wants, to become familiar with the stimuli. In a test, twelve stimulus items were presented, each 5 times, in random order (with certain uniformity constraints as given in Appendix D), leading to a total number of 60 presentations. Randomization was also provided across the tests. In a presentation, when a stimulus item was presented over the headphones, all stimuli were displayed on the subject terminal screen. Each stimulus corresponded to a key on the subject terminal keyboard. Subject responded by pressing appropriate key on the subject terminal keyboard. For each presentation, positions of response choices on the subject terminal screen were also randomized. In a test, as the presentation of the following stimulus proceeded, the response key pressed was compared with the correct response and the correct response scores were updated and were displayed on the operator screen along with stimulus and response for each presentation. For each presentation, time taken by the subject to respond was also recorded. At the end of each test, a stimulus-response confusion matrix was formed in which stimuli were represented along rows and responses were represented along columns. Each entry in the cell represents the frequency of occurrence of a stimulus-response pair. The diagonal elements give the correct responses whereas off-diagonal elements represent errors. Sum of the diagonal elements gives total number of correct responses. Percentage correct recognition score and response time statistics were also displayed along with the confusion matrix on the operator screen.

For each experimental condition, tests were repeated till the scores got stabilized. Tests were taken at least once with feedback of the correct response followed by tests without feedback. In between the tests without feedback, tests were conducted with feedback if required by the subject. Only the tests without feedback were considered for analysis. A test took about 5–8 minutes.

In the scheme of temporal splitting with step transition (TS_ST), there were a total of 20 test conditions (4 duty cycles $\times$ 5 SNR). For these 20 test conditions, each subject took about 20–25 hours for completion of the tests. Test sessions for the three subjects were spread over two months depending upon the availability of the subjects.

In the scheme of temporal splitting with trapezoidal transition (TS_TR), there were a total of 30 test conditions (5 processing conditions × 6 SNR). For these 30 test conditions, each subject took about 30 hours for completion of all the tests. Test sessions for the five subjects were spread over a span of about three months depending upon the availability of the subjects.

Listening tests with normal hearing subjects for evaluation of temporal splitting with step transition (varying overlap interval) and trapezoidal transition (fixed overlap, varying slope) will be referred to as “Experiment I” and “Experiment II” in all tables and figures.

4.4 Test results with step transition (TS_ST)

At the end of each test, responses of the subjects were stored in the form of stimulus-response confusion matrix. For each experimental condition, five tests with stabilized scores were considered for analysis of the confusion matrices and response time statistics. Since each test involved 5 presentations of each stimulus, overall analysis for each experimental condition involved 25 presentations of each stimulus.

Compilation of subject’s qualitative assessment for unprocessed and processed test stimuli, under various test conditions was carried out. Response times were used to compare the load on perception. Information transmission analysis (described briefly earlier in section 3.4, and later in Appendix B) that gives a measure, which is independent of subject’s biasing for response, was performed on the confusion matrix. 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 improvement in response times and recognition scores, information transmission analysis results for unprocessed and processed signals for each subject and averaged across the subjects are discussed in the following subsections. To determine the significance of the processing, t-tests (Snedector and Cochran, 1980) were carried out.

4.4.1 Quality assessment

Under no-noise condition subjects indicated poorer quality for the speech processed with all the processing conditions. However, with noise, subjects indicated that the quality of the processed speech was better than that of the unprocessed speech, with better quality for the duty cycle of 70%.

4.4.2 Response times

Table 4.1 gives the response times (average of five tests) for unprocessed and processed speech for the three subjects AB, BT, and SR. The table also provides relative decrease in response time (processed vs unprocessed) and significance level p from one-tailed t-test. For unprocessed speech, under no-noise condition, the response time averaged across the subjects was 2.28 s. It increased by 18, 25, 33, and 43% for 6, 3, 0, -3 dB SNR conditions respectively. This increase in response time is an indication of load on perception process due to low SNRs.

In the presence of masking noise, the response times are generally smaller in case of processed speech. Figure 4.12 shows the average of response times of the three subjects for unprocessed and processed speech under all SNR conditions for all duty cycles and Fig. 4.13 shows the percentage relative decrease in response times. For processed speech with 70% duty cycle, averaged relative decrease in response time is 1.4, 8.9, 6.8, and 6.6% for 6, 3, 0, -3 dB SNR conditions respectively. Relative decreases in response time were higher for 70% duty cycle compared to the duty cycles of 75% and 60%.

4.4.3 Recognition scores

The recognition scores for unprocessed and processed speech and percentage relative improvement for processed speech for the three subjects are given in Table 4.2. Figure 4.14 shows percentage recognition scores for unprocessed and processed speech and Fig. 4.15 shows relative improvement (%) for processed speech.

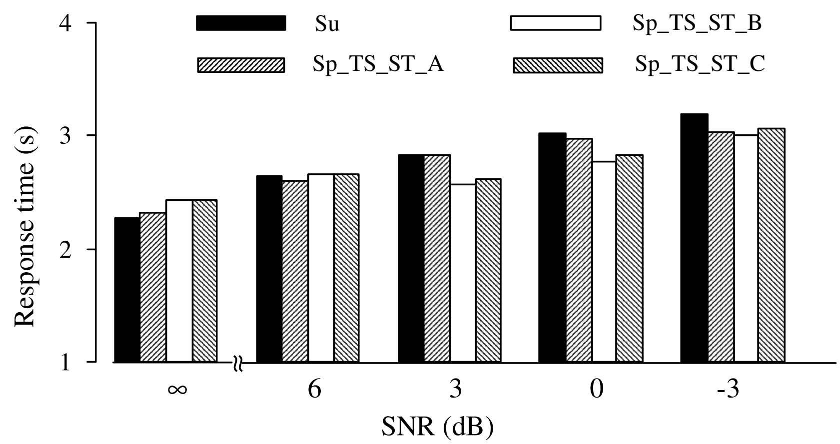


FIG. 4.12. Experiment I. (TS_ST). Averaged response times (s). Su: Unprocessed speech. Sp_TS_ST_A, Sp_TS_ST_B, Sp_TS_ST_C correspond to processed speech with 75, 70, and 60% duty cycles respectively.

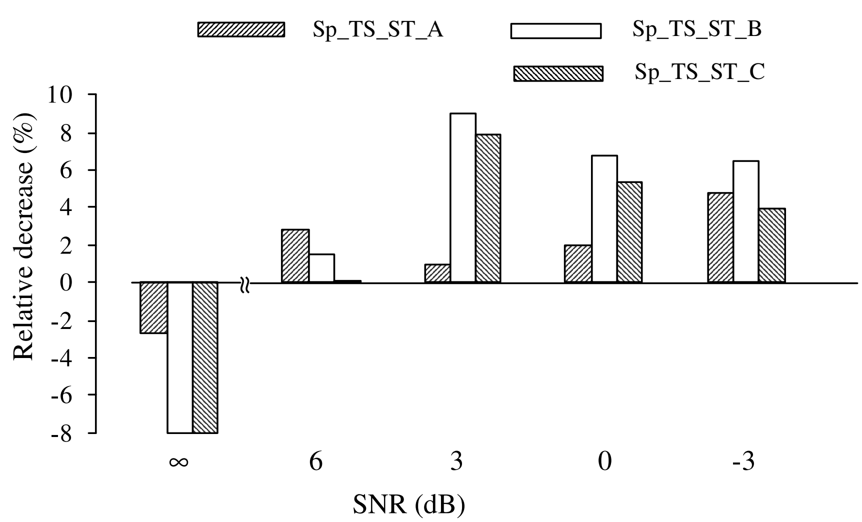


FIG. 4.13. Experiment I. (TS_ST). Averaged of relative decrease (%) in response times. Sp_TS_ST_A, Sp_TS_ST_B, Sp_TS_ST_C correspond to processed speech with 75, 70, and 60% duty cycles respectively.

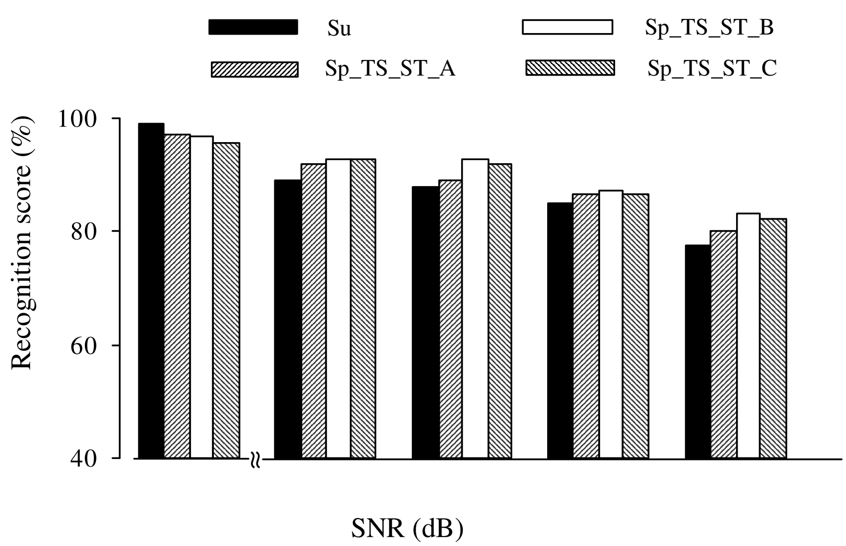


FIG. 4.14. Experiment I. (TS_ST). Averaged recognition scores (%). Su: Unprocessed speech. Sp_TS_ST_A, Sp_TS_ST_B, Sp_TS_ST_C correspond to processed speech with 75, 70, and 60% duty cycles respectively.

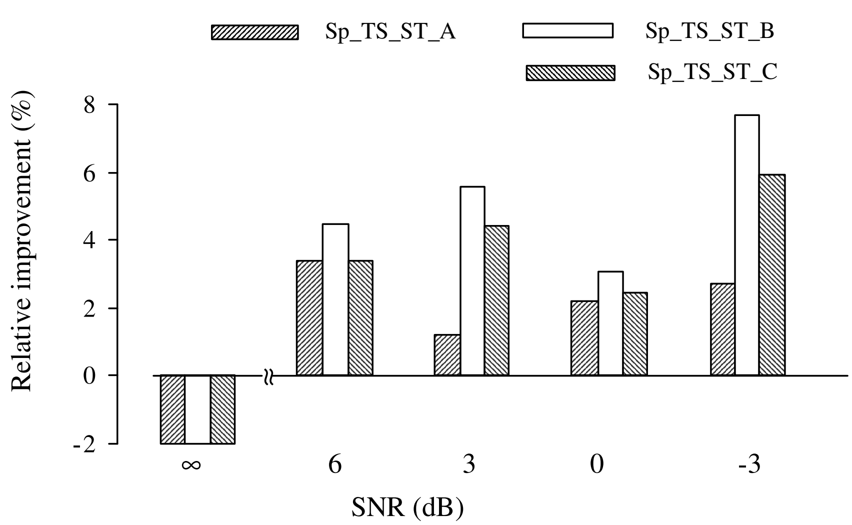


FIG. 4.15. Experiment I. (TS_ST). Averaged relative improvements (%) in recognition scores. Sp_TS_ST_A, Sp_TS_ST_B, Sp_TS_ST_C correspond to processed speech with 75, 70, and 60% duty cycles respectively.

For unprocessed speech, the scores are perfect under no-noise and decrease as the level of noise increases. The averaged score decreases from 99% for no-noise to 78% for -3 dB SNR. For processed stimuli, under no-noise condition the scores decrease with decrease in duty cycle. This indicates that, under no-noise condition, temporal splitting degrades speech perception to a small extent. However, when noise is present, the scores for the processed stimuli were higher. All the three subjects have shown improvement in score with processed speech at all the three duty cycles; scores were generally highest for 70% duty cycle. For higher levels of noise, there is more improvement in recognition score. For 70% duty cycle, the percentage relative improvement in recognition score varies from 3.7–5.5, 3.1–10.6, -1–7.0 and 1.2–12.0 for SNRs of 6, 3, 0, -3 dB respectively. Percentage relative improvement averaged across the three subjects for 70% duty cycle were 4.5, 5.6, 3.1, and 7.7 for 6, 3, 0, and -3 dB SNRs respectively.

The recognition scores of the individual subjects for different SNR conditions were subjected to t-test (one-tailed) to obtain statistical significance of the processing. As seen from Table 4.2, at -3 dB SNR, for 70% duty cycle, subject BT has shown statistically significant improvement ($p=0.04$). Subject SR has shown statistically significant improvement ($p=0.05$) at 60% duty cycle.

4.4.4 Information transmission analysis

For each test condition, combined confusion matrices for five tests (with stabilized scores) were subjected to information transmission analysis. Relative information transmitted for unprocessed and processed speech for all the subjects are given in Table 4.3. There was degradation in overall information transmission with increase in noise level for unprocessed speech. The overall information transmission, averaged across the subjects, for unprocessed speech decreases from 99% under no-noise condition to 83% at -3 dB SNR condition. Figure 4.16 shows information transmitted for all SNR conditions. With processed speech, average of relative improvement in overall information transmitted for 6, 3, 0, and -3 SNR conditions were 2.1, 3.7, 1.6, and 3.4% respectively for 70% duty cycle. For duty cycles 75% and 60%, relative improvements are lower than these values.

For a look into the contributions by various consonantal features in perception, information transmission analysis was carried out for grouping of consonants by features of voicing, place, manner, nasality, frication, and duration. These results are also given in Table 4.3.

Voicing: Relative information transmission is very high for all the subjects under all the conditions. For unprocessed speech, averaged across the subjects, relative information transmitted decreases from 100% under no-noise condition to 95.3% at -3 dB SNR condition. At -3 dB SNR condition, processing with 70% duty cycle has shown a relative improvement of 4% averaged across the three subjects, i.e. restoring it back to near-perfect transmission.

Place: Figure 4.17 shows the percentage relative information transmitted (averaged across the subjects). For unprocessed speech, relative information transmitted averaged across the subjects decreases from 97.3% under no-noise condition to 41% at -3 dB SNR condition. Processing brings improvement in this feature for all SNR conditions, with maximum improvement for 70% duty cycle. Relative improvements averaged across the subjects are 11.0%, 19.9%, 21.9% and 31.4% for 6, 3, 0, -3 dB SNR conditions respectively.

Manner: Relative information transmitted decreases from 100% under no-noise condition to 69.3% at -3 dB condition, for unprocessed speech. The processing improves the information of this feature at 6 dB and -3 dB SNR conditions and the relative improvements are 3.3% and 4.6% respectively for 70% duty cycle.

Nasality: Relative information transmission is perfect for unprocessed and processed speech for all subjects under all SNR conditions and duty cycles.

Frication: For unprocessed speech, relative information transmitted averaged across the subjects, varies from 100% under no-noise condition to 48.7% at -3 dB SNR condition. There is improvement in the transmission of this feature at 6 and -3 dB SNR conditions and the relative improvement for processing with 70% duty cycle are 6.3% and 18.8% respectively.

Duration: Figure 4.18 show the percentage relative information transmitted (averaged across the three subjects) for this feature. Relative information transmitted for unprocessed speech for decreases from 96.3% under no-noise condition to 43.3% at -3 dB SNR condition. With processing, relative information transmitted for this feature improved at all the SNR conditions, improvement is always higher for 70% duty cycle. Averaged across the three subjects, relative improvements are 12.5, 18.3, 28.8, 21.5% for 6, 3, 0, -3 dB SNR conditions respectively.

From the above analysis, it is seen that, with decrease in SNR, the information transmitted for duration and place features for unprocessed speech gets degraded most and processing helps in restoring the reception of these features. Improvements are generally highest for 70% duty cycle.

4.4.5 Summary of the results

To evaluate scheme of temporal splitting with step transition along with overlap (duty cycle > 0.5), listening tests were conducted on three normal hearing subjects with simulated sensorineural hearing impairment. Effect of processing with different overlap periods was studied under different SNR conditions. Response times, consonant recognition score, and relative information transmitted for overall and for different speech features for unprocessed and processed speech under all test conditions were analyzed.

From the qualitative assessment of unprocessed and processed speech, it was found that under adverse listening condition of low SNR, subjects indicated better quality for processed speech than for unprocessed speech.

From the response times it was seen that, for unprocessed speech, response times increased with decrease in SNR indicating that decrease in SNR results in increased load on perception. For processed stimuli, response times generally decreased. Hence the scheme of temporal splitting helped in reducing the load on perception for normal hearing persons under adverse listening conditions. Improvements were highest at 70% duty cycle.

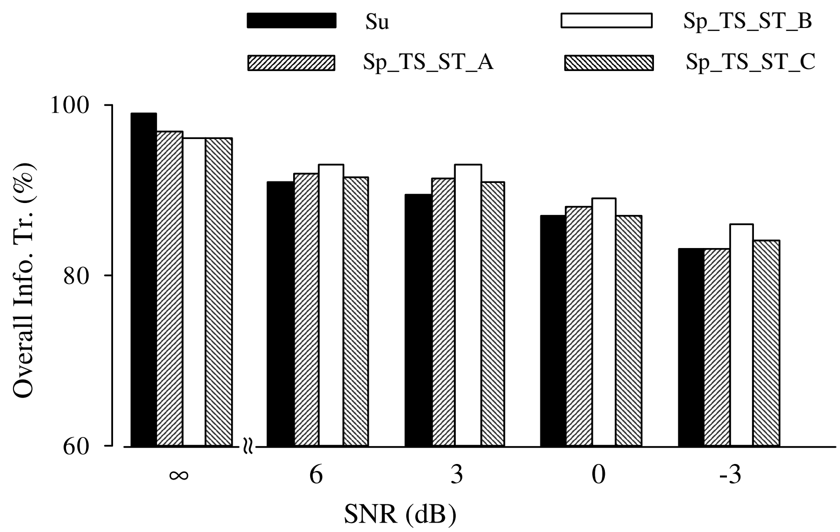


FIG. 4.16. Experiment I. (TS_ST). Averaged overall relative information transmitted (%). Su: Unprocessed speech. Sp_TS_ST_A, Sp_TS_ST_B, Sp_TS_ST_C correspond to processed speech with 75, 70, and 60% duty cycles respectively.

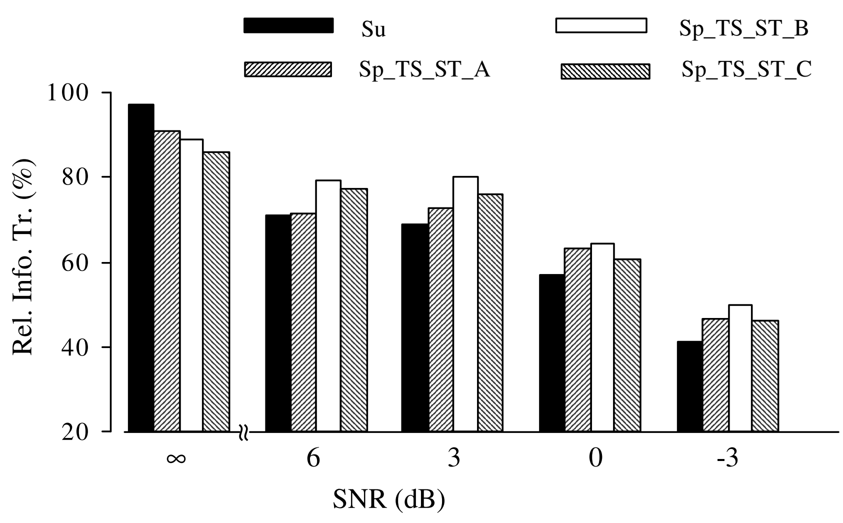


FIG. 4.17. Experiment I. (TS_ST). Averaged relative information transmitted (%) for place feature. Su: Unprocessed speech. Sp_TS_ST_A, Sp_TS_ST_B, Sp_TS_ST_C correspond to processed speech with 75, 70, and 60% duty cycles respectively.

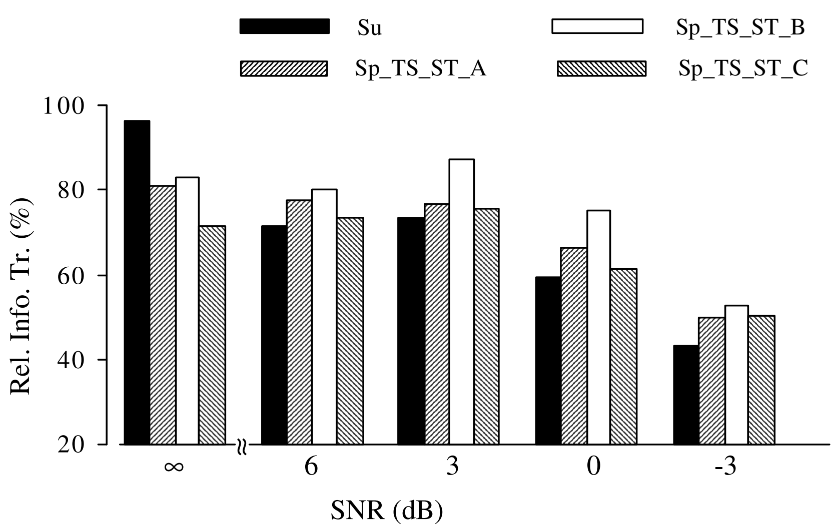


FIG. 4.18. Experiment I. (TS_ST). Averaged relative information transmitted (%) for duration feature. Su: Unprocessed speech. Sp_TS_ST_A, Sp_TS_ST_B, Sp_TS_ST_C correspond to processed speech with 75, 70, and 60% duty cycles respectively.

Consonant recognition scores for unprocessed speech decreased with increase in noise level, but these improved with processing. Processing with duty cycle of 70% has resulted in higher recognition score compared to the duty cycles of 75 and 60%.

From the analysis of information transmission, it can be observed that, overall information transmitted decreases with decrease in SNR for unprocessed speech and improves for the processed speech. With decrease in SNR, the perception of the duration feature in unprocessed speech gets degraded most. Improvement in its reception by inter-aural switching indicates that processing has reduced the effect of temporal masking.

In these results, it is seen that the duty cycle of 70% gave maximum improvement. Hence the scheme of temporal splitting with inter-aural switching along with overlap helped in improving the speech perception for normal hearing persons under adverse listening conditions by reducing the effect of temporal masking. Use of trapezoidal variation during switching may further help in improving the perception by reducing the spectral distortion due to step transition during switching.

4.5 Test results with trapezoidal transition (TS_TR)

Results of the listening tests conducted on five subjects are presented here. For ascertaining the improvement in speech quality, a compilation of subject's qualitative assessment about the set of the test stimuli, under various listening conditions, was carried out. Test results were analyzed in the same way as described in the previous section for step transition. Response times, percentage correct recognition scores, relative improvement in response times and recognition scores, paired t-test results, information transmission analysis results for unprocessed and processed signals for each subject and averaged across the subjects are discussed in the following subsections.

4.5.1 Quality assessment

Compilation of subject's qualitative assessment under different test conditions indicated, that under no-noise condition, quality deteriorates for processed speech. With noise present, subjects indicated better speech quality for processed speech. Also, quality was better for speech processed with trapezoidal transition than with step transition.

4.5.2 Response times

Table 4.4 gives the response times for processed and unprocessed speech for all SNR conditions and transition durations. It also gives the paired t-test significance levels.

For unprocessed speech, compared to the ∞ SNR condition, the response times increase moderately with decrease in SNR indicating that, decrease in SNR increases the load on the perception. Paired t-test (one-tailed) was performed on the data to measure the statistical significance of the processed data. For unprocessed speech, under no-noise condition, the response time averaged across the five subjects was 1.83 s. It increased by 1.6, 3.3, 10.4, 6.6, 4.9% for 6, 3, 0, -3, -6 dB SNR conditions respectively. This increase is response time is an indication of load on perception process due to low SNRs.

In the presence of masking noise, the response times are generally smaller in case of processed speech. Figure 4.19 shows the response times (averaged across five subjects) for unprocessed and processed speech, and Fig. 4.20 shows relative decrease in response times. Averaged across the subjects, relative decrease in response times for step transitions are 3.9, 4.9, 8.3, and 0.14% for 6, 3, 0, and -3 dB SNR conditions respectively. For transition duration of 1 ms, relative decrease are 9.6, 12.4, 10.2, and 3.5% for 6, 3, 0, -3 dB SNR conditions. For 2 ms and 3 ms transition durations, corresponding values are 10.2, 11.3, 9.6, 1.5% and 7.5, 13.6, 13.6, 8.1% respectively. Relative decrease was statistically significant for 3 dB ($p < 0.02$) and 0 dB ($p < 0.03$) for all transition durations. At -6 dB SNR condition processing with 0, 1, and 2 ms resulted in an increase in response time. However, transition duration of 3 ms has resulted in decrease of 8.0% ($p < 0.05$). Compared to step transition, trapezoidal transition results in more improvement in response times, indicating that the trapezoidal transition is more effective in reducing load on perception process.

4.5.3 Recognition scores

Recognition scores for unprocessed and processed speech and percentage relative improvement for processed speech, under all SNR conditions are given in Table 4.5. Figure 4.22 shows percentage recognition scores averaged across the subjects for unprocessed and processed speech and Fig. 4.23 shows percentage relative improvement for processed speech. For unprocessed speech, recognition scores for all the subjects decrease as the SNR degrades. For subject BT, there is relatively small variation, while for subject LT, there is a very large variation. The averaged score decreases from 100% under no-noise to 81% for -6 dB SNR condition. Processing with dichotic presentation has resulted in improvement in the recognition scores, and the extent of improvement appears to be related to the subject's susceptibility to poor SNR. Improvements were there for all the transition durations. These were generally highest for 2 ms transition. Table 4.5 also gives one-tailed t-test significance levels for recognition scores. For processed speech with step transition, averaged across the five subjects, percentage relative improvement in recognition score were 1.9, 3.4, 2.4, 5.5, and 11.3 for 6, 3, 0, -3, and -6 dB SNR conditions respectively. For the same SNR conditions the percentage relative improvements were larger with trapezoidal variation. Improvements in recognition scores were statistically significant for higher levels of noise.

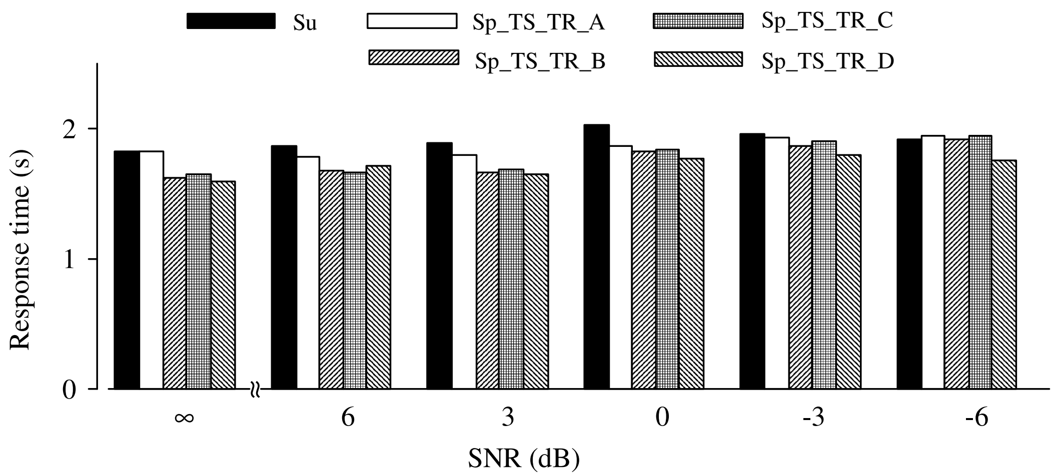


FIG. 4.19. Experiment II. (TS_TR). Averaged response times (s). Su: Unprocessed speech. Sp_TS_TR_A, Sp_TS_TR_B, Sp_TS_TR_C, Sp_TS_TR_D correspond to processed speech with 0, 1, 2, 3 ms transition durations respectively.

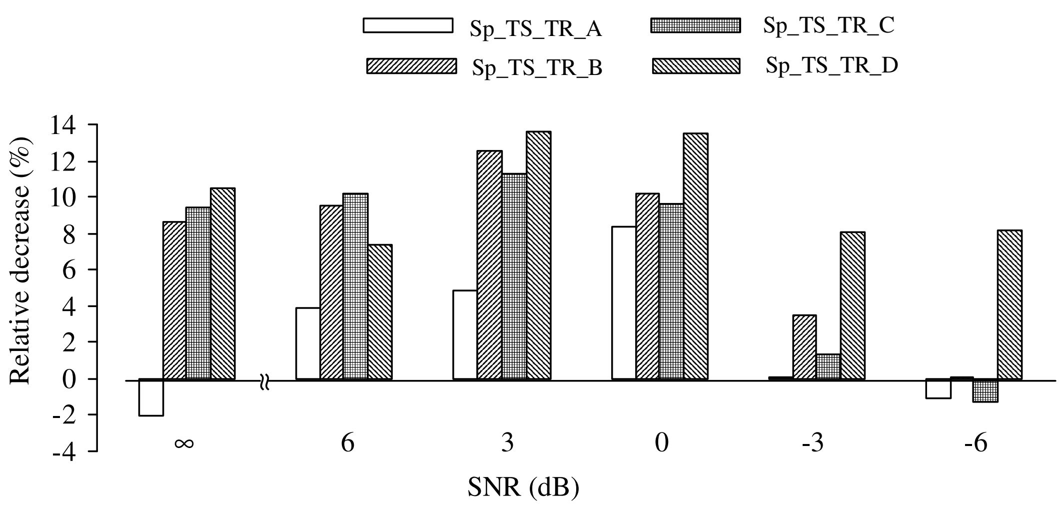


FIG. 4.20. Experiment II. (TS_TR). Averaged relative decrease (%) in response times. Sp_TS_TR_A, Sp_TS_TR_B, Sp_TS_TR_C, Sp_TS_TR_D correspond to processed speech with 0, 1, 2, 3 ms transition durations respectively.

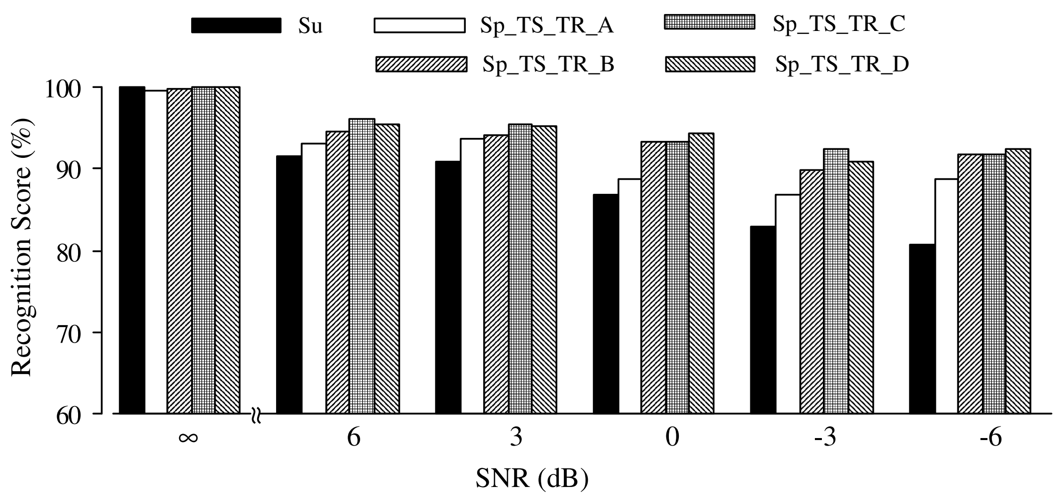


FIG. 4.21. Experiment II. (TS_TR). Averaged recognition scores (%). Su: Unprocessed speech. Sp_TS_TR_A, Sp_TS_TR_B, Sp_TS_TR_C, Sp_TS_TR_D correspond to processed speech with 0, 1, 2, 3 ms transition durations respectively.

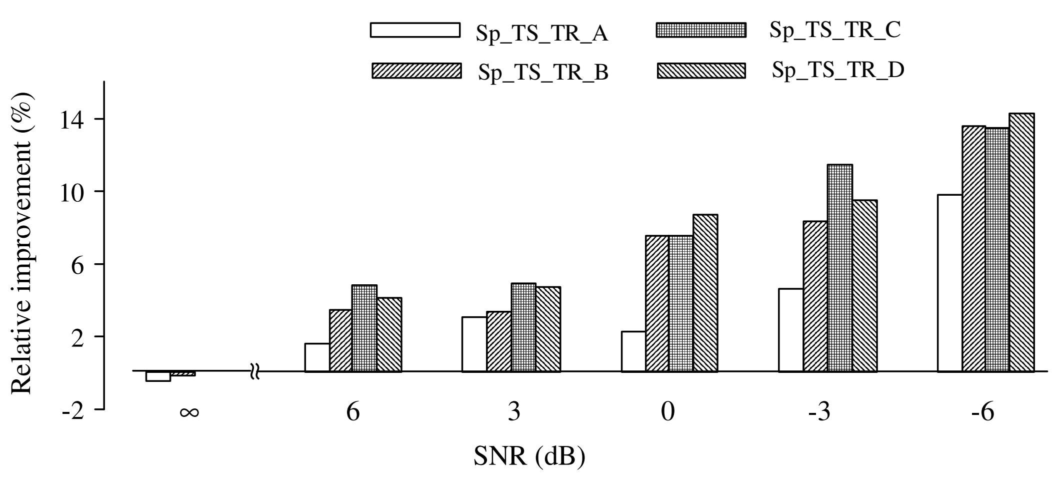


FIG. 4.22. Experiment II. (TS_TR). Averaged relative improvement (%) in recognition scores. Sp_TS_TR_A, Sp_TS_TR_B, Sp_TS_TR_C, Sp_TS_TR_D correspond to processed speech with 0, 1, 2, 3 ms transition durations respectively.

4.5.4 Information transmission analysis

Table 4.6 (a) gives the overall information transmitted for five subjects under all SNR conditions, for all processing conditions. It can be seen that, there is degradation in overall information transmission with increase in noise level for unprocessed speech. Figure 4.24 shows the percentage overall information transmitted, averaged across the subjects for unprocessed and processed signal for SNR condition. Relative improvements averaged across the subjects, with processed speech for 6, 3, 0, -3, -6 dB SNR conditions were 1.1, 3.3, 2.4, 2.8, and 4.7% respectively for step variation. Relative improvements averaged across the subjects, with trapezoidal variation for 1, 2, 3 ms transitions were 3.1, 4.0, 3.8 for 6 dB; 3.1, 4.1, 4.1 for 3 dB; 5.0, 5.3, 5.5 for 0 dB; 3.6, 6.9, 5.2 for -3 dB; 6.6, 6.6, 8.0 for -6 dB SNR conditions respectively. An interesting observation is that, for unprocessed speech, subject LT has a very low recognition score with low SNR. However, the relative information transmitted for this subject is not much lower than that for other subjects. This indicates that errors in reception by this subject are not randomly distributed. It can be seen that, dichotic presentation brings the relative information transmitted for this subject to almost the same level as for other subjects.

To obtain the contributions by various consonantal features in the perception, information transmission analysis was carried out for grouping of phonemes by features of voicing, place, manner, nasality, frication, and duration. Tables 4.6 (b), (c), (d), (e), (f), and (g) show the relative information transmitted for features of voicing, place, manner, nasality, frication, and duration respectively under all SNR conditions for all subjects. Averaged across the five subjects, percentage relative improvements in information transmission for all features are given in Table 4.7.

Voicing: For unprocessed speech, it can be observed that with decrease in SNR, reception of voicing is not affected much. Averaged across the subjects, relative improvement in information transmitted with processed speech for 0, -3, -6 dB SNR conditions were 2.8, 1.4, and 0.4% respectively for step variation. The improvements for trapezoidal transitions of 1, 2, 3 ms were 3.5, 3.5, 2.8% for 0 dB SNR; 1.6, 1.6, 1.6% for -3 dB SNR; 1.2, 0.4, 1.9% for -6 dB SNR conditions.

Place: Figure 4.25 shows the percentage relative information transmitted (averaged across the subjects) for place feature. Relative information transmitted averaged across the subjects changes from 100% to 54%, for no-noise condition to -6 dB SNR, for unprocessed speech. For 6, 3, 0, -3, -6 dB SNR conditions, for fading function with step variation, average relative improvements were 3.3, 9.2, 8.1, 20.2, 54.4%. For fading function with trapezoidal variation with 1, 2, 3 ms transition durations, the average relative improvements were 11.0, 13.6, 14.1% for 6 dB SNR; 10.0, 13.3, 13.2% for 3 dB SNR; 29.5, 29.5, 34.8% for 0 dB SNR; 29.4, 40.0, 36.2% for -3 dB SNR; 67.1, 66.5, 71.2% for -6 dB SNR.

Manner: For unprocessed speech, relative information transmission decreases from 100% at no-noise to 81% at -6 dB SNR. For this feature also, step transition does not result in much improvement. For transition duration with 1, 2, 3 ms, improvements were 0.4, 3.4, 2.1% for 6 dB SNR; -1.2, 1.9, 2.5% for 3 dB SNR; 7.1, 6.7, 5.2% for 0 dB SNR; 1.6, 5.8, 2.1% for -3 dB SNR; 8.5, 9.8, 10.2% for -6 dB SNR.

Nasality: Relative information transmission for nasality feature is perfect for unprocessed speech under all SNR conditions. It does not get affected by processing for dichotic presentation.

Frication: Relative information transmitted varies from 100% under no-noise condition to 69% at -6 dB SNR for unprocessed speech. With step transition, there is improvement in information transmission only at -6 dB SNR. For fading function with trapezoidal variation, improvements averaged across the subjects for 1, 2, 3 ms were 2.6, 7.4, 5.1% for 6 dB SNR; -1.5, 3.7, 5.2% for 3 dB SNR; 16.3, 14.5, 11.8% for 0 dB SNR; 1.4, 9.4, 2.4% for -3 dB SNR; 18.8, 22.7, 20.6% for -6 dB SNR.

Duration: Figure 4.26 shows the percentage relative information transmitted (averaged across the subjects) for duration feature. Averaged across the subjects, the information transmitted changes from 100% to 49%, for no-noise condition to -6 dB SNR for unprocessed speech. Average of relative improvement in information transmitted with processed speech for 6, 3, 0, -3, -6 dB SNR conditions were -0.6, 2.9, 17.5, 64.2, and 223.2% respectively for step variation. The relative improvements for trapezoidal transitions of 1, 2, 3 ms were 8.4, 9.7,

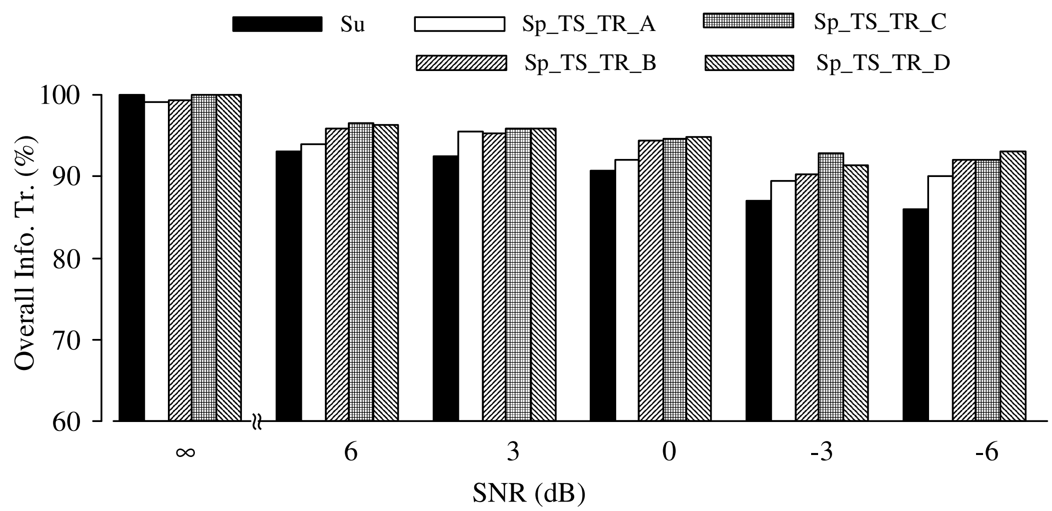


FIG. 4.23. Experiment II. (TS_TR). Averaged overall relative information transmitted (%). Su: Unprocessed speech. Sp_TS_TR_A, Sp_TS_TR_B, Sp_TS_TR_C, Sp_TS_TR_D correspond to processed speech with 0, 1, 2, 3 ms transition durations respectively.

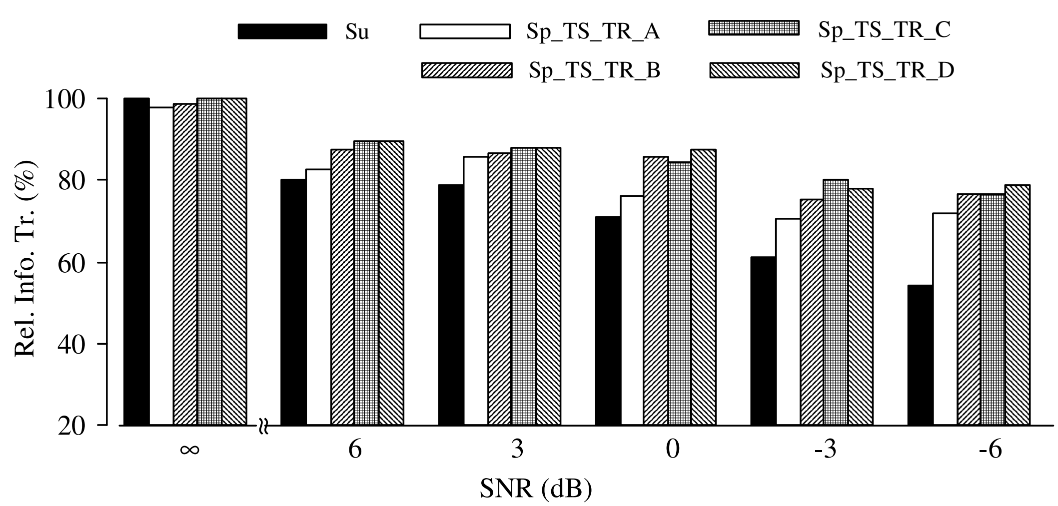


FIG. 4.24. Experiment II. (TS_TR). Averaged relative information transmitted (%) for place feature. Su: Unprocessed speech. Sp_TS_TR_A, Sp_TS_TR_B, Sp_TS_TR_C, Sp_TS_TR_D correspond to processed speech with 0, 1, 2, 3 ms transition durations respectively.

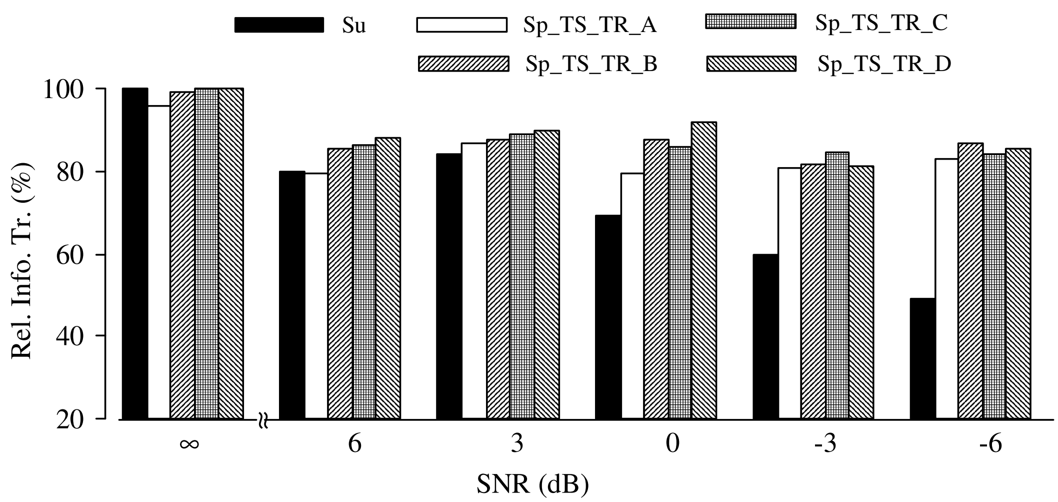


FIG. 4.25. Experiment II. (TS_TR). Averaged relative information transmitted (%) for duration feature. Su: Unprocessed speech. Sp_TS_TR_A, Sp_TS_TR_B, Sp_TS_TR_C, Sp_TS_TR_D correspond to processed speech with 0, 1, 2, 3 ms transition durations respectively.

12.4% for 6 dB SNR; 4.0, 6.2, 7.2% for 3 dB SNR; 40.3, 39.9, 55.2% for 0 dB SNR; 73.8, 70.5, 75.2% for -3 dB SNR; 258.5, 244.0, 235.5% for -6 dB SNR conditions.

It can be observed that, under the adverse listening condition of simulated loss, reception of place and duration was severely degraded. Subjects JK and LT had great difficulty in the reception of these two features under poor listening conditions. With processing, the reception of place and duration improved for all the subjects, and particularly for subjects JK and LT. The improvements for trapezoidal fading are higher than those for step fading.

It can be seen that, transition duration does not have much effect on the perception of duration feature, while longer transition duration generally results in higher improvement in the reception of place feature. It is to be noted that, the trapezoidal fading results in lesser amount of spectral distortion, and therefore it might be more helpful in place perception.

4.5.5 Summary of the results

To evaluate the scheme of temporal splitting with trapezoidal variation and overlap (duty cycle > 0.5) listening tests were conducted on five normal hearing subjects with simulated hearing loss. For each subject, in each test condition, recognition score, relative information transmitted for overall and for different consonantal features, and subject's response times for unprocessed and processed speech were analyzed.

From the response times it can be seen that under noise condition, there is decrease in response time due to processing. Decrease in response times is higher with trapezoidal transition than with step transition. Hence the scheme of temporal splitting with trapezoidal fading helps in further reducing the load on perception for normal hearing persons under adverse listening condition.

For unprocessed signal, recognition scores decreased with increase in noise level, but improved with processing. Scores were higher for trapezoidal variation than step variation. Generally the scores were higher for transition durations of 2 and 3 ms. For SNR of -6 dB, relative improvement in recognition scores of 11.3% and 15.6% were observed for step variation and trapezoidal variation. Hence the trapezoidal fading function helps to improve the speech intelligibility.

From the information transmission analysis, it was observed that there was maximum improvement for duration feature (relative improvement of 244% at -6 dB SNR) indicating that the processing helps in reducing the effect of temporal masking. For place feature, relative improvement was 71.2% at -6 dB SNR. Trapezoidal fading function with slower variation resulted in additional improvement to a small extent for duration feature, but to a large extent for place features at higher levels of noise.

Thus the temporal splitting with trapezoidal fading function along with overlap improves the speech intelligibility for normal hearing persons under adverse listening conditions.

4.6 Discussion

In this chapter, the implementation and evaluation of the schemes of temporal splitting using step and trapezoidal fading along with overlap to reduce the effect of temporal masking in speech perception is discussed. Evaluation was done by conducting listening tests on normal hearing subjects with simulated bilateral sensorineural hearing loss. Subject's response time, recognition score, relative information transmission for overall and various consonantal features were analyzed. Processing with both the schemes improved the response times, recognition scores, and transmission of consonantal features. Improvements are higher for the scheme of temporal splitting with trapezoidal transition.

Temporal splitting with both step and trapezoidal fading improved the perception of mainly the duration feature. This indicates that the temporal splitting reduces the effect of temporal masking. Temporal splitting with trapezoidal fading improved the perception of duration feature further. With trapezoidal fading there was improvement in the perception of place feature also, which can be attributed to reduction in the spectral distortions that occur in step transition.

TABLE 4.1. Experiment I. (TS_ST). Response times for Unprocessed Speech (Su) and for Processed Speech Sp_TS_ST_A, Sp_TS_ST_B, Sp_TS_ST_C corresponding to duty cycles 75, 70, and 60%. S: Subject, Tavg. = average response time (s), s.d. = standard deviation (s), R.D. = relative decrease in % with respect to unprocessed. p: significance level for one-tailed t-test (processed vs unprocessed, n = 5, df = 8).

S		∞ SNR				6 dB SNR				3 dB SNR				0 dB SNR				-3 dB SNR			
		Su	Sp_TS_ST_			Su	Sp_TS_ST_			Su	Sp_TS_ST_			Su	Sp_TS_ST_			Su	Sp_TS_ST_		
			A	B	C		A	B	C		A	B	C		A	B	C		A	B	C
AB	Tavg.	2.46	2.68	2.88	2.93	3.03	2.85	3.12	3.13	3.04	2.91	2.84	3.14	3.01	2.95	2.95	2.96	3.31	3.52	3.26	3.38
	s.d.	0.6	0.3	0.3	0.5	0.2	0.3	0.1	0.3	0.4	0.4	0.3	0.4	0.5	0.7	0.8	0.8	0.5	0.9	0.9	0.8
	R.D.		-9.3	-17.4	-19.2		6.2	-3.0	-3.2		4.5	6.6	-3.2		1.9	2.1	1.5		-6.4	1.6	-2.1
	p		0.2	0.09	0.1		0.1	0.2	0.3		0.3	0.2	0.3		0.4	0.4	0.5		0.3	0.5	0.4
BT	Tavg.	2.84	2.67	2.61	2.58	2.98	3.18	3.18	2.99	3.52	3.74	3.09	2.88	3.89	3.92	3.31	3.53	3.85	3.37	3.56	3.45
	s.d.	0.7	0.8	0.3	0.5	0.76	1.0	0.3	0.6	0.8	1.0	0.5	0.5	0.8	0.8	0.6	0.7	0.9	0.7	0.3	0.6
	R.D.		6.1	8.2	9.1		-6.6	-6.4	-0.3		-6.1	12.3	18.3		-0.7	15.1	9.3		12.3	7.5	10.3
	p		0.4	0.3	0.3		0.4	0.3	0.5		0.4	0.2	0.09		0.5	0.1	0.2		0.2	0.3	0.2
SR	Tavg.	1.55	1.62	1.81	1.81	1.95	1.78	1.68	1.88	1.98	1.89	1.82	1.81	2.15	2.05	2.08	2.04	2.45	2.24	2.19	2.36
	s.d.	0.08	0.07	0.3	0.3	0.3	0.3	0.09	0.2	0.5	0.2	0.3	0.2	0.4	0.3	0.4	0.4	0.7	0.4	0.6	0.8
	R.D.		-4.9	-16.7	-16.8		8.6	13.6	3.5		4.3	7.9	8.7		4.9	3.2	5.3		8.6	10.6	3.8
	p		0.07	0.05	0.06		0.2	0.03	0.3		0.4	0.3	0.2		0.3	0.4	0.3		0.3	0.3	0.4
Tavg.		2.28	2.32	2.43	2.44	2.65	2.60	2.66	2.67	2.85	2.85	2.58	2.61	3.02	2.97	2.78	2.84	3.20	3.04	3.00	3.06
pter 4 Avg. R.D.			-2.7	-8.6	-9.0		2.7	1.4	0.0		0.9	8.9	7.9		2.0	6.8	5.4		4.8	6.6	4.0

TABLE 4.2. Experiment I. (TS_ST). Percentage recognition scores (RS) for Unprocessed Speech (Su) and for Processed Speech Sp_TS_ST_A, Sp_TS_ST_B, Sp_TS_ST_C corresponding to duty cycles 75, 70, and 60%. 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, n = 5, df = 8).

S		∞ SNR				6 dB SNR				3 dB SNR				0 dB SNR				-3 dB SNR			
		Su	Sp_TS_ST_			Su	Sp_TS_ST_			Su	Sp_TS_ST_			Su	Sp_TS_ST_			Su	Sp_TS_ST_		
			A	B	C		A	B	C		A	B	C		A	B	C		A	B	C
AB	RS	100	98.0	97.0	94.0	91	93.0	96.0	95.0	95	94.0	98.0	97.0	91	94.0	94.0	92.0	80	85.0	88.0	86.0
	s.d.	0.0	2.0	2.5	1.5	6.8	3.1	4.9	1.7	1.7	3.3	1.7	2.6	5.9	4.5	6.3	6.7	9.1	10.9	9.4	8.1
	R.I.		-2.0	-3.0	-6.0		2.1	5.5	4.4		-1.0	3.1	2.1		3.3	3.3	1.1		6.2	10.0	7.5
	p		0.05	0.01	0.0		0.3	0.1	0.2		0.4	0.007	0.1		0.2	0.2	0.4		0.2	0.1	0.2
BT	RS	98	96.0	96.0	96.0	81	84.0	84.0	86.0	75	77.0	83.0	81.0	71	74.0	76.0	77.0	67	66.0	75.0	70.0
	s.d.	1.4	3.8	4.0	2.4	5.4	4.5	8.1	5.5	7.9	3.3	8.6	9.8	8.9	8.5	8.6	8.1	1.9	8.9	8.6	5.6
	R.I.		-2.0	-2.0	-2.0		3.7	3.7	6.1		2.6	10.6	8.0		4.2	7.0	8.4		-1.5	12.0	4.5
	p		0.1	0.07	0.07		0.2	0.30	0.1		0.3	0.1	0.2		0.3	0.2	0.2		0.4	0.04	0.2
SR	RS	100	98.0	98.0	97.0	95	99.0	99.0	97.0	95	96.0	98.0	98.0	93	92.0	92.0	91.0	86	89.0	87.0	91.0
	s.d.	0.8	2.4	1.8	3.4	3.2	0.7	0.9	2.5	2.5	2.0	1.7	2.2	5.1	2.4	5.7	3.3	5.2	3.8	6.1	1.9
	R.I.		-2.0	-2.0	-3.0		4.2	4.2	2.1		2.1	3.1	3.1		-1.0	-1.0	-2.1		3.4	1.2	5.8
	p		0.1	0.2	0.06		0.01	0.02	0.2		0.3	0.04	0.08		0.3	0.4	0.3		0.2	0.4	0.05
Chapter 4 Avg. RS		99	97.3	97.0	95.7	89	92.0	93.0	92.7	88	89.0	93.0	92.0	85	86.7	87.3	86.7	77.7	80.0	83.3	82.3
Chapter 4 Avg. R.I.			-2.0	-2.3	-3.7		3.3	4.5	4.2		1.2	5.6	4.4		2.2	3.1	2.5		2.7	7.7	5.9

TABLE 4.3. Experiment I. (TS_ST). Relative information transmitted (%) for Unprocessed Speech (Su) and for Processed Speech Sp_TS_ST_A, Sp_TS_ST_B, Sp_TS_ST_C corresponding to duty cycles 75, 70, and 60% respectively. S: Subject.

Feature	S	∞ SNR				6 dB SNR				3 dB SNR				0 dB SNR				-3 dB SNR			
		Su	Sp_TS_ST_			Su	Sp_TS_ST_			Su	Sp_TS_ST_			Su	Sp_TS_ST_			Su	Sp_TS_ST_		
			A	B	C		A	B	C		A	B	C		A	B	C		A	B	C
Overall	AB	100	98	96	95	90	92	95	94	93	94	97	95	90	92	92	90	82	85	87	87
	BT	98	96	95	96	86	86	84	85	81	83	84	81	80	80	82	79	77	73	81	75
	SR	99	98	98	96	96	99	99	96	95	97	98	97	92	92	92	92	90	90	89	91
	Avg.	99	97.3	96.3	95.7	90.7	92.3	92.7	91.7	89.7	91.3	93	91	87	88	88.7	87	83	82.7	85.7	84.3
Voicing	AB	100	100	100	100	100	100	100	100	100	100	100	100	97	97	100	100	94	97	97	100
	BT	100	97	100	100	100	97	92	97	95	97	92	97	97	100	97	100	92	88	100	92
	SR	100	100	100	97	100	100	100	100	100	100	100	97	97	100	97	100	100	100	100	100
	Avg.	100	99	100	99	100	99	97.3	99	98.3	99	97.3	98	97	99	98	100	95.3	95	99	97.3
Place	AB	100	92	88	83	76	76	85	85	81	81	91	84	70	81	78	70	41	53	63	60
	BT	94	87	84	88	54	58	57	59	42	49	58	52	28	40	44	40	26	27	37	33
	SR	98	93	94	87	83	98	96	88	84	88	92	92	73	69	71	72	56	60	55	65
	Avg.	97.3	90.7	88.7	86	71	77.3	79.3	77.3	69	72.7	80.3	76	57	63.3	64.3	60.7	41	46.7	51.7	52.7
Manner	AB	100	100	100	98	91	100	100	95	98	96	98	96	92	91	90	84	82	76	78	70
	BT	100	98	98	98	86	85	86	77	75	71	70	73	70	60	65	68	57	57	66	56
	SR	100	100	100	100	100	100	100	95	100	100	98	92	84	80	80	79	69	74	71	76
	Avg.	100	99.3	99.3	98.7	92.3	95	95.3	89	91	89	88.7	87	82	77	78.3	77	69.3	69	71.7	67.3
Nasality	AB	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	BT	100	100	100	100	100	100	96	100	100	100	96	100	100	100	100	100	100	100	100	94
	SR	100	100	100	100	100	100	100	100	100	100	100	96	100	100	100	100	100	100	100	100
	Avg.	100	100	100	100	100	100	98.7	100	100	100	98.7	98.7	100	100	100	100	100	100	100	98
Frication	AB	100	100	100	97	85	100	100	92	97	94	97	93	87	85	84	73	70	61	64	51
	BT	100	97	97	97	76	74	77	61	58	51	51	53	50	33	41	45	27	26	43	29
	SR	100	100	100	100	100	100	100	91	100	100	97	90	74	68	68	66	49	57	52	61
	Avg.	100	99	99	98	87	91.3	92.3	81.3	85	81.7	81.7	78.7	70.3	62	64.3	61.3	48.7	48	53	47
Duration	AB	100	81	81	54	76	76	85	79	83	82	96	79	60	79	87	63	43	53	55	46
	BT	94	80	83	81	68	61	64	57	62	68	76	52	51	59	70	44	49	37	58	42
	SR	95	82	85	79	70	96	92	85	76	80	89	96	67	61	70	77	38	59	45	63
	Avg.	96.3	81	83	71.3	71.3	77.7	80.3	73.7	73.7	76.7	87	75.7	59.3	66.3	75.7	61.3	43.3	49.7	52.7	50.3

TABLE 4.4. Experiment II. (TS_TR). Response times (s) for Unprocessed Speech (Su) and for Processed Speech Sp_TS_TR_A, Sp_TS_TR_B, Sp_TS_TR_C, Sp_TS_TR_D corresponding to transition durations 0, 1, 2, 3 ms. S: Subject, Tavg. = Average response time (s), s.d. = standard deviation (s), R.D. = average of relative decrease in % with respect to unprocessed. p: significance level one-tailed paired t-test (unprocessed vs processed, averaged across the subjects, n = 5, df = 4).

S	∞ SNR					6 dB SNR					3 dB SNR					0 dB SNR					-3 dB SNR					-6 dB SNR				
	Su	Sp_TS_TR_				Su	Sp_TS_TR_				Su	Sp_TS_TR_				Su	Sp_TS_TR_				Su	Sp_TS_TR_				Su	Sp_TS_TR_			
		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D
AB	1.94	1.60	1.47	1.44	1.58	1.61	1.50	1.60	1.65	1.65	1.76	1.79	1.71	1.64	1.48	1.92	1.72	1.67	1.90	1.61	2.17	2.06	1.66	2.03	1.76	1.91	1.83	1.74	1.93	1.75
BT	1.54	1.86	1.83	1.70	1.73	2.09	1.78	1.64	1.24	1.56	1.79	1.75	1.48	1.43	1.41	1.92	1.89	1.68	1.71	1.51	1.45	1.63	1.60	1.57	1.49	1.83	1.72	1.61	1.78	1.61
AC	2.43	2.20	1.85	2.27	1.71	2.27	2.23	2.02	2.26	2.20	2.30	2.26	2.19	2.22	2.25	2.54	2.49	2.48	2.45	2.52	2.72	2.60	2.70	2.53	2.74	2.55	2.68	2.66	2.57	2.31
JK	1.58	1.86	1.60	1.40	1.56	1.67	1.62	1.58	1.60	1.58	1.68	1.68	1.41	1.57	1.53	1.62	1.51	1.56	1.42	1.35	1.62	1.49	1.48	1.53	1.37	1.69	1.53	1.59	1.51	1.39
LT	1.65	1.63	1.39	1.41	1.38	1.68	1.80	1.52	1.54	1.58	1.95	1.52	1.54	1.57	1.56	2.13	1.69	1.72	1.70	1.85	1.82	1.90	1.91	1.88	1.64	1.63	1.95	2.00	1.93	1.75
Tavg.	1.83	1.83	1.63	1.64	1.59	1.86	1.79	1.67	1.66	1.71	1.9	1.80	1.67	1.69	1.65	2.03	1.86	1.82	1.84	1.77	1.96	1.94	1.87	1.91	1.80	1.92	1.94	1.92	1.94	1.76
s.d.	0.4	0.3	0.2	0.4	0.1	0.3	0.3	0.2	0.4	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.4	0.4	0.4	0.5	0.5	0.4	0.5	0.4	0.6	0.4	0.4	0.4	0.4	0.3
R.D.		-2.1	8.8	9.6	10.7		3.9	9.6	10.2	7.5		4.9	12.4	11.3	13.6		8.3	10.2	9.6	13.6		0.14	3.5	1.5	8.1		-1.0	-0.04	-1.4	8.0
P		0.5	0.1	0.08	0.1		0.2	0.03	0.14	0.1		0.2	0.01	0.02	0.01		0.05	0.02	0.03	0.008		0.4	0.3	0.2	0.07		0.4	0.5	0.4	0.05

Chapter 5

Processing with time-varying comb filters

5.1 Introduction

In the scheme of spectral splitting (reviewed in Chapter 3), sensory cells corresponding to alternate bands of the basilar membrane are always stimulated, whereas sensory cells of other bands are always relaxing in both the ears. In temporal splitting (presented in Chapter 4), all the sensory cells of the ears get relaxed alternately for some time. In the second phase of our research, a scheme for combined spectral and temporal splitting of speech signal for binaural dichotic presentation has been devised using time-varying comb filters. With the combined splitting, all the sensory cells of the basilar membrane get periodic relaxation from stimulation. The implementation and evaluation of this scheme with normal hearing subjects and simulated bilateral sensorineural loss are presented in this chapter.

5.2 Time varying comb filters

To reduce the effect of spectral and temporal masking simultaneously, Lunner *et al.* (1993) studied the combination of spectral and temporal splitting by alternately switching the odd and even bands between the two ears. Authors have reported that the scheme resulted in perception of periodic interruptions and did not improve speech perception. Sweeping the bands (gradual shifting between odd and even bands and vice versa) instead of instantaneous switching may reduce the spectral distortion, and thereby, it may improve the perception degraded due to increased temporal and spectral masking. Hence, we have investigated a scheme of sweeping of odd and even bands for binaural dichotic presentation as a possible solution to reduce the spectral distortion that deteriorates speech quality. The scheme combines the effect of both spectral splitting (separation of odd and even bands for presenting to the two ears) and temporal splitting.

For sweeping the bands, a pair of time-varying comb filters with complementary magnitude response are used. The comb filters used are based on auditory critical bandwidths (Zwicker, 1961) as used by Chaudhari and Pandey (1998a,b).

The frequency range between two alternate critical band center frequencies (e.g. between l^{th} and $(l+2)^{th}$ critical bands), was divided into m equal parts, to obtain center frequencies of the bands of the time varying comb filters. This procedure was carried out for the entire frequency range. Eighteen critical bands used for splitting and division between each alternate critical bands for obtaining sweeping of bands are given in Tables 5.1 and Table 5.2. The duration of sweeping is 20 ms (the same time as used for inter-aural switching for temporal splitting). Each set of successive pair of complementary comb filter operates for specified time (total shifting period/number of shifting, 20/m ms), repeating after every 20 ms till the entire speech signal gets processed. The implementation of the scheme of combined splitting is as shown in Fig. 5.1a. Figure 5.1b shows the sweeping of the magnitude response of a time varying comb filter.

TABLE 5.1. Eighteen critical bands for spectral splitting.

Filter for left ear		
Band	Center frequency kHz	Passband frequency kHz
1	0.13	0.01-0.20
3	0.35	0.30-0.40
5	0.57	0.51-0.63
7	0.84	0.77-0.92
9	1.17	1.08-1.27
11	1.60	1.48-1.72
13	2.16	2.00-2.32
15	2.92	2.70-3.15
17	4.05	3.70-4.40

Filter for right ear		
Band	Center frequency kHz	Passband frequency kHz
2	0.25	0.20-0.30
4	0.45	0.40-0.51
6	0.70	0.63-0.77
8	1.00	0.92-1.08
10	1.37	1.27-1.48
12	1.86	1.72-2.00
14	2.51	2.32-2.70
16	3.42	3.15-3.70
18	4.70	4.40-5.00

TABLE 5.2. Bandwidths for eighteen critical band filters with four shifting between odd-even-odd bands.

Filter for left ear			Filter for right ear		
Band	Center	Passband	Band	Center	Passband
1	**0.15**	**0.01-0.20**	2	**0.25**	**0.20-0.30**
	0.20	0.15-0.25		0.30	0.25-0.35
	0.25	0.20-0.30		0.13	0.07-0.20
	0.30	0.25-0.35		0.20	0.15-0.25
3	**0.35**	**0.30-0.40**	4	**0.45**	**0.40-0.51**
	0.40	0.35-0.46		0.51	0.45-0.57
	0.45	0.40-0.51		0.35	0.30-0.40
	0.51	0.45-0.57		0.40	0.35-0.46
5	**0.57**	**0.51-0.63**	6	**0.70**	**0.63-0.77**
	0.63	0.57-0.69		0.77	0.70-0.84
	0.70	0.63-0.77		0.57	0.51-0.63
	0.77	0.70-0.84		0.63	0.57-0.69
7	**0.84**	**0.77-0.92**	8	**1.00**	**0.92-1.08**
	0.91	0.84-0.99		1.08	1.00-1.17
	1.00	0.92-1.08		0.84	0.77-0.92
	1.08	1.00-1.17		0.91	0.84-0.99
9	**1.17**	**1.08-1.27**	10	**1.37**	**1.27-1.48**
	1.27	1.18-1.37		1.48	1.37-1.59
	1.37	1.27-1.48		1.17	1.08-1.27
	1.48	1.37-1.59		1.27	1.18-1.37
11	**1.60**	**1.48-1.72**	12	**1.86**	**1.72-2.00**
	1.72	1.60-1.85		2.00	1.85-2.15
	1.86	1.72-2.00		1.60	1.48-1.72
	2.00	1.85-2.15		1.72	1.60-1.85
13	**2.16**	**2.00-2.32**	14	**2.51**	**2.32-2.70**
	2.33	2.15-2.51		2.71	2.50-2.92
	2.51	2.32-2.70		2.16	2.00-2.32
	2.71	2.50-2.92		2.33	2.15-2.51
15	**2.92**	**2.70-3.15**	16	**3.42**	**3.15-3.70**
	3.16	2.91-3.42		3.72	3.41-4.03
	3.42	3.15-3.70		2.92	2.70-3.15
	3.72	3.41-4.03		3.16	2.91-3.42
17	**4.05**	**3.70-4.40**	18	**4.70**	**4.40-5.00**
	4.39	4.01-4.77		5.07	4.62-5.52
	4.70	4.40-5.00		4.05	3.70-4.40
	5.07	4.62-5.52		4.39	4.01-4.77

Each time varying comb filter constitutes a number of comb filters, depending on the number of shifting (represented as m). Each of these m comb filters has 9 pass bands corresponding to the auditory critical bands. At a given time, two comb filters which have complementary magnitude response, one each from the pair of time varying comb filters are used to process the speech for binaural dichotic presentation. If the comb filters in the time varying filters for the left ear are numbered in the order of sweeping as [1], [2], ..., [m/2], [m/2 +1], ..., [m], then those for the right ear will be numbered as [m/2 +1], [m/2 +2], ...[m], [1], [2], ..., [m/2]. These comb filters have magnitude responses such that the bands of comb filter [2] will be shifted along the frequency axis with respect to comb filter [1] as shown in Fig. 5.1b. In a similar way all the pass bands of each of the comb filter will be a shifted version of the corresponding pass band of the previous one. After m shifting when the cycle repeats, it looks as though each pass band merge into the next pass band. The complementary pairs are [1] and [m/2 +1], [2] and [m/2+2], ..., and [m/2] and [m] as shown by the positioning of rotating switch in Fig. 5.1a.

In the scheme of spectral splitting for binaural dichotic presentation investigated by Chaudhari and Pandey (1998a,b), reviewed earlier in Chapter 3, two complementary comb filters with pass bands corresponding to critical band auditory filters were used. In their scheme, 18 bands were used to cover the frequency range of 5 kHz. Comb filters were 128-coefficient linear phase FIR filters, designed for sharp transitions between the bands using frequency sampling technique and sampling rate of 10 kHz (Chaudhari and Pandey, 1998b). In floating point implementation, these filters have maximum pass band ripple of 4 dB, minimum stop band attenuation of 10 dB, and transition width (Δf) of 78 Hz (Cheeran *et al.* 2002). The odd and even bands which form a pair of comb filters, split the speech into two such that the spectral components that are likely to get masked are presented to different ears. 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. With the same intensity, binaurally presented components will be louder than monaurally presented components, however the loudness is generally less than double (Scharf, 1969). 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. Hence in our research, it was decided to conduct perceptual balance tests and optimize the comb filter design for minimum perceived spectral distortion.

Loudness perception tests were conducted using different types of stimuli, to determine the difference in intensity for same perception in monaural and binaural presentations. Loudness perception test results showed that the perceived levels match when the binaural level was 4–9 dB lower than monaural level (described in Appendix C). Based on these results, pairs of comb filters were designed with different levels at crossovers between adjacent bands and were tested by slowly swept sinusoidal tones. Experiments showed that the change in loudness was negligible with crossover response lying between -4 dB and -6 dB. With this constraint for response at the crossover frequencies, comb filters with responses corresponding to 18 critical bands were designed as linear phase FIR filters, using frequency sampling technique. Work has been carried out to optimize the magnitude response of the comb filters for minimizing perceived spectral distortion by having low pass band ripple, high stop band attenuation and perceptual balance at inter-band crossovers. The work was carried out jointly with my colleague A. N. Cheeran and it is described in Appendix C. Figure 5.2 shows magnitude response of optimized comb filters. These filters are 256-coefficient linear phase FIR filters with pass band ripple less than 1 dB, stop band attenuation of more than 30 dB, transition width varying from 78 Hz to 117 Hz, and signal level at transition crossovers between -4 to -6 dB. These filters have resulted in improved speech quality and speech perception (Cheeran *et al.* 2002). Implementation and evaluation of scheme of spectral splitting using comb filters with sharp transitions and comb filters with magnitude response optimized for minimum perceived spectral distortion, are given in Appendix C.

The pre-calculated set of filter coefficients were cyclically swept with m shifting (2, 4, 8, or 16) with a time period of 20 ms. After every time slot of 20/m ms a new set of coefficients for next pair of comb filters takes over. The sweeping of magnitude responses in the time varying comb filters is represented in Figs 5.3 and 5.4, for 2 and 4 shifting respectively.

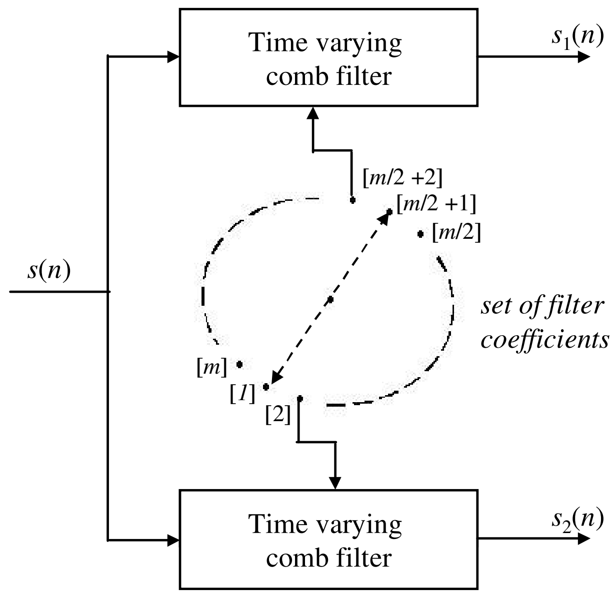
A program "`filt`" was written in C (described in Appendix C) for offline processing of the speech signal. The process of splitting was verified by obtaining the spectrograms of the processed outputs using a spectrographic analysis set-up (as described in Appendix A). Figures 5.5, 5.6 and 5.7 show the wideband (Δf = 300 Hz) spectrograms of a swept sine wave, random white noise, and speech syllable /*a*s*a*/ respectively for the scheme of combined splitting. Each figure shows spectrograms of unprocessed signal and the two processed signals having complementary symmetry (sweep duration = 20 ms and no. of shifting = 4).

5.3 Listening tests

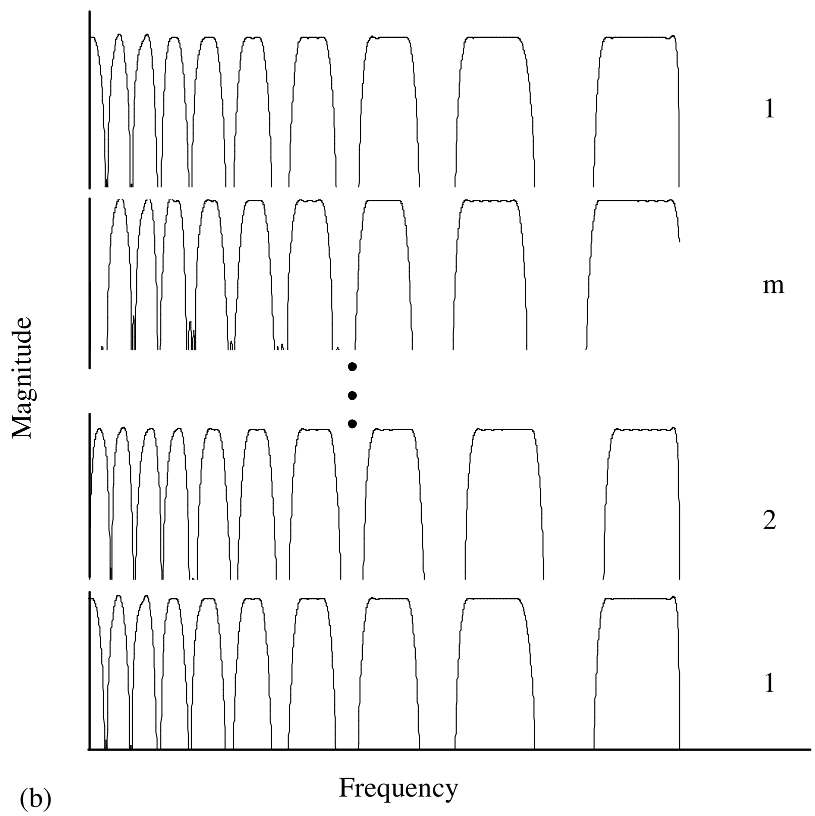
Listening tests were conducted to assess the effectiveness of the scheme of combined spectral and temporal splitting using time varying comb filters. Tests were conducted on five normal hearing subjects with simulated sensorineural hearing loss. In the tests, subjects were asked to identify a closed set of 12 English consonants in VCV context, presented over the headphones. Listening tests consisted of (a) diotic presentation of unprocessed speech, and (b) dichotic presentation of processed speech.

5.3.1 Test material and subjects

The test material was same as that used in the schemes of temporal processing (TS_ST and TS_TR). Twelve English consonants /p, b, t, d, k, g, m, n, s, z, f, v/ were used in the VCV context with vowel /*a*/ as in father. Simulation of sensorineural loss in normal hearing persons has been done for SNRs of ∞, 6, 3, 0, -3, -6, -9, -12, and -15 dB. Broadband Gaussian noise (from function generator HP 33120A) was added to the speech stimuli maintaining the SNR constant on the basis of short-time (≈ 10 ms) energy of the speech signal.



(a)



(b)

FIG. 5.1. (a) Schematic representation of combined splitting using time varying comb filters and (b) Representation of magnitude response of one of the time varying comb filters which includes a set of comb filters (magnitude responses), which are swept over one after the other cyclically.

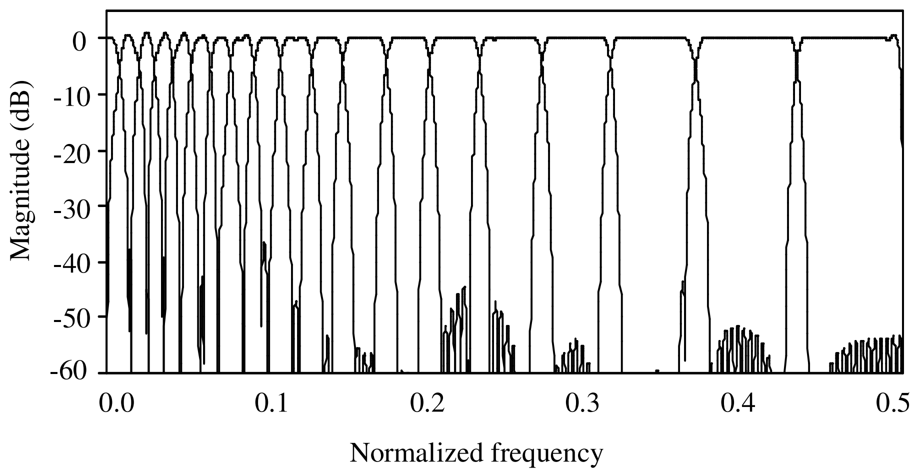


FIG. 5.2. Magnitude responses of the pair of comb filters (256-coefficients) designed for low perceived spectral distortion. Crossover lies between -4 dB and -6 dB with respect to pass band response. Minimum stop band attenuation 30 dB and maximum pass band ripple is 1 dB.

Five normal hearing subjects (AB: M 28, SS: F 24, VK: M 27, AC: F 37, JK: M 25) participated in the listening test. Subjects AC and JK had participated in the scheme of TS_TR, and subject AB had participated in both TS_ST and TS_TR experiments. All the subjects had pure tone thresholds less than 20 dB HL in the frequency range of 125 Hz to 6 kHz.

5.3.2 Experimental procedure

To evaluate the scheme of inter-aural switching with spectral and temporal splitting, and to find the near-optimal number of shifting, listening tests were conducted. Experimental set-up and experimental procedure for conducting the listening tests were same as described earlier in Chapter 4. Processed and unprocessed stimuli were added with broadband noise at different SNR conditions to simulate sensorineural hearing loss in normal hearing subjects. There were a total of 45 (5 processing conditions x 9 SNR) test conditions. Tests were conducted in an acoustically isolated chamber. Test scores were obtained in the form of confusion matrix. The confusion

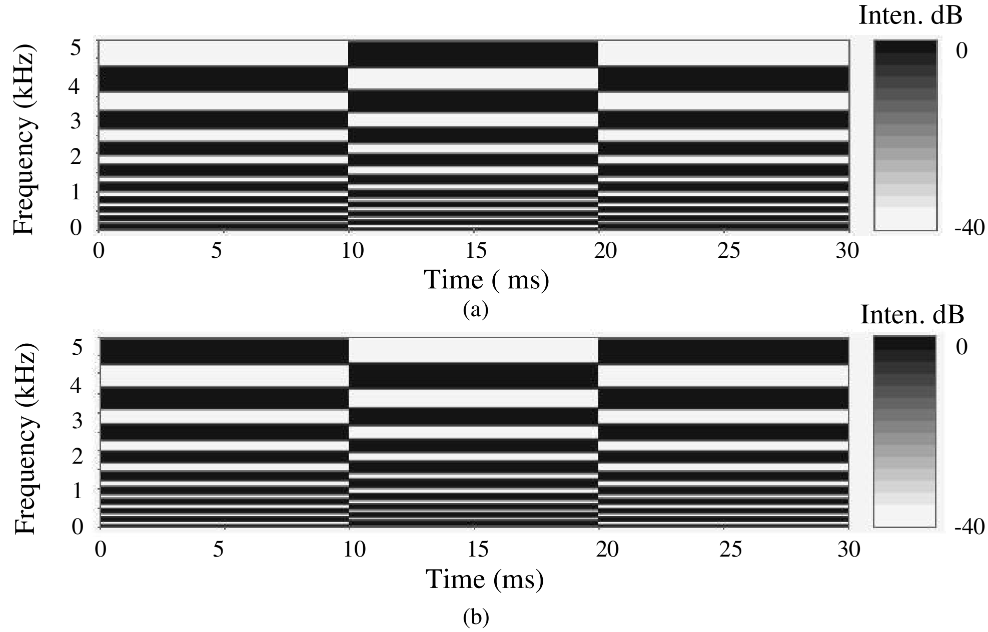


FIG. 5.3. An idealized representation of magnitude response of the pair of time varying comb filters using 2 shifting (m). After every 10 ms ($20/m$) next pair of comb filter takeover. The cycle repeats after 20 ms. (a) for left ear (b) for right ear.

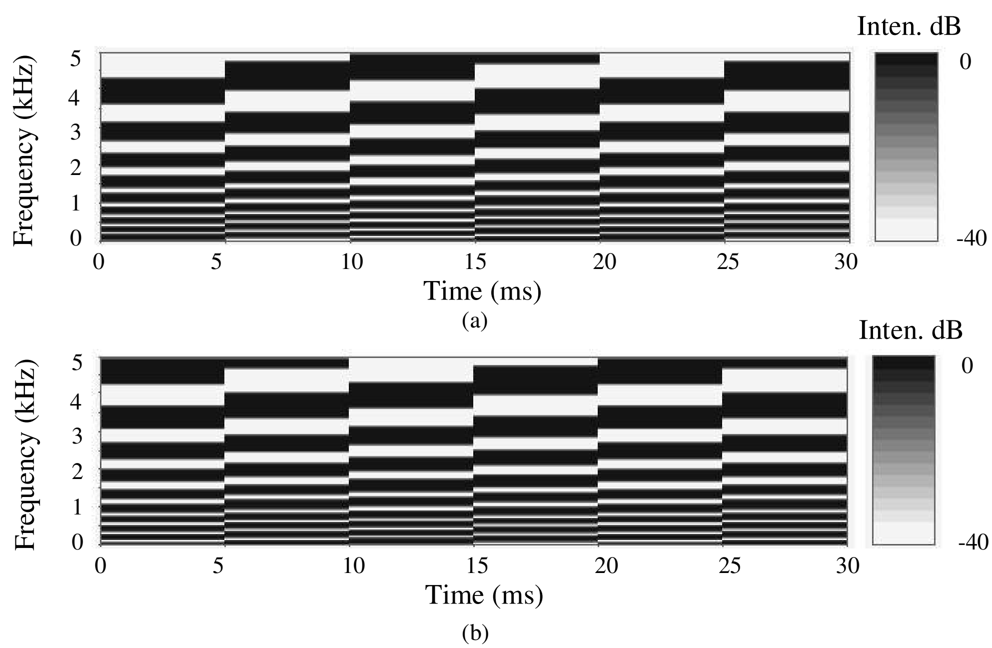


FIG. 5.4. An idealized representation of magnitude response of the pair of time varying comb filters using 4 shifting (m). After every 5 ms ($20/m$) next pair of comb filter takeover. The cycle repeats after 20 ms. (a) for left ear (b) for right ear.

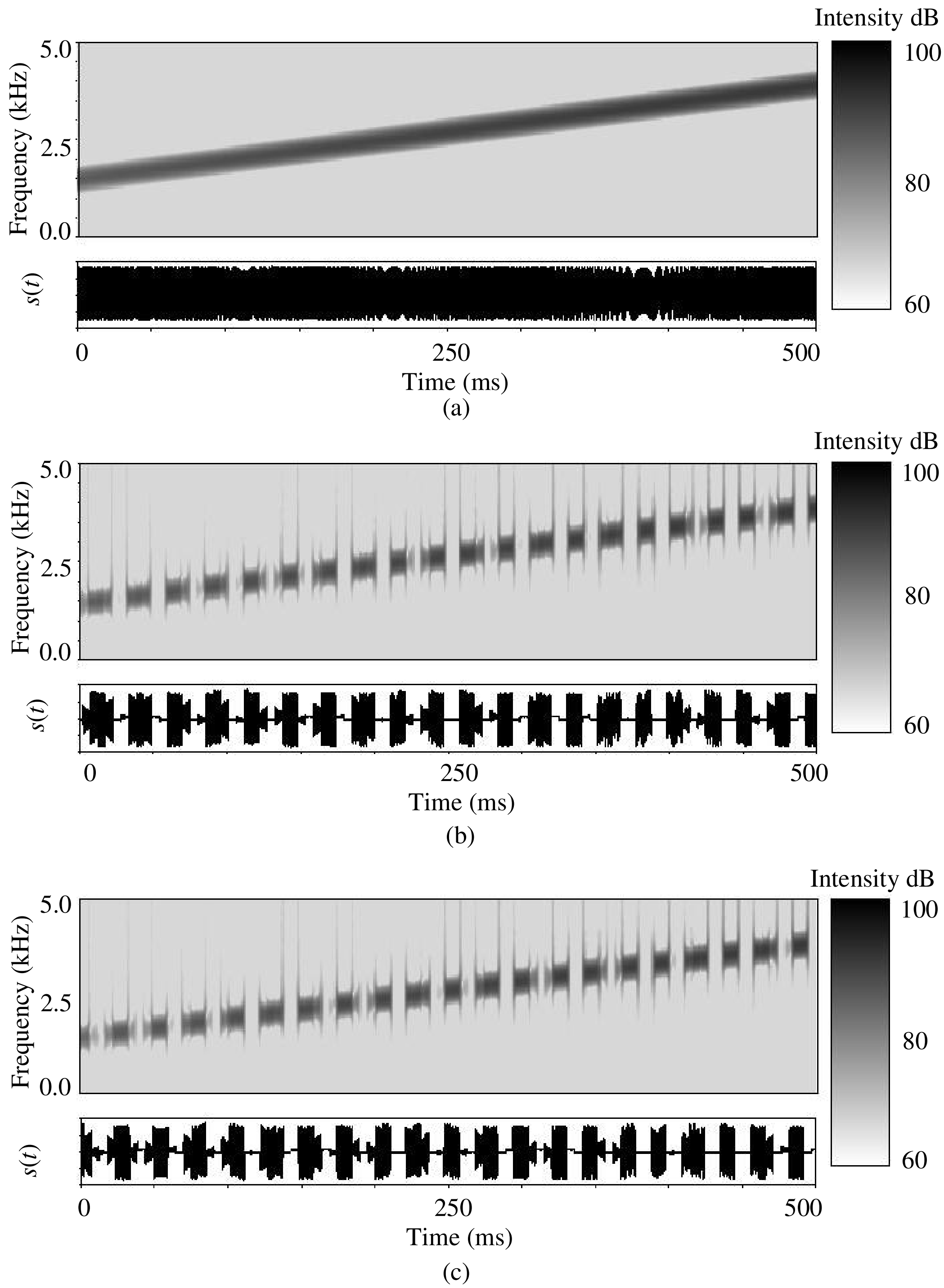


FIG. 5.5. Combined splitting (CS): Wide-band spectrograms ($\Delta f \approx 300$ Hz) of swept sine wave. Sweep period=20 ms and no. of shifting=4. (a) unprocessed (b) processed, left ear (c) processed, right ear.

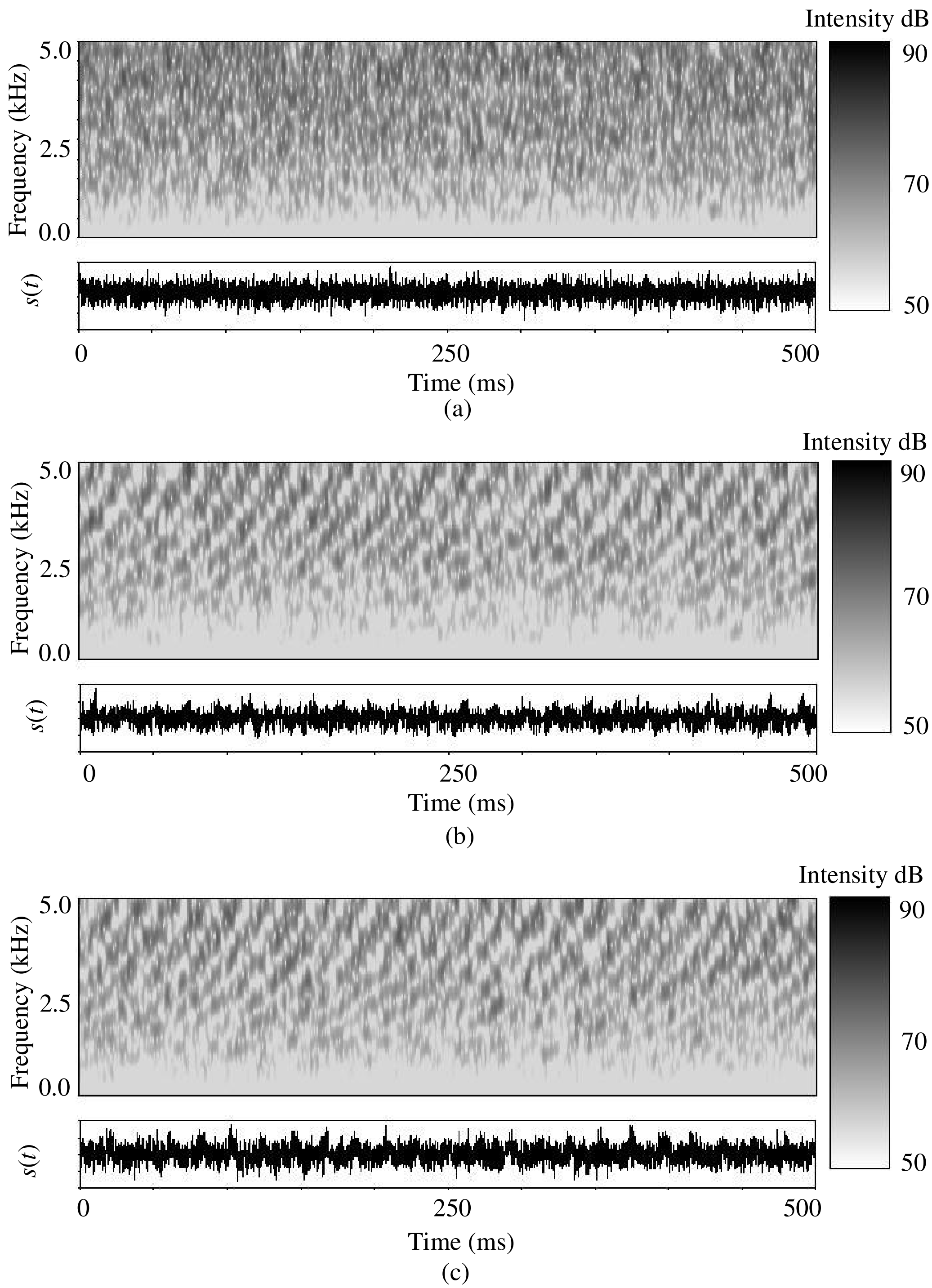


FIG. 5.6. Combined splitting (CS): Wide-band spectrograms ($\Delta f \approx 300$ Hz) of random noise. Sweep period = 20 ms and no. of shifting = 4. (a) unprocessed (b) processed, left ear (c) processed, right ear.

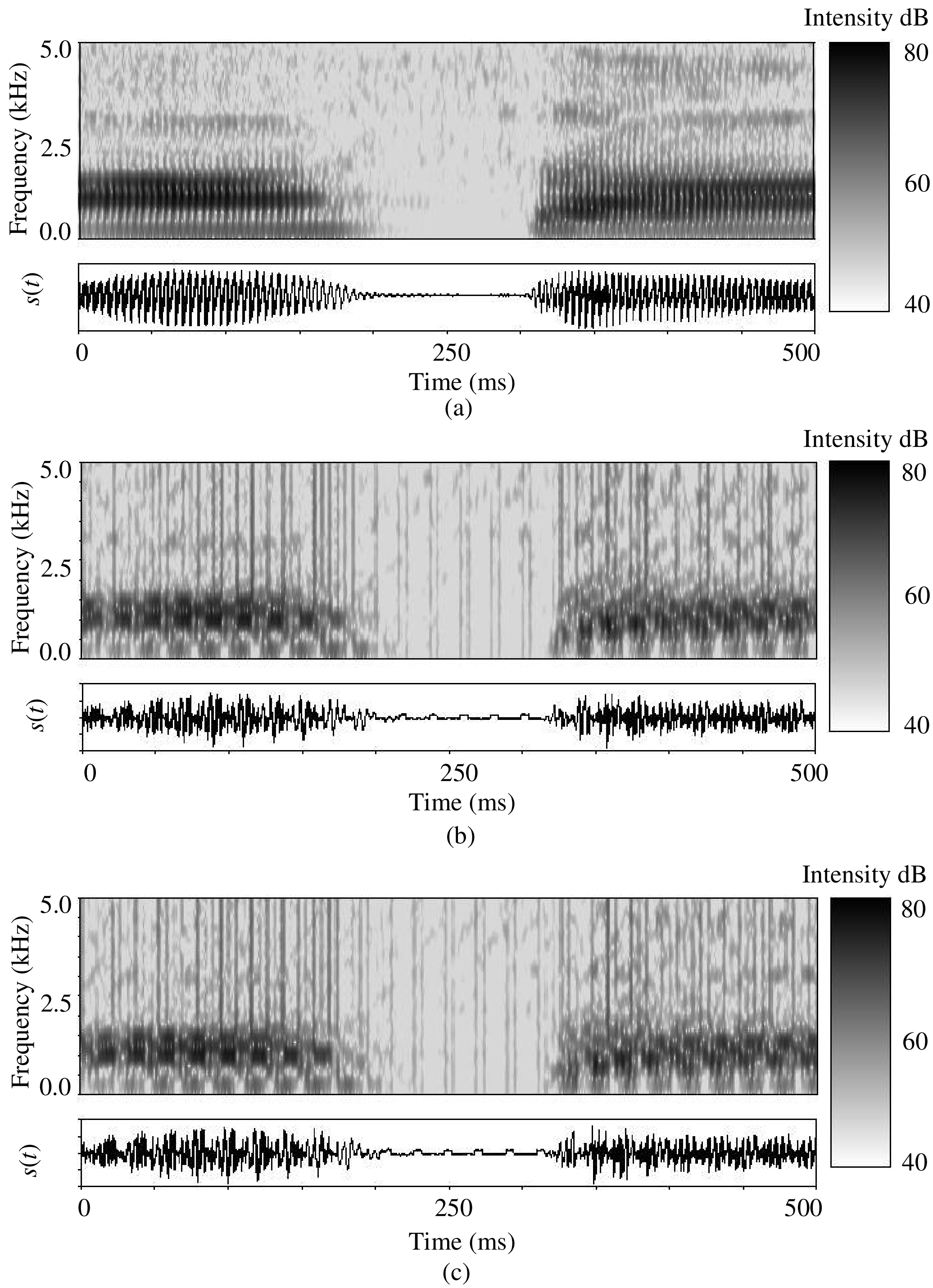


FIG. 5.7. Combined splitting (CS): Wide-band spectrograms ($\Delta f \approx 300$ Hz) of speech syllable /*asa*/. Sweep period = 20 ms and no. of shifting = 4. (a) unprocessed (b) processed, left ear (c) processed, right ear.

matrix shows the stimuli along rows and responses along columns. Each entry in the cell shows the number of times a stimulus-response pair occurs in the test.

In the test, 12 stimuli were used and each stimulus was presented 5 times in random order (as described in Appendix D). A test run takes about 5–8 minutes. Hence for 45 test conditions, each subject took about 50 hours for completion of all tests. Test sessions for the five subjects were spread over a span of about four months depending upon the availability and willingness of the subjects. The tests are referred to as "Experiment III" in the tables and figures.

5.4 Test results

Results of listening tests for unprocessed and processed speech are presented in this section. For ascertaining the improvement in speech quality, a compilation of subject's qualitative assessment of the test stimuli processed under various listening conditions was carried out. Response times for unprocessed and processed speech were obtained to assess the effectiveness of the processing scheme in reducing load on perception. Percentage correct recognition scores were obtained from confusion matrix. Paired t-test was carried out to determine the level of significance due to processing. Information transmission analysis, a measure that is not affected by subject's response bias was carried out. Information transmission analysis was also performed for speech features to determine the contribution of each feature in improving the speech perception.

5.4.1 Quality assessment

Compilation of subject's qualitative assessment indicated that under no-noise condition, subjects reported the quality of processed speech to be inferior compared to that of the unprocessed speech. However, under poor SNR conditions quality was better for processed speech signal than that for unprocessed signal. There was preference for speech processed with 4 and 8 shifting.

5.4.2 Response time

Table 5.3 gives the response times of individual subjects and averaged across the five subjects for processed and unprocessed speech, for all SNR conditions. It also gives percentage relative decrease in response time averaged across the subjects and paired t-test (one-tailed) significance levels. For unprocessed speech, the response times increased with decrease in SNR indicating that decrease in SNR increases the load on the perception process. Figure 5.8 shows the response times (averaged across five subjects) for unprocessed and processed speech and Fig. 5.9 shows the percentage relative decrease in response time. For unprocessed speech, average response time was 1.89 s under no-noise condition. It increased by 11.1, 18.5, 17.4, 23.2, 18.5, 22.75, 31.7, 38.1% for 6, 3, 0, -3, -6, -9, -12, -15 dB SNR conditions respectively. Averaged across the subjects, relative decrease in response time for 4 shifting were 6.7, 6.7, 3.4, 16.2, 10.3, 7.8, 15.8, 16.3% respectively. For 8 shifting, the corresponding values were 8.5, 8.8, 5.6, 14.6, 5.8, 14.0, 13.7, 12.9%. Relative decreases were statistically significant for 8 shifting, for 0, -3, -9, -12 dB SNRs with $p < 0.04$ and 4 shifting has resulted in statistically significant decrease ($p < 0.04$) at -6, -12, -15 dB SNR. For higher levels of noise, higher improvements were observed for shifting of 4 and 8.

5.4.3 Recognition score

Average of recognition scores of five tests, for each condition, for each subject was obtained. Table 5.4 gives the recognition scores for each subject for each test condition, averaged recognition scores, averaged relative improvement in recognition scores for processed speech, paired t-test (one-tailed) values. Figure 5.10 shows the recognition scores (averaged across five subjects) for unprocessed and processed speech and Fig. 5.11 shows relative improvement in recognition score. Recognition scores averaged across the subjects for unprocessed speech decreased from 100% under no-noise condition to 96, 94, 93, 88.7, 85.5, 81, 74, 64.5% for 6, 3, 0, -3, -6, -9, -12, -15 dB SNR conditions respectively. Relative improvements in recognition scores for processed speech with 4 shifting were 2.2, 5.3, 5.6, 9.4, 13.9, 15.3, 19.2, 32.1% for 6, 3, 0, -3, -6, -9, -12, -15 dB SNR conditions respectively.

TABLE 5.3. Experiment III. (CS). Response times (s) for Unprocessed Speech (Su) and Processed Speech, Sp_CS_A, Sp_CS_B, Sp_CS_C, Sp_CS_D corresponding to 4 shiftings 16, 8, 4, and 2 respectively. S: Subject, Tavg. = average response times (s), s.d. = standard deviation (s), R.D. = Average of relative decrease in % with respect to unprocessed. p: significance level for one-tailed paired t-test (unprocessed vs processed, averaged across the subjects, n = 5, df = 4).

S	∞ SNR					6 dB SNR					3 dB SNR					0 dB SNR					-3 dB SNR				
	Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_			
		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D
AB	1.56	1.44	1.55	1.55	1.56	1.64	1.52	1.58	1.58	1.62	1.67	1.67	1.51	1.66	1.67	1.79	1.75	1.67	1.76	1.81	1.81	1.65	1.73	1.78	1.67
SS	2.18	1.78	1.79	1.89	1.70	2.63	2.18	2.29	2.09	2.05	2.87	2.08	1.84	1.83	1.89	2.72	2.29	2.56	2.46	2.61	2.68	1.99	1.99	1.86	1.90
VK	2.73	3.20	3.00	2.46	2.41	2.62	2.50	2.32	2.26	2.81	2.79	2.17	2.62	2.69	2.74	2.58	2.50	2.29	2.56	2.59	2.85	2.38	2.42	2.27	2.54
AC	1.81	2.03	1.84	1.82	1.90	1.93	1.88	1.89	1.90	1.87	2.16	2.20	2.24	2.28	2.24	2.39	2.12	2.32	2.13	2.38	2.37	2.38	2.06	2.25	2.29
JK	1.15	1.25	1.80	1.56	1.78	1.68	1.69	1.47	1.78	1.49	1.69	1.65	1.76	1.71	1.80	1.60	1.85	1.58	1.70	1.81	1.94	1.82	1.65	1.49	1.39
Tavg.	1.89	1.94	1.99	1.86	1.87	2.10	1.96	1.91	1.92	1.97	2.24	1.95	1.99	2.03	2.07	2.22	2.10	2.08	2.12	2.24	2.33	2.04	1.97	1.93	1.96
s.d.	0.6	0.8	0.6	0.4	0.3	0.5	0.4	0.4	0.3	0.5	0.6	0.3	0.4	0.4	0.4	0.5	0.3	0.4	0.4	0.4	0.5	0.3	0.3	0.3	0.5
R.D.		-2.4	-9.9	-2.5	-5.2		6.2	8.5	6.7	6.1		10.0	8.8	6.7	5.1		3.4	5.6	3.4	-2.0		11.4	14.6	16.2	15.9
p		0.4	0.3	0.4	0.5		0.08	0.02	0.1	0.2		0.09	0.2	0.2	0.2		0.2	0.02	0.1	0.3		0.04	0.01	0.03	0.02

S	-6 dB SNR					-9 dB SNR					-12 dB SNR					-15 dB SNR				
	Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_			
		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D
AB	2.07	1.74	1.76	1.81	1.88	1.93	1.51	1.67	1.68	1.75	1.98	1.94	1.76	1.67	1.89	2.46	2.01	2.02	2.01	2.20
SS	2.30	1.83	1.76	1.85	1.82	2.20	2.08	1.85	2.25	2.24	1.98	2.29	2.10	1.93	2.15	1.97	2.07	2.18	1.96	1.85
VK	2.78	2.44	2.72	2.55	2.58	3.07	2.22	2.19	2.34	2.20	3.36	2.69	2.65	2.47	2.61	3.25	2.62	2.23	2.34	2.49
AC	2.38	2.49	2.86	2.35	2.82	2.64	2.37	2.54	2.92	2.61	2.92	2.62	2.52	2.79	2.68	2.86	3.14	3.04	2.82	3.06
JK	1.68	1.47	1.54	1.52	1.60	1.73	1.58	1.59	1.47	1.78	2.20	1.75	1.57	1.54	1.63	2.53	1.78	1.72	1.69	1.91
Tavg.	2.24	1.99	2.13	2.02	2.14	2.32	1.95	1.97	2.13	2.12	2.49	2.26	2.12	2.08	2.19	2.61	2.32	2.24	2.17	2.30
s.d.	0.4	0.5	0.6	0.4	0.5	0.5	0.4	0.4	0.6	0.4	0.6	0.4	0.5	0.5	0.5	0.5	0.6	0.5	0.4	0.5
R.D.		11.3	5.8	10.3	4.7		14.8	14.0	7.8	6.8		7.4	13.7	15.8	10.5		10.5	12.9	16.3	11.5
p		0.03	0.3	0.02	0.3		0.03	0.03	0.2	0.2		0.1	0.04	0.03	0.07		0.1	0.1	0.04	0.07

TABLE 5.4. Experiment III. (CS). Recognition scores (%) for Unprocessed Speech (Su) and Processed Speech, Sp_CS_A, Sp_CS_B, Sp_CS_C, Sp_CS_D corresponding to 4 shiftings 16, 8, 4, and 2 respectively. S: Subject, Avg. = average recognition scores, s.d. = standard deviation, R.I. = Average of relative improvement in % with respect to unprocessed. p: significance level for one-tailed paired t-test (unprocessed vs processed, averaged across the subjects, n = 5, df = 4).

S	∞ SNR					6 dB SNR					3 dB SNR					0 dB SNR					-3 dB SNR				
	Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_			
		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D
AB	100	99.3	100	100	100	99.0	100	99.7	100	100	98.3	99.3	99.7	99.7	99.3	96.0	99.0	99.3	99.7	99.3	95.3	99.7	99.7	100	98.0
SS	100	100	100	98.3	100	93.0	97.7	98.7	93.7	95.3	92.3	97.7	98.7	98.7	94.3	88.7	93.7	98.0	97.0	97.3	83.7	99.7	99.7	98.7	99.0
VK	100	99.7	100	100	100	93.3	94.0	100	100	98.3	89.0	100	100	98.3	94.3	92.3	99.7	98.7	100	97.0	83.7	98.0	99.3	98.3	96.7
AC	100	99.0	100	99.7	100	97.0	100	100	99.7	99.7	94.0	95.0	99.3	99.3	98.3	92.7	95.3	93.3	96.7	95.3	90.0	91.7	96.3	92.3	93.0
JK	100	99.0	100	98.3	99.0	99.3	99.0	99.3	98.3	99.7	96.7	99.0	97.7	98.7	99.7	96.0	98.0	97.7	98.0	99.3	91.3	93.7	97.3	95.3	95.0
Avg.	100	99.4	100	99.3	99.8	96.0	98.1	99.5	98.3	98.6	94.0	98.2	99.1	98.9	97.2	93.0	97.1	97.4	98.3	97.6	88.7	96.6	98.5	96.9	96.3
s.d.	0.0	0.4	0.0	0.9	0.5	3.0	2.5	0.6	2.7	2.0	3.7	2.0	0.9	0.6	2.7	3.0	2.6	2.4	1.5	1.7	5.1	3.7	1.6	3.1	2.4
R.I.		-0.6	0.0	-0.7	-0.2		1.9	3.4	2.2	2.4		4.6	5.5	5.3	3.4		4.3	4.6	5.6	4.9		9.1	11.2	9.4	8.8
P		0.02	-	0.07	0.2		0.06	0.04	0.1	0.02		0.05	0.03	0.01	0.007		0.007	0.03	0.006	0.007		0.03	0.009	0.02	0.03

S	-6 dB SNR					-9 dB SNR					-12 dB SNR					-15 dB SNR				
	Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_			
		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D
AB	91.3	98.7	98.0	98.0	98.3	88.7	99.7	97.6	96.7	98.0	83.3	94.6	93.0	94.3	94.3	76.7	89.3	87.0	89.7	86.0
SS	74.7	95.7	92.3	96.7	94.7	67.0	85.0	88.3	87.3	79.0	59.3	79.3	79.0	82.7	65.7	59.7	75.3	73.3	82.7	68.0
VK	84.7	95.3	96.3	98.0	93.3	73.3	93.7	94.3	93.0	87.7	70.0	84.7	86.7	85.7	82.0	51.3	76.7	88.0	86.3	78.7
AC	85.7	92.3	92.0	95.3	90.0	86.3	90.0	87.7	89.3	87.7	74.0	84.0	87.3	85.3	86.3	63.3	71.3	73.7	72.0	71.3
JK	91.3	94.7	93.0	96.7	94.3	89.3	91.0	93.7	95.3	93.3	83.7	86.3	87.0	88.3	90.7	71.7	79.3	85.7	88.3	77.3
Avg	85.5	95.3	94.3	96.9	94.1	81.0	91.9	92.3	92.3	89.1	74.0	85.8	86.6	87.3	83.8	64.5	78.4	81.5	83.8	76.3
s.d.	6.8	2.3	2.7	1.1	3.0	10.1	5.4	4.2	4.0	7.1	10.2	5.6	5.0	4.4	11.1	10.0	6.8	7.4	7.1	7.0
R.I.		12.0	10.8	13.9	10.6		14.7	15.4	15.3	10.8		17.0	18.1	19.2	13.2		23.0	28.7	32.1	20.0
p		0.02	0.02	0.009	0.02		0.02	0.03	0.02	0.01		0.007	0.005	0.006	0.0		0.006	0.01	0.006	0.02

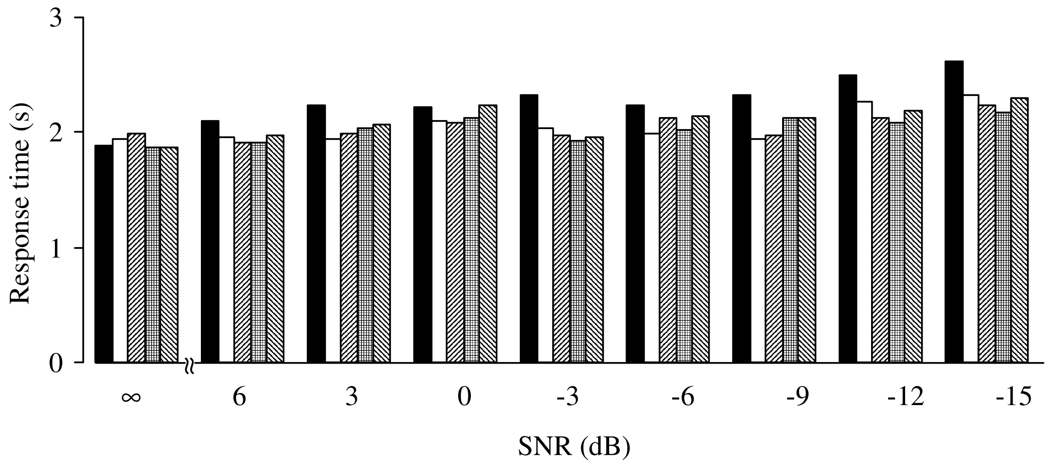


FIG. 5.8. Experiment III. (CS). Averaged response time (s). Su: Unprocessed speech. Sp_CS_A, Sp_CS_B, Sp_CS_C, Sp_CS_D correspond to processed speech with 16, 8, 4, 2 shifting respectively.

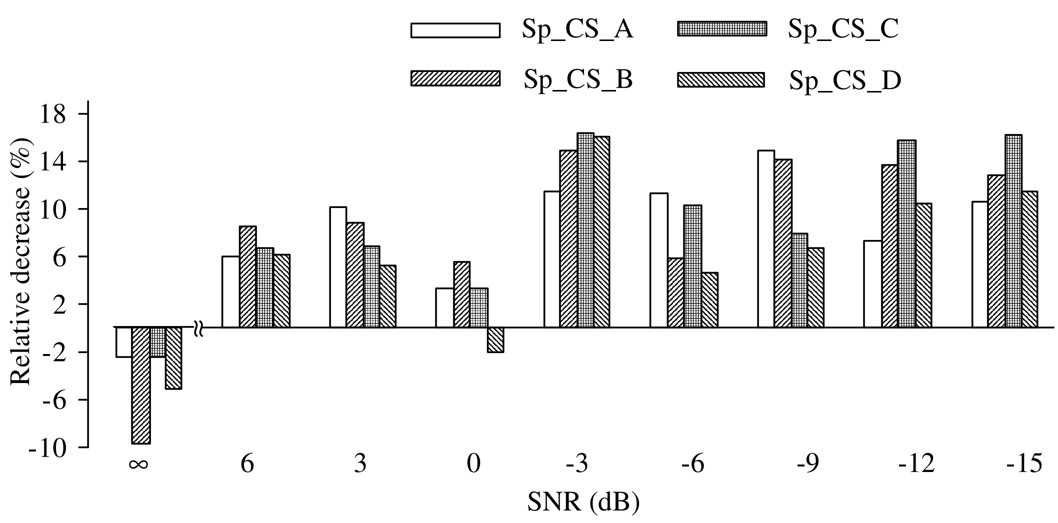


FIG. 5.9. Experiment III. (CS). Averaged relative decrease (%) in response times (s). Sp_CS_A, Sp_CS_B, Sp_CS_C, Sp_CS_D correspond to processed speech with 16, 8, 4, 2 shifting respectively.

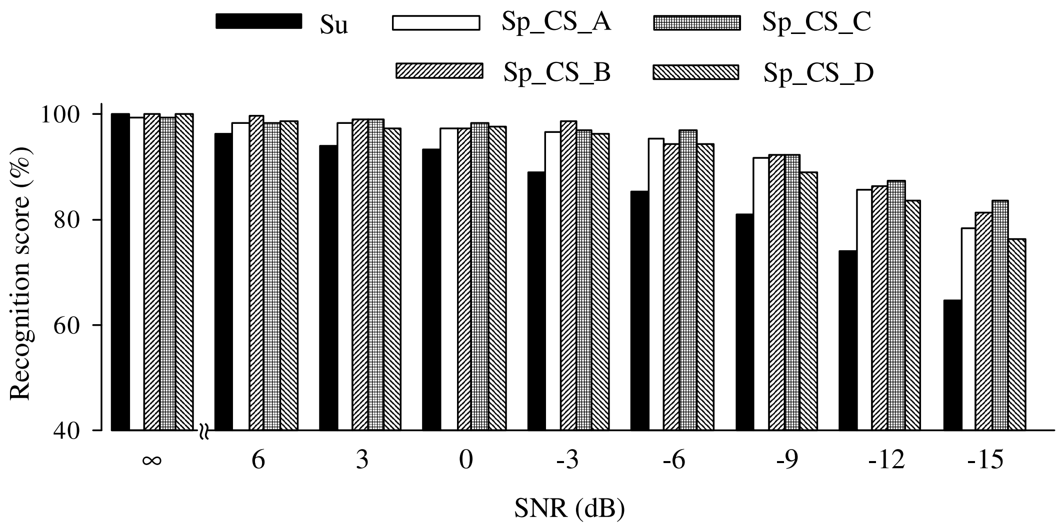


FIG. 5.10. Experiment III. (CS). Averaged recognition score (%). Su: Unprocessed speech. Sp_CS_A, Sp_CS_B, Sp_CS_C, Sp_CS_D correspond to processed speech with 16, 8, 4, 2 shifting respectively.

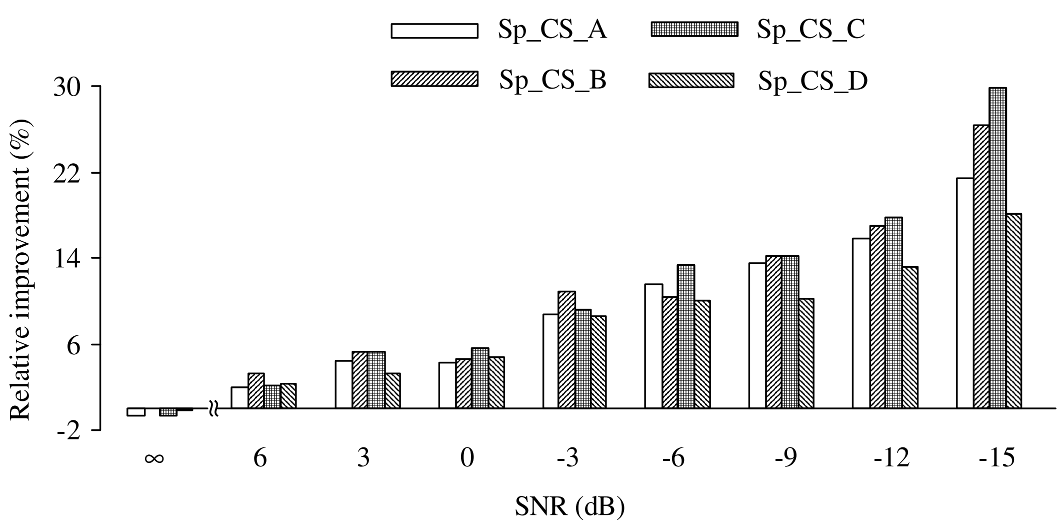


FIG. 5.11. Experiment III. (CS). Averaged relative improvement (%) in recognition score. Sp_CS_A, Sp_CS_B, Sp_CS_C, Sp_CS_D correspond to processed speech with 16, 8, 4, 2 shifting respectively.

For 8 shifting corresponding values were 3.4, 5.5, 4.6, 11.2, 10.8, 15.4, 18.1, 28.7%.

From the paired t-test (one-tailed) analysis, the improvements due to processing with shifting are statistically significant for 4 and 8 shifting under -12 and -15 dB SNR conditions.

5.4.4 Information transmission analysis

Table 5.5 gives the overall information transmitted for all subjects under all test conditions and a plot of values (averaged across five subjects) is shown in Fig. 5.12. Overall information transmitted averaged across the five subjects for unprocessed speech, varied from 100% under no-noise condition to 96, 94.6, 93, 89, 87, 83, 77, 70% for 6, 3, 0, -3, -6, -9, -12, -15 dB SNR conditions respectively. Under poor SNR conditions, processing resulted in an increase in information transmitted. Relative improvements of overall information transmitted for processed speech with 4 shifting were 2.6, 3.9, 4.8, 8.5, 10.9, 10.5, 13.2, 18.5% respectively. For 8 shifting corresponding values were 3.2, 4.3, 3.7, 10.1, 8.4, 12.0, 13.0, 16.6 %.

It is seen that, subject VK has low recognition score with low SNR. However, relative information transmitted for this subject is not lower than for other subjects. This indicates that, errors in reception by this subject are not randomly distributed. With dichotic presentation, relative information transmitted is better for this subject than for the other subjects.

Table 5.5 also gives the relative information transmitted for features of voicing, place, manner, nasality, frication, and duration.

TABLE 5.5. Experiment III. (CS). Relative information transmitted (%) for Unprocessed Speech (Su) and for Processed Speech Sp_CS_A, Sp_CS_B, Sp_CS_C, Sp_CS_D corresponding to 4 shiftings 16, 8, 4, and 2 respectively for (a) overall, and feature groupings: (b) voicing, (c) place, (d) manner, (e) nasality, (f) frication, and (g) duration. S: Subject.

(a) Overall

S	∞ SNR					6 dB SNR					3 dB SNR					0 dB SNR					-3 dB SNR					-6 dB SNR					-9 dB SNR					-12 dB SNR					-15 dB SNR				
	Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_			
		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D
AB	100	99	100	100	100	98	100	99	100	100	98	99	99	99	99	95	98	99	99	99	94	99	99	100	97	91	98	97	97	98	86	99	96	96	97	81	92	91	93	92	75	89	84	87	83
SS	100	100	100	98	100	95	97	98	96	95	95	98	98	98	96	89	96	97	97	97	84	99	99	98	99	79	96	93	96	95	74	87	89	88	80	69	80	80	84	71	68	77	73	80	69
VK	100	99	100	100	100	92	92	100	100	98	91	100	100	97	93	93	99	98	100	96	85	97	99	98	96	85	94	96	97	92	79	94	95	92	86	70	82	87	83	83	62	77	87	84	77
AC	100	98	100	99	100	97	100	100	99	99	93	95	99	99	98	94	94	93	96	95	92	92	96	92	92	88	91	91	94	89	86	89	89	87	85	80	83	86	85	84	70	73	76	74	70
JK	100	99	100	98	98	99	98	99	98	99	96	98	97	98	99	96	97	97	97	99	91	93	97	95	95	92	95	93	97	93	90	92	94	94	93	84	87	88	88	91	74	82	85	87	75
Avg.	100	99	100	99	99.6	96	97.4	99	98.6	98	94.6	98	98.6	98	97	93	97	97	98	97	89	96	98	96.6	96	87	95	94	96	93.4	83	92	93	91.4	88	77	85	86.4	87	84	70	80	81	82.4	75

(b) Feature: voicing

S	∞ SNR					6 dB SNR					3 dB SNR					0 dB SNR					-3 dB SNR					-6 dB SNR					-9 dB SNR					-12 dB SNR					-15 dB SNR				
	Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_			
		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D
AB	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	97	100	100	95	100	100	100	100	93	97	90	100	95	100	97	89	100	93
SS	100	100	100	100	100	97	100	100	97	100	100	100	100	97	100	100	100	100	100	100	97	100	100	100	100	100	100	97	100	100	97	95	100	100	95	97	100	97	94	97	86	94	97	95	93
VK	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	96	100	100	100	100	100	100	100	96	88	95	88	95	95	91	91	100	88	78
AC	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	95	100	100	100	95	100	96	100	100	95	100	82	88	82	68	82
JK	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	92	100	100	88	100	95	95	85	66	76
Avg.	100	100	100	100	100	99	100	100	99	100	100	100	100	99	100	100	100	100	100	100	99	100	100	100	100	100	99	99	100	99	98	99	100	99	98	93	98	95	94	97	91	93	91	83	84

(c) Feature: place

S	∞ SNR					6 dB SNR					3 dB SNR					0 dB SNR					-3 dB SNR					-6 dB SNR					-9 dB SNR					-12 dB SNR					-15 dB SNR				
	Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_			
		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D
AB	100	100	100	100	100	97	100	98	100	100	97	97	98	98	97	83	98	97	98	98	80	98	98	100	96	71	94	92	90	95	61	98	89	86	90	51	84	75	80	86	47	65	66	69	63
SS	100	100	100	97	100	98	98	100	100	93	95	100	98	100	83	79	98	97	100	97	63	98	98	97	95	34	85	75	90	82	24	59	62	59	38	11	49	41	51	27	14	35	39	55	28
VK	100	100	100	100	100	95	87	100	100	93	94	100	100	97	86	93	98	100	100	93	71	91	97	94	92	68	85	88	92	81	41	82	85	79	65	35	59	60	57	57	12	39	62	63	46
AC	100	98	100	98	100	89	100	100	98	98	80	82	98	98	93	77	82	77	86	85	71	74	86	75	78	63	74	73	80	66	55	70	65	67	63	34	54	64	57	58	27	42	42	38	37
JK	100	95	100	100	97	98	95	97	93	98	90	95	90	93	98	86	94	90	91	97	79	84	89	84	85	74	84	76	87	81	68	74	78	82	76	53	60	62	66	70	39	50	59	70	43
Avg.	100	98.6	100	99	99.4	95.4	96	99	98	96.4	91	95	97	97	91.4	83.6	94	92	95	94	73	89	94	90	89	62	84.4	81	88	81	50	77	76	74.6	66.4	37	61.2	60.4	62	60	29	46	54.6	59	43.4

(d) Feature: manner

S	∞ SNR					6 dB SNR					3 dB SNR					0 dB SNR					-3 dB SNR					-6 dB SNR					-9 dB SNR					-12 dB SNR					-15 dB SNR				
	Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_			
		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D
AB	100	96	100	100	100	98	100	100	100	100	96	100	100	100	100	98	96	96	100	98	98	98	98	100	92	83	96	93	95	93	86	100	95	94	95	82	89	94	97	86	61	88	82	88	75
SS	100	100	100	95	100	78	92	94	81	89	81	91	94	95	100	83	81	92	89	90	76	100	100	96	100	76	98	98	97	98	57	92	92	97	81	58	80	68	87	57	58	74	66	77	57
VK	100	98	100	100	100	81	83	100	100	100	76	100	100	93	88	80	100	94	100	92	74	100	100	98	93	74	90	97	98	85	69	92	96	91	76	60	78	79	86	71	54	84	86	76	62
AC	100	96	100	100	100	100	100	100	100	100	96	92	98	98	94	100	95	84	96	90	96	91	100	95	93	95	95	88	92	82	81	87	85	88	80	64	86	84	80	77	52	69	68	71	65
JK	100	100	100	92	98	98	100	100	100	100	98	100	100	100	100	100	97	100	100	98	86	91	100	98	95	88	100	96	100	92	88	98	98	96	92	85	85	90	87	84	64	86	85	83	71
Avg.	100	98	100	97.4	99.6	91	95	98.8	96	97.8	89	96.6	98.4	97	96.4	92	93.8	93.2	97	93.6	86	96	99.6	97.4	94.6	83	95.8	94.4	96.4	90	76	94	93	93	85	70	84	83	87.4	75	58	80	77.4	79	66

(e) Feature: nasality

S	∞ SNR					6 dB SNR					3 dB SNR					0 dB SNR					-3 dB SNR					-6 dB SNR					-9 dB SNR					-12 dB SNR					-15 dB SNR				
	Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_			
		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D
AB	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	85	96	100	95	65	89	75	81	78
SS	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	88	92	91	88	96	85	80	88	83
VK	100	100	100	100	100	97	95	100	100	100	100	100	100	100	97	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	97	88	97	97	90	100	88	88	97	97	97
AC	100	100	100	100	100	100	100	100	100	100	97	100	100	100	100	100	100	100	100	97	100	100	100	100	97	100	97	100	100	97	100	100	100	97	100	100	97	97	95	95	87	95	100	89	82
JK	100	100	100	100	100	100	100	100	100	100	97	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	97	100	100	97	100	97	97	100	97	100	95	84	88	78	71
Avg.	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	99	100	100	99	100	100	100	99	99	95	94	95	94	96	86	90	84	78	79

(f) Feature: frication

S	∞ SNR					6 dB SNR					3 dB SNR					0 dB SNR					-3 dB SNR					-6 dB SNR					-9 dB SNR					-12 dB SNR					-15 dB SNR				
	Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_			
		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D
AB	100	94	100	100	100	97	100	100	100	100	96	100	100	100	100	97	94	94	100	97	97	97	97	100	88	73	94	89	92	89	78	100	92	91	92	71	85	92	95	78	57	81	73	85	66
SS	100	100	100	92	100	62	86	90	67	81	67	84	90	92	100	71	67	86	81	82	59	100	100	94	100	61	97	97	95	97	30	86	87	95	69	29	71	52	85	39	33	66	53	66	33
VK	100	97	100	100	100	69	72	100	100	100	61	100	100	89	80	68	100	90	100	88	58	100	100	97	89	58	87	95	97	76	50	87	94	85	62	41	64	74	79	56	27	77	76	69	51
AC	100	94	100	100	100	100	100	100	100	100	94	86	97	97	90	100	92	73	94	82	93	85	100	92	89	91	92	81	87	74	69	79	74	81	68	42	78	74	69	62	32	54	53	61	50
JK	100	100	100	87	97	97	100	100	100	100	97	100	100	100	100	100	95	100	100	97	76	86	100	97	92	79	100	94	100	87	79	97	97	94	87	79	76	83	84	74	42	78	81	74	59
Avg.	100	97	100	96	99.4	85	91.6	98	93.4	96	82	94	97.4	95.6	94	87	89.6	88.6	95	89	76.6	93.6	99.4	96	91.6	72.4	94	91.2	94.2	84.6	61	90	89	89	76	52.4	75	75	82.4	62	38	71	67	71	52

(g) Feature: duration

S	∞ SNR					6 dB SNR					3 dB SNR					0 dB SNR					-3 dB SNR					-6 dB SNR					-9 dB SNR					-12 dB SNR					-15 dB SNR				
	Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_				Su	Sp_CS_			
		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D		A	B	C	D
AB	100	100	100	100	100	100	100	95	100	100	93	91	95	95	91	86	95	91	95	96	87	95	100	100	95	65	91	81	87	91	59	100	87	85	94	50	95	84	88	84	49	85	63	76	81
SS	100	100	100	100	100	95	95	100	100	94	100	100	95	100	100	84	96	91	100	96	83	96	100	100	91	47	80	77	94	73	16	65	62	72	50	8	65	42	68	31	7	54	41	64	29
VK	100	100	100	100	100	78	66	100	100	88	89	100	100	92	81	79	95	100	100	83	60	91	100	100	96	58	94	94	100	87	36	100	100	83	70	42	57	79	64	62	7	48	70	60	54
AC	100	100	100	100	100	100	100	100	100	96	85	92	96	96	96	88	89	69	94	86	86	84	96	76	89	61	82	87	85	83	58	79	79	70	67	26	70	81	81	70	23	63	46	50	52
JK	100	88	100	96	89	100	88	91	96	100	91	89	91	88	100	89	94	88	81	96	68	76	76	79	96	88	89	70	95	85	81	68	78	85	91	48	52	68	74	94	35	54	61	64	63
Avg.	100	97.6	100	99	98	95	90	97	99	95.6	92	94.4	95.4	94	93.6	85	94	88	94	91.4	77	88.4	94.4	91	93.4	64	87	82	92	84	50	82.4	81.2	79	74.4	35	68	71	75	68	24	61	56	63	56

Voicing: Perception of this feature was degraded at higher noise levels. Relative information transmitted for unprocessed speech for -12 and -15 dB SNR condition were 93 and 91%. For processed speech with 8 and 16 shifting the relative improvements were 2 and 5.6% at -12 dB SNR. At -15 dB SNR condition, relative improvement for processed speech with 16 shifting was 3%.

Place: Figure 5.13 shows the relative information transmitted (averaged across five subjects) for unprocessed and processed speech. For unprocessed speech, it varied from 100% under no-noise condition to 95.4, 91.0, 83.6, 73.0, 62.0, 50.0, 37.0, 29.0% for 6, 3, 0, -3, -6, -9, -12, -15 dB SNR conditions respectively. Relative improvements for processed speech with 4 shifting were 3.1, 7.1, 13.9, 24.7, 54.3, 64.4, 115.1, 177.0% respectively. For 8 shifting corresponding values were 3.9, 6.6, 10.4, 29.7, 39.6, 68.9, 99.3, 148.5%.

Manner: Figure 5.14 shows the information transmitted (averaged across five subjects) for manner feature, for unprocessed and processed speech. For unprocessed speech, it varied from 100% under no-noise condition to 91, 89, 92, 86, 83, 76, 70, 58% for 6, 3, 0, -3, -6, -9, -12, -15 dB SNR conditions respectively. Relative improvements for processed speech with 4 shifting were 6.2, 9.5, 6.0, 14.7, 17.0, 25.8, 27.8, 36.8% respectively. For 8 shifting corresponding values were 9.6, 11.2, 2.1, 17.4, 14.8, 25.4, 20.1, 34.2%.

Nasality: Perception of this feature was degraded at higher noise levels. Relative information transmitted for unprocessed speech was 86% at -15 dB SNR condition. For processed speech, relative improvements with 8 and 16 shifting were 3.3 and 4.6%.

Frication: Figure 5.15 shows the relative information transmitted (averaged across five subjects) for unprocessed and processed speech. For unprocessed speech, it varied from 100% under no-noise condition to 85, 82, 87, 76.6, 72.4, 61, 52.4, 38% for 6, 3, 0, -3, -6, -9, -12, -15 dB SNR conditions respectively. Relative improvements for processed speech with 4 shifting were 11.8, 18.7, 11.6, 31.2, 34.2, 67.9, 78.0, 94.3% respectively. For 8 shifting the corresponding values were19.2, 21.7, 4.67, 36.2, 30.5, 65.2, 54.1, 85.7%.

Duration: Figure 5.16 shows the relative information transmitted (averaged across five subjects) for unprocessed and processed speech. For unprocessed speech, it varied from 100% under no-noise condition to 95, 92, 85, 77, 64, 50, 35, 24% for 6, 3, 0, -3, -6, -9, -12, -15 dB SNR conditions respectively. Relative improvements for processed speech with 4 shifting were 5.9, 3.0, 10.8, 21.3, 50.7, 110.1, 228.8, 365.4%respectively. With 8 shifting corresponding values were 3.9, 4.5, 3.6, 25.1, 34.5, 109.0, 166.9, 317.7%.

Subjects SS and VK had great difficulty in perception of place and duration features under poor listening condition. With processing, the reception of these features improved for all the subjects, and particularly for subjects SS and VK.

5.5 Discussion

In this chapter, implementation of the speech processing scheme for combined spectral and temporal splitting and its evaluation has been presented. Combined splitting is achieved by using cyclically swept time-varying comb filters. Tests were conducted on five normal hearing subjects with simulated sensorineural hearing impairment. Effect of processing with different number of shifting of time-varying comb filters was studied. For each subject, in each test condition, subject’s responses were stored in the form of response time statistics and confusion matrix. Percentage correct recognition score, relative information transmitted for different features, and subject’s response times for unprocessed and processed speech were analyzed. Subjects assessment for quality of speech signal was also compiled.

From the response times, it can be seen that under masking noise condition, there is moderate decrease in response time due to processing. Decreases in response times are generally higher with 8 and 16 shifting. Hence the scheme of combined splitting helps in reducing the load on perception for normal hearing persons under adverse listening condition.

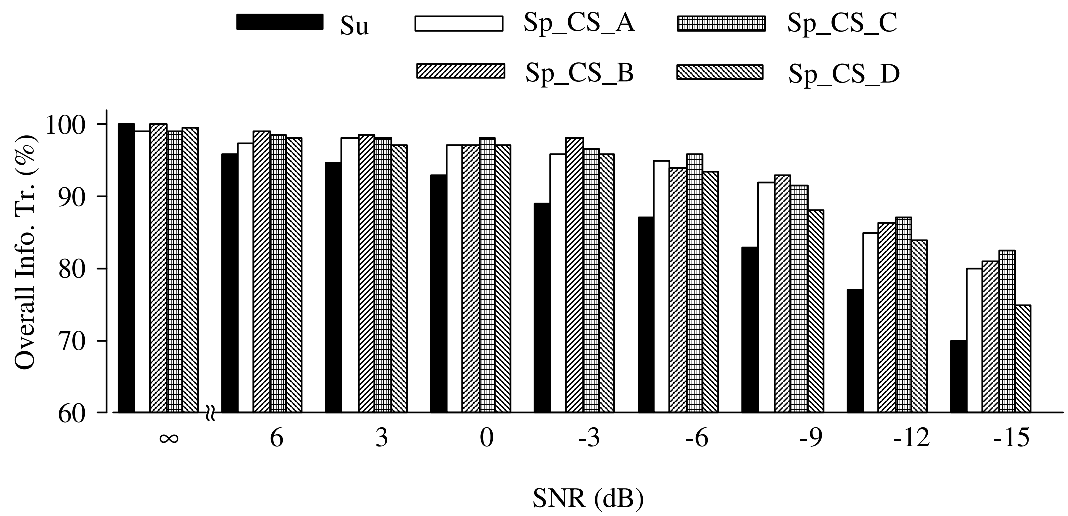


FIG. 5.12. Experiment III. (CS). Averaged overall relative information transmitted (%). Su: Unprocessed speech. Sp_CS_A, Sp_CS_B, Sp_CS_C, Sp_CS_D correspond to processed speech with 16, 8, 4, 2 shifting respectively.

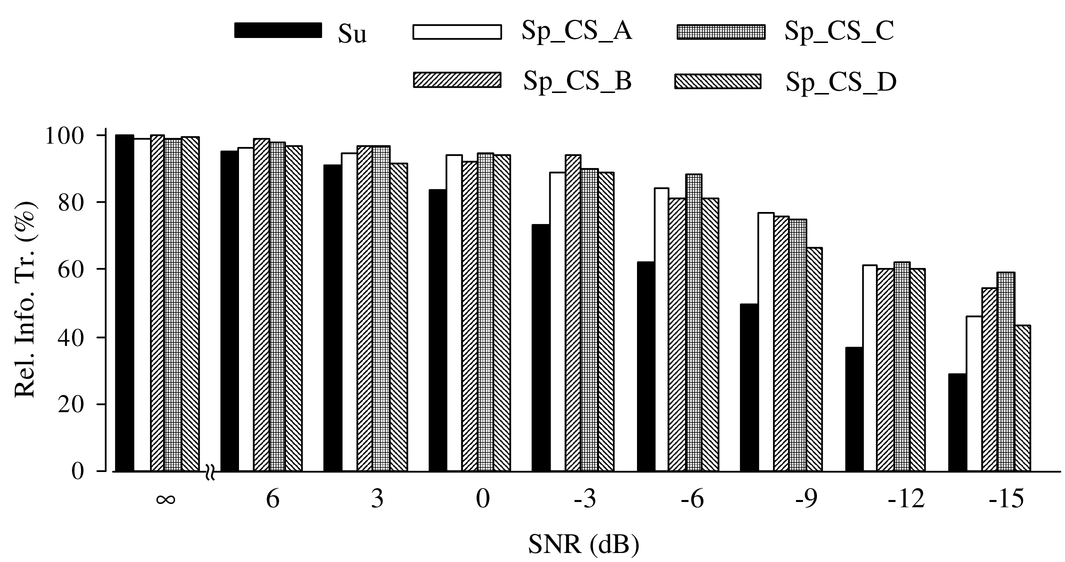


FIG. 5.13. Experiment III. (CS). Averaged relative information transmitted (%) for place feature. Su: Unprocessed speech. Sp_CS_A, Sp_CS_B, Sp_CS_C, Sp_CS_D correspond to processed speech with 16, 8, 4, 2 shifting respectively.

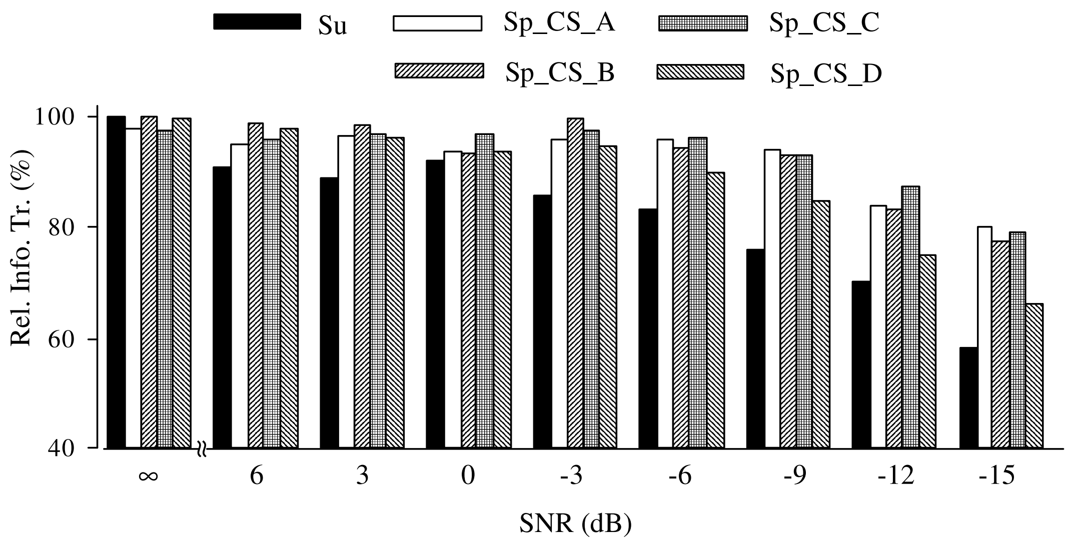


FIG. 5.14. Experiment III. (CS). Averaged relative information transmitted (%) for manner feature. Su: Unprocessed speech. Sp_CS_A, Sp_CS_B, Sp_CS_C, Sp_CS_D correspond to processed speech with 16, 8, 4, 2 shifting respectively.

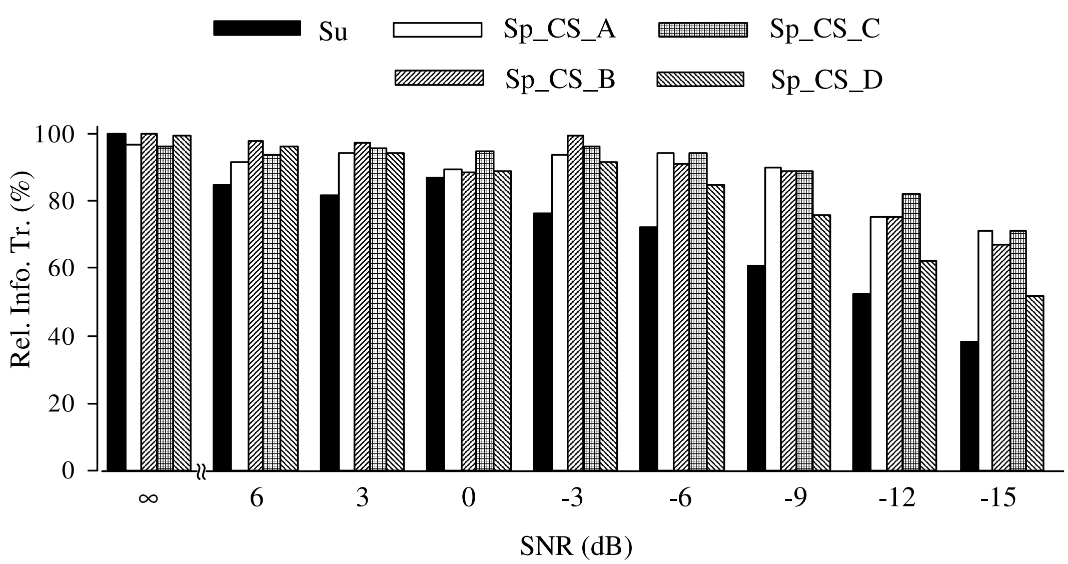


FIG. 5.15. Experiment III. (CS). Averaged relative information transmitted (%) for frication feature. Su: Unprocessed speech. Sp_CS_A, Sp_CS_B, Sp_CS_C, Sp_CS_D correspond to processed speech with 16, 8, 4, 2 shifting respectively.

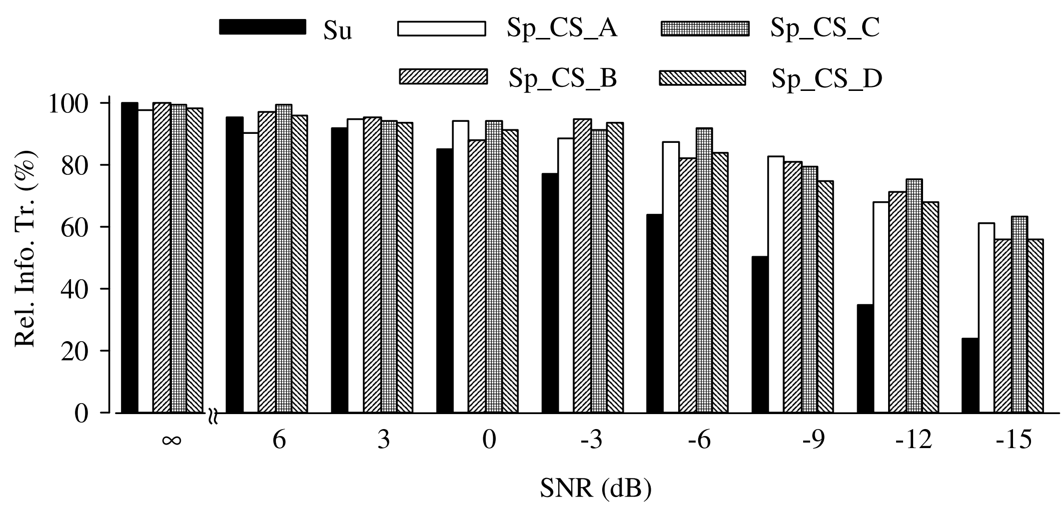


FIG. 5.16. Experiment III. (CS). Averaged relative information transmitted (%) for duration feature. Su: Unprocessed speech. Sp_CS_A, Sp_CS_B, Sp_CS_C, Sp_CS_D correspond to processed speech with 16, 8, 4, 2 shifting respectively.

From the recognition scores, it was found that as the noise level increases, scores decrease for unprocessed speech. Processing improved the recognition scores. Improvements are generally higher for 4 and 8 shifting.

Overall information transmitted and information transmitted for different speech features was obtained. Overall information transmitted is higher for 4 and 8 shifting. It is observed that, among the features, there is maximum improvement for duration and place features, indicating that the processing helps in reducing the effect of temporal and spectral masking. Highest improvement is seen for duration feature. For both place and duration features, improvements are generally higher for 4 and 8 shifting. Relative information transmitted for frication and manner features also improved with processing. For manner feature, improvements are almost same for 4, 8, and 16 shifting. For frication feature, improvements are higher for 4 shifting.

The scheme for splitting speech signal using time varying comb filters provides better speech intelligibility for normal persons with simulated hearing loss and the improvements increase under adverse listening condition. The improvement in the perception of place and duration features show that the scheme has helped in reducing the effect of spectral and temporal masking. Thus the investigation has shown that the scheme has the potential of improving speech perception for persons using binaural hearing aids. Further investigations will help in establishing optimal values of the parameters for combined splitting.

Chapter 6

Evaluation with persons having sensorineural hearing impairment

6.1 Introduction

An overall evaluation of the schemes of temporal splitting and combined spectral and temporal splitting was carried out along with the scheme of spectral splitting by conducting listening tests on subjects with sensorineural hearing loss. For the scheme of spectral splitting, a pair of complementary comb filters from the filter sets in the scheme of combined splitting was used. Each comb filter has 9 pass bands corresponding to critical band auditory filters. Magnitude response of these filters was optimized for low perceived spectral distortion by having low pass band ripple, high stop band attenuation, and perceptual balance at inter-band crossovers. For the schemes of temporal splitting and combined splitting, evaluation was done by considering the optimal parameters that gave maximum improvement for normal hearing subjects with simulated sensorineural hearing loss. These tests are referred to as "Experiment IV" in all the tables and figures.

6.2 Processing

The three processing schemes were:

a) temporal splitting with 20 ms inter-aural switching period with 70% duty cycle and trapezoidal fading with 2 ms transition duration,

b) spectral splitting with comb filters based on auditory critical bands and responses optimized for low perceived spectral distortion,

c) combined spectral and temporal splitting using comb filters with sweep cycle of 20 ms and 4 and 8 filter sets for sweeping the magnitude response.

Listening tests involved diotic presentation of unprocessed speech and dichotic presentation of processed speech. Figures 6.1, 6.2, 6.3 show the wideband (Δf = 300 Hz) spectrograms of random noise for the scheme of temporal splitting, spectral splitting, and combined splitting with 4 shifting, respectively. Figures 6.4, 6.5, 6.6 show the wideband spectrograms of speech syllable /*a*s*a*/ for the three schemes. Each figure shows spectrograms of unprocessed signal and the two processed signals having complementary symmetry.

6.3 Test material and subjects

The test material for consonantal identification was same as that used in the schemes of temporal splitting and combined splitting. Twelve English consonants /p, b, t, d, k, g, m, n, s, z, f, v/ were used in the VCV context with vowel /*a*/ as in father. For quality assessment, test material consisted of Marathi phrases of one-minute duration recorded by a male and female speaker.

Five subjects with sensorineural hearing loss (SA: M 49, BA: M 61, SK: M 41, KS: F 32, BS: F 38) participated in the listening test. All the subjects had bilateral mild to severe sensorineural hearing loss. Hearing thresholds of these subjects are given in Table 6.1. Subject SA has asymmetrical high frequency loss (less loss in one ear and more loss in the other ear). Subject BA has symmetrical low frequency hearing loss and subject SK has symmetrically sloping high frequency loss. The remaining two subjects KS and BS have more loss at high frequency.

6.4 Experimental procedure

Experimental set-up and experimental procedure for conducting the listening tests were same as described earlier in Chapter 4. In the test for consonantal identification, 12 stimuli were used and each stimulus was presented 5 times in random order (as described in Appendix D). Tests were conducted in an acoustically isolated chamber. Test results were response time statistics, stimulus-response confusion matrix, and percentage correct recognition score.

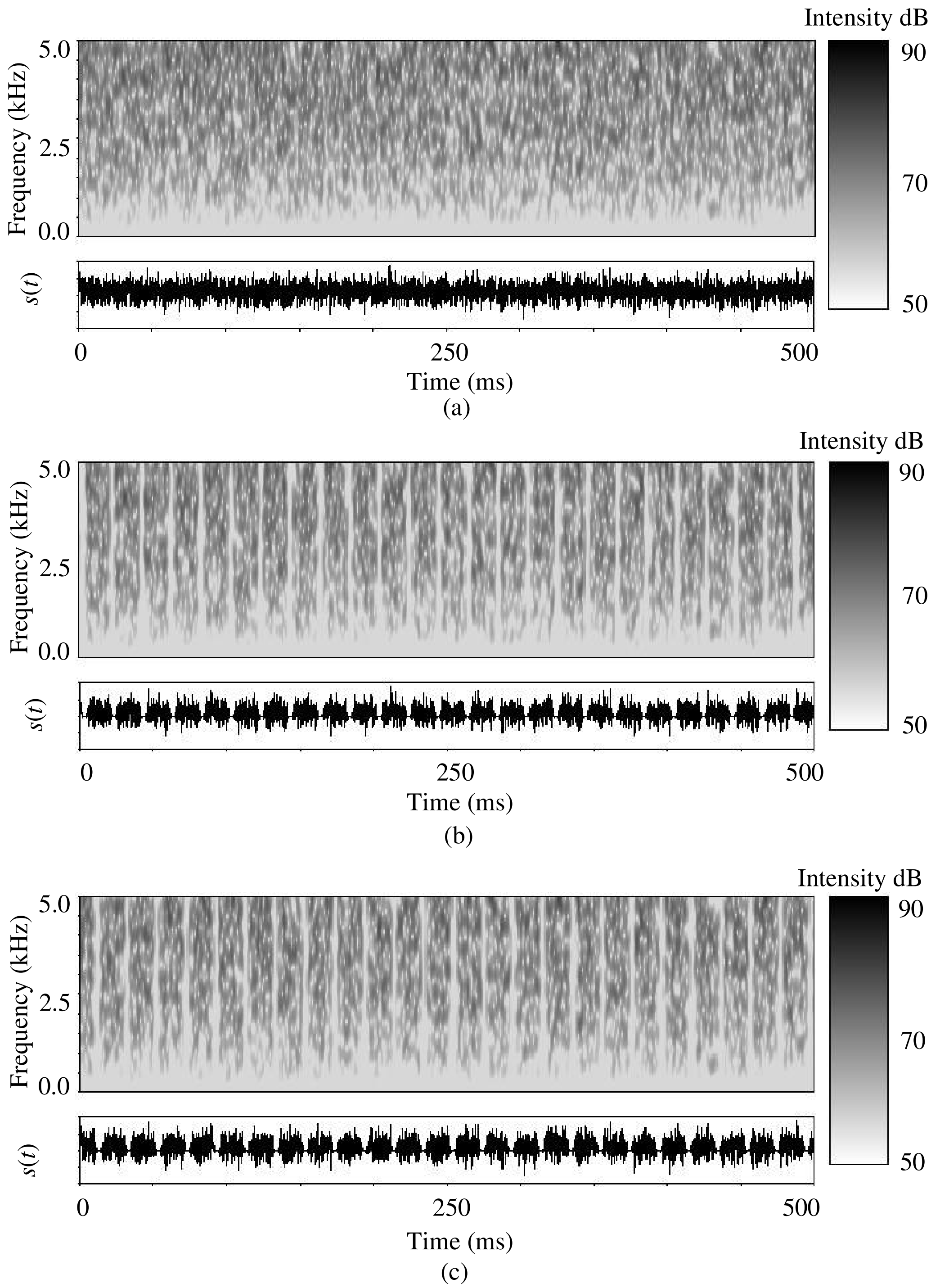


FIG. 6.1. Inter-aural switching with trapezoidal transition (TS_TR): Wide-band spectrograms ($\Delta f \approx 300$ Hz) of random noise. Inter-aural switching interval = 20 ms, duty cycle = 70% and transition duration = 2 ms. (a) unprocessed (b) processed, left ear (c) processed, right ear.

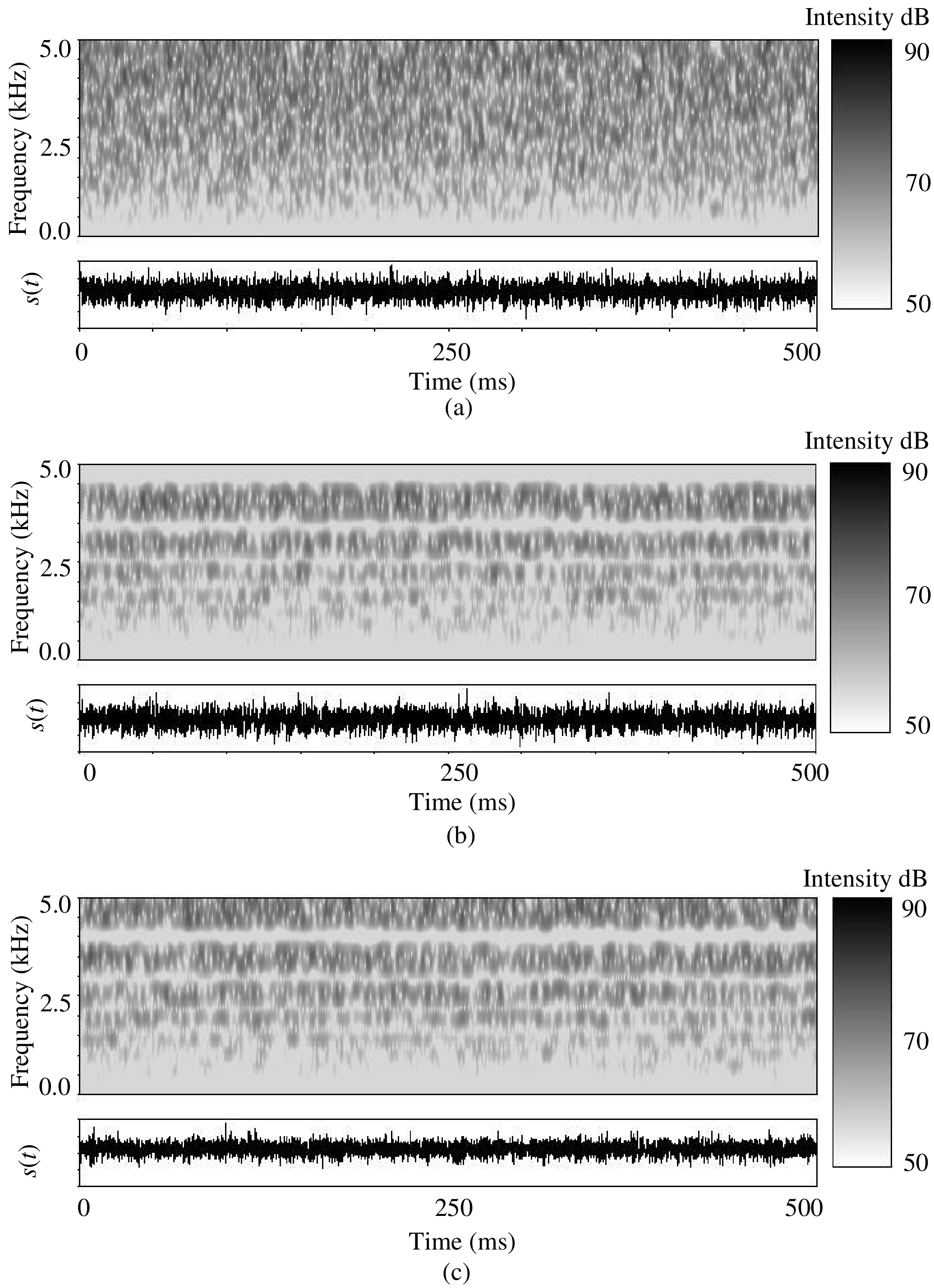


FIG. 6.2. Spectral splitting (SS): Wide-band spectrograms ($\Delta f \approx 300$ Hz) of random noise. (a) unprocessed (b) processed, left ear (c) processed, right ear.

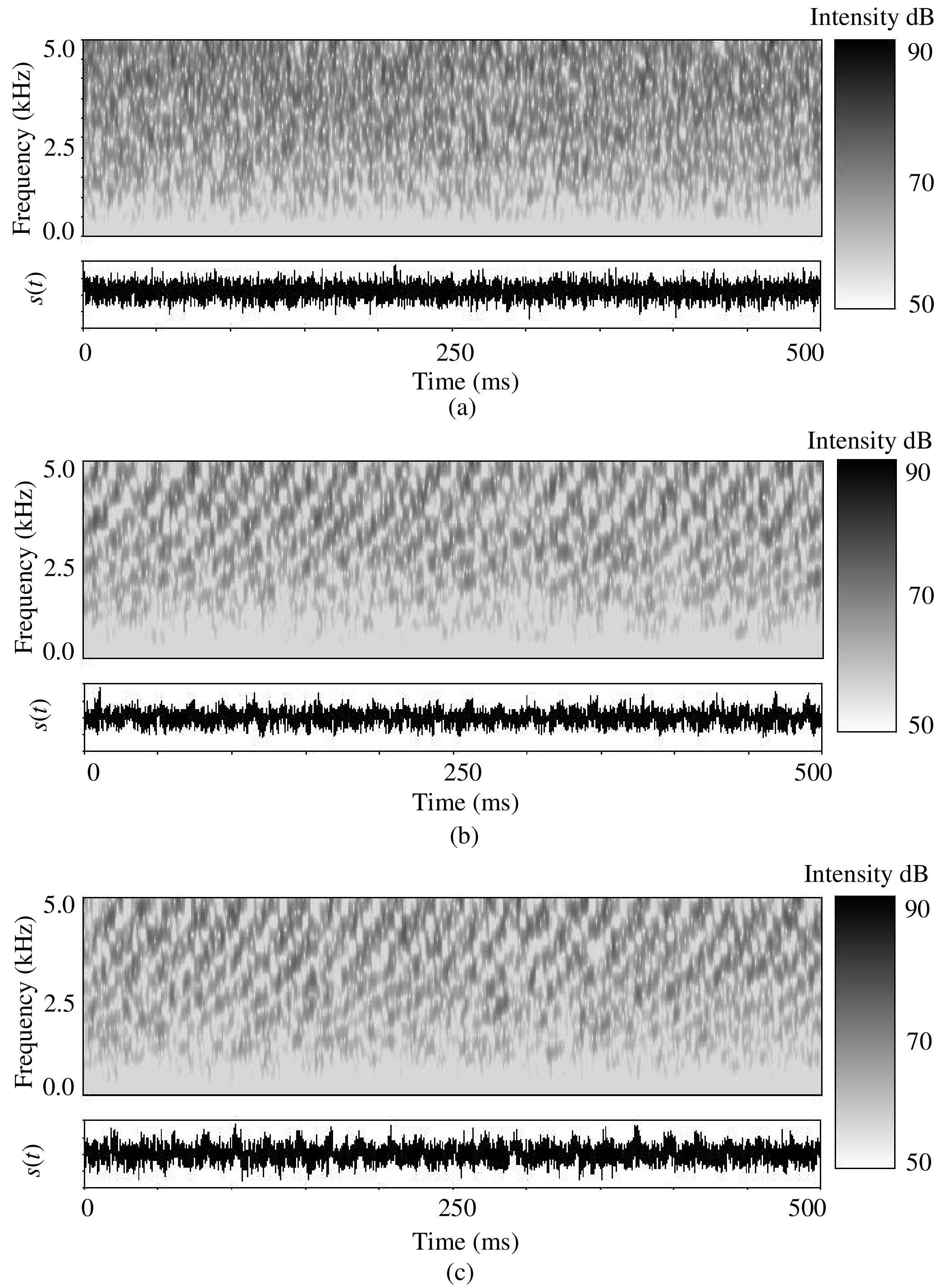


FIG. 6.3. Combined splitting (CS): Wide-band spectrograms ($\Delta f \approx 300$ Hz) of random noise. Sweep period = 20 ms and no. of shifting = 4. (a) unprocessed (b) processed, left ear (c) processed, right ear.

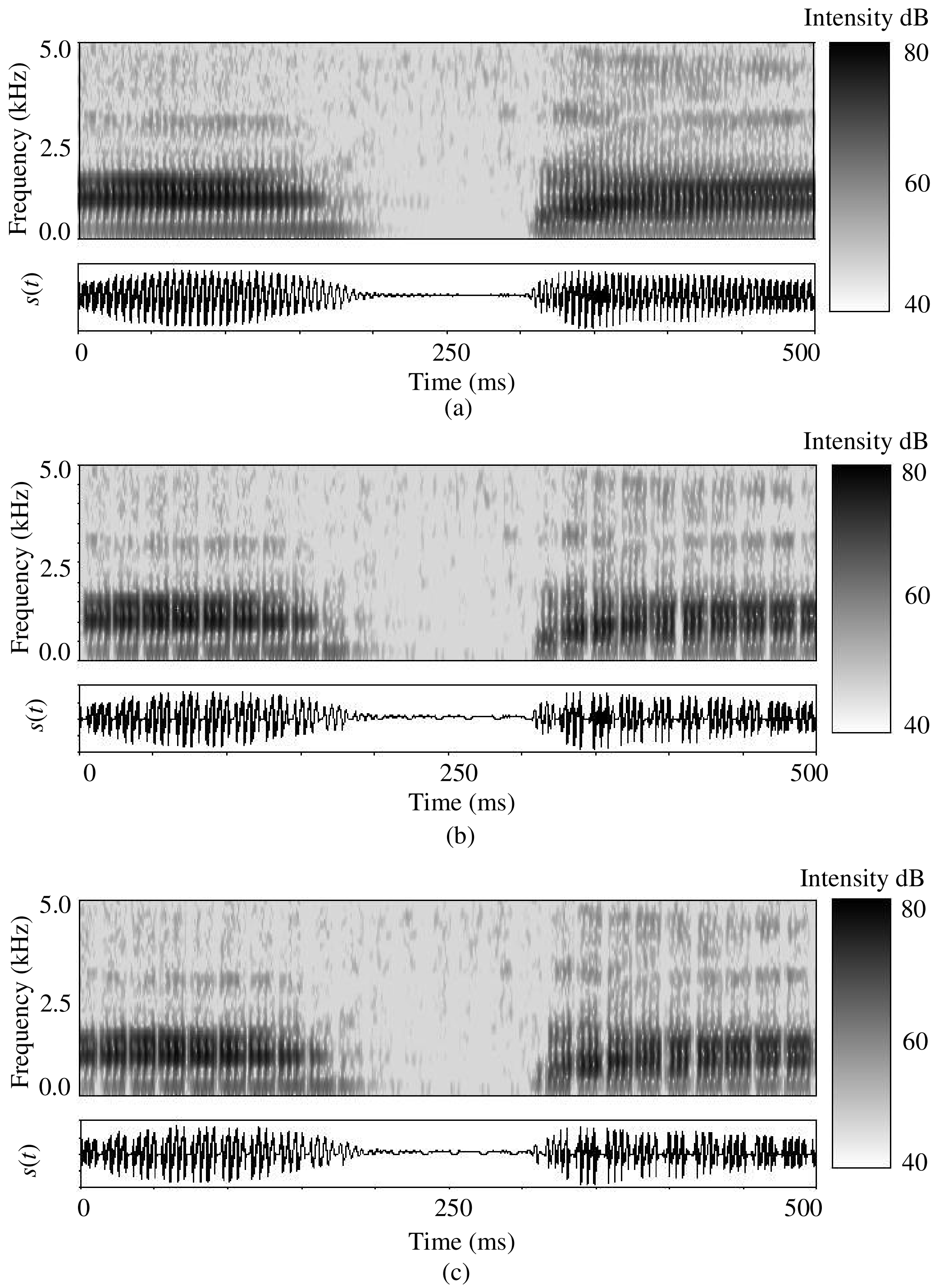


FIG. 6.4. Inter-aural switching with trapezoidal transition (TS_TR): Wide-band spectrograms ($\Delta f \approx$ 300 Hz) of speech syllable /*asa*/. Inter-aural switching interval = 20 ms, duty cycle = 70% and transition duration = 2 ms. (a) unprocessed (b) processed, left ear (c) processed, right ear.

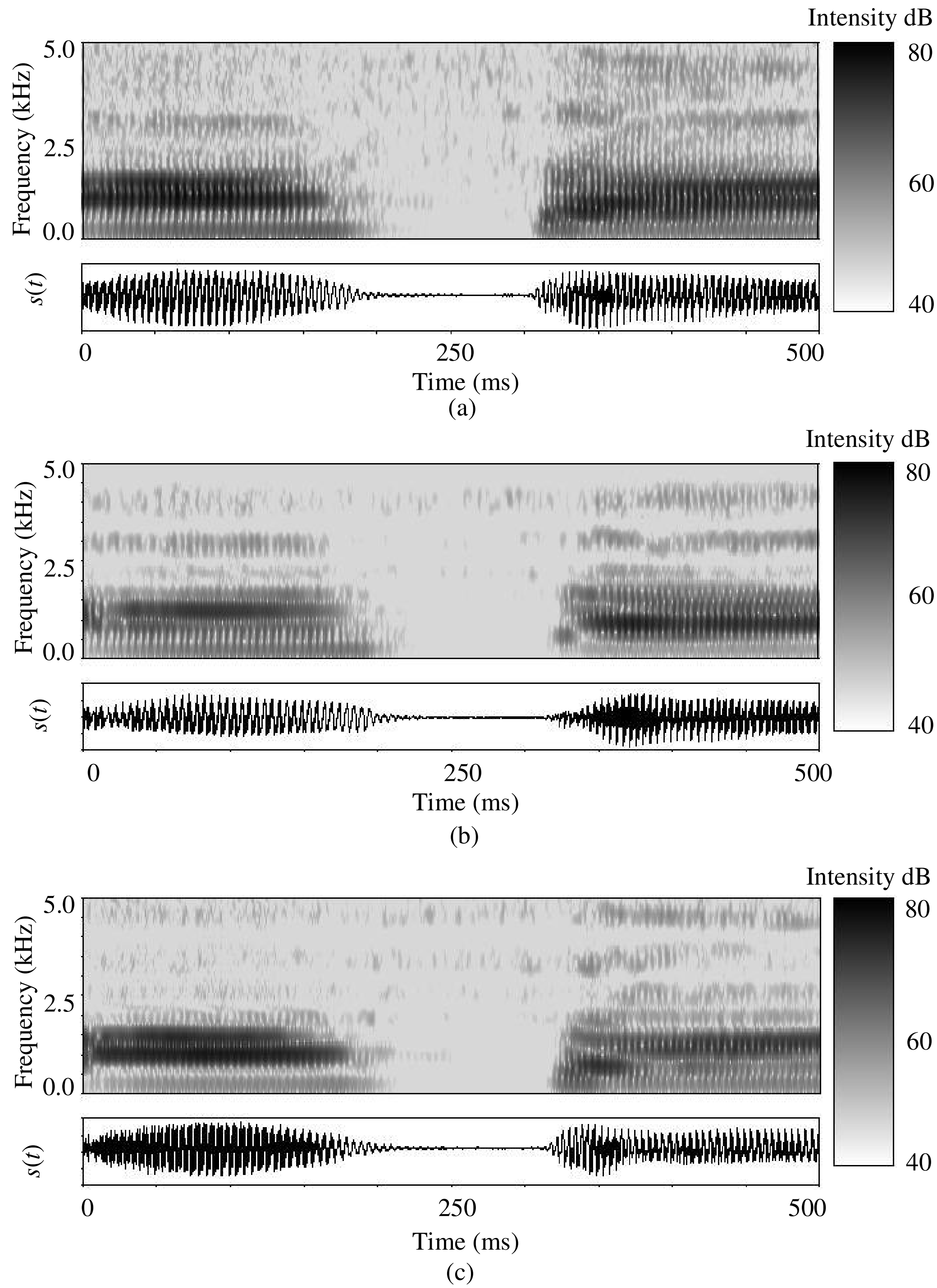


FIG. 6.5. Spectral splitting (SS): Wide-band spectrograms ($\Delta f \approx 300$ Hz) of speech syllable /*asa*/. (a) unprocessed (b) processed, left ear (c) processed, right ear.

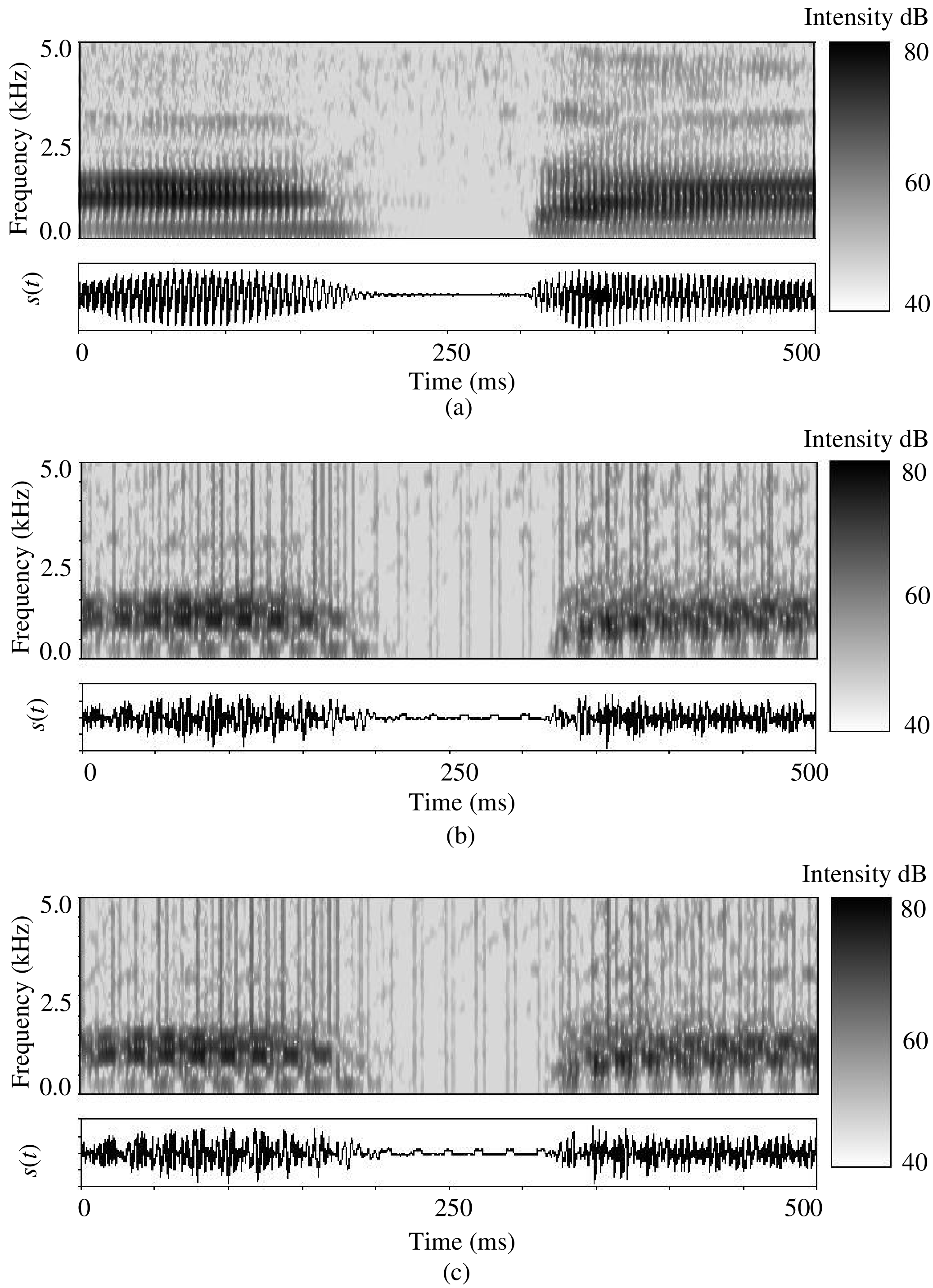


FIG. 6.6. Combined splitting (CS): Wide-band spectrograms ($\Delta f \approx 300$ Hz) of speech syllable /*asa*/. Sweep period = 20 ms and no. of shifting = 4. (a) unprocessed (b) processed, left ear (c) processed, right ear.

TABLE 6.1. Hearing thresholds of the subjects with bilateral hearing impairment. PTA: Average of pure tone thresholds in dB HL, taken at test frequencies 0.5, 1, 2 kHz. Speech presentation level measured as dB SPL for sustained /a/ set by the subject as the most comfortable level for the two ears individually.

Subject Code (Sex, Age)	Ear L=Left R=Right	Hearing thresholds (dB HL) Frequency (kHz)					PTA (dB)	Presentation level (dB SPL)
		0.25	0.5	1	2	4		
SA (M, 49)	L	50	45	45	45	35	45	111
	R	60	70	60	65	85	65	116
BA (M, 61)	L	55	65	50	40	45	52	97
	R	60	60	45	35	35	47	96
SK (M, 41)	L	75	75	85	90	100	83	110
	R	75	85	90	95	110	90	110
KS (F, 32)	L	60	45	100	110	120	85	107
	R	65	85	100	115	120	100	112
BS (F, 38)	L	75	80	105	110	120	98	114
	R	70	65	85	110	120	87	110

The confusion matrix shows the stimuli along rows and responses along columns. Each entry in the cell shows the number of times a stimulus-response pair occurs in the test. There were a total of 5 test conditions. A test run takes about 6–10 minutes. Hence for 5 test conditions, each subject took about 5–6 hours for completion of all tests. Test sessions for the five subjects were spread over a span of about one month depending upon the availability and willingness of the subjects.

For ascertaining the speech quality, a compilation of subject's qualitative assessment of the test stimuli processed with various test conditions was carried out. The test material consisted of a short passage (about 1-min duration) in Marathi (mother tongue of the subjects), recorded by a male and a female speaker. The subject's task was to rank the speech signal processed under various conditions (including the unprocessed version). A pair of sounds were presented in succession and the subject reported his/her preference. This process was repeated to obtain a ranking of all of them.

6.5 Test results

Results of listening tests for qualitative assessment and consonantal identification are presented in the following subsections. Response times were used to assess the effectiveness of the processing schemes in reducing load on perception. Paired t-test was carried out on response times and recognition scores to determine the level of significance due to processing. Information transmission analysis, a measure that is not affected by subject's response bias, was carried out. Information transmission analysis was also performed for consonantal features to determine the contribution of features in improving speech perception.

6.5.1 Quality assessment

Out of five subjects, SK was unable to provide reliable ranking for both the male and female speech. This subject had larger response time in consonantal identification test (as seen in the next subsection). Results of qualitative assessment of remaining four subjects are given in Table 6.2. From the table it can be seen that, for male speech, subjects SA, BA, and KS ranked the quality of spectral splitting as the highest, and ranked temporal and combined splitting lower than the unprocessed speech. Subject BS ranked unprocessed as the highest, followed by spectral splitting. For female speech, all the four subjects ranked the unprocessed speech as highest and combined splitting as lowest. Subjects SA, BA, and KS preferred spectral splitting next to unprocessed one. Temporal splitting was ranked as third by these subjects. For subject BS, both spectral and temporal splitting were ranked next to unprocessed speech.

6.5.2 Response time

Table 6.3 gives the response times for unprocessed and processed speech for the five subjects SA, BA, SK, KS, and BS. The table also provides relative decrease in response time

Table 6.2. Experiment IV. Quality assessment. S: Subject. Ranking of speech quality for unprocessed speech (Su), and processed speech corresponding to temporal splitting (SpT), spectral splitting (SpS), combine splitting with 4 shifting (Sp4), and combined splitting with 8 shifting (Sp8).

Speaker (M/F)	S	Su	SpT	SpS	Sp4	Sp8
M	SA	2	3	1	5	4
	BA	2	5	1	4	3
	KS	2	3	1	4	4
	BS	1	3	2	4	4
F	SA	1	3	2	4	4
	BA	1	3	2	4	5
	KS	1	3	2	4	4
	BS	1	2	2	4	4

(processed vs. unprocessed), significance level p from one-tailed t-test for individual subjects, and significance level p from one-tailed paired t-test.

Figure 6.7 shows the response times of the five subjects for unprocessed and processed speech. Figure 6.8 shows the percentage relative decrease in response times for processed speech for the five subjects. For unprocessed speech, response times varied from 3.24 s to 8.43 s. With processing response times decreased. Relative decrease in response times ranged from 3.4 to 36.1% and 4.9 to 40.5% for processed speech with temporal and spectral splitting respectively. For processed speech with combined splitting, relative decrease in response times ranged from 0.4 to 28.1% and 5.1 to 31.8% for 4 and 8 shifting respectively. Relative decreases in response times were statistically significant for the schemes of spectral splitting and combined splitting with 8 shiftings.

6.5.3 Recognition scores

The recognition scores for unprocessed and processed speech and percentage relative improvement for processed speech for the five subjects are given in Table 6.4. Figure 6.9 shows percentage recognition scores for unprocessed and processed speech. Relative improvement (%) for processed speech for the five subjects are shown in Fig. 6.10.

TABLE 6.3. Experiment IV. Response times for listening tests with hearing impaired subjects for Unprocessed Speech (Su) and Processed Speech corresponding to temporal splitting (SpT), spectral splitting (SpS), and combined splitting with 4 and 8 shifting (Sp4, Sp8). S: Subject, Tavg. = average response times (s), s.d. = standard deviation (s), R.D = relative decrease in % with respect to unprocessed. p: significance level for one-tailed t-test (unprocessed vs processed, n =5, df = 8). p (paired): significance level for one-tailed paired t-test (unprocessed vs processed, averaged across the subjects, n = 5, df = 4.

S		Su	SpT	SpS	Sp4	Sp8
SA	Tavg.	3.24	2.07	2.04	2.33	2.21
	s.d.	0.2	0.2	0.2	0.2	0.1
	R.D.		36.1	37.0	28.1	31.8
	P		0.0	0.0	0.0001	0.0
BA	Tavg.	3.83	3.33	2.82	3.43	2.87
	s.d.	0.3	0.6	0.3	0.6	0.3
	R.D.		13.1	26.4	10.4	25.1
	P		0.06	0.0003	0.1	0.0004
SK	Tavg.	8.43	7.40	5.02	6.57	6.03
	s.d.	1.6	0.7	0.5	1.5	0.4
	R.D.		12.2	40.5	22.1	28.5
	P		0.1	0.0008	0.05	0.005
KS	Tavg.	4.72	4.56	4.5	4.7	4.48
	s.d.	0.4	0.3	0.70	0.3	0.2
	R.D.		3.4	4.9	0.4	5.1
	P		0.3	0.3	0.5	0.1
BS	Tavg.	6.97	6.14	6.07	6.89	6.02
	s.d.	0.8	0.6	1.3	0.7	0.9
	R.D.		11.9	12.9	1.2	13.6
	P		0.05	0.1	0.4	0.06
Tavg.		5.4	4.7	4.1	4.8	4.3
Avg. R.D.			15.3	24.3	12.4	20.8
p (paired)			0.008	0.03	0.06	0.02

Recognition scores of the five subjects were low for unprocessed speech and varied from 73.3 to 92.0%. Processing with all the three schemes improved the scores. For temporal splitting, relative improvements in recognition scores varied from 4.1 to 17.9%, while for spectral splitting they ranged from 4.0 to 17.5%. Relative improvements for the combined splitting were higher and varied from 1.4 to 11.8% and 5.0 to 20.5% for 4 and 8 shifting respectively.

Two subjects (KS, BS) having severe high frequency loss have shown maximum relative improvement for the scheme of temporal splitting, One subject (BA) with symmetrical low frequency hearing loss, and another subject (SK) with symmetrically sloping high frequency loss have shown maximum improvement for spectral splitting and combined

TABLE 6.4. Experiment IV. Recognition scores (%) for listening tests with hearing impaired subjects for Unprocessed Speech (Su) and Processed Speech corresponding to temporal splitting (SpT), spectral splitting (SpS), and combined splitting with 4 and 8 shifting (Sp4, Sp8). RS = percentage recognition score, 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, n = 5, df = 8). p (paired): significance levels for one-tailed t-test (processed vs unprocessed, averaged across the subjects, n = 4, df = 3).

S		Su	SpT	SpS	Sp4	Sp8
SA	RS	92.0	98.98	95.66	98.68	98.98
	s.d.	1.4	0.9	4.5	1.8	0.9
	R.I.		7.6	4.0	7.3	7.6
	p		0.0	0.06	0.0001	0.0
BA	RS	88.68	92.34	96.98	91.00	98.7
	s.d.	4.2	3.6	1.8	3.2	1.4
	R.I.		4.1	9.4	2.6	11.3
	p		0.09	0.002	0.2	0.0
SK	RS	76.34	84.00	89.68	85.34	92.02
	s.d.	2.2	1.9	3.6	3.8	1.8
	R.I.		10.0	17.5	11.8	20.5
	p		0.0002	0.0001	0.0009	0.0
KS	RS	76.02	89.66	82.90	78.66	84
	s.d.	4.3	1.4	2.9	2.7	2.8
	R.I.		17.9	9.1	3.5	10.5
	p		0.0001	0.009	0.1	0.004
BS	RS	73.32	86.00	77.98	74.32	76.98
	s.d.	4.6	3.8	4.5	4.2	2.2
	R.I.		17.3	6.4	1.4	5.0
	p		0.0007	0.07	0.4	0.07
Avg. RS		81.3	90.2	88.6	85.6	90.1
Avg. R.I.			11.4	9.3	5.3	11.0
p (paired)			0.004	0.006	0.02	0.006

combined splitting with 8 shifting. A subject (SA) having asymmetrical high frequency loss (less loss in one ear and more loss in the other ear) benefits from temporal splitting and combined splitting.

The recognition scores of the individual subjects were subjected to t-test (one-tailed) to obtain statistical significance of the processing. As seen from Table 6.4, subject SA has shown highly significant improvement for temporal splitting and combined splitting with 8 shifting. Subject BA's improvements are highly significant for spectral splitting and combined splitting with 8 shifting. Subject SK's improvements are statistically significant for spectral splitting and combined splitting with 8 shifting, while those for Subjects KS and BS are significant for temporal splitting.

6.5.4 Information transmission analysis

For each test condition, combined confusion matrices for five tests were obtained and subjected to information transmission analysis. Relative information transmitted for consonantal features is given in Table 6.5 and plotted in Figs. 6.11 to 6.17.

Overall: Figure 6.11 shows the overall relative information transmitted for all the subjects. For unprocessed speech, overall information transmitted varied from 76 to 92%. Its relative improvement varied from 2 to 14%, 1 to 10%, -1 to 7%, and 4 to 12% for processed speech with temporal splitting, spectral splitting and combined splitting with 4 and 8 shifting respectively.

Voicing: Figure 6.12 shows the relative information transmitted for voicing feature. For subjects BA, SK, and KS, relative information transmitted with unprocessed speech were 80, 67, and 84% respectively. Relative improvements were highest for the scheme of combined splitting with 8 shifting and were 21, 45, and 19% respectively. For subject SA, reception of voicing is nearly perfect with unprocessed speech, and remains unaffected after processing. For subject BS, the nearly perfect reception under unprocessed condition remains nearly unaffected with the temporal and combined splitting, but show as degradation by 9% under spectral splitting.

Place: Figures 6.13 show the relative information transmitted for place feature. For unprocessed speech, information transmitted for this feature varied from 32 to 85%. For subjects KS and BS, relative improvements were highest for the scheme of temporal splitting and were 76 and 66% respectively. For the remaining three subjects, relative improvements were highest for the scheme of combined splitting with 8 shifting and they varied from 14 to 34%.

Manner: Figures 6.14 show the relative information transmitted for manner feature. For unprocessed speech, information transmitted for this feature varied from 53 to 83%. For subjects KS and BS, relative improvements were highest for the scheme of temporal splitting and were 33 and 20% respectively. For the remaining three subjects relative improvements were highest for the scheme of combined splitting with 8 shifting and varied from 18 to 60%.

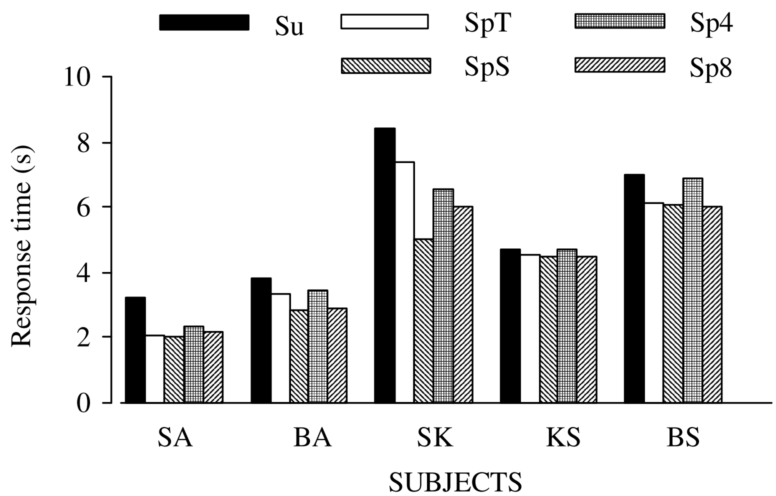


FIG. 6.7. Experiment IV. Response times (s) for five subjects with hearing impairment Su: Unprocessed speech. SpT, SpS, Sp4, Sp8 correspond to processed speech with temporal splitting, spectral splitting, combined splitting with 4 and 8 shifting respectively.

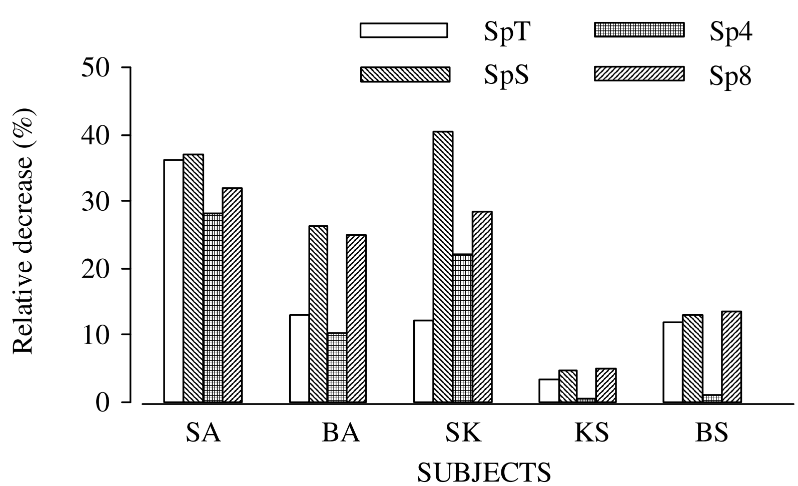


FIG. 6.8. Experiment IV. Relative decrease (%) in response times for five subjects with hearing impairment. SpT, SpS, Sp4, Sp8 correspond to processed speech with temporal splitting, spectral splitting, combined splitting with 4 and 8 shifting respectively.

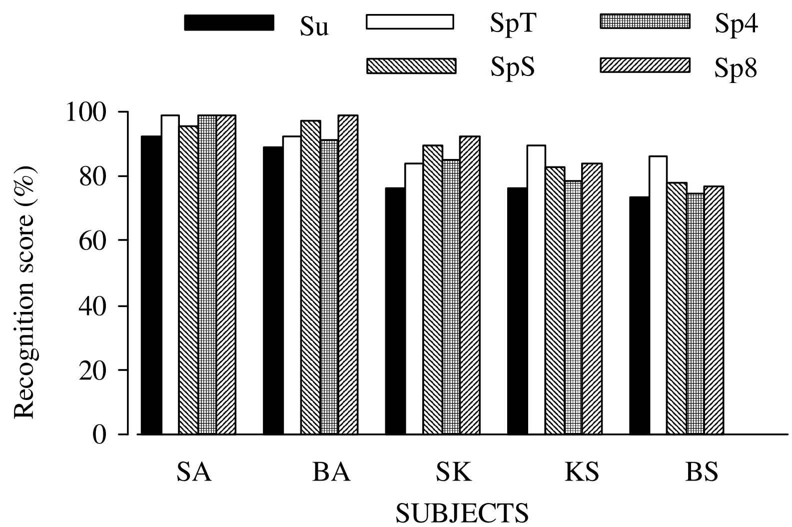


FIG. 6.9. Experiment IV. Recognition score (%) of five subjects with hearing impairment. Su: Unprocessed speech. SpT, SpS, Sp4, Sp8 correspond to processed speech with temporal splitting, spectral splitting, combined splitting with 4 and 8 shifting respectively.

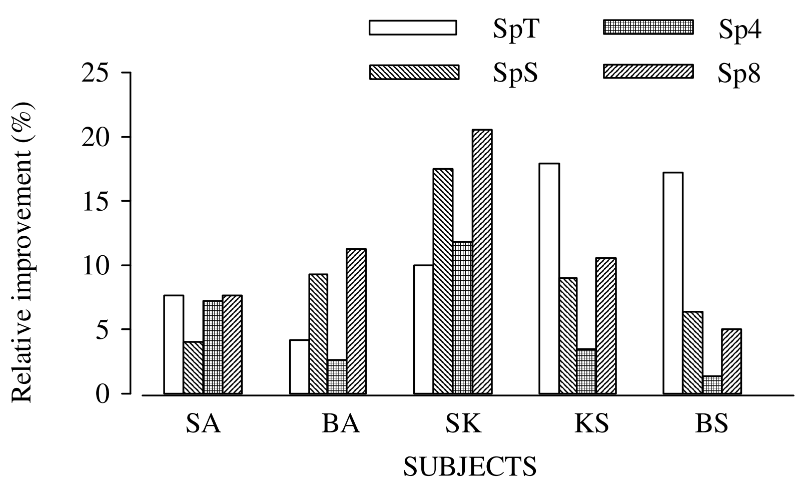


FIG. 6.10. Experiment IV. Relative improvement (%) in recognition score for five subjects with hearing impairment. SpT, SpS, Sp4, Sp8 correspond to processed speech with temporal splitting, spectral splitting, combined splitting with 4 and 8 shifting respectively.

Nasality: Figure 6.15 shows relative information transmitted for this feature. With unprocessed speech, relative information transmitted for nasality feature was low for subjects BA and SK and was 81 and 88% respectively. For subject BA, temporal splitting and combined splitting have helped in bringing the relative information transmitted to 100%, while for subject SK, spectral splitting and combined splitting have helped in bringing the relative information transmitted to 100%. For the remaining three subjects perception of this was perfect for unprocessed as well as for processed speech.

Frication: Figures 6.16 show the relative information transmitted for frication feature. For unprocessed speech, information transmitted for this feature varied from 29 to 79%. For subjects KS and BS, relative improvements were highest for the scheme of temporal splitting and are 76 and 50% respectively. For the remaining three subjects, relative improvements were highest for scheme of combined splitting with 8 shifting and varied from 27 to 159%.

Duration: Figures 6.17 show the relative information transmitted for duration feature. For unprocessed speech, information transmitted for this feature varied from 35 to 71%. For subjects KS and BS, relative improvements are high for the scheme of temporal splitting and are 108 and 35% respectively. For the remaining three subjects, relative improvements were highest for the scheme of combined splitting with 8 shifting and varied from 35 to 174%.

6.6 Discussion

An overall evaluation of the schemes of temporal and combined splitting was done along with the scheme of spectral splitting by conducting listening tests on five persons with moderate to severe bilateral sensorineural hearing loss. For this purpose, the three processing schemes were implemented with the processing parameters that resulted in maximum improvements in the earlier experiments involving tests with normal hearing subjects and simulated sensorineural hearing loss.

Listening tests were conducted using twelve English consonants in VCV context. Subject's response times, recognition scores, overall information transmitted, and information transmitted for consonantal features were analyzed. Subjects also ranked the speech quality for a short passage.

TABLE 6.5. Experiment IV. Relative information transmitted (%) for listening tests with hearing impaired subjects for Unprocessed Speech (Su) and Processed Speech corresponding to temporal splitting (SpT), spectral splitting (SpS), and combined splitting with 4 and 8 shifting (Sp4, Sp8) for (a) overall, and feature groupings: (b) voicing, (c) place, (d) manner, (e) nasality, (f) frication, and (g) duration. S: Subject.

(a) Overall

S	Su	SpT	SpS	Sp4	Sp8
SA	92	98	96	98	98
BA	88	90	96	90	98
SK	81	84	89	86	91
KS	81	92	85	80	85
BS	76	85	77	79	79
Avg.	83.6	89.8	88.6	86.6	90.2

(b) Feature: voicing

S	Su	SpT	SpS	Sp4	Sp8
SA	97	97	100	97	100
BA	80	81	94	79	97
SK	67	84	90	97	97
KS	84	90	92	100	100
BS	97	100	88	100	97
Avg.	85.0	90.4	92.8	94.6	98.2

(c) Feature: place

S	Su	SpT	SpS	Sp4	Sp8
SA	85	98	90	95	97
BA	74	82	91	92	95
SK	59	62	73	75	79
KS	42	74	55	45	54
BS	32	53	44	32	38
Avg.	58.4	73.8	70.6	67.8	72.6

(d) Feature: manner

S	Su	SpT	SpS	Sp4	Sp8
SA	83	96	92	96	98
BA	80	91	98	83	100
SK	53	71	80	68	85
KS	73	97	79	73	83
BS	66	79	62	64	67
Avg.	71.0	86.8	82.2	76.8	86.6

(e) Feature: nasality

S	Su	SpT	SpS	Sp4	Sp8
SA	100	100	100	100	100
BA	81	100	95	100	100
SK	88	91	100	100	100
KS	100	96	100	100	100
BS	100	100	100	100	100
Avg.	93.8	97.4	99.0	100	100

(f) Feature: frication

S	Su	SpT	SpS	Sp4	Sp8
SA	74	94	88	93	97
BA	79	85	100	72	100
SK	29	57	66	46	75
KS	55	97	66	56	72
BS	44	66	35	41	45
Avg.	56.2	79.8	71.0	61.6	77.8

(g) Feature: duration

S	Su	SpT	SpS	Sp4	Sp8
SA	71	95	83	96	96
BA	35	68	86	95	96
SK	39	75	61	54	80
KS	48	100	55	56	58
BS	52	70	54	62	64
Avg.	49.0	81.6	67.8	72.6	78.8

Compared with the response time with unprocessed speech, all the three processing schemes resulted in decreased response times, indicating reduction in load on perception process. Relative decrease in response time was statistically significant for the schemes of spectral splitting and combined splitting with 8 shifting. The extent of improvement was highest for spectral splitting, even for the subjects with high frequency loss. This indicates that spectral splitting is more effective in reducing perceptual load.

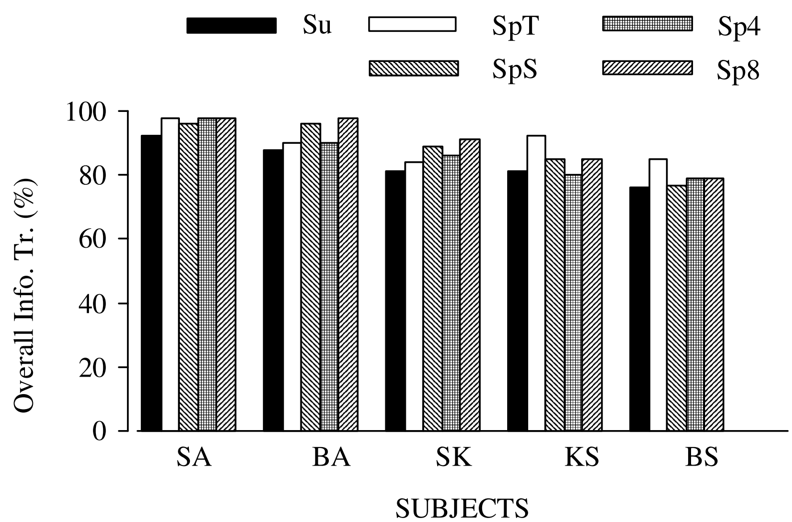


FIG. 6.11. Experiment IV. Overall relative information transmitted (%) for five subjects with hearing impairment. Su: Unprocessed speech. SpT, SpS, Sp4, Sp8 correspond to processed speech with temporal splitting, spectral splitting, combined splitting with 4 and 8 shifting respectively.

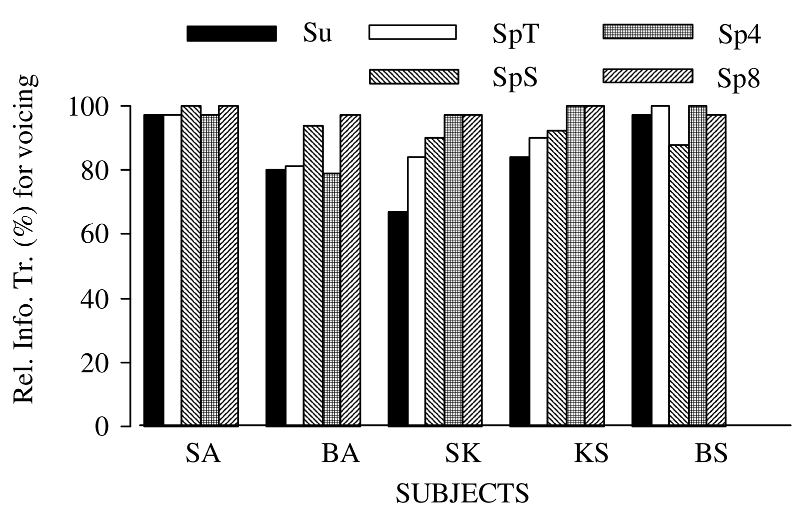


FIG. 6.12. Experiment IV. Relative information transmitted (%) for voicing feature for five subjects with hearing impairment. Su: Unprocessed speech. SpT, SpS, Sp4, Sp8 correspond to processed speech with temporal splitting, spectral splitting, combined splitting with 4 and 8 shifting respectively.

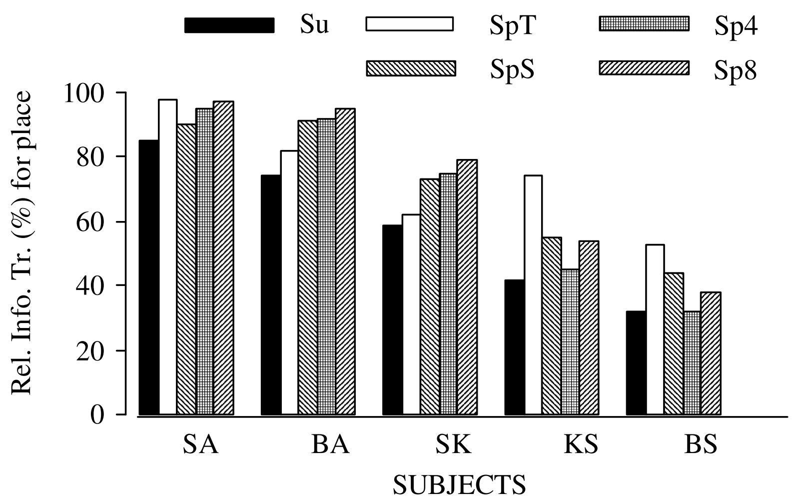


FIG. 6.13. Experiment IV. Relative information transmitted (%) for place feature for five subjects with hearing impairment. Su: Unprocessed speech. SpT, SpS, Sp4, Sp8 correspond to processed speech with temporal splitting, spectral splitting, combined splitting with 4 and 8 shifting respectively.

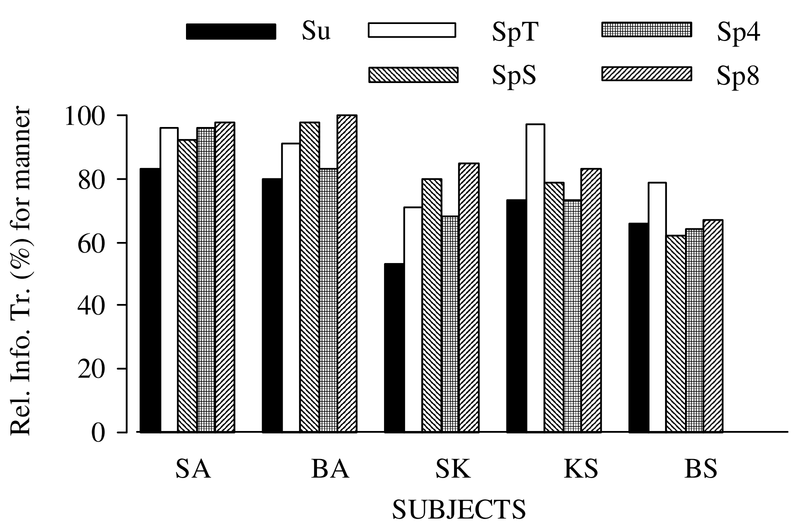


FIG. 6.14. Experiment IV. Relative information transmitted (%) for manner feature for five subjects with hearing impairment. Su: Unprocessed speech. SpT, SpS, Sp4, Sp8 correspond to processed speech with temporal splitting, spectral splitting, combined splitting with 4 and 8 shifting respectively.

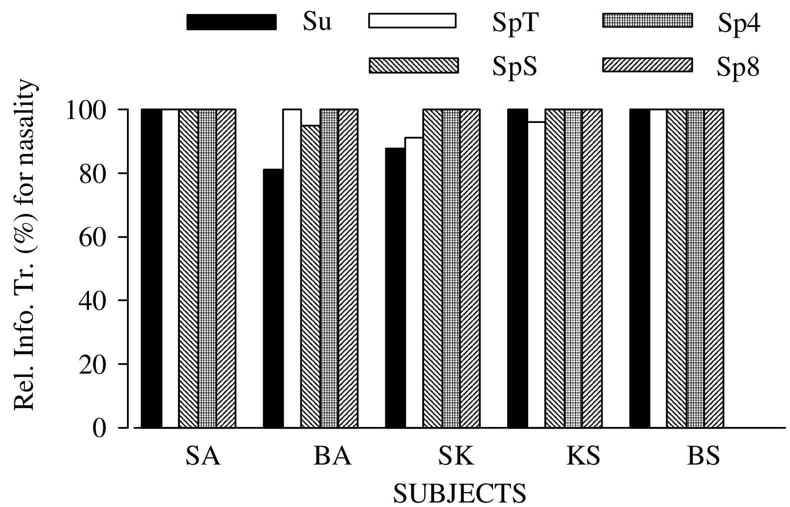


FIG. 6.15. Experiment IV. Relative information transmitted (%) for nasality feature for five subjects with hearing impairment. Su: Unprocessed speech. SpT, SpS, Sp4, Sp8 correspond to processed speech with temporal splitting, spectral splitting, combined splitting with 4 and 8 shifting respectively.

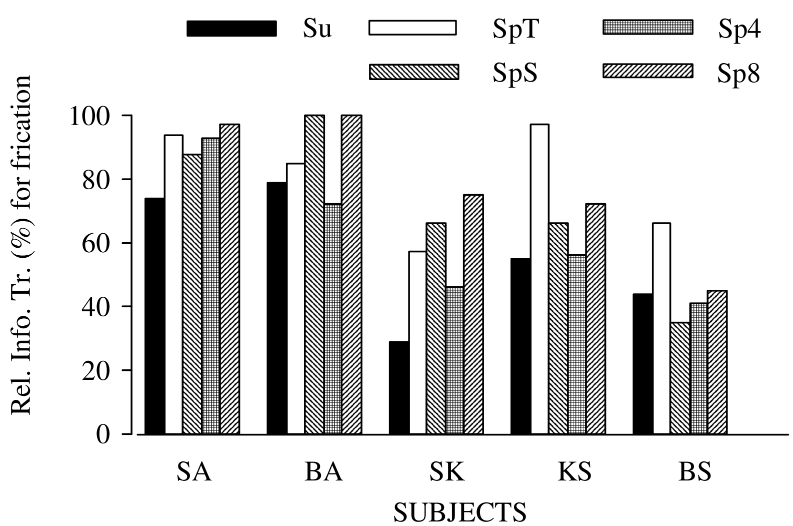


FIG. 6.16. Experiment IV. Relative information transmitted (%) for frication feature for five subjects with hearing impairment. Su: Unprocessed speech. SpT, SpS, Sp4, Sp8 correspond to processed speech with temporal splitting, spectral splitting, combined splitting with 4 and 8 shifting respectively.

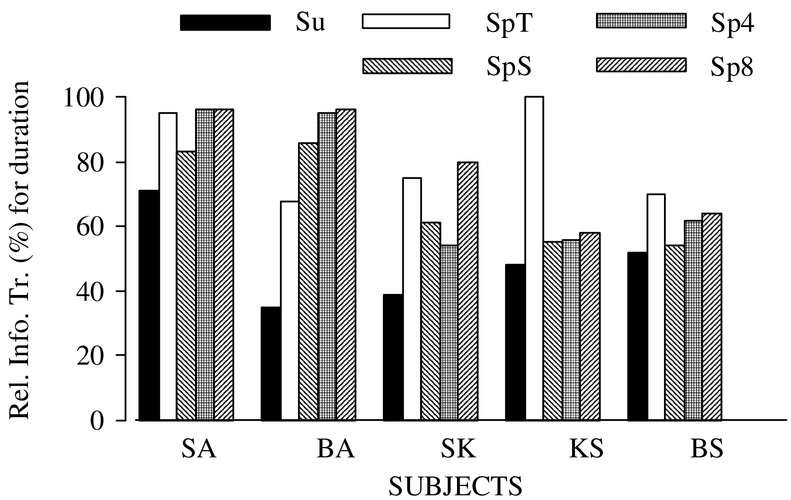


FIG. 6.17. Experiment IV. Relative information transmitted (%) for duration feature for five subjects with hearing impairment. Su: Unprocessed speech. SpT, SpS, Sp4, Sp8 correspond to processed speech with temporal splitting, spectral splitting, combined splitting with 4 and 8 shifting respectively.

Recognition scores indicate that binaural dichotic presentation improved consonantal identification, and the improvements were highest with combined splitting with 8 shifting.

Information transmission analysis shows that for unprocessed speech, relative information transmitted for place, duration, and frication features gets degraded most. Processing improved the transmission of all the consonantal features, particularly the place, duration, and frication features. Generally, temporal splitting has shown highest improvement for duration feature and spectral splitting has shown highest improvement for place feature. Combined splitting with 8 shifting further improved the perception of both duration and place features. For the two subjects KS and BS, having severe loss at high frequency, relative improvements for place, duration, and frication features were high for the scheme of temporal splitting. The remaining three subjects have shown highest improvements for the scheme of combined splitting. The two subjects with severe loss at high frequency show improvement in manner feature for the scheme of temporal splitting. The other three subjects have shown highest improvements for the scheme of combined splitting with 8 shifting. With unprocessed speech, relative information transmitted for nasality feature was low for subjects BA and SK, and was brought to 100% with combined splitting.

From the analysis of recognition scores and information transmission, it is observed that, the scheme that gives maximum benefit by reducing the effects of increased masking depends on the individual hearing loss configuration. Persons with high frequency hearing loss benefit from the scheme of temporal splitting, whereas persons with low frequency hearing loss and gradual sloping symmetrical loss benefit more with the scheme of combined splitting. Reception of the relatively robust consonantal features (voicing, manner, and nasality) also improves because of dichotic presentation. Hence the processing schemes for dichotic presentation have the potential of improving speech perception for persons using binaural hearing aids.

We see that in listening tests involving the normal hearing subjects under simulated hearing loss (Chapter 4 and Chapter 5) and the hearing impaired subjects (this chapter), the subjects are able to perceptually integrate the dichotically presented speech signal, and the presentation results in improved speech reception. As indicated by the improvements in response time, dichotic presentation also decreases the load on the perception process. For hearing impaired subjects, the improvement in consonantal reception and reduction in response time do not follow the same trend. Reduction in response time is highest with spectral splitting, indicating that perceptual integration of spectrally split information is possibly the easiest. In the tests involving quality ranking of a short passage, the unprocessed speech is ranked the best, followed by spectral splitting. Combined splitting is generally ranked the lowest in both the cases. This indicates that while the combined splitting reduces the effects of both types of masking, it puts a relatively more load on the perception process. Therefore, in order to estimate the advantages of combined splitting, extended tests with hearing impaired subjects are needed.

Appendix A

Spectrographic analysis

A.1 Introduction

A spectrogram displays time-varying spectral characteristics of the signal as a two-dimensional pattern, with vertical and horizontal axes corresponding to frequency and time respectively, and darkness of the pattern proportional to the signal energy (O'Shaughnessy, 1987; Rabiner and Schafer, 1978). Analog spectrographic devices take about 10 minutes and provide a 12 dB dynamic range. The advent of the fast Fourier transform (FFT) and digital computer with associated displays resulted in time-efficient production of digital spectrograms. Digital spectrograms apart from being fast in operation provide larger dynamic range compared to analog spectrograph (Morris, 1988; O'Shaughnessy, 1987; Thomas *et al.*, 1994).

In speech signals, often two types of spectrograms are used: wide-band and narrow-band. Wide-band spectrograms (with analysis filter bandwidth of 300 Hz) clearly show the formant structure as dark broad bands. The center of each band gives an estimate of formant frequency. Wide-band spectrograms show the pitch periods as narrow, vertical striations corresponding to the closing of vocal chords. The good time response clearly displays the formant transitions. Narrow-band spectrograms (with analog filter bandwidth = 45 Hz) provide good frequency resolution to separate the individual harmonics of the voiced speech signals. Formant frequencies may be found from the narrow-band spectrograms, by searching for the harmonics having highest amplitude (darkest band). However, due to poor time resolution, the dynamic patterns like formant transitions, voice-onset-time, etc. get heavily smeared (Oppenheim, 1970; O'Shaughnessy, 1987; Rabiner and Schafer, 1978). Several approaches have been reported to provide good time and spectral resolution in the same spectrographic display (Cheung and Lim, 1992; Thomas *et al.*, 1994) However, separate wide-band and narrow-band spectrograms are still commonly used.

A.2 Digital spectrographic analysis

A digital spectrogram can be generated by computing the time varying log magnitude spectrum using short time Fourier transform (STFT) (O'Shaughnessy, 1987; Rabiner and Schafer, 1978). The STFT can be obtained by computing the discrete Fourier transform (DFT) of the speech signal $s(n)$ and is given by

$$S(n, k) = \sum_{m=0}^{N-1} w(m)s(n-m)e^{-j\frac{2\pi mk}{N}} \qquad 0 \le k \le N-1 \qquad \text{(A.1)}$$

where n represents the discrete time samples, k the discrete frequency, N the DFT size, and $w(n)$ is Hamming window of length L samples, where $L < N$ and is given by

$$w(m) = 0.54 - 0.46\cos(2\pi m/L - 1), \quad 0 \le m \le L-1 \qquad \text{(A.2)}$$
$$0 \qquad \text{otherwise}$$

The frequency resolution of the spectrogram depends on the type of the window, window length L, and the sampling rate f_s. For Hamming window, the frequency resolution $F_{rs} \approx 1.36 f_s/L$ (Harris, 1978). For a given window, choice of duration decides the time and frequency resolution (Flanagan, 1972; Morris, 1988; Oppenheim, 1970). With f_s = 10 k Sa/s, window of length 45 samples generates wide-band (300 Hz) spectrogram, while window length of 300 samples generates narrow-band (45 Hz) spectrogram. In case of voiced speech, the magnitude spectrum has a –6 dB/octave tilt, and in order to properly display the higher frequency components in the limited dynamic range of spectrographic display, pre-emphasis is needed. This is usually done by computing a first difference of the input signal waveform before computing STFT (Oppenheim, 1970; O'Shaughnessy, 1987)

A.3 Implementation of the spectrographic analyzer

For spectrographic analysis of speech signals, a PC based package was developed at IIT Bombay (Thomas *et al*., 1994; Thomas, 1996). The signal acquisition was done through a signal acquisition card (PCL-208 from Dynalog Micro Systems, Mumbai) and spectrograms were displayed using an EGA card and monitor. The package provided an option of generating either a wide-band/narrow-band or a combined spectrogram. The package was modified to generate spectrograms at much faster speed using a PC with VGA card and a DSP card based on 16-bit fixed point TMS320C25 DSP processor operating at 40 MHz (PCL-DSP25 user's manual, 1989; TI-TMS320C2X user's guide, 1993). Signal acquisition was done through the ADC on the DSP board. Faster speed was achieved by appropriate partitioning of tasks and parallel operation of the DSP and PC processors. PC was used to display the spectrogram and DSP processor was used to obtain DFT and log magnitude (Baragi, 1996; Chaudhari and Pandey, 1999; Prasad, 1996).

With increase in speed of PC processor, the use of DSP based peripherals for computation for spectrographic analysis has become unnecessary. A PC based spectrographic analyzer has been developed by Ratanpal (2000). The system uses sound card for signal acquisition. Spectrograms are displayed using SVGA card and display. The main program "`spec99`", written in C, handles the spectral analysis and display operations, and invokes the program "`audio`" written in Visual C++, which is used for recording and playback of sound. The set-up can be used for signal acquisition, playback, and spectrographic analysis. The speech signal is displayed along the time axis. A segment can be selected using the cursors and the spectrogram is generated for this selected segment. The selected waveform segment is plotted using 500×45 pixels, and above this, the spectrogram is displayed on an area of 500×128 pixels with an intensity resolution of 64 gray levels. The gray scale of the dynamic range is displayed on the side of the spectrogram. Highest intensity is displayed as black and lowest intensity is displayed as white. Intermediate intensities are displayed with different shades of gray levels. The spectrogram has 128 frequency points corresponding to 0–f_s/2 frequency scale.

The selected segment is divided into 500 overlapping time frames. After pre-emphasis, an L-point Hamming window is applied to each frame (window length is decided by the frequency resolution). Zero padding is then done so that the magnitude spectrum can be obtained using a 256-point FFT, irrespective of the window length. The log magnitude of the 128 frequency samples in the FFT sequence is then calculated and each value is linearly mapped to 64 gray levels. After displaying the spectrogram, the magnitude of a point (time, frequency) in the display can be obtained using cursors. Spectrum for a particular time position can also be displayed. The spectrogram can be stored in the form of a Postscript file with a resolution of 256 gray levels (Adobe systems, 1988), thus providing a dynamic range of 48 dB. The selected portion of the signal can be output to the left, right or both channels of the sound card. The software works with a color/monochrome monitor and a PC with Celeron/Pentium-III processor, sound card, and SVGA display.

Examples of wide-band spectrogram for (i) swept sine wave (ii) random noise (iii) syllable /*asa*/ are shown in Fig. A.1.

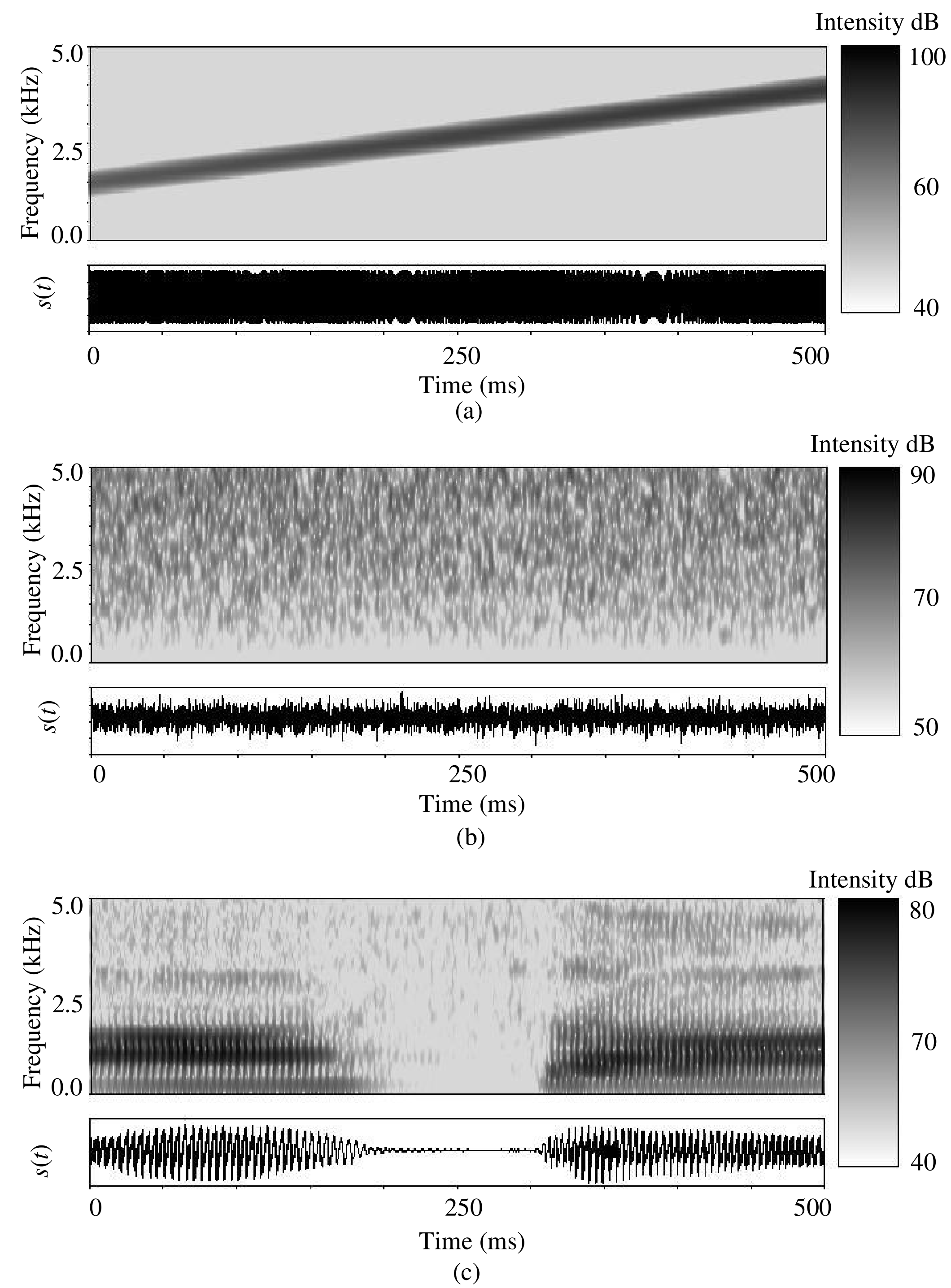


FIG. A.1. Wide-band spectrograms ($\Delta f \approx 300$ Hz) of (a) swept sine wave, (b) random noise, (c) speech syllable /*asa*/.

Appendix B

Method of performance evaluation

B.1 Information transmission analysis

Generally the recognition scores (percent correct scores) are considered for evaluation of speech processing schemes. For more detailed study of the patterns of errors, stimulus-response scores are represented in the form of confusion matrix, which shows the stimuli along rows and responses along the columns (Miller and Nicely, 1955). Each entry in the cell represents either the frequencies or probabilities of stimulus-response pair. If x and y represent sets of n stimulus items $\{x_1, x_2, \ldots x_n\}$ and responses $\{y_1, y_2, \ldots y_n\}$ respectively, and if $N(x_i)$, $N(y_j)$, 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

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

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

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

The diagonal elements correspond to correct responses, whereas off-diagonal elements correspond to errors. Sum of diagonal elements of a confusion matrix in which cell entries correspond to the probabilities of correct responses gives the recognition score R_s.

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

For studying the pattern of confusions, smaller confusion matrices are derived by combining the stimuli and responses into groups, in such a way that confusions within the groups are more likely than the confusions between the groups. Stimuli are grouped based on certain common features. Recognition scores obtained for the smaller confusion matrices can be used for studying the reception of different features.

The recognition scores do not take into account the patterning of responses i.e. distribution of errors in off-diagonal cells and may be affected by the subject's biasing for response. A method that takes into account the subject's bias and chance scoring, and provides a measure of covariance between stimuli and responses, is the information transmission analysis (Miller and Nicely, 1955). Mean logarithmic probability (MLP) measure of information is used as a measure of the information transmitted. The information measures of the stimulus x and response y, $I(x)$ and $I(y)$ respectively, are given by

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

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

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 from x to y is given by

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

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

Relative information transmission, like recognition score, can also be obtained for smaller confusion matrices derived from the original matrix based on certain common features of stimulus and response items. Thus the relative information transmission can be used to measure the transmission performance in the context of specific features.

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

$$
\begin{aligned}
p(x_i\,;\,y_j) &= R_s/n\,, \qquad & i = j \\
&= \frac{1-R_s}{n^2-n} & i \neq j \\
p(x_i) = p(y_j) &= 1/n &
\end{aligned}
\qquad \text{(B.9)}
$$

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, $I_{rel}(x;y) = \log_2(n/(n\text{-}1))/\log_2(n)$.

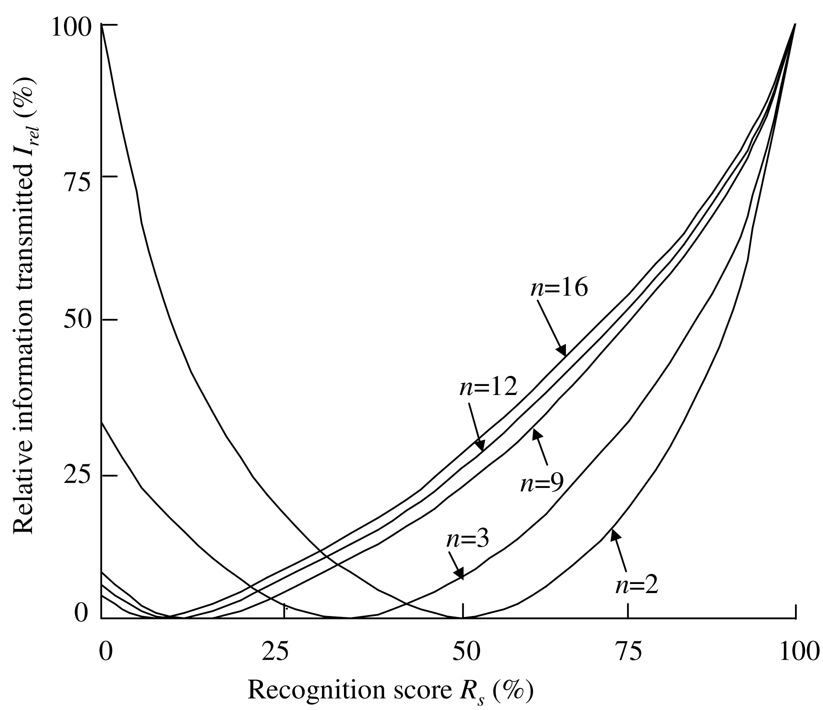


FIG. B.1. Relative information transmitted (I_{rel}) vs recognition score (R_S) for a special case when the correct responses are equally distributed among the diagonal cells and the errors are equally distributed among the off-diagonal cells in the stimulus-response confusion matrix (as given Eqn. B.9).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

B.2 Analysis programs

For an experimental condition, confusion matrices of several test runs were combined using the program "cummat". Analysis of the confusion matrix for obtaining recognition score and information transmission was done 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. The program uses another file '`infogr.dat`' that contains information about the feature groupings. 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, 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).

B.3 Sample analysis results

The confusion matrix selected for sample analysis corresponds to the test condition TS_TR, with TS (transition duration) = 2 ms, SNR = -6 dB for subject LT.

1) output of "`cummat`"

STIMULUS-RESPONSE CONFUSION MATRIX

12 300

S/R	aPa	aBa	aTa	aDa	aKa	aGa	aMa	aNa	aSa	aZa	aFa	ava	+
aPa	16	0	0	0	5	0	0	0	0	0	4	0	25
aBa	0	25	0	0	0	0	0	0	0	0	0	0	25
aTa	0	0	25	0	0	0	0	0	0	0	0	0	25
aDa	0	0	0	25	0	0	0	0	0	0	0	0	25
aKa	1	0	0	0	12	0	0	0	5	0	7	0	25
aGa	0	0	0	0	0	22	0	0	0	0	0	3	25
aMa	0	0	0	0	0	0	17	8	0	0	0	0	25
aNa	0	0	0	0	0	0	0	25	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	0	0	0	0	25	0	0	25
aFa	3	0	0	0	10	0	0	0	4	0	8	0	25
aVa	0	0	0	0	0	0	0	0	0	3	0	22	25
+	20	25	25	25	27	22	17	33	34	28	19	25	300

No. of files: 5

--- 21-01-2000 11:57:14

--- 21-01-2000 12:03:25

--- 21-01-2000 12:10:01

--- 21-01-2000 12:16:18

--- 21-01-2000 12:23:10

78.3 86.7 82.3 3.5

1.85 2.02 1.93 0.07 0.99 5.0

2) 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
```

3) 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   64    0    0    0   20    0    0    0    0    0   16    0
   aBa    0   99    0    0    0    0    0    0    0    0    0    0
   aTa    0    0   99    0    0    0    0    0    0    0    0    0
   aDa    0    0    0   99    0    0    0    0    0    0    0    0
   aKa    4    0    0    0   48    0    0    0   20    0   28    0
   aGa    0    0    0    0    0   87    0    0    0    0    0   12
   aMa    0    0    0    0    0    0   68   32    0    0    0    0
   aNa    0    0    0    0    0    0    0   99    0    0    0    0
   aSa    0    0    0    0    0    0    0    0   99    0    0    0
   aZa    0    0    0    0    0    0    0    0    0   99    0    0
   aFa   12    0    0    0   40    0    0    0   16    0   32    0
   aVa    0    0    0    0    0    0    0    0    0   12    0   87

   Correct: 82.3

* (2): VOICING

  S/R  UV   VO
   UV  100    0
   VO    0  100

    Correct: 100.0
```

```
* (2): NASALITY

  S/R  OR   NA
   OR  100    0
   NA    0  100

    Correct: 100.0

* (2): FRICATION

  S/R  ST   FR
   ST  91   10
   FR  13   87

    Correct: 89.3

* (2): DURATION

  S/R  SH   LO
   SH  96    5
   LO   0  100

    Correct: 96.0

* (3): PLACE

  S/R  FN   MD   BK
   FN  76   12   13
   MD   0  100    0
   BK  23   11   68

    Correct: 84.7

* (2): MANNER

  S/R  OS   FR   NA
   OS  88   13    0
   FR  13   87    0
   NA   0    0  100

    Correct: 100.0
```

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

```
** INFORMATION TRANSMISSION **

* (12): OVERALL
   Stimulus info    = 3.5832
   Response info    = 3.5552
   Trans info       = 2.9906
   Perc transn      = 83.5

* (2): VOICING
   Stimulus info    = 0.9796
   Response info    = 0.9796
   Trans info       = 0.9796
   Perc transn      = 100.0


* (2): NASALITY
   Stimulus info    = 0.6497
   Response info    = 0.6497
   Trans info       = 0.6497
   Perc transn      = 100.0

* (2): FRICATION
   Stimulus info    = 0.9180
   Response info    = 0.9367
   Trans info       = 0.4490
   Perc transn      = 48.9

* (2): DURATION
   Stimulus info    = 0.6497
   Response info    = 0.7348
   Trans info       = 0.5033
   Perc transn      = 77.5

* (3): PLACE
   Stimulus info    = 1.4829
   Response info    = 1.4638
   Trans info       = 0.8342
   Perc transn      = 56.3

* (3): MANNER
   Stimulus info    = 1.4587
   Response info    = 1.4690
   Trans info       = 1.0091
   Perc transn      = 69.2
```

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

NO. OF STIMULI:12

COR	ERR	IS	IR	IT	RTR	FEATURE	N
82	100	3.58	3.56	2.99	83	Overall	12
100	0	0.98	0.98	0.98	100	Voicing	2
100	0	0.65	0.65	0.65	100	Nasality	2
89	60	0.92	0.94	0.45	49	Frication	2
96	23	0.65	0.73	0.50	77	Duration	2
85	87	1.48	1.46	0.83	56	Place	3
89	60	1.46	1.47	1.01	69	Manner	3

Appendix C

Optimization of comb filter response

C.1 Introduction

In a scheme of spectral splitting investigated by Chaudhari and Pandey (1998a, b), the speech is divided into eighteen bands corresponding to critical bands of auditory filter (Zwicker, 1961). The odd and even bands which form a pair of comb filters with complementary magnitude response, were presented to the two ears to reduce spectral masking. Spectral components lying in the pass bands are presented to one ear, whereas those lying in the transition region are presented to both the ears. Hence, there may be a change in intensity perception when the spectral component is passing through the transition region, thereby reducing the speech quality. In the investigation of combined scheme of spectral and temporal splitting (CS), comb filters are designed with three considerations, reducing pass band ripple, increasing the stop band attenuation, and adjustment of magnitude response at band transitions to minimize the changes in intensity perception. Filter design was carried out jointly with my colleague Alice Cheeran (Cheeran *et al*., 2001) and is described in this appendix. Evaluation of the optimized comb filters is also given in this appendix.

C.2 Comb filters with sharp transition

In the scheme investigated by Chaudhari and Pandey (1998a), each comb filter was designed as 128-coefficient linear phase FIR filter using frequency sampling technique. A package "`spfilt`" developed by Kasthuri (1997) was used for the filter design. Filters were designed for sharp transition between bands to avoid overlapping of bands in the two ears. The implementation of the comb filters was done in real time with TI/TMS320C50 DSP processors and was evaluated by conducting listening tests on persons with sensorineural hearing impairment (Chaudhari and Pandey, 1998b, 1999).

In our investigation, we have used sampling rate of 10 k samples/s. Signal processing has been carried out off-line, with filters implemented using floating-point arithmetic, so that the effects of coefficient quantization and calculation errors are not significant. For this implementation, the magnitude response of the 128-coefficients comb filters is shown in Fig. C.1. The maximum pass band ripple and minimum stop band attenuation in the magnitude response are 4 dB and 10 dB respectively. Frequency samples in the pass band have a gain of 1 and those in the stop band have a gain of 0, with the band transition of 1 sample, corresponding to Δf =78 Hz.

As a first step towards designing better filters, comb filters were redesigned using 256 coefficients taking into consideration the present developments in DSP processors. The magnitude response of the comb filters is shown in Fig. C.2. With doubling the number of coefficients, the transition width is halved to 39 Hz. The maximum pass band ripple and minimum stop band attenuation are 3 dB and 11 dB respectively. Thus doubling the number of coefficients did not result in significant improvement in the pass band ripple and stop band attenuation. Relatively large pass band ripples may result in perceptual distortion, and the relatively small stop band attenuation may not give adequate separation between bands. Another serious problem with the filters is that for different bands, the crossover between adjacent bands are at different levels, and these may result in perceptual distortions due to decrease or increase in the perceived intensity of spectral components at the crossover frequencies. When a sine wave with its frequency slowly swept between 100 Hz and 5 kHz was processed with the set of two comb filters and presented binaurally, a change in intensity was clearly perceived at the transition between bands.

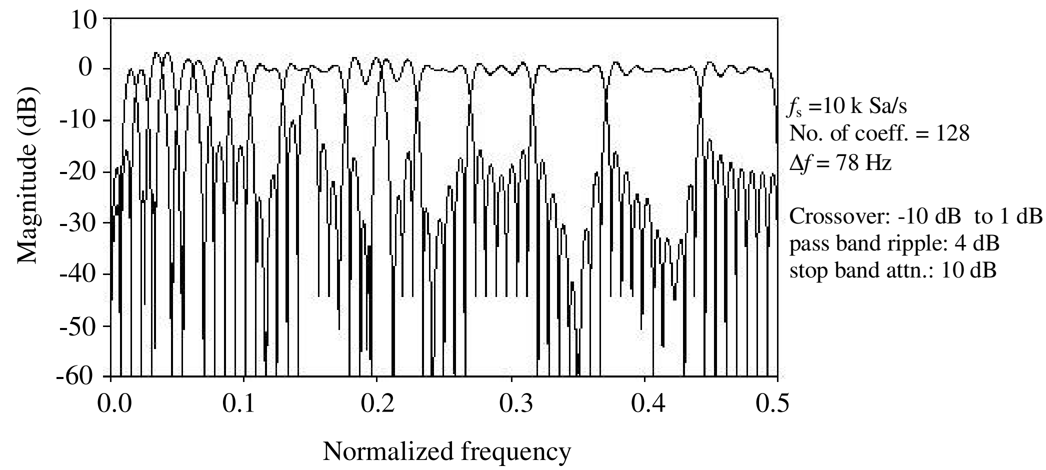


FIG. C.1. Overlapped magnitude responses of the two comb filters having sharp transitions and 128 filter coefficients.

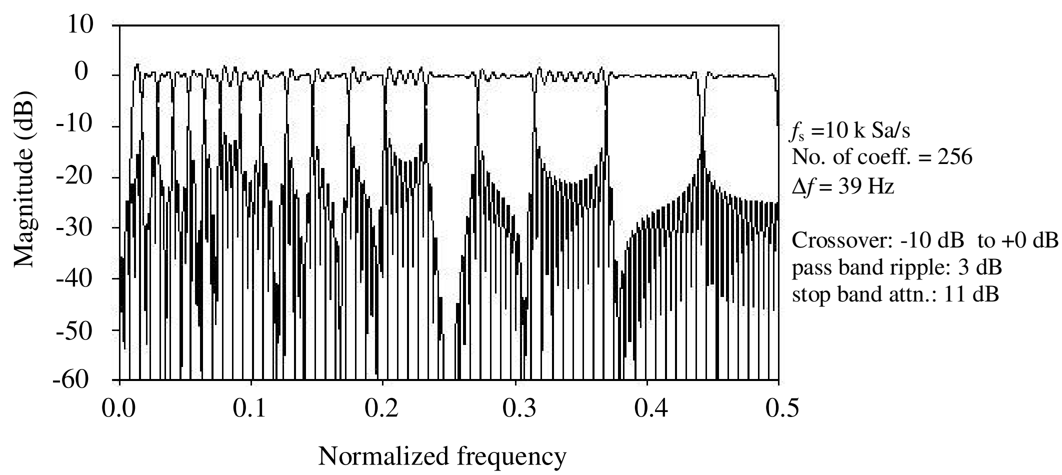


FIG. C.2. Overlapped magnitude responses of the two comb filters having sharp transitions and 256 filter coefficients.

Measurements for just noticeable difference (JND), or smallest detectable change, in intensity have been reported for a variety of stimuli by a variety of methods (Moore, 1997). The reported values are generally in the range of 0.3-1.5 dB. JND for overall intensity of a synthetic vowel is about 1.5 dB, while for first and second formants it is 1.5 dB and 3 dB respectively (Flanagan, 1972). Thus pass band ripples of 3–4 dB are likely to result in perceivable changes in processed speech signal. To see the effect of pass band ripple on the intensity perception, swept tones with frequency slowly varying from 310 Hz to 390 Hz (corresponding to pass band of 3rd critical band auditory filter) were processed and tested by listening. For the filters with pass band ripple of 4 dB, a variation in intensity was perceived. Hence it was decided to reduce the pass band ripple to less than 1 dB.

Therefore, the comb filters are redesigned with three considerations: reducing pass band ripple, increasing stop band attenuation, and adjustment of magnitude response at band transitions to minimize the changes in intensity perception.

C.3 Comb filters with improved pass band ripple and stop band attenuation

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

For adjusting magnitude response of the comb filters, 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, 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 reduce the maximum pass band ripple and increase the minimum stop band attenuation. The number of transition samples (1 or 2) was dependent on the available stop bandwidth, increasing from low frequency to high frequency.

Using the above iterative method, Ratanpal (2000) designed a pair of comb filters with 128 coefficients with one or two samples in the transition region, as described later in section C.6. The comb filters obtained for f_s = 10 k Hz have a transition band of 156 Hz at lower frequencies and 234 Hz at higher frequencies. They have a pass band ripple of less than 1 dB and stop band attenuation of 22 dB. We repeated the same design method for 128 and 256 coefficients. The number of transition samples between the bands of a comb filter is kept as 1 for lower bands and as 2 for higher bands. The magnitude of each transition sample was varied iteratively and its effects on the pass band ripple and stop band attenuation were observed. The transition sample values were iteratively adjusted in order to get an optimal set of transition values such that pass band ripple remained less than 1 dB and stop band attenuation was maximized.

The filter design with 256 coefficients and sampling rate of 10 k samples/s has a transition band of 78 Hz at lower frequencies and 117 Hz at higher frequencies. Figure C.3. shows the magnitude response of these comb filters. Pass band ripple of 1 dB and stop band attenuation of 38 dB were obtained. Listening tests showed that no change in intensity due to pass band ripple was perceived when the frequency of input sinusoidal tone was swept. When a sine 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, when the frequency passed through the transition region. Thus, even though this comb filter provided relatively flat pass bands and adequate band separation, it was necessary to modify the magnitude response at the transitions to balance the perception at all frequencies.

C.4 Comb filters adjusted with inter-band crossovers

In ideal spectral splitting, any spectral component would be presented to one ear. However, with filters having finite inter-band crossover in the magnitude response, the components lying in the pass bands are presented to one ear (with a gain of one), whereas those lying in the transition regions are presented to both ears (with gain less than one). Frequency components presented binaurally will be louder than frequency components presented monaurally, however the loudness is generally less than double (Scharf, 1969). Auditory system does not act as a linear integrator for binaural presentation with respect to intensity, and at higher levels strong non-linear compression occurs. If the magnitude response is not properly shaped at transitions, the components lying in the transition region will be perceived with different loudness, and may reduce the speech quality and may also result in degraded intelligibility. Loudness perception tests were 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. The filters were tested by processing sinusoidal tones, with frequency slowly swept over narrow ranges to cover a pair of adjacent bands. These tests and results are presented in the following two subsections.

C.4.1 Perceptual balance of monaural and binaural intensity levels

Listening tests were conducted to establish the intensity levels for monaural and binaural presentation, such that they evoke the same perceived loudness level. Stimuli used were pure tones of frequencies 250 Hz, 1 kHz, 2 kHz, and 4 kHz, broad band noise, and sustained vowel /*a*/. Five normal hearing subjects participated in these tests. The stimuli of 1 s duration were presented through headphones, monaurally and binaurally one after the other, with 1 s inter-stimulus interval. Monaural intensity level was kept constant at 85 dB and binaural intensity level was varied from 84 dB to 70 dB in steps of 1 dB, to establish the monaural versus binaural intensity balance. These listening tests were conducted using program "`mono_bin`", which is described in Appendix D.

The subject was instructed to label the binaural loudness as "high", "same", or "low" as compared to monaural loudness. For a particular binaural level, the pair of monaural and binaural stimuli, were presented repeatedly with inter-presentation gap of 2 s, until the subject responded. This process was continued with the different binaural intensities. From these observations, the median value of the binaural intensity for "same" response was obtained. For obtaining a more refined estimate of the binaural intensity for "same" loudness, listening tests were conducted for binaural intensities spread over 4 dB on both sides of this intensity level. However, this time the binaural intensity value was randomly selected, and the subject was asked to label the loudness as before. If the subject response was "same" or "high" the presentation was repeated with binaural level decreased by 2 dB, and if response was "less" the level was increased by 1 dB. This process was repeated until the same response was obtained for a binaural level at least 3 times. The median value for the "same" intensity level was taken as the intensity level for perceptual balance. The results of all the subjects were noted and the difference in intensity required for same perception in monaural and binaural presentation for a particular stimulus was tabulated. Table C.1 gives the intensity difference in monaural and binaural presentation for the same perception for different stimuli for all the subjects. From the table it may be observed that the perception is same when binaural level is 4–9 dB lower than monaural level. An interesting observation here is that for tones, there is a significant inter-subject variation. Also for each subject there is a significant inter-tone variation. However, for vowel and broadband noise, the level difference for balance is about 8 dB for all the subjects.

TABLE C.1. Difference between monaural and binaural hearing levels for same perceived loudness, for different signals.

Subject	Signal					
	Pure tone 250 Hz (dB)	Pure tone 1000 Hz (dB)	Pure tone 2000 Hz (dB)	Pure tone 4000 Hz (dB)	Vowel /*a*/ (dB)	Broadband Noise (dB)
S1	7.8	7.1	3.6	6.9	7.7	8.7
S2	7.8	5.9	4.8	6.6	7.7	7.4
S3	12.4	7.1	9.1	7.4	9.8	8.7
S4	5.0	4.8	6.0	6.4	8.7	8.5
S5	12.4	7.1	8.9	9.4	8.7	8.7
Avg.	9.1	6.4	6.5	7.3	8.5	8.4

C.4.2 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 crossover between adjacent bands and listening tests were conducted with swept sine waves. Seven pairs (for right and left) of comb filters were designed with different crossover points varying between -3 dB and -9 dB at the crossover region of pass bands of 1st and 2nd auditory critical bands. The first band (100–200 Hz) is centered at 150 Hz and second band (200–300 Hz) is centered at 250 Hz. A sine wave with frequency swept from 100 Hz to 300 Hz over an interval of 30 s was processed with these comb filters and listening tests were conducted. For the processed signal, the perception, which started at one ear, moved slowly to the other ear. During transition from one ear to the other, perception was binaural. An increase in intensity was perceived as the stimulus moved from one ear to other ear in case of swept sine waves processed with comb filters with crossover point lying above -4 dB. With crossover points below -6 dB, a decrease in loudness (dip) was perceived. For swept sine waves processed with comb filter pairs with crossover points between -4 dB and -6 dB, change in intensity perception was not noticeable as the swept sine wave moved from one ear to other ear. To verify the effect in the high frequency range, comb filter pairs were designed with the crossover point varying between -3 dB and -9 dB at the overlapping of pass bands of 15th and 16th auditory filters. Listening test was conducted using sine wave with frequency swept between 3 kHz and 3.5 kHz, so as to cover the overlapping region of these bands. Results were similar to those obtained for lower frequency bands. Hence it was concluded that for perceptual balance, the magnitude response crossover in the transition band should be kept in the range of -4 dB to -6 dB.

C.5 Filter design method

In the frequency sampling technique of designing linear phase FIR filters (Proakis and Manolakis, 1997; Rabiner *et al.*, 1970; Rabiner and Gold, 1998; Rabiner and Schafer, 1978), the desired frequency response $H_d(e^{jw})$ is sampled at a set of uniformly spaced frequencies,

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

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

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

The filter coefficients are determined as IDFT of this set of samples, i.e.

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

From the N-point FIR impulse response $h(n)$, frequency response $H(e^{jw})$ is calculated, which will coincide $H_d(e^{jw})$ at $w_k = 2\pi k/N$ and is given as

$$H(e^{jw}) = \sum_{n=0}^{N-1} h(n)e^{-jwn} \tag{C.3}$$

And therefore L uniformly spaced samples of the frequency response are given as

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

For filter design, a program "`modfilt`" developed by Kasthuri (1997) and modified by Ratanpal (2000) has been used. The program uses frequency sampling technique of linear phase FIR filter design. For N-coefficient filter, the program allows the user to graphically enter the magnitude at $N/2$ uniformly spaced frequency samples. The value of L can be selected as $N \leq L \leq 17N$. The magnitude of the sample can be varied in steps of 0.006 in linear scale. The program calculates the filter coefficients, which can be stored as '`.txt`' file. It provides the facility to observe the interpolated magnitude response in linear scale on the monitor. Interpolated frequency response in linear scale as well as in logarithmic scale may be obtained as files in '`.ps`' format. The coefficient file can be used for providing initial magnitude response for designing a filter with a modified magnitude response.

C.6 Optimization of comb filter design

Each comb filter has 18 transitions between pass band and stop band, and it is very difficult to find optimal gains for frequency samples in the transition region. Hence it was decided to find out suitable transition response values for each of the individual critical band filters. Hence it was decided to first design 18 bandpass filter and obtain the optimal value for transition frequency sample for these filters. Subsequently, these values were used as initial value of transition frequency samples for comb filters.

Thus the following steps are used to obtain transition samples for a band pass filter.

1. Specify the sampled magnitude response for a band pass filter corresponding to a perticular auditory critical band. Samples in pass band are considered as constrained samples taking a value of 1. Samples lying close to the edge of the pass band are taken as unconstrained and remaining samples of the stop band are constrained with a value 0.
2. Begin with an initial value of the unconstrained samples.
3. Obtain the filter coefficients.
4. Observe the interpolated response and check for pass band ripple and stop band attenuation. If pass band ripple and stop band attenuation are not acceptable, adjust the transition frequency sample, and go to step 3.
5. The steps 2–4 are repeated for the next transition sample.

The magnitudes of the transition samples obtained for each auditory filter are taken as initial values in designing the comb filters. For designing the comb filters with crossover adjustment, all the above mentioned steps are involved. With a constraint of 1 dB pass band ripple and cross over point within -4 to -6 dB, iterations are done for obtaining the maximum stop band attenuation possible. The comb filters for odd and even bands are designed separately and the interpolated magnitude responses are obtained. To decide the crossover point they are to be overlapped and observed. Iterations are continued till optimum response is obtained for the pair of comb filters.

In the comb filters designed, with a constraint of 1 dB pass band ripple and cross over point between -4 to -6 dB, stop band attenuation obtained was 30 dB. Figure C.4 shows magnitude response of optimized comb filters.

For future use, a program may be developed to automate the filter design process. The program should give maximum stop band attenuation, constraining 1 dB ripple in the pass band and crossover in the transition region between -4 dB to -6 dB.

C.7 Evaluation of optimized comb filters

To evaluate the optimized comb filters scheme of spectral splitting was implemented and evaluated to compare the comb filters with sharp transition and the optimized comb filters.

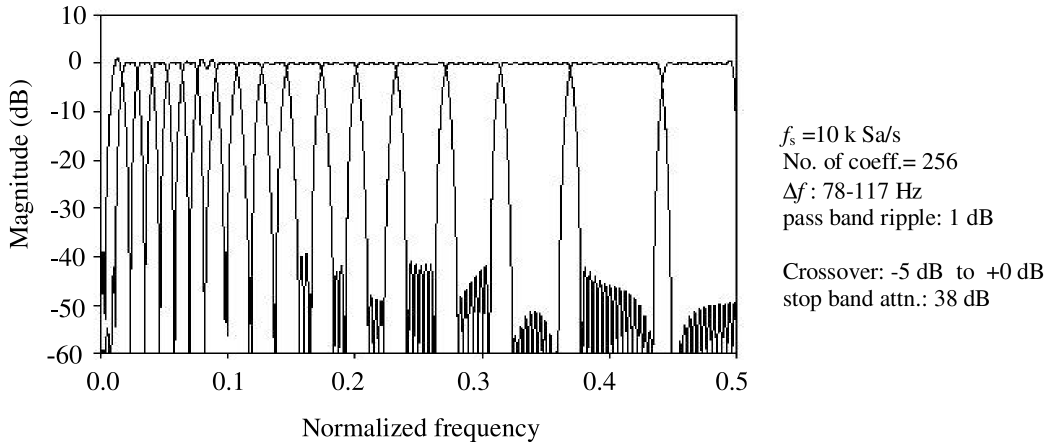


FIG. C.3. Overlapped magnitude responses of the two comb filters with 256 filter coefficients, with crossover samples introduced to constrain the pass band ripple < 1 dB and maximize the stop band attenuation.

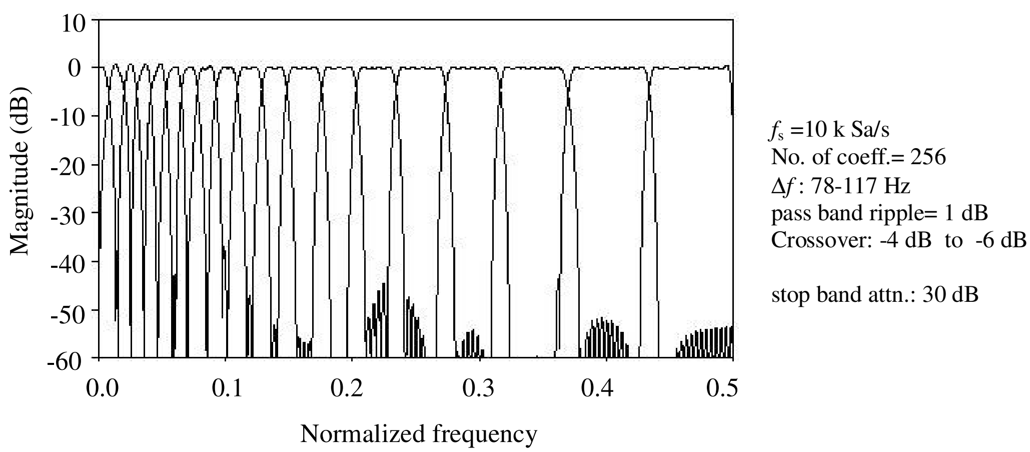


FIG. C.4. Overlapped magnitude responses of the two comb filters with 256 coefficients designed for low perceived spectral distortion.

In this section evaluation of the two schemes of spectral splitting is discussed. Listening tests were conducted to evaluate the two schemes. The tests and the results are discussed in the following sections.

C.7.1 Listening tests

In the listening tests, subjects were asked to identify a closed set of 12 English consonants in VCV context, presented over the headphones. Tests involved, diotic presentation of unprocessed speech, dichotic presentation of speech processed with i) spectral splitting with comb filters having sharp transitions, and ii) spectral splitting with optimized comb filters. These experiments are referred to as “Experiment V” in the tables and figures.

C.7.1.1 Test material and subjects

The test material was same as that used in the scheme of spectral splitting. Twelve English consonants /p, b, t, d, k, g, m, n, s, z, f, v/ in VCV context with vowel /a/. Simulation of sensorineural loss in normal hearing persons has been done for SNRs of ∞, 6, 3, 0, -3, -6, -9, -12, and -15 dB. Broadband Gaussian noise (from function generator HP 33120A) was added to the speech stimuli using the method described earlier in section 4.3.1).

Three normal hearing subjects (VK: M 27, AC: F 37, JK: M 25) participated in the listening test. Subjects AC and JK had participated in the experiment of TS-TR, All the three subjects had pure tone thresholds less than 20 dB HL in the frequency range of 125 Hz to 6 kHz.

C.7.1.2 Experimental procedure

To evaluate the schemes of spectral splitting using comb filters having sharp transition and the one using optimized comb filters, listening tests were conducted. Magnitude responses of the comb filters particularly for the two types are shown in Figs. C.2 and C.4. Experimental procedure and experimental set-up for conducting the listening tests are described in Appendix D and E respectively. Processed and unprocessed stimuli were added with broadband noise at different SNR conditions to simulate sensorineural hearing loss in normal hearing subjects. There were a total of 27 (3 processing conditions X 9 SNR) test conditions. Tests were conducted in an acoustically isolated chamber. Test scores were obtained in the form of confusion matrix. The confusion matrix shows the stimuli along rows and responses along columns. Each entry in the cell shows the number of times a stimulus-response pair occurs in the test.

In the test, 12 stimuli were used and each stimulus was presented 5 times in random order (as described in Appendix D). A test run takes about 5–8 minutes. Hence for 36 test conditions, each subject took about 40 hours for completion of all tests. Test sessions for the three subjects were spread over a span of about two months depending upon the availability and willingness of the subjects.

C.7.2 Test results

Results of listening tests for unprocessed and processed speech are presented in this section. For ascertaining the improvement in speech quality, a compilation of subject's qualitative assessment of the test stimuli processed under various listening conditions was carried out. Response times for unprocessed and processed speech were compared to assess the effectiveness of the processing scheme in reducing load on perception. Confusion matrix gives the percentage recognition score. One-tailed t-test analysis was carried out to determine the level of significance due to processing. Information transmission analysis, a measure that is not affected by subject's response bias was carried out. Information transmission analysis was also performed for speech features to determine the contribution of each feature in improving the speech perception.

C.7.2.1 Quality assessment

Compilation of subject's qualitative assessment indicated that under no-noise condition, subjects reported that the quality of processed speech, to be inferior compared to that of the unprocessed speech. However, under poor SNR conditions quality was better for processed speech signal than that for unprocessed signal. Among the two processing schemes, there was preference for speech processed with optimized comb filters.

C.7.2.2 Response time

Table C.2 gives the response times of individual subjects and averaged across the three subjects for all SNR conditions, for processed and unprocessed stimuli. Figure

C.5 shows the response times for unprocessed and processed speech. Relative improvement in response times is shown in Fig. C.6. For unprocessed speech, compared to the ∞ SNR condition, the response times increase with decrease in SNR indicating that, decrease in SNR increases the load on the perception. For unprocessed speech, average response time increased from 1.68 s under no-noise condition to 1.85, 2.05, 2.01, 2.14, 2.00, 2.19, 2.25, 2.40 s for 6, 3, 0, -3, -6, -9, -12, and -15 dB SNR conditions respectively. Averaged across the subjects, relative decrease in response times for processed speech with comb filters having sharp transitions were 0.0, 12.7, 6.6, 8.1, 0.3, 7.8, -12.4, 2.1% for 6, 3, 0, -3 -6, -9, -12, and -15dB SNR conditions respectively. For processed speech with optimized comb filters, the corresponding improvements were 1.6, 3.3, 7.5, 9.2, -2.5, 7.5, -4.0, and -1.9%.

C.7.2.3 Recognition score

Average of recognition scores of five tests, for each test condition, for each subject was obtained. t-test analysis was performed on the data to measure the level of significance in improvement due to processing (Snedector and Cochran, 1980). Table C.3 gives the recognition scores for unprocessed and processed speech, relative improvement in recognition scores for processed speech, t-test (one-tail) values. Figure C.7 shows the recognition scores (averaged across three subjects) for unprocessed and processed speech and relative improvements are shown in Fig. C.8. Recognition scores averaged across the subjects for unprocessed speech progressively decreases with decreasing SNR, from 100% at no-noise condition to 68% at -15 dB SNR conditions respectively. Relative improvements of recognition scores averaged across the subjects for processed speech with comb filters having sharp transitions were 1.3, 1.6, 2.7, 10.7, 9.3, 13.7, 18.0, 15.4% respectively for 6, 3, 0, -3, -6, -9, -12, -15 dB SNR conditions. For processed speech with optimized comb filters, the corresponding values were 1.6, 1.6, 2.9, 11.6, 13.1, 18.3, 22.0, 21.8%.

From the t-test (one-tailed) analysis, the improvements due to spectral splitting with optimized comb filters are highly significant for SNRs of -3, -6, -9, -12 and -15 dB.

C.7.2.4 Information transmission analysis

Table C.4 gives the overall information transmitted for all subjects under all test conditions. Figure C.9 shows the overall information transmitted for unprocessed and processed speech. Overall information transmitted averaged across three subjects for unprocessed speech, varied from 100% under no-noise condition to 98.7, 97.3, 96.7, 90.3, 89.3, 83, 78, 71.7% respectively for 6, 3, 0, -3, -6, -9, -12, -15 dB SNR conditions. Relative improvements (averaged across the three subjects) of overall information transmitted for processed speech with comb filters having sharp transition were 0.7, 2.1, 2.1, 8.2, 6.4, 11.1, 8.8, 9.2% respectively for 6, 3, 0, -3, -6, -9, -12, -15 dB SNR conditions. With optimized comb filters, the corresponding improvements were 1.4, 2.1, 2.8, 10.0, 9.6, 13.3, 12.4, 13.4%.

Table C.4 also gives the relative information transmitted for features of voicing, place, manner, nasality, frication and duration.

Voicing: Perception of voicing feature was degraded only at higher noise levels. Relative information transmitted for this feature were 95.7 and 86.3% at -12 and -15 dB and corresponding relative improvements with optimized comb filters were 1 and 2%.

Place: Figure C.10 shows the information transmitted for place feature. Relative information transmitted for place feature averaged across three subjects for unprocessed speech, varied from 100% under no-noise condition to 94.3, 90.7, 91.7, 73.3, 69.7, 53, 43.3, 32% respectively for 6, 3, 0, -3, -6, -9, -12, -15 dB SNR conditions. Relative improvements for processed speech with comb filters having sharp transitions were 5.7, 10.4, 7.4, 29.6, 31.5, 61.7, 53.3, 71.1% respectively for 6, 3, 0, -3, -6, -9, -12, -15 dB SNR conditions. With optimized comb filters, the corresponding improvements were 6.4, 8.9, 6.8, 34.0, 40.0, 62.3, 62.8, 88.0%.

Manner: Figure C.11 shows the information transmitted for manner feature. Relative information transmitted for this feature averaged across three subjects for unprocessed speech, varied from 100% under no-noise condition to 98.3, 96.3, 89.7, 85.7, 78.3, 65, 53.7% respectively for 3, 0, -3, -6, -9, -12, -15 dB SNR conditions. Relative improvements for processed speech with comb filters having sharp transitions were 1.0, 2.3, 9.6, 9.5, 9.9, 28.3,

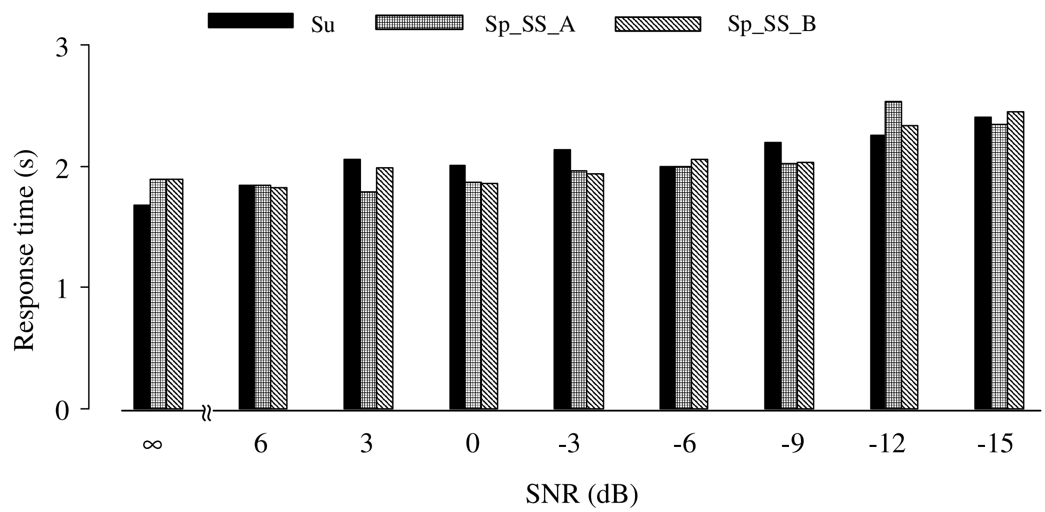


FIG. C.5. Experiment V. (SS). Averaged response time (s). Su: Unprocessed speech. Sp_SS_A and Sp_SS_B correspond to processed speech with comb filters with sharp transitions and optimized comb filters respectively.

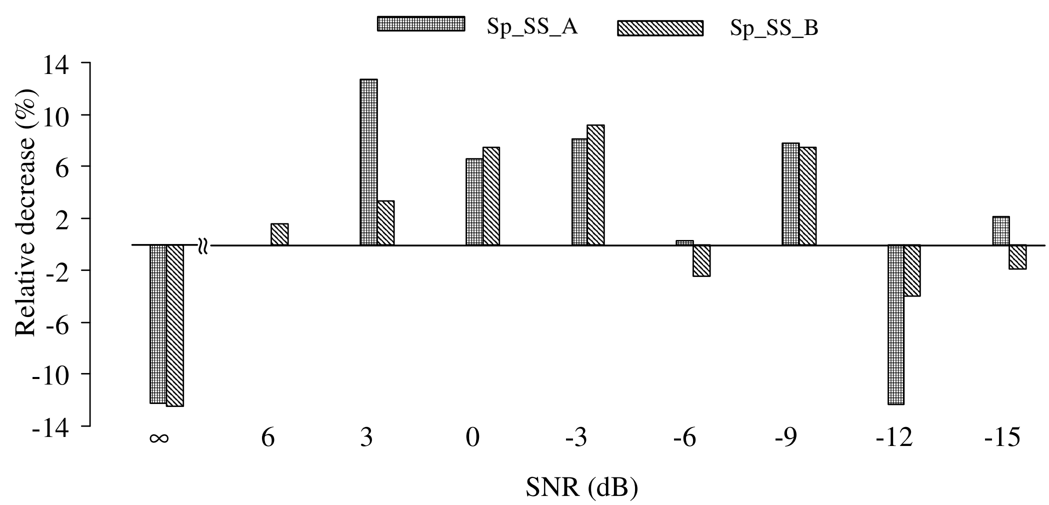


FIG. C.6. Experiment V. (SS). Averaged relative decrease (%) in response times (s). Sp_SS_A and Sp_SS_B correspond to processed speech with comb filters with sharp transitions and optimized comb filters respectively.

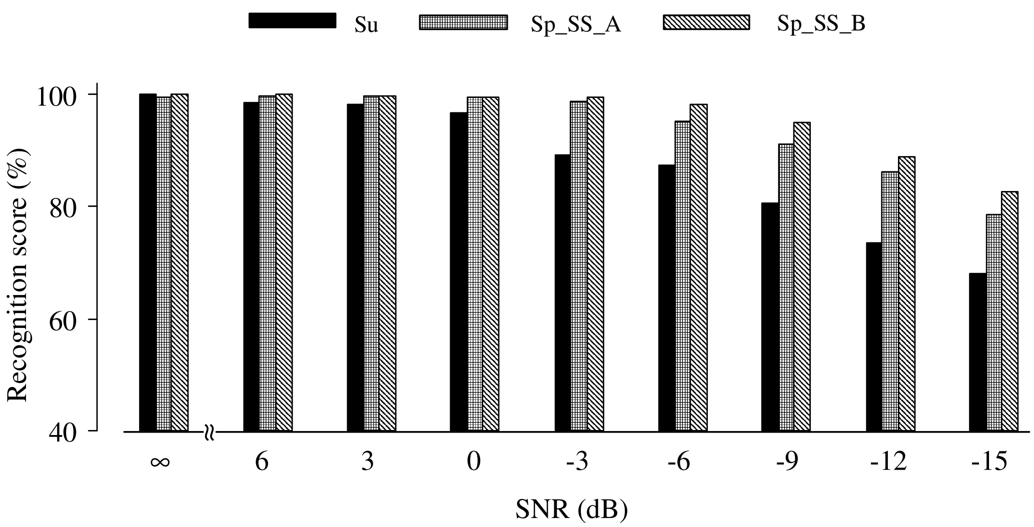


FIG. C.7. Experiment V. (SS). Averaged recognition score (%). Su: Unprocessed speech. Sp_SS_A and Sp_SS_B correspond to processed speech with comb filters with sharp transitions and optimized comb filters respectively.

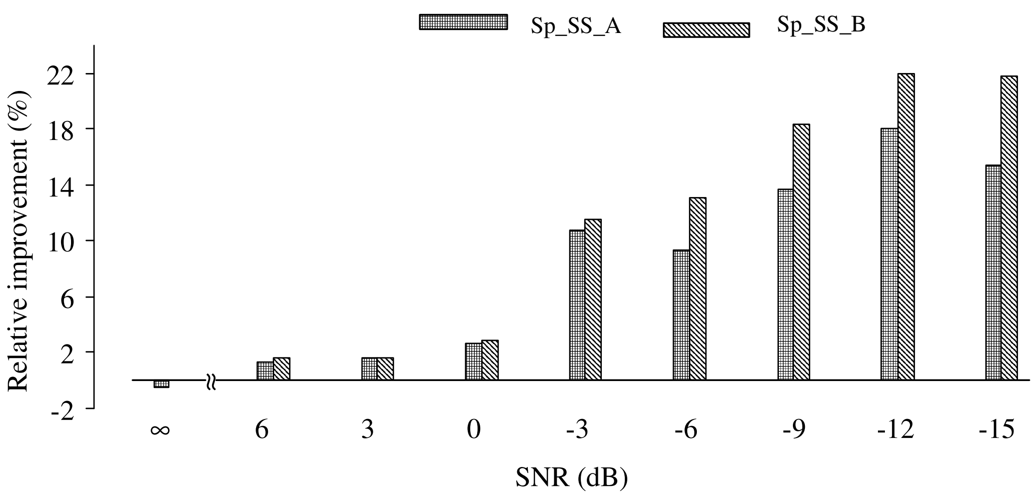


FIG. C.8. Experiment V. (SS). Averaged relative improvement (%) in recognition score. Sp_SS_A and Sp_SS_B correspond to processed speech with comb filters with sharp transitions and optimized comb filters respectively.

25.0% respectively for 3, 0, -3, -6, -9, -12, -15 dB SNR conditions. For processed speech with optimized comb filters, the corresponding improvements were 0.3, 2.6, 11.9, 17.8, 28.5, 32.1, 47.0%.

Nasality: Perception of nasality feature was degraded at low SNR conditions. Relative information transmitted for this feature at -12 and -15 dB were 85 and 58%. With optimized comb filters relative improvements were 13 and 63%.

Frication: Figure C.12 shows the information transmitted for frication feature. Relative information transmitted for this feature averaged across three subjects for unprocessed speech, varied from 100% under no-noise condition to 97.3, 95, 84.7, 77.7, 67, 49.3, 42% respectively for 3, 0, -3, -6, -9, -12, -15 dB SNR conditions. Relative improvements for processed speech with comb filters having sharp transitions were 1.8, 3.2, 16.5, 16.2, 26.8, 58.5, 24.0% respectively for 3, 0, -3, -6, -9, -12, -15 dB SNR conditions. With optimized comb filters, corresponding improvements were 0.7, 3.5, 18.8, 31.6, 43.9, 59.8, 62.7%.

Duration: Figure C.13 shows the information transmitted for duration feature. Relative information transmitted for duration feature averaged across three subjects for unprocessed speech, varied from 100% under no-noise condition to 88, 93, 91.7, 75.0, 73.7, 61.3, 48.3, 40.7% respectively for 6, 3, 0, -3, -6, -9, -12, -15 dB SNR conditions. Relative improvements for processed speech with comb filters having sharp transitions were 17.1, 7.6, 5.5, 24.5, 18.8,38.1, 44.9, -7.3% respectively for 6, 3, 0, -3, -6, -9, -12, -15 dB SNR conditions. With optimized comb filters, the corresponding improvements were 18.8, 7.6, 5.1, 30.3, 34.4, 44.6, 44.6, 13.4%.

C.7.3 Discussion

The comparison of the two schemes of processing; dichotic presentation of speech processed with comb filter pair having sharp transitions and those with optimized comb filters were done by conducting listening tests. The two schemes were evaluated by conducting listening tests on three normal hearing subjects with simulated sensorineural hearing loss. Subject's response times, percentage correct recognition score, relative information transmitted for different speech features for unprocessed and processed speech (for two processing conditions) were analyzed.

From the response times, it can be seen that under low SNR conditions, there is moderate decrease in response time due to processing. Decreases in response times are generally higher with spectral splitting using optimized comb filters. Hence the speech processing with optimized comb filters may be considered to be somewhat more effective in reducing the load on perception under adverse listening condition.

From the recognition scores, it was found that as the noise level increases, scores decrease for unprocessed speech. Processing improved the recognition scores. Improvements are generally higher for scheme of spectral splitting with optimized comb filters.

Overall relative information transmitted and relative information transmitted for different speech features was obtained. Overall relative information transmitted is higher for scheme of spectral splitting with optimized comb filters. It is observed that, among the features, there is maximum improvement for place features, at SNRs less than -3 dB, indicating that the processing helps in reducing the effect of spectral masking. Relative information transmitted for frication and manner features also improved with processing and the improvements are higher with optimized comb filters.

Thus we conclude that comb filters with responses optimized for reducing the perceived spectral distortion are more effective than the ones with sharp transitions.

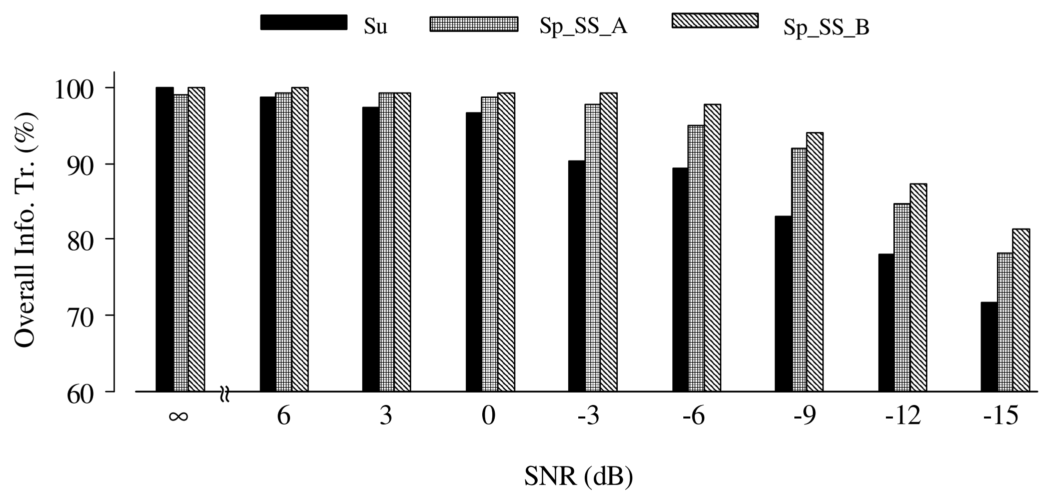


FIG. C.9. Experiment V. (SS). Averaged overall relative information transmitted (%). Su: Unprocessed speech. Sp_SS_A and Sp_SS_B correspond to processed speech with comb filters with sharp transitions and optimized comb filters respectively.

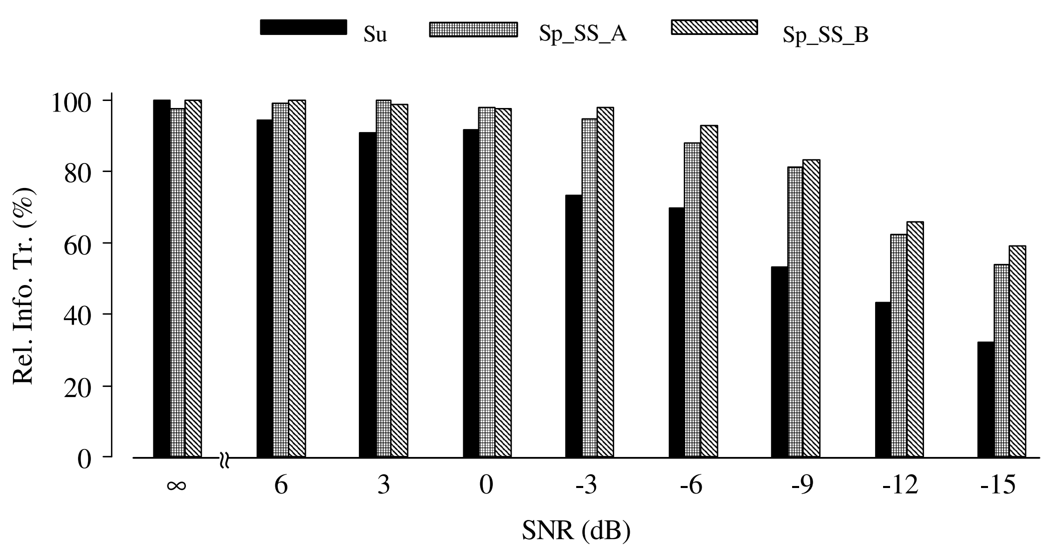


FIG. C.10. Experiment V. (SS). Averaged relative information transmitted (%) for place feature. Su: Unprocessed speech. Sp_SS_A and Sp_SS_B correspond to processed speech with comb filters with sharp transitions and optimized comb filters respectively.

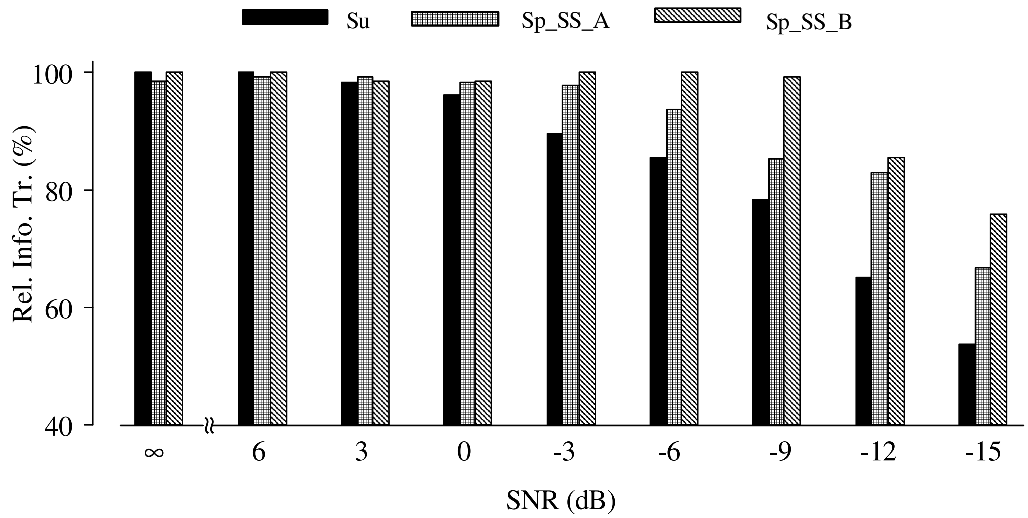


FIG. C.11. Experiment V. (SS). Averaged relative information transmitted (%) for manner feature. Su: Unprocessed speech. Sp_SS_A and Sp_SS_B correspond to processed speech with comb filters with sharp transitions and optimized comb filters respectively.

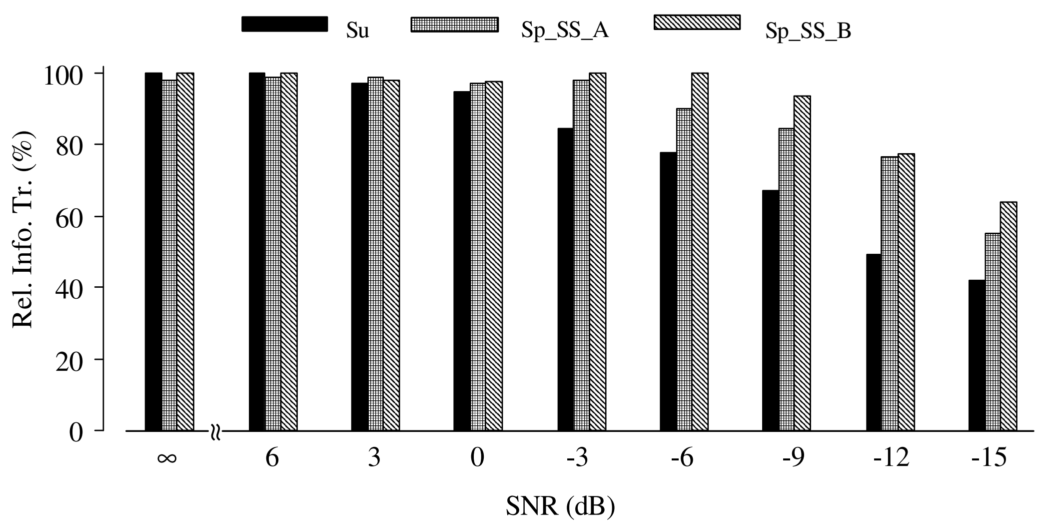


FIG. C.12. Experiment V. (SS). Averaged relative information transmitted (%) for frication feature. Su: Unprocessed speech. Sp_SS_A and Sp_SS_B correspond to processed speech with comb filters with sharp transitions and optimized comb filters respectively.

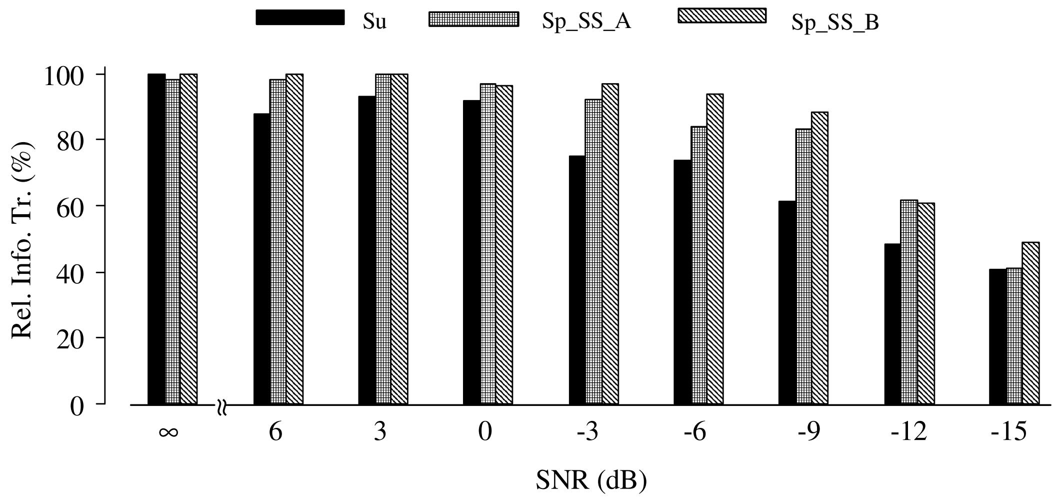


FIG. C.13. Experiment V. (SS). Averaged relative information transmitted (%) for duration feature. Su: Unprocessed speech. Sp_SS_A and Sp_SS_B correspond to processed speech with comb filters with sharp transitions and optimized comb filters respectively.

TABLE C.2. Experiment V. (SS). Response time (s) for Unprocessed Speech (Su) and for Processed Speech corresponding to spectral splitting with (i) comb filters with sharp transition between bands (Sp_SS_A) and (ii) optimized comb filters (Sp_SS_B). S: Subject.

S	∞ SNR			6 dB SNR			3 dB SNR			0 dB SNR			-3 dB SNR			-6 dB SNR			-9 dB SNR			-12 dB SNR			-15 dB SNR		
	Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_	
		A	B		A	B		A	B		A	B		A	B		A	B		A	B		A	B		A	B
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	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	1.50	1.58	1.31	1.79	1.48	1.48	1.68	1.58	1.46	1.79	1.83	1.55	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	2.09	2.15	2.04	2.25	2.01	2.10	2.37	1.83	2.02	2.34	2.02	2.27	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.	1.68	1.89	1.89	1.85	1.85	1.82	2.05	1.79	1.98	2.01	1.87	1.86	2.14	1.96	1.94	2.00	2.00	2.05	2.19	2.02	2.03	2.25	2.53	2.34	2.40	2.35	2.45

TABLE C.3. Experiment V. (SS). Recognition score (%) for Unprocessed Speech (Su) and Processed Speech corresponding to spectral splitting with (i) comb filters with sharp transition between bands (Sp_SS_A) and (ii) optimized comb filters (Sp_SS_B). 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) with n (number of tests) = 5, df = 8.

S		∞ SNR			6 dB SNR			3 dB SNR			0 dB SNR			-3 dB SNR			-6 dB SNR			-9 dB SNR			-12 dB SNR			-15 dB SNR		
		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_	
			A	B		A	B		A	B		A	B		A	B		A	B		A	B		A	B		A	B
VK	RS	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	92.7	98.3	99.0	82.3	99.0	98.7	73.3	87.3	90.7	75.7	94.3	94.0
	s.d.	0	3.0	0	2.54	0	0	1.39	0	0.76	2.35	0.76	0	1.82	2.24	0.93	3.26	1.65	0.93	3.04	1.48	1.39	3.1	5.83	2.53	6.29	2.51	1.91
	R.I.		-1.3	0.0		4.2	4.2		1.3	1.0		3.1	3.4		9.6	10.0		6.0	6.8		20.3	19.9		19.1	23.7		24.6	24.2
	p		0.173	-		0.004	0.004		0.031	0.097		0.013	0.006		0	0		0.004	0.001		0	0		0.001	0		0.0001	0.0001
JK	RS	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	93.0	98.3	99.7	87.7	87.3	96.7	83.3	87.7	89.0	71.7	73.3	77.7
	s.d.	0	0	0	0	0	0	1.81	0	0	1.48	0.76	0	3.04	1.82	0	1.42	1.17	0.76	3.47	3.44	2.64	4.08	3.04	1.56	3.35	2.04	3.47
	R.I.		0.0	0.0		0.0	0.0		1.3	1.3		0.7	1.0		7.7	9.9		5.7	7.2		-0.5	10.3		5.3	6.8		2.2	8.4
	P		-	-		-	-		0.07	0.07		0.2	0.085		0.001	0		0	0		0.443	0		0.046	0.01		0.183	0.012
AC	RS	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	76.7	89.0	96.0	72.0	87.3	89.7	64.0	83.0	86.7	57.0	68.0	75.7
	s.d.	0	0	0	0.93	1.48	0	1.89	0.93	0.76	1.42	2.17	1.39	1.39	0.93	0.93	2.64	1.91	2.23	1.37	2.8	2.17	4.51	3.18	2.35	3.6	3.43	4.01
	R.I.		-0.3	0.0		-0.3	0.7		2.1	2.5		4.2	4.2		14.7	15.1		16.0	25.2		21.3	24.6		29.7	35.5		19.3	32.8
	P		-	-		0.346	0.07		0.034	0.017		0.004	0.001		0	0		0	0		0	0		0	0		.0006	0
Avg. RS		100.0	99.5	100.0	98.4	99.7	100.0	98.2	99.8	99.8	96.8	99.4	99.6	89.2	98.7	99.5	87.5	95.2	98.2	80.7	91.2	95.0	73.5	86.0	88.8	68.1	78.5	82.5
Avg. R.I.			-0.5	0.0		1.3	1.6		1.6	1.6		2.7	2.9		10.7	11.6		9.3	13.1		13.7	18.3		18.0	22.0		15.4	21.8

TABLE C.3. Experiment V. (SS). Relative information transmitted (%) for Unprocessed Speech (Su) and for Processed Speech corresponding to spectral splitting with (i) comb filters with sharp transition between bands (Sp_SS_A) and (ii) optimized comb filters (Sp_SS_B) for (a) overall, and feature groupings: (b) voicing, (c) place, (d) manner, (e) nasality, (f) frication, and (g) duration. S: Subject.

(a) Overall

S	∞ SNR			6 dB SNR			3 dB SNR			0 dB SNR			-3 dB SNR			-6 dB SNR			-9 dB SNR			-12 dB SNR			-15 dB SNR		
	Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_	
		A	B		A	B		A	B		A	B		A	B		A	B		A	B		A	B		A	B
VK	100	98	100	97	100	100	98	100	99	97	99	100	90	98	99	93	98	99	84	98	98	76	85	89	76	92	91
JK	100	100	100	100	100	100	98	100	100	98	99	100	93	97	100	93	98	99	88	90	96	85	87	88	73	70	78
AC	100	99	100	99	98	100	96	98	99	95	98	98	88	98	99	82	89	95	77	88	88	73	82	85	66	73	75
Avg.	100	99	100	98.7	99.3	100	97.3	99.3	99.3	96.7	98.7	99.3	90.3	97.7	99.3	89.3	95	97.7	83	92	94	78	84.7	87.3	71.7	78.3	81.3

(b) Feature: voicing

S	∞ SNR			6 dB SNR			3 dB SNR			0 dB SNR			-3 dB SNR			-6 dB SNR			-9 dB SNR			-12 dB SNR			-15 dB SNR		
	Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_	
		A	B		A	B		A	B		A	B		A	B		A	B		A	B		A	B		A	B
VK	100	100	100	100	100	100	100	100	100	100	100	100	100	97	100	100	100	100	97	100	100	100	95	100	100	97	100
JK	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	78	100	90	87	92	68	70	69
AC	100	100	100	100	97	100	97	97	100	97	100	100	97	100	100	100	100	97	97	97	100	97	83	97	91	86	94
Avg.	100.0	100.0	100.0	100.0	99.0	100.0	99.0	99.0	100.0	99.0	100.0	100.0	99.0	99.0	100.0	100.0	100.0	99.0	98.0	91.7	100.0	95.7	88.3	96.3	86.3	84.3	87.7

(c) Feature: place

S	∞ SNR			6 dB SNR			3 dB SNR			0 dB SNR			-3 dB SNR			-6 dB SNR			-9 dB SNR			-12 dB SNR			-15 dB SNR		
	Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_	
		A	B		A	B		A	B		A	B		A	B		A	B		A	B		A	B		A	B
VK	100	95	100	87	100	100	87	100	98	97	98	100	76	96	97	85	92	95	59	97	94	40	58	70	37	82	82
JK	100	100	100	100	100	100	94	100	100	95	98	100	76	91	100	76	94	98	64	72	87	60	69	66	37	39	51
AC	100	98	100	96	98	100	91	100	98	83	98	93	68	97	97	48	78	86	36	75	69	30	60	61	22	41	45
Avg.	100	97.7	100	94.3	99.3	100	90.7	100	98.7	91.7	98	97.7	73.3	94.7	98	69.7	88	93	53	81.3	83.3	43.3	62.3	65.7	32	54	59.3

(d) Feature: manner

S	∞ SNR			6 dB SNR			3 dB SNR			0 dB SNR			-3 dB SNR			-6 dB SNR			-9 dB SNR			-12 dB SNR			-15 dB SNR		
	Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_	
		A	B		A	B		A	B		A	B		A	B		A	B		A	B		A	B		A	B
VK	100	96	100	100	100	100	97	100	98	91	100	100	89	98	100	88	100	100	80	96	100	65	87	90	69	91	87
JK	100	100	100	100	100	100	100	100	100	100	100	100	96	98	100	94	98	100	88	82	98	71	82	88	54	56	72
AC	100	100	100	100	98	100	98	98	98	98	95	96	84	98	100	75	83	100	67	78	100	59	80	79	38	53	69
Avg.	100	98.7	100.0	100.0	99.3	100.0	98.3	99.3	98.7	96.3	98.3	98.7	89.7	98.0	100.0	85.7	93.7	100.0	78.3	85.3	99.3	65	83.0	85.7	53.7	66.7	76.0

(e) Feature: nasality

S	∞ SNR			6 dB SNR			3 dB SNR			0 dB SNR			-3 dB SNR			-6 dB SNR			-9 dB SNR			-12 dB SNR			-15 dB SNR		
	Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_	
		A	B		A	B		A	B		A	B		A	B		A	B		A	B		A	B		A	B
VK	100	100	100	100	100	100	100	100	100	100	100	100	100	96	100	100	100	100	95	100	100	82	100	100	92	100	100
JK	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	59	100	78	78	91	28	50	61
AC	100	100	100	100	100	100	100	100	100	96	100	100	95	100	100	95	100	100	96	100	100	95	96	95	54	95	88
Avg.	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	98.7	100.0	100.0	98.3	98.7	100.0	98.3	100.0	100.0	97.0	86.3	100.0	85.0	91.3	95.3	58.0	81.7	83.0

(f) Feature: frication

S	∞ SNR			6 dB SNR			3 dB SNR			0 dB SNR			-3 dB SNR			-6 dB SNR			-9 dB SNR			-12 dB SNR			-15 dB SNR		
	Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_	
		A	B		A	B		A	B		A	B		A	B		A	B		A	B		A	B		A	B
VK	100	94	100	100	100	100	95	100	97	85	100	100	82	100	100	80	100	100	72	94	100	55	79	83	54	85	79
JK	100	100	100	100	100	100	100	100	100	100	100	100	94	97	100	90	97	100	80	97	97	56	83	81	48	57	62
AC	100	100	100	100	97	100	97	97	97	100	92	93	78	97	100	63	73	100	49	63	84	37	68	68	24	23	51
Avg.	100	98	100	100.0	99.0	100.0	97.3	99.0	98.0	95.0	97.3	97.7	84.7	98.0	100.0	77.7	90.0	100.0	67	84.7	93.7	49.3	76.7	77.3	42.0	55.0	64.0

(e) Feature: duration

S	∞ SNR			6 dB SNR			3 dB SNR			0 dB SNR			-3 dB SNR			-6 dB SNR			-9 dB SNR			-12 dB SNR			-15 dB SNR		
	Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_	
		A	B		A	B		A	B		A	B		A	B		A	B		A	B		A	B		A	B
VK	100	94	100	64	100	100	91	100	100	91	95	100	78	92	91	96	85	91	72	100	100	43	59	65	49	79	94
JK	100	100	100	100	100	100	96	100	100	96	100	100	80	85	100	70	92	100	68	79	100	72	63	57	43	33	28
AC	100	100	100	100	95	100	92	100	100	88	95	89	67	100	100	55	75	91	44	70	65	30	63	61	30	12	25
Avg.	100	98	100	88.0	98.3	100.0	93.0	100.0	100.0	91.7	96.7	96.3	75.0	92.3	97.0	73.7	84.0	94.0	61.3	83.0	88.3	48.3	61.7	61.0	40.7	41.3	49.0

TABLE C.4. Experiment V. (SS). Relative information transmitted (%) for Unprocessed Speech (Su) and for Processed Speech corresponding to spectral splitting with (i) comb filters with sharp transition between bands (Sp_SS_A) and (ii) optimized comb filters (Sp_SS_B) for (a) overall, and feature groupings: (b) voicing, (c) place, (d) manner, (e) nasality, (f) frication, and (g) duration. S: Subject.

(b) Overall

S	∞ SNR			6 dB SNR			3 dB SNR			0 dB SNR			-3 dB SNR			-6 dB SNR			-9 dB SNR			-12 dB SNR			-15 dB SNR		
	Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_	
		A	B		A	B		A	B		A	B		A	B		A	B		A	B		A	B		A	B
VK	100	98	100	97	100	100	98	100	99	97	99	100	90	98	99	93	98	99	84	98	98	76	85	89	76	92	91
JK	100	100	100	100	100	100	98	100	100	98	99	100	93	97	100	93	98	99	88	90	96	85	87	88	73	70	78
AC	100	99	100	99	98	100	96	98	99	95	98	98	88	98	99	82	89	95	77	88	88	73	82	85	66	73	75
Avg.	100	99	100	98.7	99.3	100	97.3	99.3	99.3	96.7	98.7	99.3	90.3	97.7	99.3	89.3	95	97.7	83	92	94	78	84.7	87.3	71.7	78.3	81.3

(b) Feature: voicing

S	∞ SNR			6 dB SNR			3 dB SNR			0 dB SNR			-3 dB SNR			-6 dB SNR			-9 dB SNR			-12 dB SNR			-15 dB SNR		
	Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_	
		A	B		A	B		A	B		A	B		A	B		A	B		A	B		A	B		A	B
VK	100	100	100	100	100	100	100	100	100	100	100	100	100	97	100	100	100	100	97	100	100	100	95	100	100	97	100
JK	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	78	100	90	87	92	68	70	69
AC	100	100	100	100	97	100	97	97	100	97	100	100	97	100	100	100	100	97	97	97	100	97	83	97	91	86	94
Avg.	100.0	100.0	100.0	100.0	99.0	100.0	99.0	99.0	100.0	99.0	100.0	100.0	99.0	99.0	100.0	100.0	100.0	99.0	98.0	91.7	100.0	95.7	88.3	96.3	86.3	84.3	87.7

(c) Feature: place

S	∞ SNR			6 dB SNR			3 dB SNR			0 dB SNR			-3 dB SNR			-6 dB SNR			-9 dB SNR			-12 dB SNR			-15 dB SNR		
	Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_	
		A	B		A	B		A	B		A	B		A	B		A	B		A	B		A	B		A	B
VK	100	95	100	87	100	100	87	100	98	97	98	100	76	96	97	85	92	95	59	97	94	40	58	70	37	82	82
JK	100	100	100	100	100	100	94	100	100	95	98	100	76	91	100	76	94	98	64	72	87	60	69	66	37	39	51
AC	100	98	100	96	98	100	91	100	98	83	98	93	68	97	97	48	78	86	36	75	69	30	60	61	22	41	45
Avg.	100	97.7	100	94.3	99.3	100	90.7	100	98.7	91.7	98	97.7	73.3	94.7	98	69.7	88	93	53	81.3	83.3	43.3	62.3	65.7	32	54	59.3

(d) Feature: manner

S	∞ SNR			6 dB SNR			3 dB SNR			0 dB SNR			-3 dB SNR			-6 dB SNR			-9 dB SNR			-12 dB SNR			-15 dB SNR		
	Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_	
		A	B		A	B		A	B		A	B		A	B		A	B		A	B		A	B		A	B
VK	100	96	100	100	100	100	97	100	98	91	100	100	89	98	100	88	100	100	80	96	100	65	87	90	69	91	87
JK	100	100	100	100	100	100	100	100	100	100	100	100	96	98	100	94	98	100	88	82	98	71	82	88	54	56	72
AC	100	100	100	100	98	100	98	98	98	98	95	96	84	98	100	75	83	100	67	78	100	59	80	79	38	53	69
Avg.	100	98.7	100.0	100.0	99.3	100.0	98.3	99.3	98.7	96.3	98.3	98.7	89.7	98.0	100.0	85.7	93.7	100.0	78.3	85.3	99.3	65	83.0	85.7	53.7	66.7	76.0

(e) Feature: nasality

S	∞ SNR			6 dB SNR			3 dB SNR			0 dB SNR			-3 dB SNR			-6 dB SNR			-9 dB SNR			-12 dB SNR			-15 dB SNR		
	Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_	
		A	B		A	B		A	B		A	B		A	B		A	B		A	B		A	B		A	B
VK	100	100	100	100	100	100	100	100	100	100	100	100	100	96	100	100	100	100	95	100	100	82	100	100	92	100	100
JK	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	59	100	78	78	91	28	50	61
AC	100	100	100	100	100	100	100	100	100	96	100	100	95	100	100	95	100	100	96	100	100	95	96	95	54	95	88
Avg.	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	98.7	100.0	100.0	98.3	98.7	100.0	98.3	100.0	100.0	97.0	86.3	100.0	85.0	91.3	95.3	58.0	81.7	83.0

(f) Feature: frication

S	∞ SNR			6 dB SNR			3 dB SNR			0 dB SNR			-3 dB SNR			-6 dB SNR			-9 dB SNR			-12 dB SNR			-15 dB SNR		
	Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_	
		A	B		A	B		A	B		A	B		A	B		A	B		A	B		A	B		A	B
VK	100	94	100	100	100	100	95	100	97	85	100	100	82	100	100	80	100	100	72	94	100	55	79	83	54	85	79
JK	100	100	100	100	100	100	100	100	100	100	100	100	94	97	100	90	97	100	80	97	97	56	83	81	48	57	62
AC	100	100	100	100	97	100	97	97	97	100	92	93	78	97	100	63	73	100	49	63	84	37	68	68	24	23	51
Avg.	100	98	100	100.0	99.0	100.0	97.3	99.0	98.0	95.0	97.3	97.7	84.7	98.0	100.0	77.7	90.0	100.0	67	84.7	93.7	49.3	76.7	77.3	42.0	55.0	64.0

(e) Feature: duration

S	∞ SNR			6 dB SNR			3 dB SNR			0 dB SNR			-3 dB SNR			-6 dB SNR			-9 dB SNR			-12 dB SNR			-15 dB SNR		
	Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_		Su	Sp_SS_	
		A	B		A	B		A	B		A	B		A	B		A	B		A	B		A	B		A	B
VK	100	94	100	64	100	100	91	100	100	91	95	100	78	92	91	96	85	91	72	100	100	43	59	65	49	79	94
JK	100	100	100	100	100	100	96	100	100	96	100	100	80	85	100	70	92	100	68	79	100	72	63	57	43	33	28
AC	100	100	100	100	95	100	92	100	100	88	95	89	67	100	100	55	75	91	44	70	65	30	63	61	30	12	25
Avg.	100	98	100	88.0	98.3	100.0	93.0	100.0	100.0	91.7	96.7	96.3	75.0	92.3	97.0	73.7	84.0	94.0	61.3	83.0	88.3	48.3	61.7	61.0	40.7	41.3	49.0

Appendix D

Programs for signal processing and listening experiments

D.1 Introduction

To reduce the effect of temporal masking associated with degraded temporal resolution in sensorineural hearing loss, a scheme of temporal splitting for binaural dichotic presentation has been developed. In the scheme, step and trapezoidal fading functions along with overlap have been used. These schemes of temporal splitting ars referred as TS_ST and TS_TR (temporal splitting with step fading and temporal splitting with trapezoidal fading). To reduce the effect of spectral masking associated with broad auditory filters in sensorineural hearing loss, a scheme of spectral splitting using a pair of comb filters having complementary magnitude response is used and is referred as SS. To reduce the effect of both increased spectral and temporal masking, combined scheme of spectral and temporal splitting (CS) has been investigated by using time varying comb filters. Programs for these schemes and listening tests are given in the following sections.

D.2 Temporal splitting (TS)

The temporal processing of speech involves temporally splitting the speech into various segments for presenting adjacent segments to the two ears (binaural dichotic presentation). For offline processing of digitized speech signal, a program "`split`" is developed in C. Sampling rate f_s is set at 10 k samples/s. Input parameters for processing are

Inter-aural switching period, T_p: 10 to 50 ms

On-period, T_{on}: $T_p/2$ to T_p

Transition duration, T_t: 0 to $0.5(T_p - T_{on})$

If trapezoidal fading (i.e. $T_t \neq 0$), fading curve: linear/logarithmic

Inter-aural switching period is the repetition period for switching between the two ears. With reference to Eqns. 4.1 and 4.2 and Figs. 4.2 and 4.3, the number of samples in the switching period, on-interval, and transition interval are

$N = T_p f_s$,

$L = T_{on} f_s$, and

$M = T_t f_s$

As shown in the Figs 4.2 and 4.3, fading function is symmetrical. Transition duration provides slope for the trapezoidal fading function. In trapezoidal fading, variation can be linear or logarithmic. The duty cycle for presentation to a particular ear is T_{on}/T_p, i.e., L/N.

The program accepts input file with signal samples as 16-bit integers in text format, and generates two output files in text format. Program "`t2b`" is used for converting the signal files from text format to binary format, in order to have better storage efficiency and access speed.

D.3. Spectral splitting (SS)

In spectral splitting, speech is split into odd and even bands using a pair of comb filters with complementary magnitude response, each with 9 pass bands corresponding to critical band auditory filters (Zwicker, 1961). The comb filter responses are obtained through two linear phase FIR filters, designed as described in Appendix C. The two filters are realized as part of one program for off-line processing of digitized speech signal. This program "`dico`" is a modified version of a program "`dicho`" developed earlier by Ratanpal (2000).

The program asks for names of the input '.bin' speech signal file, the two coefficient files in '.txt' format, and the two '.bin' output files for the output signals from the two comb filters. 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} .

D.4 Combined splitting (CS) using time varying comb filters

In the scheme of spectral and temporal splitting, time varying comb filters are used to sweep the odd and even bands between the two ears. The frequency range between two alternate critical band center frequencies, was divided into m equal parts, to obtain the center frequencies of the pass bands of successive pair of comb filter in a time varying comb filter with m shiftings. This procedure was carried out for obtaining the center frequencies of the critical bands for the entire frequency range for each successive pair of comb filter. Filter coefficients were obtained for each shifted pair of complementary comb filters using frequency sampling technique of linear phase FIR filter as described in Appendix C. The precalculated set of coefficients were cyclically swept with m shiftings with a sweeping period of T_p ms. After every time slot of T_p /m ms, a new set of coefficients for next pair of comb filters takes over.

For off-line processing of digitized speech signal, a program "`filt`" is written in C, jointly with my colleague A. N. Cheeran. Each pair of comb filter processes the signal for a time period equal to shifting time period divided by the number of shiftings. Program asks for the name of input speech signal input file in binary format and number of shiftings of comb filters. For a given number of shiftings, program reads corresponding coefficient files. For m shiftings, each pair of comb filter will process the signal for $T_p f_s/m$ samples, where T_p (ms) and f_s (samples/s) represent respectively sweeping (switching) period and sampling frequency. Program has facility for recording and playing the recorded signal or processed signals.

D.5 Monaural-binaural hearing level difference

To obtain the level difference in monaural and binaural perception, listening tests were conducted with signals of various types. Signals are presented monaurally followed by binaurally. The procedure is given in detail in Appendix C.

Program “`mono_bin`” has been written in C to present the signals, through two 12-bit DAC channels of PC based PCL-208 data acquisition card (Dynalog Micro Systems, Bombay, 1989), for monaural and binaural hearing alternately. The program outputs two signals alternately with a short time gap of about 100 ms, one signal for monaural and the other for binaural hearing. This pair of presentation is repeated with a time gap of 2 seconds.

Programs prompts the user to enter

* two file names in binary format for monaural and binaural presentation.
* time gap between monaural and binaural presentation
* time gap between repeated presentation for monaural and binaural hearing

D.6 Simulation of sensorineural hearing loss

The speech processing schemes were evaluated by conducting listening tests on normal hearing subjects with simulated hearing loss. Simulation of sensorineural hearing loss in normal hearing subjects was obtained by adding broad band noise to the speech signal. A program “`snr`”, initially developed by Prasad (1996) and later modified by Chaudhari (2000) was used to add the noise to the speech signal at various SNR conditions. In the experiments SNRs of 6, 3, 0, -3, -6, -9 -12, and -15 were used. The program adds the noise to the signal in such a way that SNR was kept constant on the basis of short-time (10 ms) energy of the signal. The program accepts the noise file ‘`noise.bin`’ from the working directory and asks for the binary file names of input and output signals, and the SNR (dB) value k. A block of 100 samples (equal to 100 samples for sampling rate of 10 k samples/sec) from input signal $s(n)$ and noise $e(n)$ are read and the output samples are calculated and are stored in output file. This procedure is repeated for successive set of 100 samples of the input signal and noise. The output $x(n)$ is obtained by adding signal $s(n)$ with new $e(n)$ scaled by a factor β such that SNR = k dB, on the basis of signal and noise energy in the block.

$$x(n) = s(n) + \beta\, e(n) \tag{D.1}$$

where $\beta^2 = \sigma_s^{\,2} / (\sigma_e^{\,2} \times 10^{0.1k})$

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

since the noise is stationary, σ_e remains almost same in all the blocks, and therefore the weighting factor β depends on the short-time energy of the input signal σ_s.

D.7 Listening test

For conducting listening tests, the automated listening test set-up used a program "`test`". Program "`tlist`" was used to randomize the presentation of stimuli and positions of response choices and "`info`" was used for information transmission analysis. These programs were initially developed using Fortran by Pandey (1987) and are successively modified in Pascal and 'C' by Thomas (1996), Chaudhari (2000), and the author. These programs are described in this section. The program "`test`" was used for outputting the signal on single channel and was modified to present output on two channels, one for left and other for right. The program used PCL-208 data acquisition card (Dynalog Micro Systems, Bombay, 1989) for outputting the signal at 10 k samples/s through a 12-bit DAC and com2 port for communication to the subject terminal. Subject terminal was used for displaying the test instructions to the subject in the beginning of each test run, alerting the subject while presenting the stimuli and for displaying the response choices on the screen. Subject responded through the terminal keyboard.

Program prompts the user to enter 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:

In each experimental run, the stimuli were presented in a random order with certain uniformity constraints to avoid biasing due to learning effect. Following uniformity constraints were used in randomizing the order of stimulus presentation (Pandey, 1987):

1) overall uniformity. For n stimulus items and N presentations in an experiment, each stimulus should be presented N/n times.
2) mid-range uniformity. In M consecutive presentations, any stimulus should not occur more than M/n +2 times.

3) short-range uniformity. An item should not be presented more than three times consecutively.

A program "`tlist`" was used for generating randomized test list for presenting test stimuli in the experiment. The program reads the information from a file '`list.dat`' in the following format:

Number of test stimuli *n*,

Number of presentations *N*,

Two line title for the experiment,

Test item names on separate lines.

,

The program outputs a file '`slist.dat`' that will be used by the computerized test administration program. The stimulus and response items are represented by order number. The list file contains the item to be presented as well as the order in which response choices should be displayed. The order of the response choices and the position of correct response are also randomized with the same uniformity constraints as described above, to avoid biasing in responses due to the order in which response choices are provided to the subject. To provide randomization across the test runs, 25 files (`slist1`, `slist2`, - - - `slist25`) were generated.

The test program executes the following tasks:

1) displays the message about the test stimuli and test instructions to the subject.
2) present the stimuli, record the subject response and response time, display the scores etc.
3) record the scores as a confusion matrix, display and store the test data for further analysis.

The test information to the program is provided in a file '`sigfil.dat`' in the following format:

Sampling frequency in Hz,

Number of items in the test set,

Names of test items (to be displayed as responses choices),

Names of data files containing signal samples for the stimuli.

D.8 Presentation procedure

At the beginning of the test sessions, the subject is instructed about the test material and method of listening tests (as detailed in Appendix F). At the beginning of each test run, subject can listen to the test materials as many number of times as he/she wants. In a presentation, when a stimulus item is presented over the headphones, all stimuli are displayed on the subject terminal screen. Each stimulus corresponds to a key on the subject terminal keyboard. At each presentation, subject is alerted by a message "Listen…" followed by the stimulus presentation to the headphones. Subject responds by pressing appropriate key on the subject terminal keyboard. Test will not proceed until response key is pressed. In a test run, as the presentation of the following other stimuli proceeds, the response key pressed is compared with the correct response and the correct response scores are updated. For each presentation, time taken by the subject to respond is also noted and is cumulated with each presentation. As the test progresses, PC monitor gives an update of the keys being pressed by the subject, the time taken, cumulative scores, etc. Tests can be with and without feedback of the correct response. In the feedback mode, after the response key is pressed, message 'Ok' or 'No' will be displayed depending on whether the response key pressed is correct or wrong, followed by the display of correct response for wrong pressing. In tests without feedback mode, only response key pressed will be displayed. Feedback mode is used during practice sessions and tests without feedback are considered for evaluation. At the end of each test run, a stimulus-response confusion matrix is formed in which stimuli and responses are represented along rows and columns respectively. Each entry in the cell represents the frequency of occurrence of a stimulus-response pair. The diagonal elements give the correct responses, whereas off-diagonal elements represent errors made. Sum of the diagonal elements gives total number of correct responses. Percentage correct recognition score and response time statistics are also displayed along with the confusion matrix. These results are stored as '`sssttt.res`', where sss gives subject's identification, ttt gives test number. The format of the stored results is as follows:

Number of test items, number of presentations,
Confusion matrix,
Test number, subject and test identification,
Date and time of test,

Number of presentations, number of responses, percent scores,

Minimum, maximum, mean, and standard deviation of response time in sec, total time taken for a test run in minutes,

Slist file and file containing test information (‘`sigfil.dat`’).

Appendix D

Programs for signal processing and listening experiments

D.1 Introduction

To reduce the effect of temporal masking associated with degraded temporal resolution in sensorineural hearing loss, a scheme of temporal splitting for binaural dichotic presentation has been developed. In the scheme, step and trapezoidal fading functions along with overlap have been used. These schemes of temporal splitting ars referred as TS_ST and TS_TR (temporal splitting with step fading and temporal splitting with trapezoidal fading). To reduce the effect of spectral masking associated with broad auditory filters in sensorineural hearing loss, a scheme of spectral splitting using a pair of comb filters having complementary magnitude response is used and is referred as SS. To reduce the effect of both increased spectral and temporal masking, combined scheme of spectral and temporal splitting (CS) has been investigated by using time varying comb filters. Programs for these schemes and listening tests are given in the following sections.

D.2 Temporal splitting (TS)

The temporal processing of speech involves temporally splitting the speech into various segments for presenting adjacent segments to the two ears (binaural dichotic presentation). For offline processing of digitized speech signal, a program "`split`" is developed in C. Sampling rate f_s is set at 10 k samples/s. Input parameters for processing are

Inter-aural switching period, T_p: 10 to 50 ms

On-period, T_{on}: $T_p/2$ to T_p

Transition duration, T_t: 0 to $0.5(T_p - T_{on})$

If trapezoidal fading (i.e. $T_t \neq 0$), fading curve: linear/logarithmic

Inter-aural switching period is the repetition period for switching between the two ears. With reference to Eqns. 4.1 and 4.2 and Figs. 4.2 and 4.3, the number of samples in the switching period, on-interval, and transition interval are

$N = T_p f_s$,

$L = T_{on} f_s$, and

$M = T_t f_s$

As shown in the Figs 4.2 and 4.3, fading function is symmetrical. Transition duration provides slope for the trapezoidal fading function. In trapezoidal fading, variation can be linear or logarithmic. The duty cycle for presentation to a particular ear is T_{on}/T_p, i.e., L/N.

The program accepts input file with signal samples as 16-bit integers in text format, and generates two output files in text format. Program "`t2b`" is used for converting the signal files from text format to binary format, in order to have better storage efficiency and access speed.

D.3. Spectral splitting (SS)

In spectral splitting, speech is split into odd and even bands using a pair of comb filters with complementary magnitude response, each with 9 pass bands corresponding to critical band auditory filters (Zwicker, 1961). The comb filter responses are obtained through two linear phase FIR filters, designed as described in Appendix C. The two filters are realized as part of one program for off-line processing of digitized speech signal. This program "`dico`" is a modified version of a program "`dicho`" developed earlier by Ratanpal (2000).

The program asks for names of the input '.bin' speech signal file, the two coefficient files in '.txt' format, and the two '.bin' output files for the output signals from the two comb filters. 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} .

D.4 Combined splitting (CS) using time varying comb filters

In the scheme of spectral and temporal splitting, time varying comb filters are used to sweep the odd and even bands between the two ears. The frequency range between two alternate critical band center frequencies, was divided into m equal parts, to obtain the center frequencies of the pass bands of successive pair of comb filter in a time varying comb filter with m shiftings. This procedure was carried out for obtaining the center frequencies of the critical bands for the entire frequency range for each successive pair of comb filter. Filter coefficients were obtained for each shifted pair of complementary comb filters using frequency sampling technique of linear phase FIR filter as described in Appendix C. The precalculated set of coefficients were cyclically swept with m shiftings with a sweeping period of T_p ms. After every time slot of T_p /m ms, a new set of coefficients for next pair of comb filters takes over.

For off-line processing of digitized speech signal, a program "`filt`" is written in C, jointly with my colleague A. N. Cheeran. Each pair of comb filter processes the signal for a time period equal to shifting time period divided by the number of shiftings. Program asks for the name of input speech signal input file in binary format and number of shiftings of comb filters. For a given number of shiftings, program reads corresponding coefficient files. For m shiftings, each pair of comb filter will process the signal for $T_p f_s/m$ samples, where T_p (ms) and f_s (samples/s) represent respectively sweeping (switching) period and sampling frequency. Program has facility for recording and playing the recorded signal or processed signals.

D.5 Monaural-binaural hearing level difference

To obtain the level difference in monaural and binaural perception, listening tests were conducted with signals of various types. Signals are presented monaurally followed by binaurally. The procedure is given in detail in Appendix C.

Program "`mono_bin`" has been written in C to present the signals, through two 12-bit DAC channels of PC based PCL-208 data acquisition card (Dynalog Micro Systems, Bombay, 1989), for monaural and binaural hearing alternately. The program outputs two signals alternately with a short time gap of about 100 ms, one signal for monaural and the other for binaural hearing. This pair of presentation is repeated with a time gap of 2 seconds.

Programs prompts the user to enter

* two file names in binary format for monaural and binaural presentation.
* time gap between monaural and binaural presentation
* time gap between repeated presentation for monaural and binaural hearing

D.6 Simulation of sensorineural hearing loss

The speech processing schemes were evaluated by conducting listening tests on normal hearing subjects with simulated hearing loss. Simulation of sensorineural hearing loss in normal hearing subjects was obtained by adding broad band noise to the speech signal. A program "`snr`", initially developed by Prasad (1996) and later modified by Chaudhari (2000) was used to add the noise to the speech signal at various SNR conditions. In the experiments SNRs of 6, 3, 0, -3, -6, -9 -12, and -15 were used. The program adds the noise to the signal in such a way that SNR was kept constant on the basis of short-time (10 ms) energy of the signal. The program accepts the noise file '`noise.bin`' from the working directory and asks for the binary file names of input and output signals, and the SNR (dB) value k. A block of 100 samples (equal to 100 samples for sampling rate of 10 k samples/sec) from input signal $s(n)$ and noise $e(n)$ are read and the output samples are calculated and are stored in output file. This procedure is repeated for successive set of 100 samples of the input signal and noise. The output $x(n)$ is obtained by adding signal $s(n)$ with new $e(n)$ scaled by a factor β such that SNR = k dB, on the basis of signal and noise energy in the block.

$$x(n) = s(n) + \beta\, e(n) \tag{D.1}$$

where $\beta^2 = \sigma_s^{\,2} / (\sigma_e^{\,2} \times 10^{0.1k})$

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

since the noise is stationary, σ_e remains almost same in all the blocks, and therefore the weighting factor β depends on the short-time energy of the input signal σ_s.

D.7 Listening test

For conducting listening tests, the automated listening test set-up used a program "`test`". Program "`tlist`" was used to randomize the presentation of stimuli and positions of response choices and "`info`" was used for information transmission analysis. These programs were initially developed using Fortran by Pandey (1987) and are successively modified in Pascal and 'C' by Thomas (1996), Chaudhari (2000), and the author. These programs are described in this section. The program "`test`" was used for outputting the signal on single channel and was modified to present output on two channels, one for left and other for right. The program used PCL-208 data acquisition card (Dynalog Micro Systems, Bombay, 1989) for outputting the signal at 10 k samples/s through a 12-bit DAC and com2 port for communication to the subject terminal. Subject terminal was used for displaying the test instructions to the subject in the beginning of each test run, alerting the subject while presenting the stimuli and for displaying the response choices on the screen. Subject responded through the terminal keyboard.

Program prompts the user to enter 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:

In each experimental run, the stimuli were presented in a random order with certain uniformity constraints to avoid biasing due to learning effect. Following uniformity constraints were used in randomizing the order of stimulus presentation (Pandey, 1987):

1) overall uniformity. For n stimulus items and N presentations in an experiment, each stimulus should be presented N/n times.
2) mid-range uniformity. In M consecutive presentations, any stimulus should not occur more than M/n +2 times.

3) short-range uniformity. An item should not be presented more than three times consecutively.

A program "`tlist`" was used for generating randomized test list for presenting test stimuli in the experiment. The program reads the information from a file '`list.dat`' in the following format:

Number of test stimuli *n*,

Number of presentations *N*,

Two line title for the experiment,

Test item names on separate lines.

,

The program outputs a file '`slist.dat`' that will be used by the computerized test administration program. The stimulus and response items are represented by order number. The list file contains the item to be presented as well as the order in which response choices should be displayed. The order of the response choices and the position of correct response are also randomized with the same uniformity constraints as described above, to avoid biasing in responses due to the order in which response choices are provided to the subject. To provide randomization across the test runs, 25 files (`slist1`, `slist2`, - - - `slist25`) were generated.

The test program executes the following tasks:

1) displays the message about the test stimuli and test instructions to the subject.
2) present the stimuli, record the subject response and response time, display the scores etc.
3) record the scores as a confusion matrix, display and store the test data for further analysis.

The test information to the program is provided in a file '`sigfil.dat`' in the following format:

Sampling frequency in Hz,

Number of items in the test set,

Names of test items (to be displayed as responses choices),

Names of data files containing signal samples for the stimuli.

D.8 Presentation procedure

At the beginning of the test sessions, the subject is instructed about the test material and method of listening tests (as detailed in Appendix F). At the beginning of each test run, subject can listen to the test materials as many number of times as he/she wants. In a presentation, when a stimulus item is presented over the headphones, all stimuli are displayed on the subject terminal screen. Each stimulus corresponds to a key on the subject terminal keyboard. At each presentation, subject is alerted by a message “Listen…” followed by the stimulus presentation to the headphones. Subject responds by pressing appropriate key on the subject terminal keyboard. Test will not proceed until response key is pressed. In a test run, as the presentation of the following other stimuli proceeds, the response key pressed is compared with the correct response and the correct response scores are updated. For each presentation, time taken by the subject to respond is also noted and is cumulated with each presentation. As the test progresses, PC monitor gives an update of the keys being pressed by the subject, the time taken, cumulative scores, etc. Tests can be with and without feedback of the correct response. In the feedback mode, after the response key is pressed, message ‘Ok’ or ‘No’ will be displayed depending on whether the response key pressed is correct or wrong, followed by the display of correct response for wrong pressing. In tests without feedback mode, only response key pressed will be displayed. Feedback mode is used during practice sessions and tests without feedback are considered for evaluation. At the end of each test run, a stimulus-response confusion matrix is formed in which stimuli and responses are represented along rows and columns respectively. Each entry in the cell represents the frequency of occurrence of a stimulus-response pair. The diagonal elements give the correct responses, whereas off-diagonal elements represent errors made. Sum of the diagonal elements gives total number of correct responses. Percentage correct recognition score and response time statistics are also displayed along with the confusion matrix. These results are stored as ‘`sssttt.res`’, where sss gives subject’s identification, ttt gives test number. The format of the stored results is as follows:

Number of test items, number of presentations,
Confusion matrix,
Test number, subject and test identification,
Date and time of test,

Number of presentations, number of responses, percent scores,

Minimum, maximum, mean, and standard deviation of response time in sec, total time taken for a test run in minutes,

Slist file and file containing test information (‘`sigfil.dat`’).

Appendix E

Hardware description

E.1 Signal acquisition set-up

As shown in Fig 4.10, the front end of the signal acquisition system consists of a microphone (B&K 4176) attached to sound level meter (B&K 2235). The sound level meter has built-in preamplifier, an input attenuater, a frequency weighting filter, and a buffer amplifier, and it gives ac signal as output. The microphone (½ inch and prepolarized), has a capacitance of 13 pF and sensitivity of 50 mV/Pa. Frequency weighting filter has A, C, and Lin. choices. During signal acquisition, 'A' weighting has been used. The sound level meter produces 1 V rms for sinusoidal acoustic input of 90 dB SPL for the range setting of 20-90 dB SPL. The analog signal from the meter is connected to an amplifier to increase the incoming signal level. The other blocks in the set-up, are anti-aliasing low pass filter and PC based DSP card with analog-to-digital converter (ADC).

Amplifier (output voltage range = ± 5 V, with a gain variable over 10-40 range) was used to increase the level of the incoming ac signal from the sound level meter to fit it within the input range of ADC. Sampling rate of 10 k samples/s of the ADC necessitates that the input analog signal should be band limited to a frequency of 5 kHz. Hence an anti-aliasing low pass filter was used. Low pass filter is an active anti-aliasing elliptic filter of 7^{th} order with cut-off frequency of 4.6 kHz, pass band ripple of 0.3 dB, stop band attenuation of 40 dB and transition band 4.6–5 kHz (Pandey, 1987; Shah, 1995). It has a gain of 1.6 (4.1 dB). Figure E.1 gives the magnitude response of the filter. During recording, the sound level meter is set in 20-90 dB SPL range and A weighting filter. The speaker is instructed to speak at normal conversational level, and the distance of the microphone distance is adjusted such that sound level meter shows 70-75 dB SPL. The amplifier is used at gain setting of 10. After the cascaded gain 1.6 of the anti-aliasing filter, 70 dB SPL reading of the sound level meter results in value 1.6 V rms.

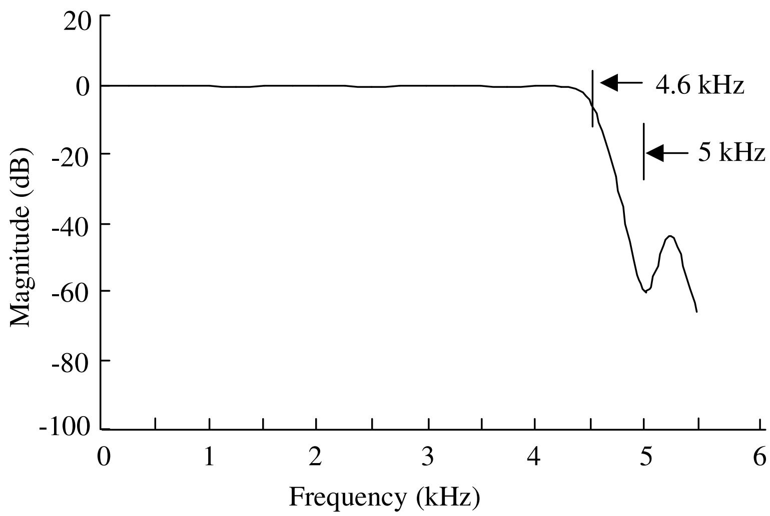


FIG. E.1. Magnitude response of the 7th order active elliptic low pass filter The pass band ripples < 0.3 dB. The stop attenuation > 40 dB.

Output signal from the low pass filter is fed to 12-bit/16-bit ADC of TI/TMS320C25 DSP board PCL-DSP 25 (Dynalog Microsystems, Bombay, 1993). TMS320C25 is 16-bit fixed point processor operating at 40 MHz. The board is interfaced to the PC bus and it has on-board ADC and DAC. Both ADC and DAC are used with 16-bit resolution and have a range of ±5 V. The acquired signal can be stored for analysis and processing, as discussed in Appendix A.

E.2 Listening test set-up

Figure 4.11 shows the experimental set-up used for conducting the listening tests. The set-up consists of a PC interfaced through RS232C asynchronous serial port, to the subject terminal which, is placed in an acoustically isolated chamber and a pair of smoothing low pass filters followed by audio amplifiers. PC was used for controlling the entire test (accepting the subject's identification and details of test data files, sending information to subject terminal, presenting the speech signals to the headphones, displaying the stimuli and responses along with response times for each presentation) and the subject terminal was used for displaying the response choices and obtaining the subject's responses from its keyboard. PC communicates with subject terminal through RS-232 serial port. PC, amplifiers, and filters were located outside the acoustically isolated chamber to keep the power dissipation and the noise from the equipment in the subject's room to a minimum. The terminal along with the headphones for stimuli presentation were located in the acoustically isolated chamber.

PC based data acquisition card PCL-208 (Dynalog Micro Systems, Bombay, 1989) was used to output the signals at 10 k samples/sec, through its two independent 12-bit digital-to-analog channels with 0–5 V output range (for fixed reference voltage of -5 V). The signals from the DAC were ac coupled to a pair of 7th order smoothing low pass elliptic filters. Filters have the characteristics same as that of the antialiasing filter used in signal acquisition card (section E.1). From the filters, signals passed through a pair of audio amplifiers and were given to the two calibrated headphones (Telephonics TDH-39P). Audio amplifiers are class B push-pull amplifiers. The input to the amplifier was fed through an attenuater using 10 kΩ logarithmic potentiometer, which varies the magnitude of the applied voltage, and helps in intensity control for the power amplifier circuit. Amplifiers were provided with input voltage of range 0 to 7.5 V. The audio amplifier board also provides outputs of ± 300 mV for feeding to the tape recorder, ± 25 mV for the microphone, and ± 5 V as test signal. Audio amplifier boards and low pass filters had been earlier developed in the laboratory [Gavankar (1995) and Shah (1995)].

E.3 Calibration of headphones

We have used TDH-39P headphones in the experiments. In listening tests, it is necessary to present the test stimuli at a calibrated sound level. Figure E.2 shows the set-up used to obtain the electroacoustic characteristics and calibration of the headphones. The set-up consists of a function/arbitrary waveform generator (HP33120A), an audio amplifier, artificial ear (B&K 4153) fitted with microphone (B&K 4176), a sound level meter (B&K 2235), headphone (TDH-39P) to be calibrated, and digital multimeter (HP34401A).

Function generator with audio amplifier provided the input for driving the headphone. Headphone was placed on the artificial ear. A force of 600 mg was applied on the headphone using spring tension mechanism of the artificial ear, to hold the headphone in place. The output from the headphone was sensed by the microphone (housed in the artificial ear) and the signal level was read as SPL on the sound level meter. Input voltage (dBm) to the headphone was measured by the digital multimeter.

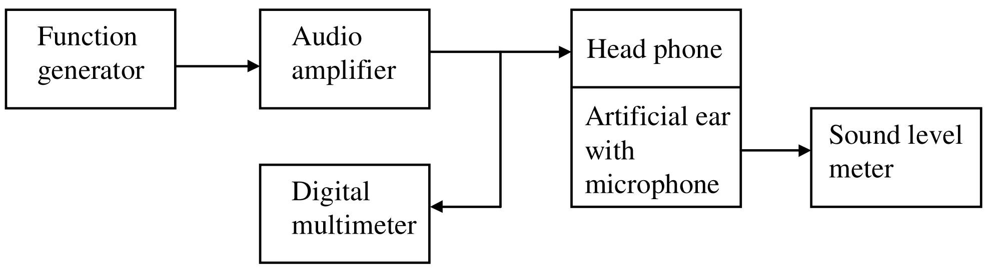


FIG. E.2. Block diagram for calibration of headphones.

To obtain the electroacoustic characteristic of the headphones, input voltage in dBm (with reference to 1 mW power in a 50 ohm load resistance, i.e. Vref = 0.224 V rms) required to produce sound pressure level of 100 dB SPL in the artificial ear was measured at different frequencies. Figure E.3 shows the electroacoustic characteristic of the two headphones TDH-39P used in the listening tests. It can be seen that the two headphones have very similar characteristics. The response varies over 10 dB over the frequency range of 125 Hz–5 kHz.

During listening experiments, presentation level should be maintained over the various test sessions. The acoustic level at the headphones can be given in terms of electric input voltage in dBm measured by digital multimeter. For this purpose, a relationship between acoustic output and electric input voltage is obtained. The most comfortable listening level 75 dB corresponds to an input voltage of -38 dBm for sustained vowel /*a*/. (For 1 kHz sinusoidal input 75 dB SPL corresponds to -40 dBm). At the beginning of each test session, the gain of the audio amplifier was adjusted such that the headphone input voltage for a sustained synthesized vowel /*a*/ was -38 dBm. The rms value of the samples of the synthesized /*a*/ and the rms value of the vowel in each of the test stimuli are kept approximately equal, to ensure that the calibration remains valid for the test stimuli. This is done by appropriate digital scaling of the sample values of the test stimuli.

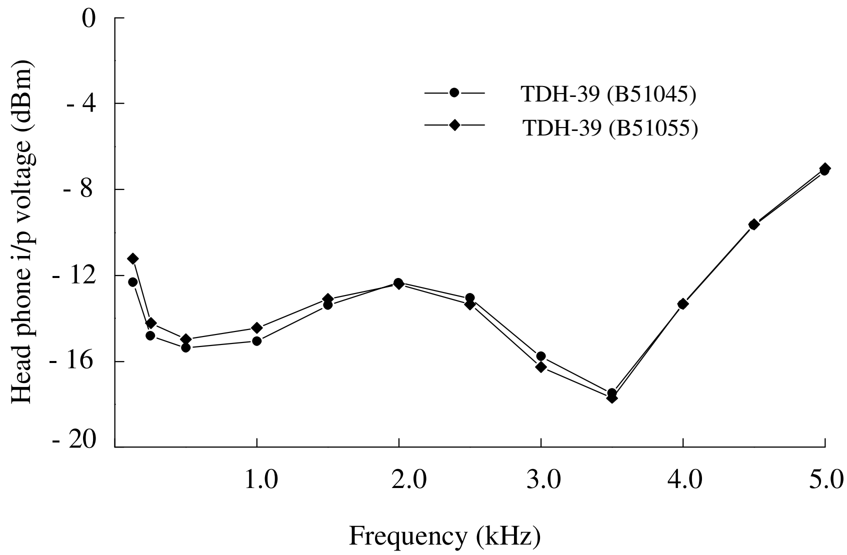


FIG. E.3. Electroacoustic characteristics of the headphones. Input voltage (dBm) versus frequency for 100 dB SPL output.

Appendix G

Simulation of sensorineural loss by broad band noise

Conducting listening tests to evaluate the schemes for speech processing and different processing conditions involves tedious procedure and is time consuming. Hence it will be difficult to test different processing conditions directly on the hearing impaired subjects. Also, there are difficulties in having large number of hearing impaired subjects with bilateral hearing loss, willing to participate in these experiments.

The issues related to simulation of hearing loss and evaluation methods have been reviewed earlier in Chapter 3 (Section 3.4). In our research, speech processing schemes were evaluated through listening tests involving normal hearing subjects with sensorineural hearing loss. The loss was simulated by addition of broad band Gaussian noise to the speech signal, with SNR kept constant on the basis of short-time (10 ms) energy of the signal.

Here we present an analysis of the results from listening tests to see the effect of SNR for loss simulation on various performance measures and to study the effectiveness of simulation for temporal and spectral masking. For this, results presented in Tables 5.3–5.5 of experiment III in Chapter 5 are used here. Listening tests involved identification of 12 consonants in VCV context with vowel /*a*/. Test material was added with broad band Gaussian noise to obtain SNRs of 6, 3, 0, -3, -6, -9, -12, and -15 dB to simulate sensorineural loss of varying degree. Test material was presented at most comfortable listening level of the individual subject. Five subjects with age between 20 to 37 years participated in the test. The performance measures used were response times, recognition score, and relative information transmission for various consonantal features. This appendix presents an analysis of results from listening tests for unprocessed speech, in order to have a basis for comparison with results obtained with different processing schemes and parameters.

Figure G.1 shows the response times of individual subjects. It is seen for no-noise condition, the average response time ranges from 1.15 seconds to 2.7 seconds which is indicative of variation in reaction time of the subjects. Under simulated loss condition, the response time increases and slopes of increase for different subjects are similar. Figure G.2 shows response times averaged across the five subjects for different SNR levels. Average value increases from 1.89 seconds under no noise condition to 2.61 seconds at -15 dB SNR. This indicates that, addition of noise increases load on hearing perception process. Increase in response time indicates that, under adverse listening condition, persons experience greater difficulty in understanding the auditory message.

Figure G.3 shows recognition scores for five subjects. Under no-noise condition, all the subjects have 100% recognition. With simulated loss, the scores decrease nearly monotonically with decreasing SNR. Figure G.4 shows the scores averaged across five subjects. Averaged score decreases from 100% at no-noise condition to 64.5% at -15 dB SNR condition. At this SNR condition, recognition scores for individual subjects range from 51.3 to 76.7%.

Overall relative information transmitted (%) for five subjects is shown in Fig. G.5 and averaged across the subjects is shown in Fig. G.6. Overall information transmitted varies from 100% at no-noise condition to 70% at -15 dB SNR condition. We see that the spread in variation for information transmission measure is much less than that for recognition scores. This indicates that reception errors are not randomly distributed, but may be distributed in accordance with feature groupings, hence it will be interesting to look at the information transmission for various consonantal features, as given in Table 3.1.

Figure G.7 shows averaged across the subjects relative information transmitted (%) for consonantal features of place, duration, manner, frication, voicing, and nasality. For comparison, the figure also shows recognition scores (%), and overall relative information transmission (%). It can be observed from Fig. G.7, percentage recognition scores, overall information transmitted (%), and relative information transmitted for all the features decrease

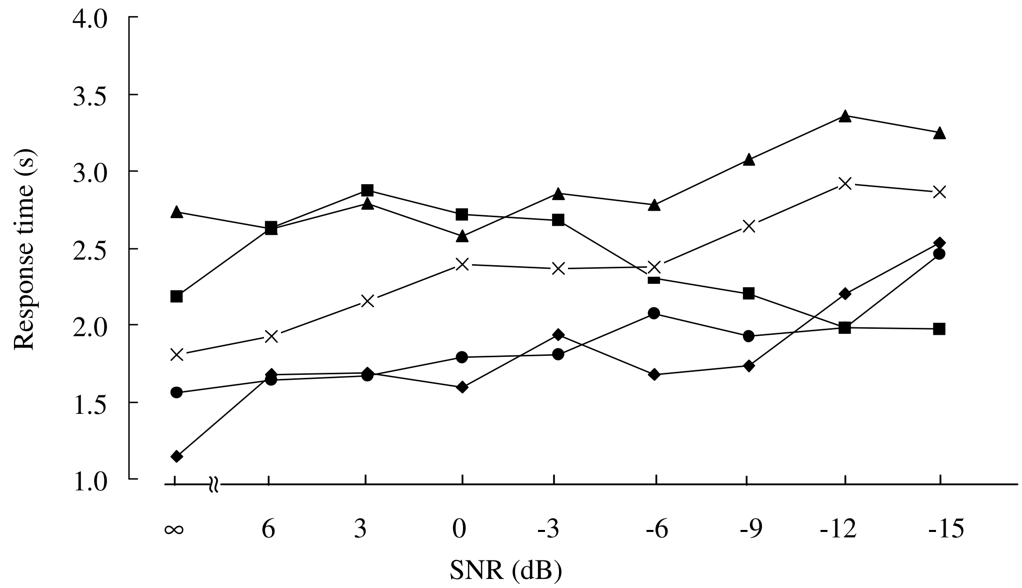


FIG. G.1. Response time (s) for five subjects under different SNRs for binaural presentation of unprocessed speech (tests conducted as part of CS experiments).

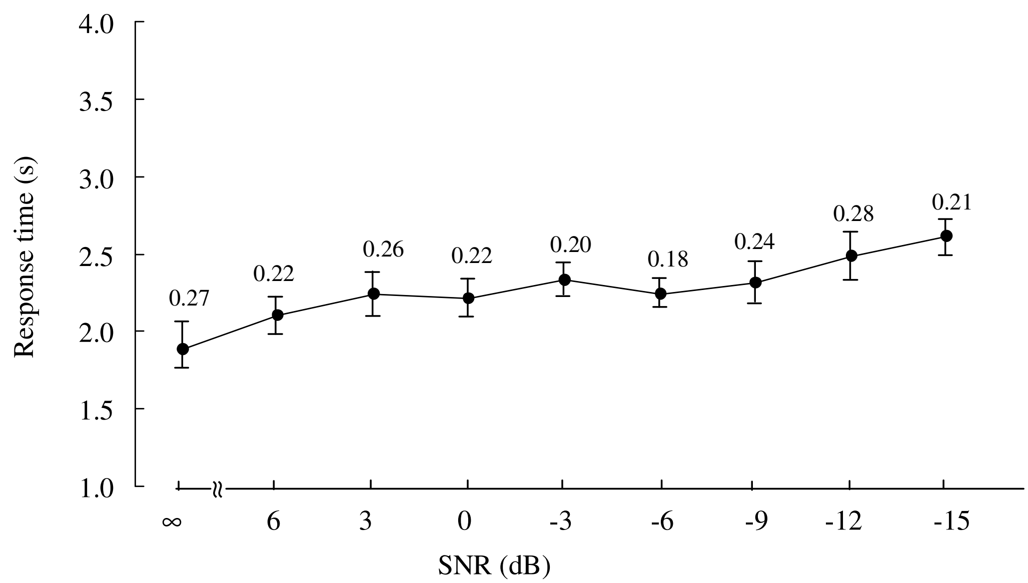


FIG. G.2. Response times of Fig. G.1 averaged across the five subjects under different SNRs. Standard error is indicated by the length of the vertical bars.

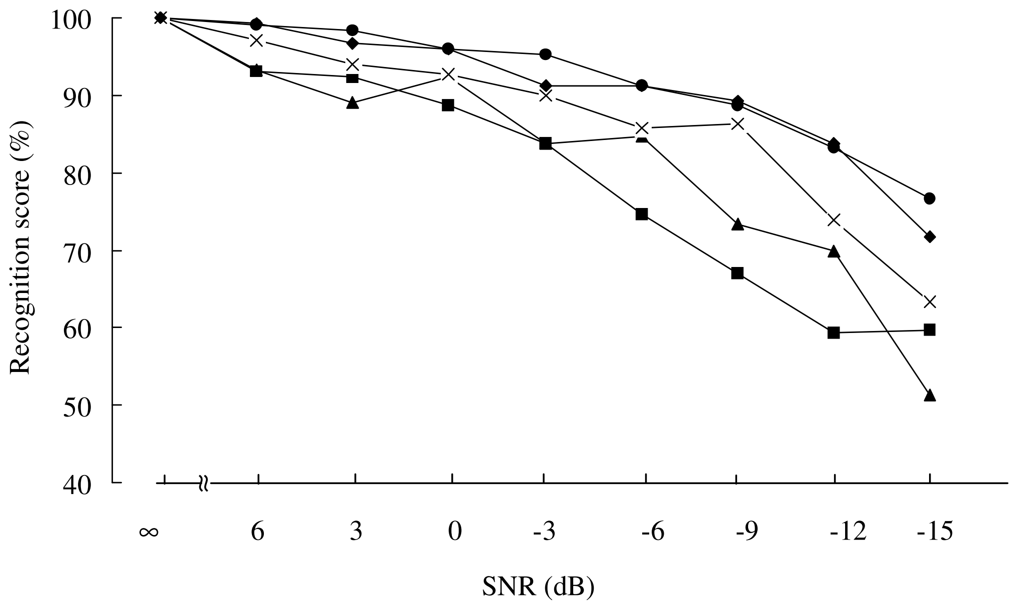


FIG. G.3. Recognition scores for five subjects under different SNRs for binaural presentation of unprocessed speech (tests conducted as part of CS experiments).

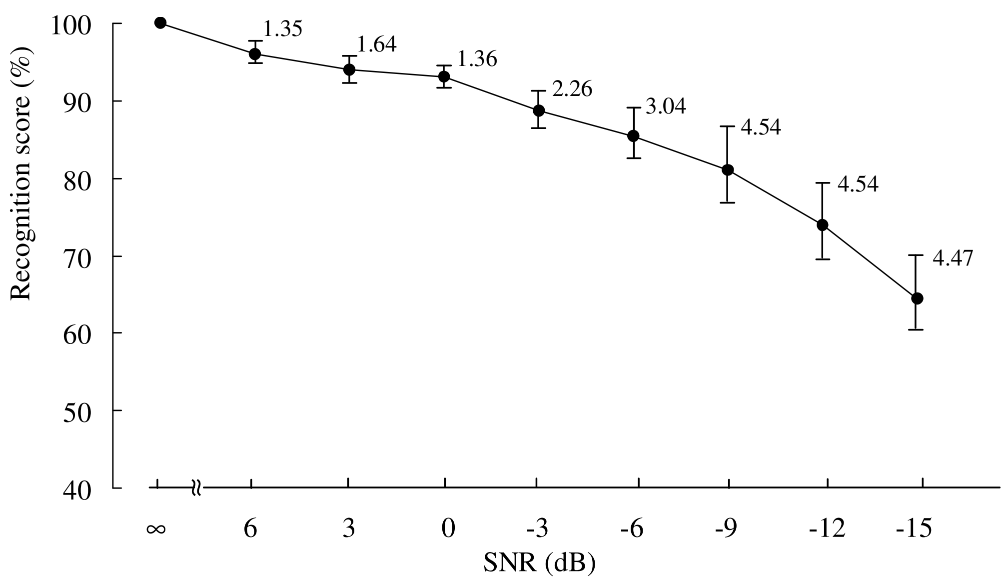


FIG. G.4. Recognition scores of Fig. G.3 averaged across five subjects under different SNRs. Standard error is indicated by the length of the vertical bars.

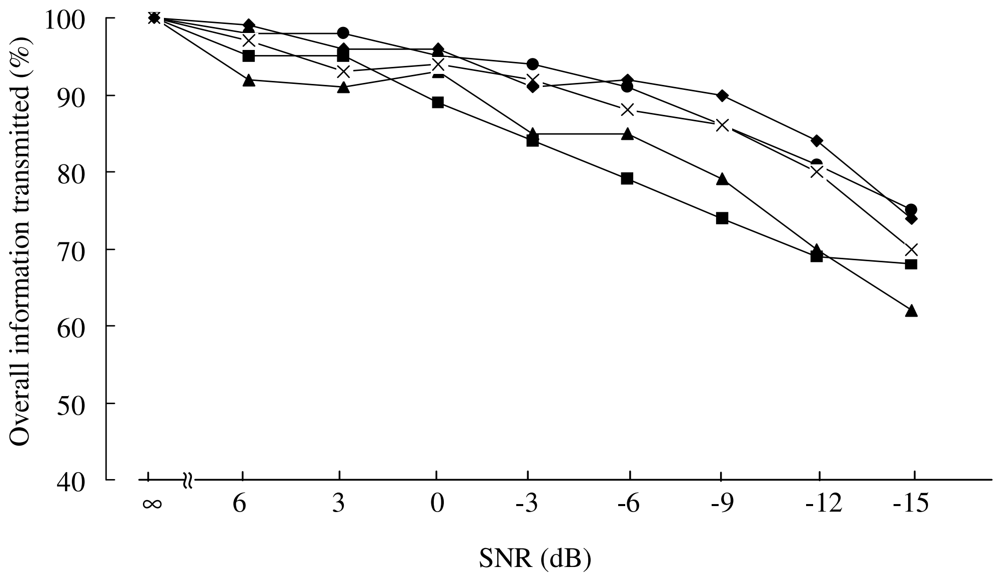


FIG. G.5. Overall information transmitted for five subjects under different SNRs for binaural presentation of unprocessed speech (tests conducted as part of CS experiments).

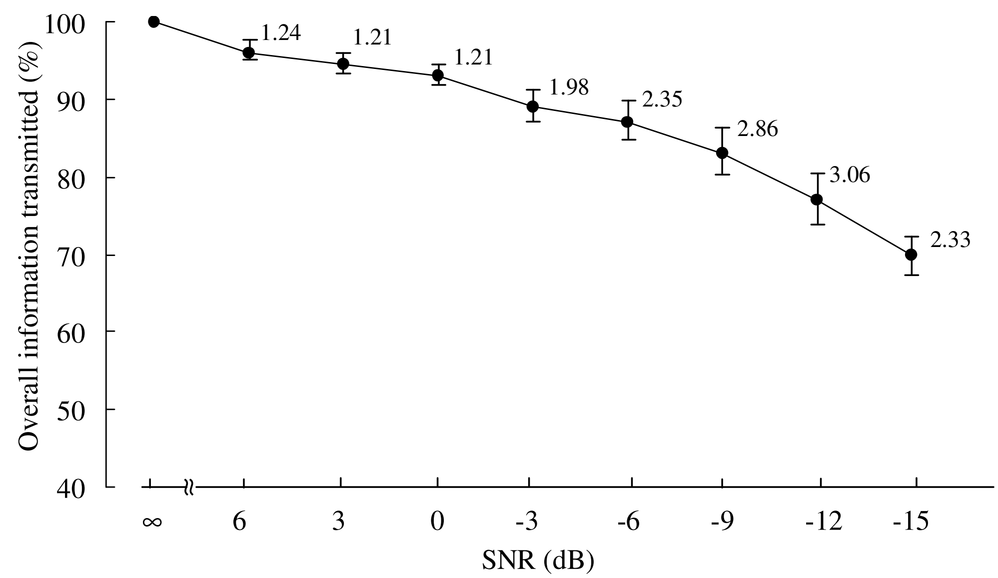


FIG. G.6. Overall information transmitted of Fig. G.5 averaged across five subjects under different SNRs. Standard error is indicated by the length of the vertical bars.

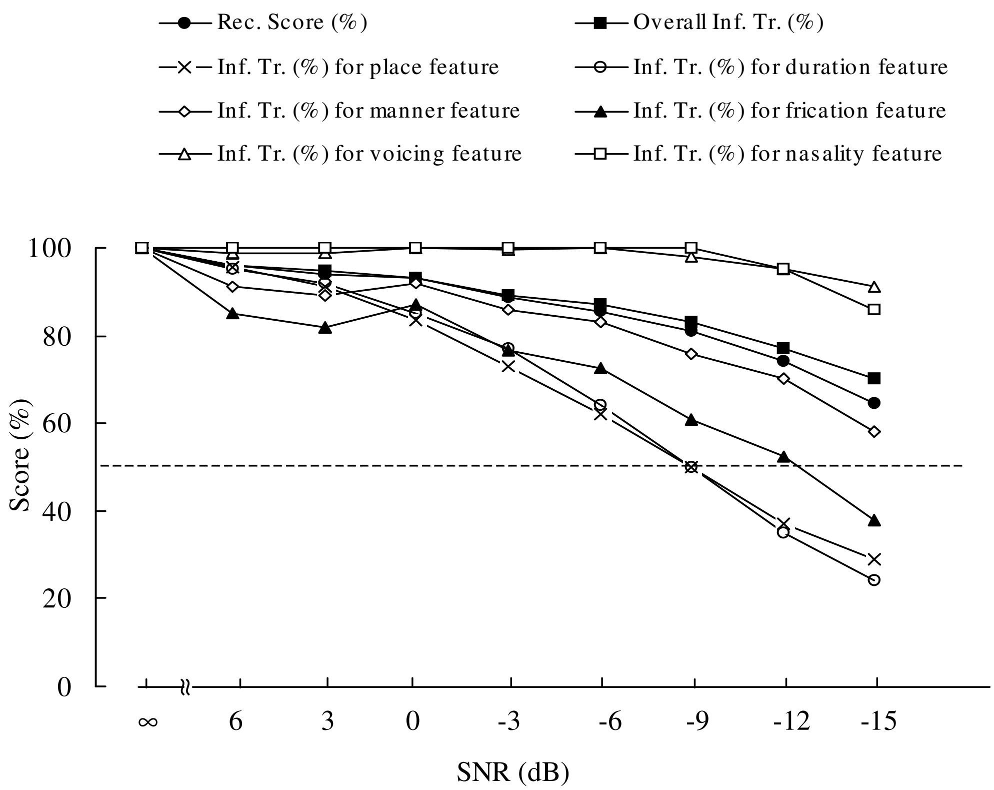


FIG. G.7. Recognition score and relative information transmission for overall and different features.

with decrease in SNR, indicating that the addition of noise to the speech degrades the quality and hence the perception. Perception of voicing and nasality features is nearly perfect and is modestly affected at higher levels of simulated loss. This is in conformity with earlier investigations that have reported these two as the most robust of the consonantal features. Compared to voicing, perception of manner (stop/frication/nasality) is more adversely affected, coming down to 58% at -15 dB SNR. There is larger decrease for frication, place and duration features at low SNRs. At -15 dB SNR, the relative transmission for place and duration features comes down to 29% and 24% respectively. From the plots, we see that SNR for 50% relative information transmission for both these features is -9 dB. Degradation in perception of place feature indicates that spectral cues required for the place feature identification get masked by simulated loss. From degradation in perception of duration feature, one may infer that the temporal cues required for the duration feature identification are masked due to temporal masking.

The values obtained here show a pattern similar to that reported by Miller and Nicely (1955) for normal hearing subjects, with 16 VCV syllables as stimuli presented at different SNR conditions of masking noise. Voicing and nasality were less affected by the noise. For place and duration features, SNR's corresponding to relative information transmission of 50% are +3 dB and -3 dB. In our study, relative information transmission of 50% corresponds to -9 dB SNR for these features. The variation may be attributed to the test material and subjects.

In listening tests with hearing impaired subjects (Experiment IV, presented in chapter 6), recognition scores for unprocessed speech, averaged across the five subjects varied from 73 to 92 % with an average of 81 %. Relative information transmitted averaged across the five subjects for place, duration, frication are 58, 49, and 56 %. These values closely match with the values obtained with simulation of hearing loss with SNR of -9 dB. The recognition scores and relative information transmitted for different consonantal features were matched (by using linear interpolation) with the corresponding average scores for simulated loss, as given in Table G.1 and shown in Fig. G.7, to obtain an SNR equivalent to hearing loss. The results are given in Table G.2. It is seen that simulation was not effective in simulating the effect of hearing loss on reception of voicing and only moderately effective for nasality and manner. However, these features are known to be not susceptible to the adverse effects of temporal and spectral masking. The equivalent SNR for other features fall within a relatively small range for an individual subject and vary across the subjects. Hence we can infer that addition of broad band noise simulates the effects of increased temporal and spectral masking and level of SNR varies with the severity of masking.

Table G.1. Response time, recognition score, and relative information transmitted for listening tests with normal hearing subjects and simulated loss (Experiment III) for unprocessed speech. S: Subject.

(a) Response time (s)

S	SNR (dB)								
	∞	6	3	0	-3	-6	-9	-12	-15
AB	1.56	1.64	1.67	1.79	1.81	2.07	1.93	1.98	2.46
SS	2.18	2.63	2.87	2.72	2.68	2.3	2.2	1.98	1.97
VK	2.73	2.62	2.79	2.58	2.85	2.78	3.07	3.36	3.25
AC	1.81	1.93	2.16	2.39	2.37	2.38	2.64	2.92	2.86
JK	1.15	1.68	1.69	1.60	1.94	1.68	1.73	2.20	2.53
Avg.	1.89	2.10	2.24	2.22	2.33	2.24	2.32	2.49	2.61

(b) Recognition score (%)

S	SNR (dB)								
	∞	6	3	0	-3	-6	-9	-12	-15
AB	100	99.0	98.3	96.0	95.3	91.3	88.7	83.3	76.7
SS	100	93.0	92.3	88.7	83.7	74.7	67.0	59.3	59.7
VK	100	93.3	89.0	92.3	83.7	84.7	73.3	70.0	51.3
AC	100	97.0	94.0	92.7	90.0	85.7	86.3	74.0	63.3
JK	100	99.3	96.7	96.0	91.3	91.3	89.3	83.7	71.7
Avg.	100	96.0	94.0	93.0	88.7	85.5	81.0	74.0	64.5

(c) Relative information transmitted (%): Overall

S	SNR (dB)								
	∞	6	3	0	-3	-6	-9	-12	-15
AB	100	98	98	95	94	91	86	81	75
SS	100	95	95	89	84	79	74	69	68
VK	100	92	91	93	85	85	79	70	62
AC	100	97	93	94	92	88	86	80	70
JK	100	99	96	96	91	92	90	84	74
Avg.	100	96	95	93	89	87	83	77	70

(d) Relative information transmitted (RIT, %) for different features, averaged across the five subjects

Feature	SNR (dB)								
	∞	6	3	0	-3	-6	-9	-12	-15
Vo	100	99	99	100	99	100	98	95	91
Pl	100	95	91	84	73	62	50	37	29
Mn	100	91	89	92	86	83	76	70	58
Na	100	100	100	100	100	100	100	95	86
Fr	100	85	82	87	77	72	61	52	38
Du	100	95	92	85	77	64	50	35	24

TABLE G.2. Equivalent simulated loss SNR for different features for the hearing impaired subjects. Score: % recognition score and relative information transmitted, SNR: Simulated loss SNR (dB) corresponding to the score obtained by linear interpolation.

S	Score/ SNR	RS (%)	Relative information transmitted (%)						
			Ov	Vo	Pl	Mn	Na	Fri	Du
SA	Score	92	92	97	85	83	100	74	71
	SNR	**-1**	**-1**	**-10**	**0**	**-6**	**-6**	**-5**	**-4**
BA	Score	89	88	80	74	80	81	79	35
	SNR	**-3**	**-5**	**< -15**	**-3**	**-7**	**< -15**	**-2**	**-12**
SK	Score	76	81	67	59	53	88	29	39
	SNR	**-11**	**-10**	**<-15**	**-7**	**<-15**	**-14**	**< -15**	**-11**
KS	Score	76	81	84	42	73	100	55	48
	SNR	**-11**	**-10**	**<-15**	**-11**	**-11**	**-6**	**-11**	**-9**
BS	Score	73	76	97	32	66	100	44	52
	SNR	**-12**	**-12**	**< -15**	**-14**	**-13**	**-6**	**-14**	**-9**
Avg.	Score	81	84	85	58	71	94	56	49
	SNR	**-9**	**-8**	**< -15**	**-7**	**-12**	**-12**	**-11**	**-9**

References

Abel, S. M. (**1971**). “Discrimination of temporal gaps,” J. Acoust. Soc. Am. **52**(**2**), 519–524.

Adobe Systems Inc. (**1988**). *Postscript Language Reference Manual* (Addision-Wesley, Reading, Massachusetts).

Apoux, F., Crouzet, O., and Lorenzi, C. (**2000b**). “Consonant recognition in noise with temporal cues: II. Effects of temporal envelope enhancement on response times,” J. Acoust. Soc. Am. **107**(**5**), 2913.

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

Apoux, F., Lorenzi, C., and Berthommier, F. (**2000a**). “Consonant recognition in noise with temporal cues: I. Effects of temporal envelope enhancement on identification performance,” J. Acoust. Soc. Am. **107**(**5**), 2913.

Arlinger, S., and Dryselius, H. (**1990**). “Speech recognition in noise, temporal and spectral resolution in normal and impaired hearing,” Acta Otolarygol. **469**, 30–37.

Asano, F., Suzuki, Y., Sone, T., Kakehata, S., Satake, M., Ohyama, K., Kobayashi, T., and Takasaka, T. (**1991**). “A digital hearing aid that compensates loudness for sensorineural hearing listeners,” *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, 3625–3628.

Babsky, E., Khodorov, B., Kositsky, G., and Zubkov, A. (**1970**). *Human Physiology*. Ed. E. Babsky, Vol. 2 (Mir Publishers, Moscow).

Baragi, Ashok B. N. (**1996**). “A speech training aid for the deaf,” M. Tech. dissertation, Department of Electrical Engineering, IIT Bombay, India.

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

Brobeck, J. R. (**1973**). *Best and Taylor’s Physiological Basis of Medical Practice*. 9th Ed. (Williams and Wilkins Company, Baltimore).

Brooks, D. N. (**1984**). “Binaural benefit-when and how much?” Scand. Audiol. **13**, 237–241.

Brüel and Kjær (**1985**). *Instruction Manual of Precision Sound Level Meter Type 2235* (Brüel and Kjær, Nærum, Denmark).

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

CHABA. (**1988**). “Speech understanding and aging,” J. Accost. Soc. Am. **83**(**3**), 859–895.

CHABA. (**1991**). “Speech-perception aids for the hearing-impaired people: Current status and needed research,” J. Acoust. Soc. Am. **90**, 637–685.

Chaudhari, D. S. (**2000**). “Dichotic presentation of speech signal for improving speech perception for the bilateral sensorineural hearing impairment,” Ph.D. dissertation, School of Biomedical Engineering, IIT Bombay.

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

Chaudhari, D. S., and Pandey, P.C. (**1998b**). “Real time speech processing for dichotic presentation for binaural hearing aids,” *Proc. NCBME ’98*, Manipal, Karnataka, India.

Cheeran, A. N., Pandey, P. C., and Jangamashetti, D. S. (**2002**). “Design of comb filters based on auditory filters for dichotic presentation for persons with bilateral sensorineural hearing impairment,” *Proc. of 14th Int. Conf. on Digital Signal Processing (DSP2002), Santorini, Greece*, 1145-1148.

Cheeran, A. N., Pandey, P. C., and Jangamashetti, D. S. (**2001**). “Comb filters for dichotic presentation to improve speech perception by persons with bilateral sensorineural hearing impairment,” J. Acoust. Soc. Am., vol. **110** (**5**), Pt. 2, p. 2705.

Cheung, S., and Lim, J. S. (**1992**). “Combined multiresolution (wide-band/narrow-band) spectrogram,” IEEE Trans. Signal Processing **40**, 957–977.

Coleridge Smith, P. D., and Scurr, J. H. (**1988**). Eds. *Mircocomputers in Medicine*. (Springer-Verlag, London).

Crandall, I. B. (**1917**). “The composition of speech,” Phys. Rev. **10**(**2**). in *Speech Intelligibility and Speaker Recognition,* edited by M. E. Hawley (Dowden Hutchinson Ross, Stroudsburg, PA), pp. 26–28.

Danaher, E. M., and Picket, J. M. (**1975**). “Some masking effects produced by low-frequency vowel formants in persons with sensorineural hearing loss,” J. Speech Hear. Res. **18**, 261–271.

Danaher, E. M., Wilson, M., and Picket, J. M. (**1978**). “Backward and forward masking in listeners with severe sensorineural hearing loss,” J. Audilogy. **17**, 324–338.

Dolan, T. G., and Small, A. M. (**1984**). “Frequency effects in backward masking,” J. Acoust. Soc. Am. **75**(**3**), 932–935.

Donaldson, G. S., and Nelson, D. A. (**2000**). “Place-pitch sensitivity and its relation to consonant recognition by cochlear implant listeners using the MPEAK and SPEAK speech processing strategies,” J. Acoust. Soc. Am. **107**(**3**), 1645–1658.

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., Marton, K., Hannley, M. T., and Lindholm, J. M. (**1985**). “Phonetic identification by elderly normal and hearing-impaired listeners,” J. Acoust. Soc. Am. **77**, 664–670.

Dorman, M. F., Soli, S., Dankowski, K., Smith, L. M., McCandless, G., and Parkin, J. (**1990**). “Acoustic cues for consonant identification by patients who use the Ineraid cochlear implant,” J. Acoust. Soc. Am. **88**(**5**), 2074–2079.

Duan, M. L., and Canlon, B. (**1996**). “Outer hair cell activity is not required for the generation of the forward masking curve,” Audiol Neurootol. **1**, 309–319.

Dubno, J. R., Dirks, D. D., and Langhofer, L. R. (**1982**). “Evaluation of hearing-impaired using the nonsense syllable test:II,” J. Speech Hear. Res. **25**, 141–148.

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., and Levit, H. (**1981**). “Predicting consonant confusions from acoustic analysis,” J. Acoust. Soc. Am. **69**(**1**), 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**(**4**). Pt.1., 2110–2121.

Durlach, N. I., Gabriel, K. J., Colburn, H. S., and Trahiotis, C. (**1986**). “Interaural correlation discrimination: relation to binaural unmasking,” J. Acoust. Soc. Am. **79**(**5**), 1548–1557.

Dynalog Microsystems (**1989**). *User's Manual, PCL-208 (PCL-718) Data Acquisition Card* (Dynalog Microsystems, Bombay).

Dynalog MicroSystems (**1993**). *User's Manual, PCL-DSP25 TMS 320C25 Digital Signal Processor Board* (Dynalog Microsystems, Bombay).

Elliot, L. L. (**1962**). "Backward and forward masking of probe tones of different frequencies," J. Acoust. Soc. Am. **34**(**8**), 1116–1117.

Elliot, L. L. (**1975**). "Temporal and masking phenomena in persons with sensorineural hearing loss," Audilogy **14**, 336–353.

Faulkner, A., and Rosen, S. (**1999**). "Contribution of temporal encodings of voicing, voicelessness, fundamental frequency, and amplitude variation to audio-visual and auditory speech perception," J. Acoust. Soc. Am. **106**(**4**), 2063–2073.

Faulkner, A., Rosen, S., and Smith, C. (**2000**). "Effects of the salience of pitch and periodicity information on the intelligibility of four-channel vocoded speech: Implications for cochlear implants," J. Acoust. Soc. Am. **108**(**4**), 1877–1887.

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 Communication in Hearing*. (Huntington, New York, Krieger).

Florentine, M., and Buus, S. (**1984**). "Temporal gap detection in sensorineural hearing impairment," J. Speech Hear. Res. **27**, 449–455.

Florentine, M., Buus, S., Scharf, B., and Zwicker, E. (**1980**). "Frequency selectivity in normally-hearing and hearing-impaired observers," J. Speech Hear. Res. **23**, 646–669.

Foust, K. O., and Gengel, R. W. (**1973**). "Speech discrimination by sensorineural hearing-impaired persons using a transposer hearing aid," Scand. Audiol. **2**, 161–170.

Freyman, R. L., and Nelson, D. A. (**1986**). "Frequency discrimination as a function of tonal duration and excitation-pattern slopes in normal and hearing-impaired listeners," J. Acoust. Soc. Am. **79**(**4**), 1034–1044.

Fu, Q. J., and Shannon R. V. (**1999**). "Effects of electrode location and spacing on phoneme recognition with the Nucleus-22 cochlear implant," Ear Hear. **20**(**4**), 321–331.

Fu, Q. J., Shannon, R. V., and Wang, X. (**1998**). "Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing," J. Acoust. Soc. Am. **104**, 3586–3596.

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

Ganong, W. F. (**1991**). *Review of Medical Physiology*. 15th Ed. (Lange Medical Publications, Los Altos, California).

Gatehouse, S., and Gordon, J. (**1990**). “Response times to speech stimuli as measures of benefit from amplification,” Br. J. Audiol, **26**, 63–68.

Gehr, S. E., and Sommers, M. S. (**1999**). “Age difference in backward masking,” J. Acoust. Soc. Am. **106**(**5**), 2793–2799.

Gelfand, S. A. (**1990**). *Hearing-an Introduction toPsychological and Physiological Acoustics.* 2nd Ed. (Marcle Dekker, New York).

Gelfand, S. A., Piper, N., and Silman, S. (**1985**). “Consonant recognition in quite as a function of aging among normal hearing subjects,” J. Acoust. Soc. Am. **78**(**4**), 1198–1206.

Geurts, L., and Wouters, J. (**1999**). “Enhancing the speech envelope of continuous interleaved sampling processors for cochlear implants,” J. Acoust. Soc. Am. **105**(**4**), 2476–2484.

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., Patterson, R. D., and Nimmo-Smith, I. (**1984**). “Dynamic range and asymmetry of the auditory filter,” J. Acoust. Soc. Am. **76**(**2**), 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., 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**, 1546–1556.

Gordon-Salant, 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-Salant, S. (**1987**). “Effects of acoustic modification on consonant recognition by elderly hearing-impaired subjects,” J. Acoust. Soc. Am. **81**, 1199–1202.

Gordon-Salant, S., and Fitzgibbons, P. J. (**1993**). “Temporal factors and speech recognition performance in young and elderly listeners,” J. Speech Hear. Res. **36**, 1276–1285.

Green, D. M., Birdsall, T. G., and Tanner, W. P. (**1957**). “Signal detection as a function of signal intensity and duration,” J. Acoust. Soc. Am. **29**(**4**), 523–531.

Gulick, W. L. (**1971**). *Hearing, Physiology, and Psychophysics*. (Oxford University Press, New York).

Guyton, A. C. (**1986**). *Textbook of Medical Physiology* (Saunders, Philadelphia, PA).

Hannley, M., and Dorman, M. F. (**1983**). "Susceptibility to intraspeech spread of masking in listeners with sensorineural hearing loss," J. Acoust. Soc. Am. **74** (**1**), 40–51.

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

Hawkins, D. B., and Yacullo, W. S. (**1984**). "Signal-to-noise ratio advantage of binaural hearing aids and directional microphones under different levels of reverberation," J. Speech Hear. Disord. **49**, 278–286.

Hirsh, I. J. (**1950**). "Binaural hearing aids: a review of some experiments," J. Speech Hear. Disord. **15**, 114–123.

Hou, Z., and Pavlovic, C. V. (**1994**). "Effects of temporal smearing on temporal resolution, frequency selectivity, and speech intelligibility," J. Acoust. Soc. Am. **96**(**3**), 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**(**3**), 765–773.

Jamieson, D.G., Ponton, W., and Espinoza-Varas, B. (**1985**). "Reduction in formant bandwidth improves vowel identification with sensorineural impairment," J.Acoust. Soc. Am. Suppl. 1 **77**, S8.

Janssen, T., Boege, P., Oestreicher, E., and Arnold, W. (**2000**), "Tinnitus and 2f1-f2 distortion product otoacoustic emissions following salicylate overdose," J. Acoust. Soc. Am. **107**(**3**) 1790–1792.

Jerger, J., Carhart, R., and Dirks, D. (**1961**). "Binaural hearing aids and speech intelligibility," J. Speech Hear. Res. **4**, 137–148.

Jesteadt, W., Bacon, S. P., and Lehman, J. R. (**1982**). "Forward masking as a function of frequency, masker level, and signal delay," J. Acoust. Soc. Am. **71**, 950–962.

Johannson, B. (**1966**). "The use of the transposer for the management of the deaf child," in *Sensory Aids for the Hearing Impaired*, edited by Levitt, H., Pickett, J. M., and Houde, R. A. (IEEE Press, New York), pp. 195–204.

Kandel, E. R., Schwartz, J. H., and Jessel, T. M. (**1991**). *Principles of Neural Sciences*. (Elsevier, New York).

Kasturi, P. (**1997**). “Real-time Filter Design with Tunable Frequency Response using TMS320C50 Processor,” Project report, Signal Processing and Instrumentation Lab., Deparment of Electrical Engineering, IIT Bombay, India.

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**(**2**), 1098–1114.

Kewley-Port, D. (**1982**). “Measurement of formant transitions in naturally produced stop consonant-vowel syllables,” J. Acoust. Soc. Am. **72**(**2**), 379–389.

Kidd, G., Mason, C. R., and Feth, L. L. (**1984**). “Temporal integration of forward masking in listeners having sensorineural hearing loss,” J. Acoust. Soc. Am. **75**(**3**), 937–944.

Kiefer, J., von. Ilberg, C., Hubner-Egner, J., Rupprecht, V., Knecht, R. (**2000**). “Optimized speech understanding with the continuous interleaved sampling speech coding strategy in patients with cochlear implants: effect of variations in stimulation rate and number of channels,” Ann. Otol. Rhinol. Laryngol. **109**, 1009–1020.

Kiessling, J., and Steffens, T. (**1991**). “Clinical evaluation of a programmable three-channel automatic gain control amplification system,” Audiology **30**, 70–81.

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

Larsby, B., and Arlinger, S. (**1998**). “A method for evaluating temporal, spectral and combined temporal-spectral resolution in hearing,” Scand. Audiol. Suppl. **27**, 3–12.

Leek, M. R., and Summers, V. (**1996**). “Reduced frequency selectivity and the preservation of spectral contrast in noise,” J. Acoust. Soc. Am. **100**(**3**), 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.

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

Liberman, A. M. (**1957**). “Some results of research on speech perception,” J. Acoust. Soc. Am. **29**(**1**), 117–123.

Lochner, J. P. A., and Burger, J. F. (**1961**). “Form of loudness function in the presence of masking noise,” J. Acoust. Soc. Am. **33**(**12**), 1705–1707.

Lorenzi, C., Berthommier, F., Apoux, F., Bacri, N. (**1999**). “Effects of envelope expansion on speech recognition,” Hear. Res. **136**, 131–138.

Loizou, P.C., Dorman, M., Poroy, O., Spahr, T. (**2000**). "Speech recognition by normal-hearing and cochlear implant listeners as a function of intensity resolution," J. Acoust. Soc. Am **108**(**5**), 2377–2387.

Lunner, T. (**1997**). *"A Digital Filterbank Hearing Aid,"* Ph.D. dissertation, Department. of Neuroscience and Locomotion, Division of Audiology, Linkoping University, 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.

Meftah, M., and Boudelaa, Sami (**1996**)."How facilitatory can lexical information be during word recognition? Evidence from moroccon arabic," *Proc. of Int. Conference on Spoken Language Processing (ICSLP)*, Philapdelphia, USA. **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**(**2**), 338–352.

Møller, A. R. (**2000**). *Hearing: Its Physiology and Pathophysiology.* (Academic, Texas).

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. (**1982**). *An Introduction to the Psychology of Hearing*. 2nd Ed. (Academic, London).

Moore, B. C. J. (**1997**). *An Introduction to the Psychology of Hearing*. 4th Ed. (Academic, London).

Moore, B. C. J. (**1998**). "Psychoacoustics of cochlear hearing impairment and the design of hearing aids*," Proc. 16th Int. Cong. Acoustics (ICA)*, Seattle, WA. 2105–2108.

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 quite and in a background of speech," J. Acoust. Soc. Am. **94**(**4**), 2050–2062.

Moore, B. C. J., Peters, R. W., and Glasberg. B. R. (**1992**). "Detection of temporal gaps in sinusoids by elderly subjects with and without hearing loss," J. Acoust. Soc. Am. **92**(**4**), Pt. 1, 1923–1932.

Morris, L. R. (**1988**). "A PC-based digital spectrograph," IEEE Micro. **8**, 68–85.

Munson, W. A. (**1947**). “The growth of auditory sensation,” J. Acoust. Soc. Am. **19**(**4**), 584–591.

Nejime, Y., and Moore, B. C. J. (**1998**). “Evaluation of the effect of speech-rate slowing on speech intelligibility in noise using a simulation of cochlear hearing loss,” J. Acoust. Soc. Am. **103**(**1**), 572–576.

Nelson, D. A., Chargo, S. J. Kopun, J. G., and Freyman, R. (**1990**). “Effects of stimulus level on forward masked psychophysical tuning curves in quite and noise,” J. Acoust. Soc. Am. **88**, 2143–2151.

Nilson, 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**(**2**), 1085–1099.

Ono, H., Okasaki, T., Nakai, S., and Harasaki, H. (**1982**). “Identification of an emphasized consonant of a monosyllable in hearing-impaired and its application to a hearing aid,” J. Acoust. Soc. Am. **71**, S58.

Oppenheim, A. V. (**1970**). “Speech spectrograms using the fast Fourier transform,” IEEE Spectrum, **7**, 57–62.

O’Shaughnessy, D. (**1987**). *Speech Communication: Human and Machine* (Addison-Wesley, Reading, Massachuetts).

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.

Pandey, P. C. (**1987**). “Speech Processing for Cochlear Prostheses,” Ph. D dissertation, Department of Electrical Engineering, Univ. of Toronto, Canada.

Pandey, P. C., Kunov, H., and Abel, S. M. (**1987**). “A speech processor providing fricative and low-frequency periodicity information for single channel cochlear prosthesis,” *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing* **34.3**.

Patterson, R. D. (**1976**). “Auditory filter shapes derived with noise stimuli,” J. Acoust. Soc. Am. **59**(**3**), 640–654.

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. **2**, 96–103.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1986**). "Speaking clearly for the hard of hearing II: intelligibility differences between clear and conversational speech," J. Speech Hear. Res. **29**, 434–446.

Pickles, J. O. (**1982**). *An Introduction to the Physiology of Hearing* (Academic, London).

Plack, C. J., and Oxenham, A. J. (**1998**). "Basilar-membrane nonlinearity and the growth of forward masking," J. Acoust. Soc. Am. **103**(**3**), 1598–1608.

Plomp, R. (**1964**). "Rate of decay of auditory sensation," J. Acoust. Soc. Am. **36**(**2**), 227–282.

Plomp, R., and Bouman, M. A. (**1959**). "Relation between hearing threshold and duration for tone pulses," J. Acoust. Soc. Am. **31**(**6**), 749–758.

Pols, L. C. W., and Schouten, M. E. H. (**1978**). "Identification of deleted consonants," J. Acoust. Soc. Am. **64**, 1333–1337.

Prasad, V. V. R. (**1996**). "Speech processing for single channel aid," M. Tech. dissertation, Dept. of Elect. Engg, IIT Bombay, India.

Previte, J. J. (**1983**). *Human Physiology*. (Mc Graw-Hill Book Company, NewYork).

Proakis, J. G., and Manolakis, D. G. (**1997**). *Digital Signal Processing Principles, Algorithms, and Applications* (Prentice Hall of India, New Delhi).

Punch, J. L., Montgomery, A. A., Schwartz, D. M., Walden, B. E., Prosek, R. A., and Howard, M. T. (**1980**). "Multidimensional scaling of quality judgments of speech signals processed by hearing aids," J. Acoust. Soc. Am. **68**(**2**), 458–466.

Raab, D. H. (**1961**). "Forward and backward masking between acoustic clicks," J. Acoust. Soc. Am. **33**(**2**), 137–139.

Rabiner, L. R., and Gold, B. (**1998**). *Theory and Application of Digital signal processing,* (Prentice Hall of India, New Delhi).

Rabiner, L. R., Gold, B. and McGonegal, C. A. (**1970**). "An approach to the approximation problem for nonrecursive digital filters," IEEE Trans. Audio Electroacoustics **18,** 83–106.

Rabiner, L. R., and Schafer, R. W. (**1978**). *Digital Processing of the Speech Signals* (Prentice Hall, Englewood Cliffs, NJ).

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-passed filtering and pitch-invariant frequency lowering," J. Acoust. Soc. Am. **74**, 409–419.

Resnick, S. B., Dubno, J. R., Hoffnung, S. and Levitt, H. (**1975**). "Phoneme errors on a nonsense syllable test," J. Acoust. Soc. Am. **58**, Suppl. No.1, S114.

Revoile, S., Holden-Pitt, L., Pickett, J. M, and Bandt, F. (**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., Holden-Pitt, L., Pickett, J. M, and Bandt, 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.

Scharf, B. (**1969**). "Dichotic summation of loudness," J. Acoust. Soc. Am. **45**, 1193-1205.

Schneider, B. A., Pichora-Fuller, M. K., Kowalchuk, D., and Lamb, M. (**1994**). "Gap detection and the precedence effect in young and old adults," J. Acoust. Soc. Am. **95**(**2**), 980–991.

Severns, M. L. (**1985**). "A computer controlled attenuator for audiological testing," J. Clinical Engineering **10**(**4**) 317-321.

Shah, N. (**1995**). "A sensory aid for the deaf," M. Tech. dissertation, Department of Electrical Engineering, IIT Bombay, India.

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 band limited noise: Effects of variations in bandwidth and signal to masker ratio," J. Acoust. Soc. Am. **77**(**2**), 635–639.

Sheely, E. C., and Bilger, R. C. (**1964**). "Temporal integration as a function of frequency," J. Acoust. Soc. Am. **36**(**10**), 1850–1857.

Simon, C. (**1978**). "On the use of comfortable listening levels in speech experiments," J. Acoust. Soc. Am. **64**, 744–751.

Skinner, M. W. (**1980**). "Speech intelligibility in noise-induced hearing loss: Effects of high-frequency compensation," J. Acoust. Soc. Am. **67**(**1**), 306–317.

Skinner, M. W., Holden, L. K., Holden, T. A., Demorest, M. E., and Fourakis, M. S. (**1997**). "Speech recognition at simulated soft, conversational, and raised-to-loud vocal efforts by adults with cochlear implants," J. Acoust. Soc. Am. **101**, 3766–3782.

Smiarowski, R. A., and Carhart, R. (**1975**). "Relations among temporal resolution, forward masking and simultaneous masking," J. Acoust. Soc. Am. **57**(**5**), 1169–1173.

Snedecor, G. W., and Cochran, W. G. (**1980**). *Statistical Methods* (The Iowa State University Press, Ames, Iowa).

Snell, K.B. (**1997**). "Age-related changes in temporal gap detection," J. Acoust. Soc. Am. **101**, 2214–2220.

Stevens, S. S. (**1965**). "Power-group transformations under glare, masking, and recruitment," J. Acoust. Soc. Am. **39**(**4**)**,** 725–735.

Stone, M. A., and Moore, B. C. J. (**1992**). "Syllabic compression: effective compression ratios for signals modulated at different rates," Br. J. Audiol, **26**, 351–361.

Strouse, A., Ashmead, D. H., Ohde, R. N., and Grantham, W. (**1998**). "Temporal processing in the aging auditory system," J. Acoust. Soc. Am. **104**(**4**), 2385–2399.

Summers, V., and Leek, M. R. (**1997**). "Intraspeech spread of masking in normal-hearing and impaired-hearing listeners," J. Acoust. Soc. Am. **101**(**5**), 2866–2875.

Thomas, T. G. (**1996**). "Experimental evaluation of improvement in speech perception with consonantal intensity and duration modification," Ph.D. dissertation, Dept. of Elect. Engg, IIT Bombay, India.

Thomas, T. G., Pandey, P. C., and Agashe, S. D. (**1994**). "A PC-based multiresolution spectrograph," J. IETE (India), **40**, 104–108.

Thomas, T. G., Pandey, P. C., and Agashe, S. D. (**1996**). "Effect of consonantal 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., Dorman, M. F., Franks, J. R., and Summerfield, Q. (**1980**). "Identification of synthetic /b d g/ by hearing-impaired listeners under monotic and dichotic formant presentation," J. Acoust. Soc. Am. **67**(**3**), 1031–1040.

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**(**5**), 2822–2825.

Tye-Murray, N., Spencer, L., and Gilbert-Bedia, E. (**1995**). "Relationship between speech production and speech perception skills in young cochlear-implant users," J. Acoust. Soc. Am. **98**(**5**), 2454–2460.

Tyler, R. S., and Moore, B. C. J. (**1992**). "Consonant recognition by some of the better cochlear-implant patients," J. Acoust. Soc. Am. **92**(**6**), 3068–3077.

Tyler, R. S., Wood, E. J., and Fernandes, M. A. (**1983**). "Frequency resolution and discrimination of constant and dynamic tones in normal and hearing-impaired listeners," J. Acoust. Soc. Am. **74**(**4**), 1190–1199.

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.

Van den Brink, G. (**1970**). "Two experiments on pitch: diplacusis of harmonic AM signals and pitch of inharmonic AM signals," J. Acoust. Am. **48**(**6**), (part **2**), 1355–1365.

Van der Horst, R., Leeuw, A. R., Dreschler, W. A. (**1999**). "Importance of temporal-envelope cues in consonant recognition," J. Acoust. Soc. Am. **105**(**3**), 1801–1809.

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**(**2**), Pt.1, 885–894.

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).

Walker, G., Byrne, D., and Dillon, H. (**1984**). "The effects of multichannel compression/expansion amplification on the intelligibility of nonsense syllables in noise," J. Acoust. Soc. Am. **76**(**3**), 746–757.

Wang, M. D., and Bilger, R. C. (**1973**). "Consonant confusions in noise: a study of perceptual features," J. Acoust. Soc. Am. **54**, 1248–1266.

Ward, W. D. (**1963**). "Diplacusis and auditory theory," J. Acoust. Soc. Am. **35**(**14**), 1746–1747.

Wilson, R. H., and Carhart, R. (**1971**). "Forward and backward masking: interactions and additivity," J. Acoust. Soc. Am. **49**(**4**), 1254–1263.

Yund, E. W., and Buckles, K. M. (**1995**). "Multichannel compression hearing aids: effect of number of channels on speech discrimination in noise," J. Acoust. Soc. Am. **97**(**2**), 1206–1223.

Yund, E. W., and Buckles, K. M. (**1995**). "Enhanced speech perception at low signal-to-noise ratios with multichannel compression hearing aids," J. Acoust. Soc. Am. **97**(**2**), 1224–1240.

Zemlin, W. R. (**1998**). *Speech and Hearing Science: anatomy and physiology* (Allyn and Bacon, Boston).

Zurek, P. M., and Delhorne, L. A. (**1987**). "Consonant reception in noise by listeners with mild and moderate sensorineural hearing impairment," J. Acoust. Soc. Am **82**(**5**), 1548–1559.

Zwicker, E. (**1961**). "Subdivision of audible frequency range into critical bands (Frequenz-gruppen)," J. Acoust. Soc. Am. **33**, 248.

Zwicker, E., and Schorn, K. (**1978**). "Psychoacoustical tuning curves in audiology," Audiology **17**, 120–140.

List of publications

1. Jangamashetti, D. S., and Pandey, P. C., (**2000**). "Dichotic presentation with inter-aural switching for reducing the effect of temporal masking due to sensorineural hearing loss," *Proc. National Conf. on Biomedical Engg.*, Roorkee, India, 346–353.
2. Jangamashetti, D. S., and Pandey, P. C., (**2000**). "Inter-aural switching with different fading functions for binaural dichotic presentation to reduce the effect of temporal masking in sensorineural hearing loss," *Proc. 4th World Multi Conf. on Systemics, Cybernetics, and Informatics (SCI'2000)*, Orlando, Florida, U.S.A, 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 Engineering (ICMBE)*, Bangalore, India, 258–263.

4. 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., Vol. **110**(**5**), Pt. 2, 2705.
5. Cheeran, A. N., Pandey, P. C., and Jangamashetti, D. S. (**2001**) "Comb filters for binaural dichotic presentation to improve speech perception by persons with bilateral sensorineural hearing impairment," J. Acoust. Soc. Am., Vol. **110**(**5**), Pt. 2, 2705.
6. Cheeran, A. N., Pandey, P. C., and Jangamashetti, D. S. (**2002**). "Design of comb filters based on auditory filter bandwidths for binaural dichotic presentation for persons with sensorineural hearing impairment," *Proc. 14th Int. Conf on Digital Signal Processing (DSP2002)*, Santorini, Greece, 1145–1148.
7. Cheeran, A. N., Pandey, P. C., and Jangamashetti, D. S. (**2002**). "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**(**5**), Pt.2(2), 2426.

Acknowledgements

I express my deep sense of gratitude and heartfelt devotion to my research supervisor Dr. P. C. Pandey, for his valuable guidance and continuous motivation during the course of investigation and thesis preparation.

I am sincerely thankful to the professors I have interacted with, particularly Dr. T. Anjaneyulu, Dr. S. D. Agashe, Dr. V. M. Gadre, and Dr. Preeti Rao for their observations, comments, and suggestions during the research work. I extend my thanks to the Q.I.P. Coordinator and office staff for their help in many ways.

I sincerely acknowledge the help and support given by Alice, Parveen, Milind, Dr. Devendra Chaudhari, Dinesh, Vidyadhar, Munish, Pratibha, Rajashree, Shubha, Uma, Deepa, Vinayak Shet, G. R. Doddagoudar, other friends and colleagues at IIT Bombay. My special thanks to Dr. Harshal Nemade for taking time out of his research work to help me in setting up the hardware part of the research work. My thanks to Sri Muntode for his timely help in the laboratory.

I have immensely benefited by having access to library and interaction with staff and students of Ali Yavar Jung National Institute for Hearing Handicapped, Mumbai. I am particulary grateful to Dr. R. Rangasayee, Director, Dr. M. N. Nagaraja, Dy. Director, Dr. Vijayalakshmi Basavaraj, Head, Audiology, and Mr. Rajiv Jelvi, Lecturer, Audiology, for all the support and help.

I thank Bharat, Javed, Savita, Lakshmi, Sri Surve, Sri Bendre, Sri Kambli, Mrs. Bharati, Mrs. Jagdeepkaur, and Mrs. Kalpana for participating in the listening tests.

For sponsoring my doctoral studies and research under Q.I.P. scheme, I wholeheartedly thank the Directorate of Technical Education, Karnataka, Sri V. C. Charantimath, Chairman, B.V.V. Sangha, Bagalkot, and Dr. S. Rajashekharaiah, Principal, Basaveshwar Engineering College, Bagalkot. I am indebted to Prof. Eshappa Setra, Principal, and Prof. V. Yogachandra, H.O.D., E&E, for granting me study leave and giving constant encouragement. I thank all the faculty members and office staff of my parent institute for their support. My colleagues Sri Mukartihal, Suchitra, Nanda, Diwakar, Shashidhar, Ravindra, Raghuram, and Sanjay are thanked for their encouragement and sharing the workload. I also express my thanks to Prof. Basavaraj Anami, Raju Injaganeri, Umesh Bhushi, Prakash Bangarshetti, Sunil Manvi, Basavaprabhu Sheeparmatti, Shridhar K., Pandurang Kulkarni, for their good wishes and help.

I am highly obliged to the families of Mr. B. M. Nandishaiah and Mr. H. R. Prabhakar, for extending homely feeling during my stay at IIT Bombay.

I shall be failing in my duty if I do not remember the sacrifice made by my parents, in-laws, sister Renuka, her husband Basavaraj, their children, brothers Anand and Basavaraj and sister Leela. I am ever grateful to them for their moral support and encouragement. They were very eager to see my research complete successfully. I am especially thankful to my in-laws, brother-in-law Prakash & his family, sister-in-law Renuka & her family for taking the lion's share in managing our little daughter.

This work would not have seen the light of the day but for the understanding and unfailing support of my husband Suresh. He definitely needs a special word of appreciation for his patient toleration of my disposition during the period of my research work. His love, affection, and encouragement helped to see the work finish smoothly.

Last but not the least my lovely daughter Akhila needs a special word of appreciation for her courage and patience in staying away from me during the course of the work. She had to put up long years of my absence from home and deserves sincere word of thanks for her understanding and cooperation.

Dakshayani S. Jangamashetti

References

Abel, S. M. (**1971**). "Discrimination of temporal gaps," J. Acoust. Soc. Am. **52**(**2**), 519–524.

Adobe Systems Inc. (**1988**). *Postscript Language Reference Manual* (Addision-Wesley, Reading, Massachusetts).

Apoux, F., Crouzet, O., and Lorenzi, C. (**2000b**). "Consonant recognition in noise with temporal cues: II. Effects of temporal envelope enhancement on response times," J. Acoust. Soc. Am. **107**(**5**), 2913.

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

Apoux, F., Lorenzi, C., and Berthommier, F. (**2000a**). "Consonant recognition in noise with temporal cues: I. Effects of temporal envelope enhancement on identification performance," J. Acoust. Soc. Am. **107**(**5**), 2913.

Arlinger, S., and Dryselius, H. (**1990**). "Speech recognition in noise, temporal and spectral resolution in normal and impaired hearing," Acta Otolarygol. **469**, 30–37.

Asano, F., Suzuki, Y., Sone, T., Kakehata, S., Satake, M., Ohyama, K., Kobayashi, T., and Takasaka, T. (**1991**). "A digital hearing aid that compensates loudness for sensorineural hearing listeners," *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, 3625–3628.

Babsky, E., Khodorov, B., Kositsky, G., and Zubkov, A. (**1970**). *Human Physiology*. Ed. E. Babsky, Vol. 2 (Mir Publishers, Moscow).

Baragi, Ashok B. N. (**1996**). "A speech training aid for the deaf," M. Tech. dissertation, Department of Electrical Engineering, IIT Bombay, India.

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

Brobeck, J. R. (**1973**). *Best and Taylor's Physiological Basis of Medical Practice*. 9th Ed. (Williams and Wilkins Company, Baltimore).

Brooks, D. N. (**1984**). “Binaural benefit-when and how much?” Scand. Audiol. **13**, 237–241.

Brüel and Kjær (**1985**). *Instruction Manual of Precision Sound Level Meter Type 2235* (Brüel and Kjær, Nærum, Denmark).

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

CHABA. (**1988**). “Speech understanding and aging,” J. Accost. Soc. Am. **83**(**3**), 859–895.

CHABA. (**1991**). “Speech-perception aids for the hearing-impaired people: Current status and needed research,” J. Acoust. Soc. Am. **90**, 637–685.

Chaudhari, D. S. (**2000**). “Dichotic presentation of speech signal for improving speech perception for the bilateral sensorineural hearing impairment,” Ph.D. dissertation, School of Biomedical Engineering, IIT Bombay.

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

Chaudhari, D. S., and Pandey, P.C. (**1998b**). “Real time speech processing for dichotic presentation for binaural hearing aids,” *Proc. NCBME ’98*, Manipal, Karnataka, India.

Cheeran, A. N., Pandey, P. C., and Jangamashetti, D. S. (**2002**). “Design of comb filters based on auditory filters for dichotic presentation for persons with bilateral sensorineural hearing impairment,” *Proc. of 14th Int. Conf. on Digital Signal Processing (DSP2002), Santorini, Greece*, 1145-1148.

Cheeran, A. N., Pandey, P. C., and Jangamashetti, D. S. (**2001**). “Comb filters for dichotic presentation to improve speech perception by persons with bilateral sensorineural hearing impairment,” J. Acoust. Soc. Am., vol. **110** (**5**), Pt. 2, p. 2705.

Cheung, S., and Lim, J. S. (**1992**). “Combined multiresolution (wide-band/narrow-band) spectrogram,” IEEE Trans. Signal Processing **40**, 957–977.

Coleridge Smith, P. D., and Scurr, J. H. (**1988**). Eds. *Mircocomputers in Medicine*. (Springer-Verlag, London).

Crandall, I. B. (**1917**). “The composition of speech,” Phys. Rev. **10**(**2**). in *Speech Intelligibility and Speaker Recognition,* edited by M. E. Hawley (Dowden Hutchinson Ross, Stroudsburg, PA), pp. 26–28.

Danaher, E. M., and Picket, J. M. (**1975**). "Some masking effects produced by low-frequency vowel formants in persons with sensorineural hearing loss," J. Speech Hear. Res. **18**, 261–271.

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

Dolan, T. G., and Small, A. M. (**1984**). "Frequency effects in backward masking," J. Acoust. Soc. Am. **75**(**3**), 932–935.

Donaldson, G. S., and Nelson, D. A. (**2000**). "Place-pitch sensitivity and its relation to consonant recognition by cochlear implant listeners using the MPEAK and SPEAK speech processing strategies," J. Acoust. Soc. Am. **107**(**3**), 1645–1658.

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., Marton, K., Hannley, M. T., and Lindholm, J. M. (**1985**). "Phonetic identification by elderly normal and hearing-impaired listeners," J. Acoust. Soc. Am. **77**, 664–670.

Dorman, M. F., Soli, S., Dankowski, K., Smith, L. M., McCandless, G., and Parkin, J. (**1990**). "Acoustic cues for consonant identification by patients who use the Ineraid cochlear implant," J. Acoust. Soc. Am. **88**(**5**), 2074–2079.

Duan, M. L., and Canlon, B. (**1996**). "Outer hair cell activity is not required for the generation of the forward masking curve," Audiol Neurootol. **1**, 309–319.

Dubno, J. R., Dirks, D. D., and Langhofer, L. R. (**1982**). "Evaluation of hearing-impaired using the nonsense syllable test:II," J. Speech Hear. Res. **25**, 141–148.

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., and Levit, H. (**1981**). "Predicting consonant confusions from acoustic analysis," J. Acoust. Soc. Am. **69**(**1**), 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**(**4**). Pt.1., 2110–2121.

Durlach, N. I., Gabriel, K. J., Colburn, H. S., and Trahiotis, C. (**1986**). "Interaural correlation discrimination: relation to binaural unmasking," J. Acoust. Soc. Am. **79**(**5**), 1548–1557.

Dynalog Microsystems (**1989**). *User's Manual, PCL-208 (PCL-718) Data Acquisition Card* (Dynalog Microsystems, Bombay).

Dynalog MicroSystems (**1993**). *User's Manual, PCL-DSP25 TMS 320C25 Digital Signal Processor Board* (Dynalog Microsystems, Bombay).

Elliot, L. L. (**1962**). "Backward and forward masking of probe tones of different frequencies," J. Acoust. Soc. Am. **34**(**8**), 1116–1117.

Elliot, L. L. (**1975**). "Temporal and masking phenomena in persons with sensorineural hearing loss," Audilogy **14**, 336–353.

Faulkner, A., and Rosen, S. (**1999**). "Contribution of temporal encodings of voicing, voicelessness, fundamental frequency, and amplitude variation to audio-visual and auditory speech perception," J. Acoust. Soc. Am. **106**(**4**), 2063–2073.

Faulkner, A., Rosen, S., and Smith, C. (**2000**). "Effects of the salience of pitch and periodicity information on the intelligibility of four-channel vocoded speech: Implications for cochlear implants," J. Acoust. Soc. Am. **108**(**4**), 1877–1887.

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 Communication in Hearing*. (Huntington, New York, Krieger).

Florentine, M., and Buus, S. (**1984**). "Temporal gap detection in sensorineural hearing impairment," J. Speech Hear. Res. **27**, 449–455.

Florentine, M., Buus, S., Scharf, B., and Zwicker, E. (**1980**). "Frequency selectivity in normally-hearing and hearing-impaired observers," J. Speech Hear. Res. **23**, 646–669.

Foust, K. O., and Gengel, R. W. (**1973**). "Speech discrimination by sensorineural hearing-impaired persons using a transposer hearing aid," Scand. Audiol. **2**, 161–170.

Freyman, R. L., and Nelson, D. A. (**1986**). "Frequency discrimination as a function of tonal duration and excitation-pattern slopes in normal and hearing-impaired listeners," J. Acoust. Soc. Am. **79**(**4**), 1034–1044.

Fu, Q. J., and Shannon R. V. (**1999**). "Effects of electrode location and spacing on phoneme recognition with the Nucleus-22 cochlear implant," Ear Hear. **20**(**4**), 321–331.

Fu, Q. J., Shannon, R. V., and Wang, X. (**1998**). "Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing," J. Acoust. Soc. Am. **104**, 3586–3596.

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

Ganong, W. F. (**1991**). *Review of Medical Physiology*. 15th Ed. (Lange Medical Publications, Los Altos, California).

Gatehouse, S., and Gordon, J. (**1990**). "Response times to speech stimuli as measures of benefit from amplification," Br. J. Audiol, **26**, 63–68.

Gehr, S. E., and Sommers, M. S. (**1999**). "Age difference in backward masking," J. Acoust. Soc. Am. **106**(**5**), 2793–2799.

Gelfand, S. A. (**1990**). *Hearing-an Introduction toPsychological and Physiological Acoustics.* 2nd Ed. (Marcle Dekker, New York).

Gelfand, S. A., Piper, N., and Silman, S. (**1985**). "Consonant recognition in quite as a function of aging among normal hearing subjects," J. Acoust. Soc. Am. **78**(**4**), 1198–1206.

Geurts, L., and Wouters, J. (**1999**). "Enhancing the speech envelope of continuous interleaved sampling processors for cochlear implants," J. Acoust. Soc. Am. **105**(**4**), 2476–2484.

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., Patterson, R. D., and Nimmo-Smith, I. (**1984**). "Dynamic range and asymmetry of the auditory filter," J. Acoust. Soc. Am. **76**(**2**), 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., 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**, 1546–1556.

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

Gordon-Salant, S., and Fitzgibbons, P. J. (**1993**). "Temporal factors and speech recognition performance in young and elderly listeners," J. Speech Hear. Res. **36**, 1276–1285.

Green, D. M., Birdsall, T. G., and Tanner, W. P. (**1957**). "Signal detection as a function of signal intensity and duration," J. Acoust. Soc. Am. **29**(**4**), 523–531.

Gulick, W. L. (**1971**). *Hearing, Physiology, and Psychophysics*. (Oxford University Press, New York).

Guyton, A. C. (**1986**). *Textbook of Medical Physiology* (Saunders, Philadelphia, PA).

Hannley, M., and Dorman, M. F. (**1983**). "Susceptibility to intraspeech spread of masking in listeners with sensorineural hearing loss," J. Acoust. Soc. Am. **74** (**1**), 40–51.

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

Hawkins, D. B., and Yacullo, W. S. (**1984**). "Signal-to-noise ratio advantage of binaural hearing aids and directional microphones under different levels of reverberation," J. Speech Hear. Disord. **49**, 278–286.

Hirsh, I. J. (**1950**). "Binaural hearing aids: a review of some experiments," J. Speech Hear. Disord. **15**, 114–123.

Hou, Z., and Pavlovic, C. V. (**1994**). "Effects of temporal smearing on temporal resolution, frequency selectivity, and speech intelligibility," J. Acoust. Soc. Am. **96**(**3**), 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**(**3**), 765–773.

Jamieson, D.G., Ponton, W., and Espinoza-Varas, B. (**1985**). "Reduction in formant bandwidth improves vowel identification with sensorineural impairment," J.Acoust. Soc. Am. Suppl. 1 **77**, S8.

Janssen, T., Boege, P., Oestreicher, E., and Arnold, W. (**2000**), "Tinnitus and 2f1-f2 distortion product otoacoustic emissions following salicylate overdose," J. Acoust. Soc. Am. **107**(**3**) 1790–1792.

Jerger, J., Carhart, R., and Dirks, D. (**1961**). "Binaural hearing aids and speech intelligibility," J. Speech Hear. Res. **4**, 137–148.

Jesteadt, W., Bacon, S. P., and Lehman, J. R. (**1982**). "Forward masking as a function of frequency, masker level, and signal delay," J. Acoust. Soc. Am. **71**, 950–962.

Johannson, B. (**1966**). "The use of the transposer for the management of the deaf child," in *Sensory Aids for the Hearing Impaired*, edited by Levitt, H., Pickett, J. M., and Houde, R. A. (IEEE Press, New York), pp. 195–204.

Kandel, E. R., Schwartz, J. H., and Jessel, T. M. (**1991**). *Principles of Neural Sciences*. (Elsevier, New York).

Kasturi, P. (**1997**). "Real-time Filter Design with Tunable Frequency Response using TMS320C50 Processor," Project report, Signal Processing and Instrumentation Lab., Deparment of Electrical Engineering, IIT Bombay, India.

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**(**2**), 1098–1114.

Kewley-Port, D. (**1982**). "Measurement of formant transitions in naturally produced stop consonant-vowel syllables," J. Acoust. Soc. Am. **72**(**2**), 379–389.

Kidd, G., Mason, C. R., and Feth, L. L. (**1984**). "Temporal integration of forward masking in listeners having sensorineural hearing loss," J. Acoust. Soc. Am. **75**(**3**), 937–944.

Kiefer, J., von. Ilberg, C., Hubner-Egner, J., Rupprecht, V., Knecht, R. (**2000**). "Optimized speech understanding with the continuous interleaved sampling speech coding strategy in patients with cochlear implants: effect of variations in stimulation rate and number of channels," Ann. Otol. Rhinol. Laryngol. **109**, 1009–1020.

Kiessling, J., and Steffens, T. (**1991**). "Clinical evaluation of a programmable three-channel automatic gain control amplification system," Audiology **30**, 70–81.

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

Larsby, B., and Arlinger, S. (**1998**). "A method for evaluating temporal, spectral and combined temporal-spectral resolution in hearing," Scand. Audiol. Suppl. **27**, 3–12.

Leek, M. R., and Summers, V. (**1996**). "Reduced frequency selectivity and the preservation of spectral contrast in noise," J. Acoust. Soc. Am. **100**(**3**), 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.

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

Liberman, A. M. (**1957**). "Some results of research on speech perception," J. Acoust. Soc. Am. **29**(**1**), 117–123.

Lochner, J. P. A., and Burger, J. F. (**1961**). "Form of loudness function in the presence of masking noise," J. Acoust. Soc. Am. **33**(**12**), 1705–1707.

Lorenzi, C., Berthommier, F., Apoux, F., Bacri, N. (**1999**). "Effects of envelope expansion on speech recognition," Hear. Res. **136**, 131–138.

Loizou, P.C., Dorman, M., Poroy, O., Spahr, T. (**2000**). "Speech recognition by normal-hearing and cochlear implant listeners as a function of intensity resolution," J. Acoust. Soc. Am **108**(**5**), 2377–2387.

Lunner, T. (**1997**). *"A Digital Filterbank Hearing Aid,"* Ph.D. dissertation, Department. of Neuroscience and Locomotion, Division of Audiology, Linkoping University, 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.

Meftah, M., and Boudelaa, Sami (**1996**)."How facilitatory can lexical information be during word recognition? Evidence from moroccon arabic," *Proc. of Int. Conference on Spoken Language Processing (ICSLP)*, Philapdelphia, USA. **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**(**2**), 338–352.

Møller, A. R. (**2000**). *Hearing: Its Physiology and Pathophysiology.* (Academic, Texas).

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. (**1982**). *An Introduction to the Psychology of Hearing*. 2nd Ed. (Academic, London).

Moore, B. C. J. (**1997**). *An Introduction to the Psychology of Hearing*. 4th Ed. (Academic, London).

Moore, B. C. J. (**1998**). "Psychoacoustics of cochlear hearing impairment and the design of hearing aids*," Proc. 16th Int. Cong. Acoustics (ICA)*, Seattle, WA. 2105–2108.

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 quite and in a background of speech," J. Acoust. Soc. Am. **94**(**4**), 2050–2062.

Moore, B. C. J., Peters, R. W., and Glasberg. B. R. (**1992**). "Detection of temporal gaps in sinusoids by elderly subjects with and without hearing loss," J. Acoust. Soc. Am. **92**(**4**), Pt. 1, 1923–1932.

Morris, L. R. (**1988**). "A PC-based digital spectrograph," IEEE Micro. **8**, 68–85.

Munson, W. A. (**1947**). "The growth of auditory sensation," J. Acoust. Soc. Am. **19**(**4**), 584–591.

Nejime, Y., and Moore, B. C. J. (**1998**). "Evaluation of the effect of speech-rate slowing on speech intelligibility in noise using a simulation of cochlear hearing loss," J. Acoust. Soc. Am. **103**(**1**), 572–576.

Nelson, D. A., Chargo, S. J. Kopun, J. G., and Freyman, R. (**1990**). "Effects of stimulus level on forward masked psychophysical tuning curves in quite and noise," J. Acoust. Soc. Am. **88**, 2143–2151.

Nilson, 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**(**2**), 1085–1099.

Ono, H., Okasaki, T., Nakai, S., and Harasaki, H. (**1982**). "Identification of an emphasized consonant of a monosyllable in hearing-impaired and its application to a hearing aid," J. Acoust. Soc. Am. **71**, S58.

Oppenheim, A. V. (**1970**). "Speech spectrograms using the fast Fourier transform," IEEE Spectrum, **7**, 57–62.

O'Shaughnessy, D. (**1987**). *Speech Communication: Human and Machine* (Addison-Wesley, Reading, Massachuetts).

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.

Pandey, P. C. (**1987**). "Speech Processing for Cochlear Prostheses," Ph. D dissertation, Department of Electrical Engineering, Univ. of Toronto, Canada.

Pandey, P. C., Kunov, H., and Abel, S. M. (**1987**). "A speech processor providing fricative and low-frequency periodicity information for single channel cochlear prosthesis," *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing* **34.3**.

Patterson, R. D. (**1976**). "Auditory filter shapes derived with noise stimuli," J. Acoust. Soc. Am. **59**(**3**), 640–654.

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. **2**, 96–103.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1986**). "Speaking clearly for the hard of hearing II: intelligibility differences between clear and conversational speech," J. Speech Hear. Res. **29**, 434–446.

Pickles, J. O. (**1982**). *An Introduction to the Physiology of Hearing* (Academic, London).

Plack, C. J., and Oxenham, A. J. (**1998**). "Basilar-membrane nonlinearity and the growth of forward masking," J. Acoust. Soc. Am. **103**(**3**), 1598–1608.

Plomp, R. (**1964**). "Rate of decay of auditory sensation," J. Acoust. Soc. Am. **36**(**2**), 227–282.

Plomp, R., and Bouman, M. A. (**1959**). "Relation between hearing threshold and duration for tone pulses," J. Acoust. Soc. Am. **31**(**6**), 749–758.

Pols, L. C. W., and Schouten, M. E. H. (**1978**). "Identification of deleted consonants," J. Acoust. Soc. Am. **64**, 1333–1337.

Prasad, V. V. R. (**1996**). "Speech processing for single channel aid," M. Tech. dissertation, Dept. of Elect. Engg, IIT Bombay, India.

Previte, J. J. (**1983**). *Human Physiology*. (Mc Graw-Hill Book Company, NewYork).

Proakis, J. G., and Manolakis, D. G. (**1997**). *Digital Signal Processing Principles, Algorithms, and Applications* (Prentice Hall of India, New Delhi).

Punch, J. L., Montgomery, A. A., Schwartz, D. M., Walden, B. E., Prosek, R. A., and Howard, M. T. (**1980**). "Multidimensional scaling of quality judgments of speech signals processed by hearing aids," J. Acoust. Soc. Am. **68**(**2**), 458–466.

Raab, D. H. (**1961**). "Forward and backward masking between acoustic clicks," J. Acoust. Soc. Am. **33**(**2**), 137–139.

Rabiner, L. R., and Gold, B. (**1998**). *Theory and Application of Digital signal processing,* (Prentice Hall of India, New Delhi).

Rabiner, L. R., Gold, B. and McGonegal, C. A. (**1970**). "An approach to the approximation problem for nonrecursive digital filters," IEEE Trans. Audio Electroacoustics **18,** 83–106.

Rabiner, L. R., and Schafer, R. W. (**1978**). *Digital Processing of the Speech Signals* (Prentice Hall, Englewood Cliffs, NJ).

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-passed filtering and pitch-invariant frequency lowering," J. Acoust. Soc. Am. **74**, 409–419.

Resnick, S. B., Dubno, J. R., Hoffnung, S. and Levitt, H. (**1975**). “Phoneme errors on a nonsense syllable test,” J. Acoust. Soc. Am. **58**, Suppl. No.1, S114.

Revoile, S., Holden-Pitt, L., Pickett, J. M, and Bandt, F. (**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., Holden-Pitt, L., Pickett, J. M, and Bandt, 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.

Scharf, B. (**1969**). “Dichotic summation of loudness,” J. Acoust. Soc. Am. **45**, 1193-1205.

Schneider, B. A., Pichora-Fuller, M. K., Kowalchuk, D., and Lamb, M. (**1994**). “Gap detection and the precedence effect in young and old adults,” J. Acoust. Soc. Am. **95**(**2**), 980–991.

Severns, M. L. (**1985**). “A computer controlled attenuator for audiological testing,” J. Clinical Engineering **10**(**4**) 317-321.

Shah, N. (**1995**). “A sensory aid for the deaf,” M. Tech. dissertation, Department of Electrical Engineering, IIT Bombay, India.

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 band limited noise: Effects of variations in bandwidth and signal to masker ratio,” J. Acoust. Soc. Am. **77**(**2**), 635–639.

Sheely, E. C., and Bilger, R. C. (**1964**). “Temporal integration as a function of frequency,” J. Acoust. Soc. Am. **36**(**10**), 1850–1857.

Simon, C. (**1978**). “On the use of comfortable listening levels in speech experiments,” J. Acoust. Soc. Am. **64**, 744–751.

Skinner, M. W. (**1980**). “Speech intelligibility in noise-induced hearing loss: Effects of high-frequency compensation,” J. Acoust. Soc. Am. **67**(**1**), 306–317.

Skinner, M. W., Holden, L. K., Holden, T. A., Demorest, M. E., and Fourakis, M. S. (**1997**). “Speech recognition at simulated soft, conversational, and raised-to-loud vocal efforts by adults with cochlear implants,” J. Acoust. Soc. Am. **101**, 3766–3782.

Smiarowski, R. A., and Carhart, R. (**1975**). “Relations among temporal resolution, forward masking and simultaneous masking,” J. Acoust. Soc. Am. **57**(**5**), 1169–1173.

Snedecor, G. W., and Cochran, W. G. (**1980**). *Statistical Methods* (The Iowa State University Press, Ames, Iowa).

Snell, K.B. (**1997**). "Age-related changes in temporal gap detection," J. Acoust. Soc. Am. **101**, 2214–2220.

Stevens, S. S. (**1965**). "Power-group transformations under glare, masking, and recruitment," J. Acoust. Soc. Am. **39**(**4**)**,** 725–735.

Stone, M. A., and Moore, B. C. J. (**1992**). "Syllabic compression: effective compression ratios for signals modulated at different rates," Br. J. Audiol, **26**, 351–361.

Strouse, A., Ashmead, D. H., Ohde, R. N., and Grantham, W. (**1998**). "Temporal processing in the aging auditory system," J. Acoust. Soc. Am. **104**(**4**), 2385–2399.

Summers, V., and Leek, M. R. (**1997**). "Intraspeech spread of masking in normal-hearing and impaired-hearing listeners," J. Acoust. Soc. Am. **101**(**5**), 2866–2875.

Thomas, T. G. (**1996**). "Experimental evaluation of improvement in speech perception with consonantal intensity and duration modification," Ph.D. dissertation, Dept. of Elect. Engg, IIT Bombay, India.

Thomas, T. G., Pandey, P. C., and Agashe, S. D. (**1994**). "A PC-based multiresolution spectrograph," J. IETE (India), **40**, 104–108.

Thomas, T. G., Pandey, P. C., and Agashe, S. D. (**1996**). "Effect of consonantal 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., Dorman, M. F., Franks, J. R., and Summerfield, Q. (**1980**). "Identification of synthetic /b d g/ by hearing-impaired listeners under monotic and dichotic formant presentation," J. Acoust. Soc. Am. **67**(**3**), 1031–1040.

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**(**5**), 2822–2825.

Tye-Murray, N., Spencer, L., and Gilbert-Bedia, E. (**1995**). "Relationship between speech production and speech perception skills in young cochlear-implant users," J. Acoust. Soc. Am. **98**(**5**), 2454–2460.

Tyler, R. S., and Moore, B. C. J. (**1992**). "Consonant recognition by some of the better cochlear-implant patients," J. Acoust. Soc. Am. **92**(**6**), 3068–3077.

Tyler, R. S., Wood, E. J., and Fernandes, M. A. (**1983**). "Frequency resolution and discrimination of constant and dynamic tones in normal and hearing-impaired listeners," J. Acoust. Soc. Am. **74**(**4**), 1190–1199.

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.

Van den Brink, G. (**1970**). "Two experiments on pitch: diplacusis of harmonic AM signals and pitch of inharmonic AM signals," J. Acoust. Am. **48**(**6**), (part **2**), 1355–1365.

Van der Horst, R., Leeuw, A. R., Dreschler, W. A. (**1999**). "Importance of temporal-envelope cues in consonant recognition," J. Acoust. Soc. Am. **105**(**3**), 1801–1809.

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**(**2**), Pt.1, 885–894.

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).

Walker, G., Byrne, D., and Dillon, H. (**1984**). "The effects of multichannel compression/expansion amplification on the intelligibility of nonsense syllables in noise," J. Acoust. Soc. Am. **76**(**3**), 746–757.

Wang, M. D., and Bilger, R. C. (**1973**). "Consonant confusions in noise: a study of perceptual features," J. Acoust. Soc. Am. **54**, 1248–1266.

Ward, W. D. (**1963**). "Diplacusis and auditory theory," J. Acoust. Soc. Am. **35**(**14**), 1746–1747.

Wilson, R. H., and Carhart, R. (**1971**). "Forward and backward masking: interactions and additivity," J. Acoust. Soc. Am. **49**(**4**), 1254–1263.

Yund, E. W., and Buckles, K. M. (**1995**). "Multichannel compression hearing aids: effect of number of channels on speech discrimination in noise," J. Acoust. Soc. Am. **97**(**2**), 1206–1223.

Yund, E. W., and Buckles, K. M. (**1995**). "Enhanced speech perception at low signal-to-noise ratios with multichannel compression hearing aids," J. Acoust. Soc. Am. **97**(**2**), 1224–1240.

Zemlin, W. R. (**1998**). *Speech and Hearing Science: anatomy and physiology* (Allyn and Bacon, Boston).

Zurek, P. M., and Delhorne, L. A. (**1987**). “Consonant reception in noise by listeners with mild and moderate sensorineural hearing impairment,” J. Acoust. Soc. Am **82**(**5**), 1548–1559.

Zwicker, E. (**1961**). “Subdivision of audible frequency range into critical bands (Frequenz-gruppen),” J. Acoust. Soc. Am. **33**, 248.

Zwicker, E., and Schorn, K. (**1978**). “Psychoacoustical tuning curves in audiology,” Audiology **17**, 120–140.

List of publications

1. Jangamashetti, D. S., and Pandey, P. C., (**2000**). "Dichotic presentation with inter-aural switching for reducing the effect of temporal masking due to sensorineural hearing loss," *Proc. National Conf. on Biomedical Engg.*, Roorkee, India, 346–353.
2. Jangamashetti, D. S., and Pandey, P. C., (**2000**). "Inter-aural switching with different fading functions for binaural dichotic presentation to reduce the effect of temporal masking in sensorineural hearing loss," *Proc. 4^{th} World Multi Conf. on Systemics, Cybernetics, and Informatics (SCI'2000)*, Orlando, Florida, U.S.A, 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 Engineering (ICMBE)*, Bangalore, India, 258–263.

4. 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., Vol. **110**(**5**), Pt. 2, 2705.
5. Cheeran, A. N., Pandey, P. C., and Jangamashetti, D. S. (**2001**) "Comb filters for binaural dichotic presentation to improve speech perception by persons with bilateral sensorineural hearing impairment," J. Acoust. Soc. Am., Vol. **110**(**5**), Pt. 2, 2705.
6. Cheeran, A. N., Pandey, P. C., and Jangamashetti, D. S. (**2002**). "Design of comb filters based on auditory filter bandwidths for binaural dichotic presentation for persons with sensorineural hearing impairment," *Proc. 14^{th} Int. Conf on Digital Signal Processing (DSP2002)*, Santorini, Greece, 1145–1148.
7. Cheeran, A. N., Pandey, P. C., and Jangamashetti, D. S. (**2002**). "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**(**5**), Pt.2(2), 2426.

Acknowledgements

I express my deep sense of gratitude and heartfelt devotion to my research supervisor Dr. P. C. Pandey, for his valuable guidance and continuous motivation during the course of investigation and thesis preparation.

I am sincerely thankful to the professors I have interacted with, particularly Dr. T. Anjaneyulu, Dr. S. D. Agashe, Dr. V. M. Gadre, and Dr. Preeti Rao for their observations, comments, and suggestions during the research work. I extend my thanks to the Q.I.P. Coordinator and office staff for their help in many ways.

I sincerely acknowledge the help and support given by Alice, Parveen, Milind, Dr. Devendra Chaudhari, Dinesh, Vidyadhar, Munish, Pratibha, Rajashree, Shubha, Uma, Deepa, Vinayak Shet, G. R. Doddagoudar, other friends and colleagues at IIT Bombay. My special thanks to Dr. Harshal Nemade for taking time out of his research work to help me in setting up the hardware part of the research work. My thanks to Sri Muntode for his timely help in the laboratory.

I have immensely benefited by having access to library and interaction with staff and students of Ali Yavar Jung National Institute for Hearing Handicapped, Mumbai. I am particulary grateful to Dr. R. Rangasayee, Director, Dr. M. N. Nagaraja, Dy. Director, Dr. Vijayalakshmi Basavaraj, Head, Audiology, and Mr. Rajiv Jelvi, Lecturer, Audiology, for all the support and help.

I thank Bharat, Javed, Savita, Lakshmi, Sri Surve, Sri Bendre, Sri Kambli, Mrs. Bharati, Mrs. Jagdeepkaur, and Mrs. Kalpana for participating in the listening tests.

For sponsoring my doctoral studies and research under Q.I.P. scheme, I wholeheartedly thank the Directorate of Technical Education, Karnataka, Sri V. C. Charantimath, Chairman, B.V.V. Sangha, Bagalkot, and Dr. S. Rajashekharaiah, Principal, Basaveshwar Engineering College, Bagalkot. I am indebted to Prof. Eshappa Setra, Principal, and Prof. V. Yogachandra, H.O.D., E&E, for granting me study leave and giving constant encouragement. I thank all the faculty members and office staff of my parent institute for their support. My colleagues Sri Mukartihal, Suchitra, Nanda, Diwakar, Shashidhar, Ravindra, Raghuram, and Sanjay are thanked for their encouragement and sharing the workload. I also express my thanks to Prof. Basavaraj Anami, Raju Injaganeri, Umesh Bhushi, Prakash Bangarshetti, Sunil Manvi, Basavaprabhu Sheeparmatti, Shridhar K., Pandurang Kulkarni, for their good wishes and help.

I am highly obliged to the families of Mr. B. M. Nandishaiah and Mr. H. R. Prabhakar, for extending homely feeling during my stay at IIT Bombay.

I shall be failing in my duty if I do not remember the sacrifice made by my parents, in-laws, sister Renuka, her husband Basavaraj, their children, brothers Anand and Basavaraj and sister Leela. I am ever grateful to them for their moral support and encouragement. They were very eager to see my research complete successfully. I am especially thankful to my in-laws, brother-in-law Prakash & his family, sister-in-law Renuka & her family for taking the lion's share in managing our little daughter.

This work would not have seen the light of the day but for the understanding and unfailing support of my husband Suresh. He definitely needs a special word of appreciation for his patient toleration of my disposition during the period of my research work. His love, affection, and encouragement helped to see the work finish smoothly.

Last but not the least my lovely daughter Akhila needs a special word of appreciation for her courage and patience in staying away from me during the course of the work. She had to put up long years of my absence from home and deserves sincere word of thanks for her understanding and cooperation.

Dakshayani S. Jangamashetti