DICHOTIC PRESENTATION FOR IMPROVING SPEECH PERCEPTION BY PERSONS WITH BILATERAL SENSORINEURAL HEARING IMPAIRMENT

Thesis

submitted in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy

by

Devendra S. Chaudhari

School of Biomedical Engineering
Indian Institute of Technology, Bombay
Powai, Mumbai 400 076, India

March 2000

INDIAN INSTITUTE OF TECHNOLOGY, BOMBAY

THESIS APPROVAL

Thesis entitled : Dichotic presentation for improving speech perception by persons with bilateral sensorineural hearing impairment

by **Devendra S. Chaudhari**

is approved for the award of the degree of DOCTOR OF PHILOSOPHY.

Supervisor	: PCPandey	(Prof. P. C. Pandey)
Internal Examiner	: S.D. Agashe	(Prof. S. D. Agashe)
External Examiner	: SSanyal	(Prof. S. Sanyal)
Chairman	: BSeth	(Prof. B. Seth)

Date: March 16, 2000

Dedicated to my parents

Smt. Sulochana and Shri. Sona Chaudhari

my first teachers

Devendra S. Chaudhari. *Dichotic Presentation for Improving Speech Perception by Persons with Bilateral Sensorineural Hearing Impairment*, Ph.D. Thesis, Biomedical Engineering IIT Bombay, Powai, Mumbai, 2000, Supervisor: Dr. P.C. Pandey

ABSTRACT

Sensorineural impairment of the hearing mechanism is associated with decrease in frequency resolving capacity of the auditory system due to spread of spectral masking along the cochlear partition. As the consonantal place feature is cued by spectral differences, the hearing impaired persons may find difficulties particularly in identifying this feature. Splitting the speech signal by filtering it with a filter bank and adding signals from alternate bands for presenting to the two ears is likely to reduce the effect of spectral masking and thus help in improving the speech intelligibility. The present research deals with implementation and evaluation of a scheme for binaural dichotic presentation by splitting speech into two signals with complementary short-time spectra by using filters with the magnitude response based on critical bands (corresponding to auditory filters) and linear phase response. The scheme uses 18 critical bands over a 5 kHz frequency range. For experimental evaluation, the test material consisted of nonsense syllables with twelve English consonants in vowel-consonant-vowel and consonant-vowel contexts.

In the first set of experiments, the scheme was implemented for off-line processing; using a cascade combination of band reject filters (linear phase FIR filters with 255 coefficients). Listening tests were carried out, on normal hearing subjects with simulated hearing loss and on subjects with bilateral sensorineural hearing loss, for comparing the dichotic presentation of processed signals with diotic presentation of unprocessed signals. The scheme resulted in improving speech quality, response time, recognition scores, and transmission of features, particularly the place feature, indicating the usefulness of the scheme for better reception of spectral characteristics.

In the second set of experiments, the scheme was implemented for real-time processing using two DSP boards. The critical band based comb filter response for each channel was realized as a 128-point linear phase FIR filter using frequency sampling technique. Listening tests were carried out, on subjects with bilateral sensorineural hearing loss, for comparing the dichotic presentation of processed signals with diotic presentation of unprocessed signals. Test results were similar to those obtained with off-line processing. In a second implementation, the two filter magnitude responses were altered within ± 3 dB, as a partial compensation for the frequency dependence of the hearing loss of the individual subjects. The additional improvements were found to be related to the extent of variation in hearing loss with frequency for the individual subjects. Thus, shaping of the magnitude response can be coupled with splitting of speech signal for the dichotic presentation.

CONTENTS

Chapters

Appendices

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVATIONS

Symbol	Designation
$A_a(f)$	adjustable filter magnitude response, in dB
$A_c(f)$	gain, in dB, for constant gain implementation
$\alpha(f)$	interpolated value of hearing loss, in dB, at frequency f
α_{max}	maximum value of hearing loss, in dB
α_{min}	minimum value of hearing loss, in dB
f	signal frequency
Δf	bandwidth
f_p	pass-band edge frequency
f_s	stop-band edge frequency
$I(r)$	information measure of response r, in bits
$I(s)$	information measure of stimulus s, in bits
$I(s;\ r)$	information transmitted from stimulus s to response r in bits
$I_{trel}(s;\ r)$	information transmitted relative to stimulus information
p	significance level, probability
$p(r_j)$	probability of response r_j
$p(s_i)$	probability of stimulus s_i
$p(s_i;r_j)$	probability of stimulus response pair (s_i,r_j)
R_s	recognition or articulation score
$s(t)$	analog signal
$s(n)$	digital signal

Abbreviation	Explanation
AC	alternating current
ADC	analog-to-digital converter
AG	adjustable gain
AGC	automatic gain control
AIC	analog interface circuit
ALU	arithmetic and logic unit
AVC	automatic volume control
Avg.	average

B&K	Brüel and Kjær
CB	critical band
CD	carrier detect
CG	constant gain
CTS	clear to send data
C/V	consonant-to-vowel
CV	consonant-vowel
CVR	consonant-to-vowel ratio
DAC	digital-to-analog converter
dB	decibel
DC	direct current
DEC	Digital Equipment Corporation
df	degree of freedom
DL	difference limen
DMA	direct memory access
DSP	digital signal processing
DSR	data set ready
DTR	data terminal ready
FFT	fast Fourier transform
Fig.	figure
FIR	finite impulse response
HL	hearing level
HLL	high level language
HP	Hewlett-Packard
HTL	hearing threshold level
Hz	Hertz
IEC	International Electrotechnical Commission
IHC	inner hair cell
INTEN.	intensity
LDL	loudness discomfort level
MLP	mean logarithmic probability
NS	not significant
OHC	outer hair cell
PC	personal computer
PCL	PC-LabCard
PS	processed speech
PS-CG	processed speech with constant gain filter

PS-AG	processed speech with adjustable gain filter
PTA	pure tone threshold average
RAM	random access memory
RS	recognition score
RTS	request to send data
RX	receive
s.d.	standard deviation
SNR	signal-to-noise ratio
SPL	sound pressure level
SR	sampling rate
TI	Texas Instruments
TX	transmit
US	unprocessed speech
VCV	vowel-consonant-vowel

Chapter 1

INTRODUCTION

1.1 Problem overview

Sensorineural hearing impairment is characterized by increase in threshold of hearing, reduction in dynamic range of hearing, degradation of temporal resolution and increase in temporal masking, and degraded frequency selectivity due to increase in spectral masking. Many hearing aids incorporate frequency compensation and amplitude compression. Some of the techniques currently being investigated are based on signal processing schemes such as spectral transposition, and speech enhancement using the properties of "clear" speech. These are likely to increase the performance of hearing aids for persons with residual hearing as well as that of other sensory aids like cochlear prostheses and vibro-tactile aids used by profoundly hearing impaired.

The increase in hearing loss is associated with widening of auditory filters along the cochlear partition (Dubno and Dirks, 1989; Moore, 1997). The above mentioned processing techniques do not adequately address the problem of degradation of speech perception caused by broadening of critical bands corresponding to the auditory filters. The hearing impaired persons have uncertainty in the recognition of the transition of formants, and frequency band of the noise burst due to spectral masking. For example, in the perception of the unvoiced stops /p, t, k/ with vowel /*a*/, the high frequency noise of /t/ distinguishes it from the low frequency noise of /p/ and /k/. Stop /k/ is characterized by slightly higher frequency noise as compared to /p/ (Flanagan, 1972). The consonantal segments in which formant transitions and the noise bands are not widely different will have perceptual confusions due to spectral masking.

A possible solution for this problem is a binaural dichotic presentation in which speech is split into two signals with complementary spectra. Thus, two adjacent bands that are likely to mask each other get presented to different ears. The human ability to perceptually combine the binaurally received signals from the two ears for improving speech reception under adverse listening conditions has been established earlier (Moore, 1982). Since the scheme is aimed at reducing the effect of spectral masking, it should result in improvement in reception of consonantal "place" feature without adversely affecting the reception of other features.

1.2 Research objective

The hearing aids, based on speech processing for binaural dichotic (different signal in each ear) presentation can reduce the effect of spectral masking. The objective of the present study is to investigate a scheme of splitting speech into two signals, with complementary spectra, for binaural dichotic presentation, in order to lessen the effect of reduced auditory frequency selectivity and thereby to improve speech intelligibility. The scheme is implemented for off-line processing, and is experimentally evaluated through listening tests, involving normal hearing subjects with simulated hearing impairment and subjects with bilateral sensorineural hearing impairment. The scheme is also implemented for real-time processing and listening tests are carried out with subjects having bilateral sensorineural impairment.

1.3 Thesis outline

Chapter 2 gives a review of the auditory system, types of hearing impairment, some of the perceptual effects of sensorineural hearing loss, sensory aids, and some of the speech processing techniques for improving speech reception for the hearing impaired. Chapter 3 provides a literature review on binaural dichotic presentation, description of proposed scheme of investigation, and discussion of evaluation techniques to be used. The results obtained from experimental evaluation of the scheme in off-line and real-time processing are presented and discussed in

Chapters 4 and 5 respectively. Chapter 6 provides the summary of the work done, conclusions drawn from the study, and suggestions for future work.

Appendices provide supplementary information. Appendix A provides a description of spectrographic analyzer set-up. The hardware and software for experiments are described in Appendix B and C, respectively. Appendix D outlines the procedure followed for the electroacoustic calibration of the headphones that were used in the listening tests. Information transmission analysis and analysis of sample listening test results are described in Appendix E and F, respectively. The hearing-impaired subject data are covered in Appendix G. Test instructions to the subjects and forms are included in Appendix H. Spectrograms of the unprocessed and processed speech stimuli are given in a separate supplement.

Chapter 2

SENSORINEURAL HEARING IMPAIRMENT

2.1 Introduction

This chapter provides an overview of auditory system, various types of hearing impairment, some of the perceptual effects of sensorineural hearing loss, and sensory aids. It also reviews some of the speech processing techniques for improving speech reception in case of sensorineural hearing loss.

2.2 The auditory system

Fig. 2.1 shows a structure of the peripheral auditory system (Flanagan, 1972; Pickles, 1982; Thibodeau, 1992). The sound waves received by external ear pass through ear canal and cause vibrations of the tympanic membrane (or eardrum). The middle ear consists of a cavity with a delicate chain of three tiny bones (the malleus, the incus, and the stapes). These bones couple the vibration of the tympanic membrane to the inner ear. The inner ear consists of a fluid filled bony spiral of two and half turn, called cochlea. Fig. 2.2 shows the structure of the inner ear and a transverse section of the cochlea with its three chambers: scala vestibuli, scala media, and scala tympani. At upper side, the scala media is separated from scala vestibuli by Reissner's membrane. It is separated from scala tympani by basilar membrane and bony shelf at lower side. The two outer scalae and inner scala contain perilymph and endolymph respectively. The organ of Corti sits on the basilar membrane, and contains hair cells (about 30,000 in number). The hair cells are arranged as two rows of "ïnner" hair cells and three-to-five rows of "outer" hair cells. The organ of Corti is covered by tectorial membrane (Pickles, 1982).

Fig. 2.3 shows transmission of sound waves through cochlear longitudinal cross section (Thibodeau, 1992). Incidence of sound waves on tympanic membrane causes it to vibrate. This energy reaches the oval window and gets transmitted to the fluid in the cochlear chambers (perilymph and endolymph) setting up traveling waves, and causing basilar membrane to vibrate. The basilar membrane, with stiffness highest near the oval window and progressively reducing along the length, behaves like a dispersive transmission medium in which the traveling waves lose the high frequency energy while propagating towards apical end. Differential spatial activity takes place in response to these traveling waves of different frequencies. The highest frequencies in the audible range affect the region near the oval window, whereas the lowest frequencies affect the far end, the helicotrema (Moore, 1982; Pickles, 1982; O'Shaughnessy, 1987).

Fig. 2.4 indicates patterns of vibration on the basilar membrane for low frequency sounds. The low frequency sounds set relatively higher length of basilar membrane into vibration as compared to high frequency sounds. This might be the reason for high sensitivity of the ear to low frequency sounds (Moore, 1982). The upward and downward movement of basilar membrane results in upward-inward and downward-outward movement of reticular lamina respectively, as shown in the Fig. 2.5. The inward and outward bending of the hair cell cause generation of electric potential difference which stimulates the cochlear nerve endings resulting in nerve impulses (Guyton, 1986; O'Shaughnessy, 1987). These impulses are transmitted over the cochlear nerve, a part of vestibulo-chochlear nerve (eighth cranial nerve), to the higher processes of the brain (Wilson, 1990; Thibodeau, 1992). The information travels through the cochlear nucleus, the superior olivary complex, the inferior colliculus, and the medial geniculate body, ending at left and right hemisphere of the auditory cortex (Moore, 1982; Pickles, 1982). The fibers in the pathways undergo considerable amount of convergence and divergence at many stages (Levitt *et al.*, 1980).

2.3 Hearing impairment

Hearing loss may be divided into four types: conductive loss, sensorineural loss, central loss, and functional loss.

Conductive loss results, due to the defects of the ear canal, eardrum, or middle ear cavity. This results in less acoustic energy reaching the cochlea. It may be due to wax in the external auditory meatus, due to fluid, pus, or infections in the middle ear (otitis media), or due to otosclerosis disease (fixation of stapes in oval window). These conductive losses are greater for frequencies below 1 kHz (Wall, 1995; Webster, 1988). In conductive hearing loss, the hearing thresholds get increased, but frequency selectivity does not degrade (Zwicker and Schorn, 1978).

Sensorineural loss is caused by defects in cochlea (sensory), or auditory nerve (neural). Its causes include congenital or hereditary factors, diseases (e.g., Meniere's syndrome: destruction of organ of Corti), basilar membrane discontinuity, degeneration of neurons in the auditory nerve in ascending pathway, tumors, long term exposure to industrial noise, acoustic trauma, the action of the toxic agents, presbyacusis (degeneration of hair cells due to aging), interruption of blood supply to inner ear, viral infection spread from middle ear, etc. (Wall, 1995; Webster, 1988; Wilson, 1990). In general, this loss is not medically curable, and it becomes progressively worse with time. Two other phenomena associated with sensorineural loss are diplacusis and tinnitus. In case of diplacusis, a pure tone is perceived as sound with more than one pitch, or as a harsh or buzzing sound. Also, different pitches may be heard at the two ears for the same tone. Its causes may include mild injury to organ of Corti, or long-term exposure to very intense sound. Tinnitus, or ringing in ears is a commonly occurring auditory disorder. It may be caused by spontaneous discharge of hair cells or auditory nerve and may also be induced by exposure to intense sound. It occurs with many types of sensorineural impairment and can contribute significantly to the disruption of speech comprehension in severe cases (Levitt *et al.*, 1980; Pickles, 1982).

Central impairments are usually caused by damage to auditory cortex, inflammation of the membranes covering the brain and spinal cord (meningitis), skull trauma, or congenital defects. It may result in decreased speech comprehension ability, even though hearing thresholds may not have increased (Levitt *et al.*, 1980). Increase in the hearing threshold of one sound in the presence of another sound is known as masking. In central masking, the threshold of signal presented to one ear is raised due to a masking sound presented to another ear. This type of masking should arise beyond the peripheral auditory system, where the signals from the two ears get combined (Moore, 1982). The causes for the functional loss are psychological factors rather than physiological disorders.

In audiological clinics, the hearing threshold level (HTL) as a function of pure tone frequency, is routinely measured and plotted as audiogram. On the basis of hearing thresholds, the impairment can be graded into several categories like mild, moderate, severe, and profound (Wall, 1995).

2.4 Sensorineural loss

Sensorineural loss can be divided into two categories, sensory (i.e. cochlear) and neural (i.e. retro-cochlear). The cochlear loss could result due to a number of factors, and typically deviated audiogram configurations indicate the various pathologies (Webster, 1988; Biswas, 1995). Flat audiogram is noticed due to salicylate poisoning. Damage in the organ of Corti by ototoxic drug, stiffening of basilar membrane, and damage to hair cells and supporting Deiter's cells manifest as audiogram with hearing thresholds increasing with frequency. Ototoxic drugs, through metabolically created enzymes, damage the hair cells in the organ of Corti. A very high threshold at a particular frequency between 3-6 kHz usually suggests an acoustic trauma (loss of hair cells due to exposure to loud sounds). In case of early endolymphatic hydrops, hearing thresholds are found to be high at low tone frequencies. The hearing thresholds are relatively elevated for the frequencies in the 1-4kHz range in the congenital sensorineural damage. Various diseases like typhoid, meningitis, mumps, etc. indicate a peculiar pattern in audiogram, e.g., moderate to severe bilateral sensorineural loss is seen in typhoid. Retro-cochlear impairments result due to damage of auditory nerve fibers. Its causes may include tumor and hemorrhage (bleeding). Viral infection may results in primary degeneration of auditory neurons (relatively uncommon). Secondary degeneration involves the peripheral processes of auditory neurons constituting loss of hair cells, mostly in the first half of basal turn. Generally, neurons in apical turn are more likely to survive (Johnsson, 1985).

In addition to shift in the hearing thresholds, the sensorineural loss is characterized by reduction in dynamic range of hearing, degradation of temporal resolution and increase in temporal masking, and degraded frequency selectivity and increase in spectral masking. The loudness discomfort level (LDL) is the level at which a tone becomes uncomfortably loud. It is about 100-110 dB SPL for normal hearing persons, and it may be fairly similar for persons with sensorineural hearing loss (CHABA, 1991; Moore, 1982). The dynamic range is the difference between the loudness discomfort and hearing threshold levels. In conductive loss, the hearing thresholds and discomfort levels both get increased, with no significant change in dynamic range. In the case of sensorineural loss, the hearing thresholds increase, with no corresponding change in the discomfort levels, and thus the dynamic range may get drastically reduced. In such cases, the relationship between sound level and perceived loudness, known as "loudness growth curve," gets altered. This abnormal increase in perceived loudness with increase in sound level is known as "recruitment."

Temporal resolution is the minimum detectable gap between two successive signals. It is generally measured by means of gap detection method, in which the listener has to detect the gap between the two test signals (may be octave band noises) in the presence of continuous background broadband notched noise. The testing on normal and hearing impaired subjects indicated that, the temporal resolution for normal listener is 2-3 ms whereas, in case of sensorineural loss, it is reported to degrade up to 8 ms (Fitzgibbons and Wightman, 1982; Florentine and Buus, 1984).

Intense sounds have a masking effect on preceding and following weak sounds, known as backward and forward masking respectively. This phenomenon gets very severe in case of sensorineural impairment. Most forward and backward masking occurs within 100 ms either at onset, or ending of masking sound (Moore, 1982). The effect of forward masking is most effective within 10 ms of the masker and becomes insignificant after 100-200 ms. The backward masking at most extends over 20 ms before the masker. Both the masking effects exist due to temporal overlap of cochlear responses, in addition, the backward masking is associated with higher auditory processes (Fitzgibbons and Wightman, 1982; Moore, 1982; O'Shaughnessy, 1987).

Frequency discrimination refers to the ability to distinguish two tones presented one after another. In such a situation, the frequency difference limens (DL, or smallest detectable change in frequency) can be very small, possibly 0.2 % or 0.3%. The frequency discrimination possibly results due to detection of a shift in the neuron firing patterns (Pickles, 1982). The discrimination ability suffers in cases of sensorineural loss (Tyler *et al*., 1983).

Frequency selectivity refers to the ability to detect one frequency component in a signal consisting of many frequency components (Pickles, 1982). Fletcher (1938) suggested that the peripheral system consist of a bank of bandpass filters, referred as 'auditory filters.' In one of the techniques of characterizing the psychophysical auditory filters, a probe of fixed frequency (a tone) at low fixed intensity (10 dB above hearing threshold) is presented. The masker can be either a tone or a narrowband noise. For various frequencies of the masker around the probe, its intensity to just mask the probe is determined. Plot of the masker intensity versus frequency is known as psychophysical tuning curve of the auditory filter centered at the probe frequency.

A tuning curve resembles the magnitude response of a bandpass filter with a rounded top and sloping edges. The effective bandwidths of these tuning curves are known as 'critical bands' (CB), and these are much larger than the frequency difference limens for tone frequency discrimination (Pickles, 1982; Moore, 1982). The earlier estimates of the CB are between 15-20 % of the center frequency above 1 kHz and nearly constant below 500 Hz (Zwicker, 1961; Moore, 1982). Zwicker & Schron (1978) have reported similarity between psychoacoustical critical bands measured on humans and neurophysiologically measured tuning curves for animals.

The tuning curve shapes and CB estimates obtained by various researchers using different types of maskers and experimental techniques have differed somewhat. The problems could have been due to the effect of modulation between the probe tone and the masker. As a solution to this, Patterson (1976) described a method in which masker is a broad band noise with notch centered at the probe frequency. Fig. 2.6 shows the traditional value of the critical bandwidth and the more recent estimates, as obtained using Patterson's method, of the equivalent rectangular bandwidth of the auditory filter as a function of frequency (Moore, 1997). The bandwidths are between 11-17 % of the center frequency, and therefore for frequencies above 1 kHz these are quite close to the CB estimates obtained earlier. The values at the lower frequencies are smaller than the traditional estimates.

The shapes of the auditory filters are nearly symmetric at moderate sound level. These become asymmetric at high level, with shallower slope on the low frequency side. Sensorineural loss is associated with widening of the auditory filter bandwidths and the filter slopes becoming shallower (Tyler *et al.*, 1983; Glasberg and Moore, 1986; Dubno and Dirks, 1989; Moore, 1997).

Researcher have tried to understand the characteristics of sensorineural hearing loss in terms of the role of the hair cells in the transduction mechanism and role of auditory fibers for presenting the information to the higher brain processes. It has been reported that basilar membrane does exhibit sharp tuning curves for tone stimuli of different frequencies (Robles *et al.*, 1986; Moore, 1998). The basilar membrane vibrations at a particular location produce synchronized activity in the auditory nerve fibers innervating the corresponding hair cells. Inner hair cells (IHCs) act as the transducers for the vibrations. The outer hair cells (OHCs) 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. They also play a role in sharpening the tuning curves of the basilar membrane (Pickles, 1982; Guyton, 1996; Moore, 1998). Damage to IHCs reduces their transduction sensitivity for basilar membrane vibrations. Hearing thresholds get increased, but the loudness growth curve and the dynamic range do not get much affected. Frequency selectivity does not get very much degraded. Damage to the OHCs drastically impairs the "active" control role played by them. The sensitivity at low sound levels gets reduced, resulting in recruitment, i.e. loudness growth curve becomes more linear, and consequently dynamic range gets very much reduced. Further, the tuning curves become much broader resulting in severe spectral masking (Moore, 1998). Damage to the auditory nerve fibers alters the loudness growth curve, depending on the damage pattern. This damage generally reduces the frequency selectivity due to impairment in the coding of the basilar membrane vibrations (Pickles, 1982). The forward masking may take place due to reduction in sensitivity of recently stimulated cells. In case of backward masking, the processing may involve central activity in which the more intense signal (masker) may overtake the processing activity of earlier less intense signal (Moore, 1982).

2.5 Effects of sensorineural loss on speech perception

In sensorineural hearing loss, most apparent symptom is that in addition to hearing thresholds getting increased, the dynamic range for hearing gets reduced with an abnormal increase in perceived loudness with increase in sound level.

Normal conversational speech covers a range of at least 30 dB (Dunn and White, 1940) and is well within the dynamic range for normal hearing, but larger than the dynamic range of hearing in most cases of sensorineural loss. The amplification of low level sounds to make them audible causes intense sounds uncomfortably loud. The residual hearing area may get damaged due to continuous exposure to the sounds above the loudness discomfort level.

The degradation in temporal resolution and increase in temporal masking also severely affect speech reception. Proper reception of consonants requires adequate temporal resolution of subphonic segments like noise bursts and formant frequency transition. Further, in speech signals, vowels are generally more intense while the consonants carrying much of the information, have lower intensities (Crandall, 1917). Therefore, there is a possibility of masking of consonantal segment by vowels. Typically, stop bursts are 30 dB weaker than the following vowel and appearing in the range of backward masking, and hence may get masked, resulting in poor discrimination of stop consonants (Florentine and Buus, 1984).

Speech signal generally can be described as having a dynamically varying broad band spectrum. The important acoustic characteristics include the amplitude, frequency, and bandwidth of formants (spectral resonances specific to the vocal tract configuration, nature of excitation (voiced/unvoiced), pitch (fundamental frequency of vibration of vocal chords) in voiced segments, duration and frequency band of noise bursts, closure duration, and transition of formants. Psychoacoustic measurements of the smallest detectable change, or difference limen (DL), for some of these parameters have been reported (Flanagan, 1972). Flanagan has discussed the DLs for formant amplitude, frequency and bandwidths for synthesized vowels. The amplitude DL for first and second formants are 1.5 and 3 dB respectively. The DL for overall intensity of vowels and fricatives are about 1.5 and 0.4 dB respectively. The DL for synthetic vowel fundamental frequency is 0.3-0.5%. DL for formant frequency and bandwidths are 3-5% and 20-40% respectively.

Fricatives are characterized by broadband noise with spectral peaks and notches corresponding to the poles and zeros of the vocal tract filters. The detectable spectral peaks and notches should have Q's (ratio of center frequency to bandwidth) greater than 5 and 8 respectively. Normal listeners can identify vowels with spectral peaks only 2 dB above the spectral valleys, whereas moderate hearing impaired listeners needs at least 7 dB indicating the effect of spectral masking on perception. Though this is the case, the identification of vowels by hearing impaired is good because the vowels in the natural speech are characterized by spectral peak-to-valley differences of at least 8 to 10 dB (Leek *et al.*, 1987).

There may be uncertainty in the recognition of the transition of formants, and frequency band of the noise burst due to spectral masking. In the case of the unvoiced stops /p, t, k/ with vowel /*a*/, the high frequency noise of /t/ distinguishes it from the low frequency noise of /p/ and /k/. Stop /k/ is characterized by slightly higher frequency noise as compared to /p/, and therefore the discrimination will become difficult due to spectral masking. For the voiced stop consonants /b, d, g/, the second formant of vowel /*a*/ rises initially for /b/, and falls for /d/ or /g/. As the formant frequency is relatively clearly audible in case of /b/, it is less confused with /d/ or /g/, but the /d/ and /g/ are difficult to distinguish. As compared to stop consonants, the nasals /m, n/ are somewhat more intense and slightly longer, whereas the fricatives /s, z/ are intense and longest. The fricatives /s, z/ have energy concentration at high frequency as compared to fricatives /f, v/. The second formant transition at the end of the vowel has much influence on the discrimination of nasals /m, n/ (Miller and Nicely, 1955; Flanagan, 1972). From this discussion, one can say that the consonantal segments in which formant transitions and the noise bands are not widely different, will have perceptual confusions due to spectral masking.

The nasality and duration features are associated with amplitude/temporal cues. Voicing has periodicity/aperiodicity differences, whereas the feature of frication and place are associated with presence of aperiodic noise and spectral differences respectively (Tyler and Moore, 1992). Studies on spectral smearing has indicated that the poor speech reception by persons with cochlear damage may be due to reduced frequency selectivity (Moore, 1998). As the place feature is cued by spectral differences, the hearing impaired persons may find difficulties particularly in identifying this feature.

2.6 Speech processing schemes for hearing aids

Increase in hearing thresholds, reduction in the dynamic range, degradation of temporal resolution and increase in temporal masking, and degraded frequency selectivity and increase in spectral masking are the main characteristics of sensorineural impairments.

Since early times, sound amplification has been used as an aid in relieving the consequence of hearing impairment. The early aids included cupping one's hand behind the ear or the ear trumpet. Invention of thermionic valve and development of practical electronic amplifier assisted in producing the audiometer (the hearing loss measuring electronic instrument) and wearable electronic hearing aid having high gain. Hearing aids were among the first products incorporating the miniaturized vacuum tube, the transistor, integrated circuits, and miniature electret microphones (Levitt *et al.*, 1980). Typically, the conventional electroacoustic hearing aids incorporate a microphone, electronic filter, controls for adjusting the amplification and overall shape of the frequency response, circuits for limiting the amplified signal to a comfortable level, an earphone, and power source. In the past three decades, these aids have been greatly improved in a variety of ways. In addition to the body worn hearing aids, we have behind-the-ear and in-the-canal hearing aids. People suffering from hearing loss have difficulties in understanding the speech in noisy environment, and a lot of work has been done on reducing the background noise and improving intelligibility. The signal processing technique for reducing background noise include noise spectrum subtraction and adaptive noise cancellation (Levitt, 1994).

Frequency-selective amplification is primary aim of most hearing aids for overcoming the problem of elevated hearing thresholds. In linear amplification hearing aids with fixed gain irrespective of signal level, the gain can not be properly selected for fully compensating the hearing loss due to reduced dynamic range associated with loudness recruitment. To avoid this, most of the hearing aids incorporate some form of limiting of the output signal level, e.g. peak clipping, in which the signal is linearly amplified till a preset value and clipped after this value. Though an effective method, it adds unwanted distortion to the signal (Wall, 1995; Moore, 1998). In a hearing aid with automatic gain control (AGC) or "compressor," the amplification decreases with intensity, such that the wide dynamic range of input signal gets compressed in a smaller dynamic range of the output.

Compression-amplification can be classified into three types: compression limiting, automatic volume control (AVC), and syllabic compression (Levitt, 1994; Lunner, 1997; Moore, 1998). Compression limiting is for protecting the listener from intense sounds by rapid compression of low duration sounds (< 5 ms). Compression is relatively slow in AVC, such that long-term variations (> 200 ms) in input signal level above the compression threshold, produce relatively smaller variations in output signal. In syllabic compression system, the gain of the system varies with a response time of 20-100 ms i.e., a typical duration of syllable. Neuman *et al.* (1994) compared the three forms of compression on the basis of perceived sound quality. The study was conducted with sensorineural hearing impaired subjects (ten subjects with dynamic range less than or equal to 30 dB and ten with dynamic range greater than 30 dB). In this study, AVC has shown maximum potential as far as the matching of dynamic range of impaired auditory system is concerned. It produces least amount of distortion. However, very few commercial hearing aids incorporate this, may be because, in such an aid, a brief intense signal decreases the gain which remains low for some time thereafter and during this interval the hearing aid appears as inactive (Moore, 1998).

The compression-amplification may be single channel or multi-channel. In single channel compression devices, the total signal power can be used for determining the gain of amplifier. In these devices, when the low frequency components of speech signal are relatively intense, the level of important high frequency cues of low amplitude gets reduced (Villchur, 1973). In multi-channel (multiband) compressors, the signal is filtered into number of bands, and the power in a band is used to determine the compression ratio for the particular band; the output signal being sum of signals from all the bands. Some studies have reported improvement in intelligibility with multi-channel compression aids (Levitt, 1994). In multi-band compression, the number of bands up to two or three may be sufficient to make up for recruitment. Since the multi-band compression reduces the spectral contrast in complex stimuli, use of large number of bands for persons with reduced frequency selectivity may not be helpful. Number of bands up to eight may be beneficial to a small extent (Moore, 1997, Section 2A).

In a study by Asano *et al.* (1991), octave band spectrum over 250 Hz - 8 kHz range was calculated by using FFT for 8 ms segments of input signal, and this spectrum was used for calculating the frequency-gain characteristics by considering the loudness compensation function (relation between the loudness for normal subjects and that for hearing impaired). A digital filter with sampled frequency response as the calculated frequency-gain characteristics was used for processing the input segment. The experimental evaluation based on this scheme, indicated an improvement in recognition score (as compared to linear amplification) for 9 of the 13 sensorineural hearing impaired subjects, over wide range of input levels.

Tejero *et al.* (1995) used a different technique for multiband compression. Short-time segments of length 25 ms with an overlap of 5 ms were used for calculating magnitude and phase spectra, using FFT. The magnitude spectrum was modified in accordance with loudness compensation function. By taking inverse FFT of the new spectrum (modified magnitude spectrum and original phase spectrum) for overlapping segments, output signal was synthesized. This technique was implemented in real-time using TMS320C30 DSP based board and a PC. The scheme was tested on ten hearing impaired subjects, using a phonetically balanced 25-word list, at different levels of input signal. Comparing the test scores obtained with and without the use of the compression, all subjects benefited from compression and about half the subjects were able to get 100% score. Tejero *et al.* have pointed out the possibility of confusions among phonemes (e.g., /p/ versus /t/, /m/ versus /n/), but a detailed study of confusion pattern was not carried out by them.

In sensorineural loss, it is quite common to find some hearing in low frequency (up to 1 or 2 kHz). The high frequency inaudible speech energy can be transposed into the low frequency region as distinct audible cue. Johannson (1966) experimentally evaluated frequency transposer system on profoundly hearing impaired subjects and found improvement in discrimination of fricatives and some other phonemes. Pandey *et al.* (1987) tested the speech processor providing fricative and low frequency periodicity information for single channel cochlear prosthesis. They found that, provision of high-frequency fricative noise burst information didn't adversely affect the reception of voicing (provided with low frequency periodicity) and place (mainly through lipreading) information. Many other studies have been conducted for evaluating the effectiveness of enhancing specific features of speech signal. The technique of speech processing using the properties of "clear" speech is discussed in the following section.

2.7 Speech processing using the properties of "clear" speech

Speech used in normal, everyday situations is known as conversational speech. Speech for improving communication in difficult situations, like speaking in noisy environment or to a hearing impaired person, is called "clear" speech (Picheny 1981). The clarity in the speech may be introduced not only by changing the conversational context, sentence structure, vocabulary, speaking rate and stress, and pronunciations of individual words but also by bringing about appropriate changes in vocal efforts. Picheny *et al.* (1986) studied the acoustic characteristics associated with clear speech and conducted experiments for evaluating improvement in speech intelligibility by the use of clear speech for the hearing impaired. In clear speech, the speaking rate decreases; the increase in duration is contributed by increase in pauses as well as by increase in the duration of some phonemes. In conversational speech, vowels get very often reduced and stop bursts may not get released. These do not happen in case of clear speech. Further, in clear speech, acoustic segments corresponding to consonantal duration and consonant-to-vowel intensity ratio increase, leading to an improvement in speech quality. The testing was done on five persons with sensorineural hearing loss using nonsense clear and conversational sentences, the average increase in scores for clear speech was 17%.

In general, consonant-to-vowel (C/V) intensity ratio refers to the difference of power or energy in decibels of the consonant and that of the neighboring (preceding or following) vowel (Freyman and Nerbonne, 1989). In the experiments reported by Picheny *et al.* (1985, 1986), the speech of the talker who was most easily understood by five hearing-impaired listeners, had the highest average consonant amplitude. About 10 to 12 percent improvement in intelligibility, for consonants amplified by 10 to 21.5 dB above their natural level relative to vowels in a word, were reported in a study carried out with subjects having mild-to-moderate hearing impairment (Montgomery and Edge, 1988). In another study, the listening tests carried out with young and elderly normal hearing subjects, in the presence of background noise, also showed almost similar improvement in intelligibility (Gordon-Salant, 1986).

Freyman and Nerbonne (1991) have studied the effect of C/V ratio modification in normal hearing subjects with simulated profound hearing loss. Profound hearing loss was simulated by using pink noise. In each utterance, the consonant part was amplified by 10 dB. The results indicated an improvement in recognition performance for stops and unvoiced fricatives, whereas it was not significant for affricates and nasals. Further, recognition performance was reduced for glides and fricatives.

Thomas *et al.* (1996) conducted experiments for studying the effect of increasing C/V ratio (CVR) with synthesized test stimuli sets involving three unvoiced stop consonants with cardinal vowels /*a,* i, u/ and six stop consonants with vowel /*a*/ in the CV (consonants-vowel) and VC (vowel-consonants) context, with CVR modifications of 9 and 6 dB respectively. For simulating hearing impairment, each stimulus was mixed with synthesized broad band noise under three SNR conditions of no masking, 12 dB SNR, and 6 dB SNR. In presence of masking broad band noise, the recognition score was found to increase with increasing CVR. Recognition scores were higher for VC than CV position for all CVRs. From this, it appears that increasing CVR suppresses forward masking of the consonant by vowel more effectively than backward masking. Some vowel confusions are observed in the vowel-consonant (VC) context at higher CVRs, but up to 10 dB modification in CVR, there is no adverse effect on the recognition of vowels. The subjects' average response times were found to improve with increasing CVR.

Fairbanks *et al.* (1954) introduced time compression for increasing speaking rate by periodically discarding intervals of speech of approximately 10-30 ms duration or time expansion for decreasing speaking rate by periodically repeating 10-30 ms segment of speech. In tests conducted on normal hearing listeners, trained for 50 words vocabulary, the compression and expansion did not affect intelligibility. But in case of elderly hearing-impaired listeners, there was a decrease in intelligibility scores for both time compression and time expansion. Montgomery and Edge (1988) studied the effect of consonant lengthening by 30 ms, with and without consonant amplification for a group of hearing impaired subjects. Results indicated 5 % improvement in stimuli recognition at 95 dB SPL, whereas at 65 dB SPL the hearing impaired subjects didn't show the benefits of the scheme. Further, the indeterminate effect of consonant lengthening on the consonant recognition performance, might be due to subject background, stimuli differences, lengthening of consonants, and signal processing effect, etc.

In a study for testing the different aspects of increasing consonant duration, Thomas (1996) synthesized six stop consonants in the CV context of the vowel /*a*/ by modifying burst duration, formant transition duration, and voice onset time. Using three stimuli, he conducted the study on four normal-hearing subjects with simulated hearing loss. Each stimulus was mixed with synthesized broad band noise under three SNR conditions (i.e., ∞, 12, and 6 dB). The tests were conducted for unprocessed speech and processed speech by keeping the presentation level to the subject's comfortable listening level. It was observed that, at higher noise levels, for obtaining better performance a formant transition duration modification of up to 50 % may be combined with the burst duration. Voice onset time did not appear to be a suitable parameter for improving the performance.

2.8 Binaural dichotic presentation

Speech processing based on the properties of clear speech, as reviewed in the previous section, helps in reducing the effect of temporal masking. Another factor responsible for poor speech reception in cases of sensorineural hearing impairment is the decrease in frequency resolving capacity of the ear due to the spread of spectral masking along the cochlear partition, as discussed in Section 2.4.

Most of the speech processing techniques for hearing aids involve monaural listening which refers to sound presentation to one ear only, whereas binaural listening involves both the ears. Binaural listening could be "diotic" with the same signal presented to both the ears or it could be "dichotic" with different signals presented to the two ears. (Moore, 1982; Duda, 1996). Binaural listening offers better overall sound quality, clear speech intelligibility and more relaxed listening, and it helps in source localization (Pickles, 1982; Moore, 1982; Kollmeier *et al.*, 1992).

The phenomena of spectral masking is considered to be mainly at the cochlear level. It may be possible to improve speech reception, by filtering the speech signal using a bank of critical band filters and adding outputs from alternate bands to obtain two signals with complementary spectra for binaural dichotic presentation. The next chapter reviews earlier studies and outlines a scheme for reducing the effect of spectral masking. Implementation and testing of this scheme are presented in the subsequent chapters.

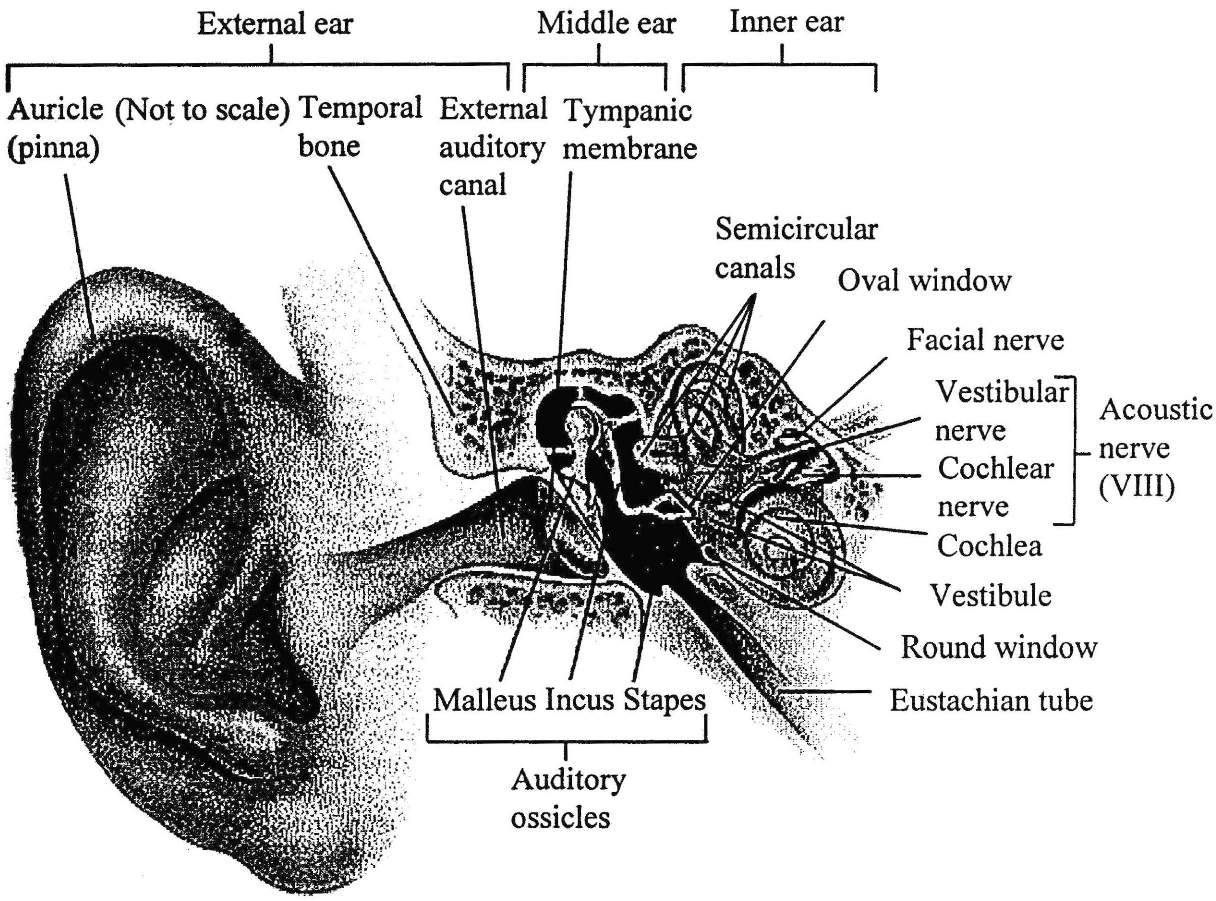


FIG. 2.1 Illustration of the structure of the peripheral auditory system showing the external, middle, and inner ear. Source: Thibodeau (1992), Fig. 8.3.

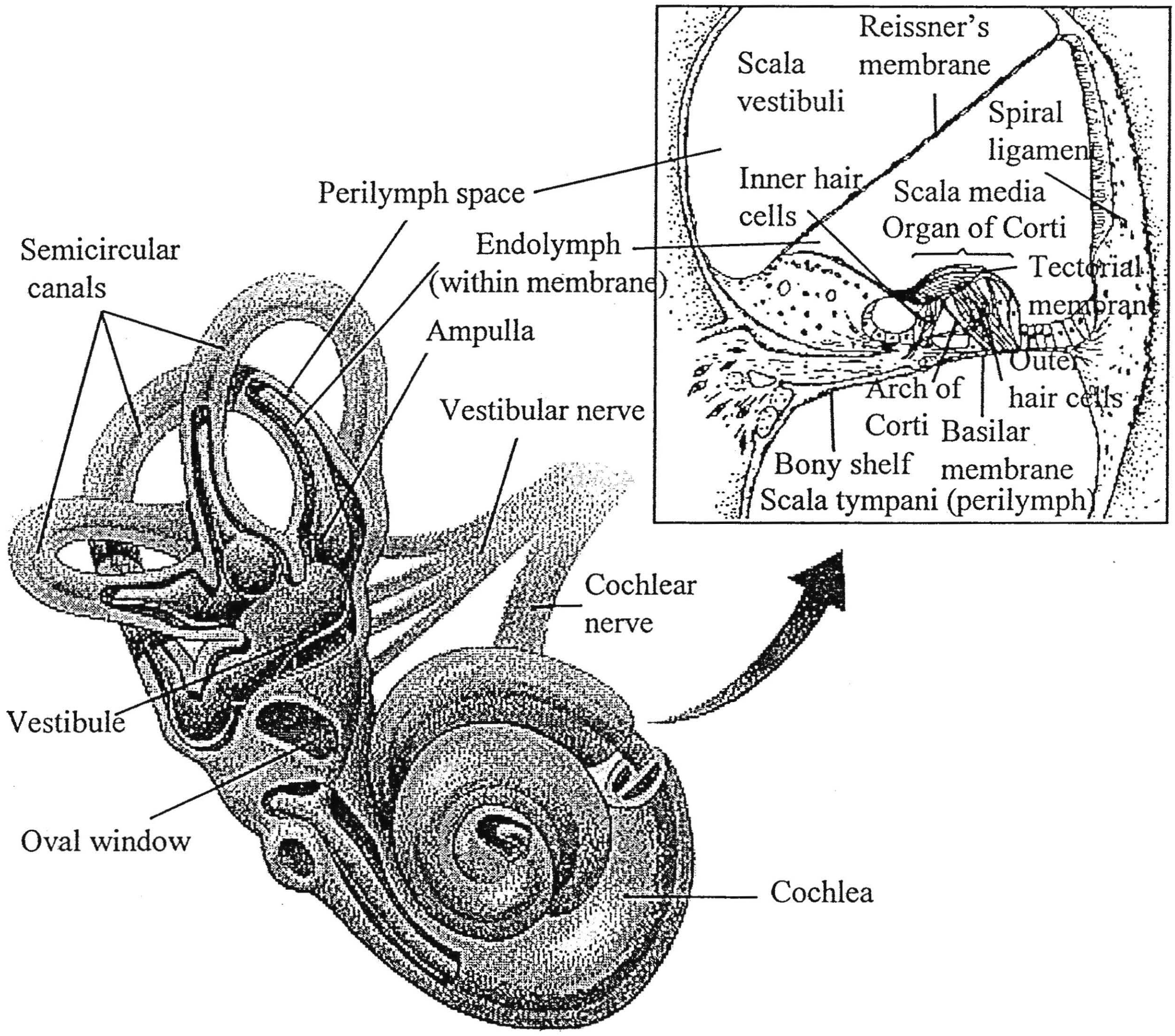


FIG. 2.2 The inner ear and transverse section of cochlea. Adapted from Thibodeau (1992), Fig. 8.4 and Levitt et al (1980), Fig 2-b.

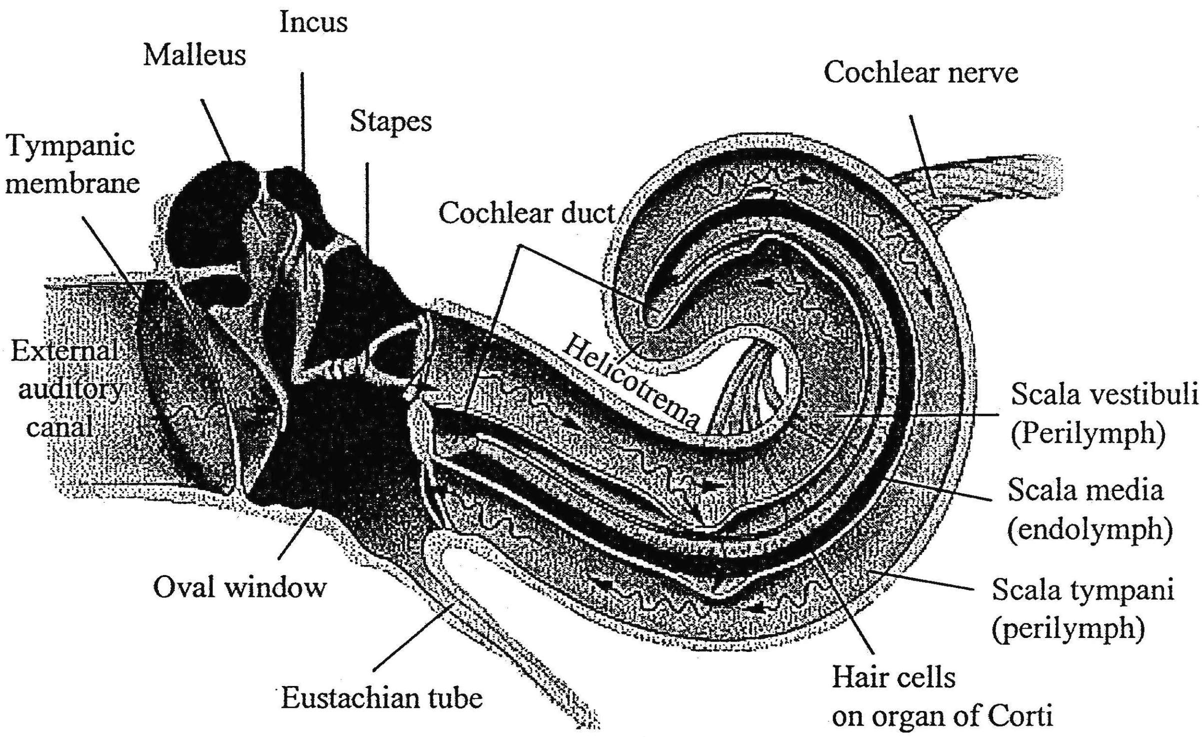


FIG. 2.3 Transmission of sound waves through cochlear longitudinal cross section. Adapted from Thibodeau (1992), Fig. 8.5.

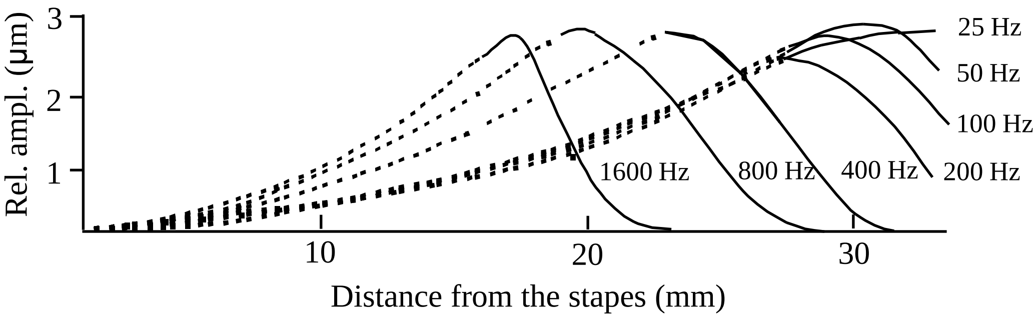


FIG 2.4 Vibrational pattern of basilar membrane. Adapted from Moore (1982), Fig. 1.8.

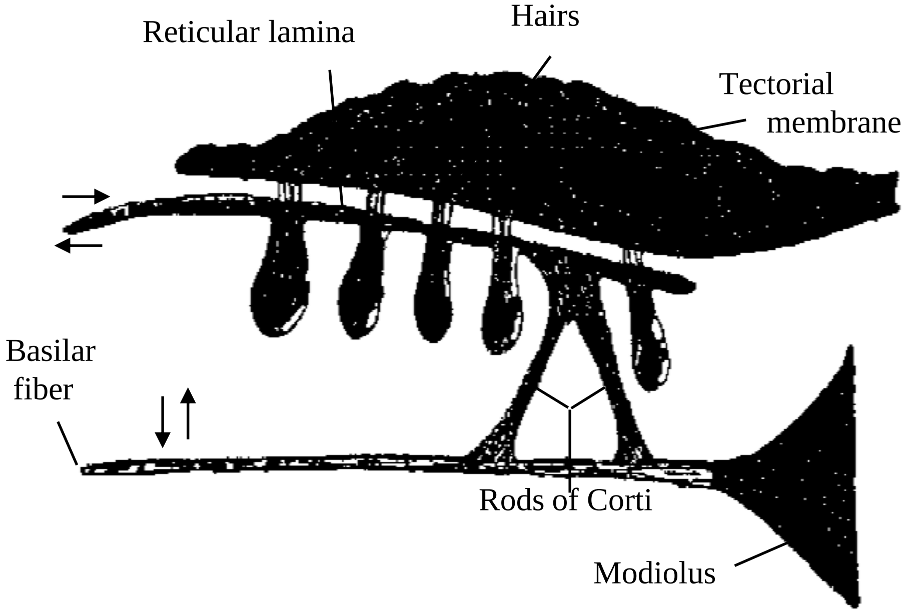


FIG. 2.5 Stimulation of hair cells by up-down movement of the basilar membrane. Source: Guyton (1986), Fig. 61-8.

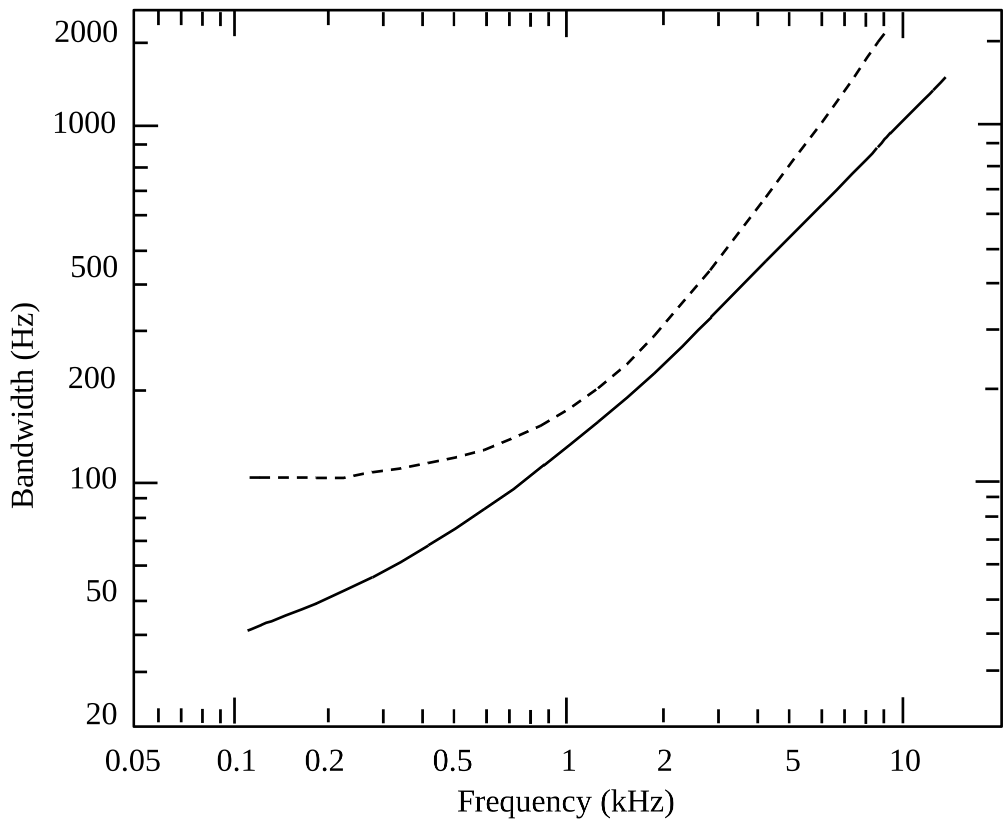


FIG. 2.6 Critical bandwidth as a function of center frequency of auditory filter. The dashed curve shows the traditional estimate (Zwicker, 1961), the solid curve gives more recently obtained estimate of equivalent rectangular bandwidth. Adapted from Moore 1997 Fig. 3.10.

Chapter 3

BINAURAL DICHOTIC PRESENTATION

3.1 Introduction

Sensorineural hearing impairment is characterized by increase in thresholds of hearing, reduction in dynamic range, degradation of temporal resolution and increase in temporal masking, and degradation of frequency selectivity due to increase in spectral masking. Some of the studies for solving these problems, particularly those due to reduction in dynamic range, degradation of temporal resolution and increase in temporal masking have been reviewed in the previous chapter.

The loss of frequency selectivity due to broadening of critical bands results in poor speech perception. Several researchers have studied the effectiveness of splitting of the speech signal on the basis of frequency bands, for binaural presentation, in order to reduce the effect of spectral masking (Lyregaard, 1982; Ho, 1987; Lunner *et al.*, 1993; Mithal, 1996). In this chapter, after reviewing these studies, a scheme is proposed for investigation. This is followed by discussion of the evaluation techniques. Implementation of the scheme, experiments for evaluation, and discussion of the test results are presented in the two following chapters.

3.2 Binaural hearing aids

The ability to perceptually combine the binaurally received signals from the two ears improves speech reception under adverse listening conditions. Binaural listening offers better overall sound quality, clear speech intelligibility, more relaxed listening, and it helps in source localization (Moore, 1982; Pickles, 1982; Kollmeier *et al.*, 1992). Many studies have reported improvement in recognition score with the use of binaural hearing aids under different listening conditions (Dirks and Carhart, 1962; Cox and Bisset, 1984; Hawkins and Yacullo, 1984, Zelnick, 1970; Nebelek and Robinson, 1982; Jerger *et al.*, 1961; Nebelek and Pickett, 1974; Moncur and Dirks, 1967).

In a study by Dirks and Carhart (1962), most of the surveyed hearing aid users indicated a liking for binaural hearing aid over monaural hearing aid under various listening situations. Balfour and Hawkins (1992) conducted experiments involving a paired comparison to elicit sound quality judgments, for monaural/binaural hearing aid processed signals, under several listening conditions. Fifteen adult subjects, with bilaterally symmetrical mild and/or moderate sensorineural hearing loss, participated in these experiments and the results indicated a preference for all perceptual dimensions of sound quality. A study by DeSimio *et al.* (1996) indicated that the binaural system offers more perceptual robustness in noise than the monaural one. The binaural performance at 15 dB SNR was found approximately equal to the monaural performance at 20 dB SNR.

Franklin (1981) studied the use of two types of binaural hearing aids: (i) extended low frequency responses in both ears, (ii) high frequency emphasis in one hearing aid and low frequency emphasis in the other. The scheme was tested on infants (in the age group of 6-28 months) and the results were compared on the basis of hearing thresholds. The results indicated hearing threshold improvement of 12-18 dB in high-low emphasized hearing aid fitting over the extended low frequency one. The author has argued that the information coming from both the ears gets combined at higher levels of auditory processes, and that it would be better to put low frequency information in the left ear, possibly due to superiority of the right hemisphere of the brain to suprasegmental (i.e. stress, rhythm, and intonation) features, and high frequency information in the right ear, since the left hemisphere appears to be dominant in discriminating consonants.

3.3 Review of dichotic presentation

One of the possible ways of improving speech perception degraded due to loss of frequency selectivity (caused by masking at the peripheral level), would be to split the speech signal on the basis of its short time spectrum into two complementary spectra for presentation to the two ears. The splitting of the information should be done in such a way that the two adjacent strong spectral components that are likely to mask each other, get presented to different ears. Several schemes may be employed for splitting speech for binaural dichotic presentation.

Considering the perception of formants to be really important, a processing system can be devised such that alternate formants are presented to different ears, and therefore do not contribute to masking of each other. Ho (1987) carried out listening tests using synthesized vowels, in which, first formant is presented to one ear and the second, to the other. It was observed that even though the two formants are presented to different ears, they result in proper perception of the vowels, possibly due to fusion of the information at higher levels in the auditory system. In order to implement and test this scheme for speech, a formant based analysis/synthesis system is required.

In one of the earliest reported work on dichotic presentation, Lyregaard (1982) studied a speech processing scheme to improve upon speech discrimination in noise due to reduced frequency selectivity. The scheme used comb filters, realized using an analog delay. The filtering set-up can be schematically represented as shown in Fig. 3.1. The bandwidth of the comb filter can be varied by adjusting the delay τ. Experiments were conducted with three values of the bandwidth: 200, 500, and 800 Hz. Three subjects with binaural hearing loss of 50 dB and two normal hearing subjects participated. The listening tests were conducted with dichotic presentation as well as diotic (unfiltered signals to both the ears). The test material consisted of two lists of 25 words each (presumably one list used for dichotic, and the other for diotic listening). Subject's responses were used for obtaining phoneme intelligibility scores. The listening level was set by the test subject. Speech intelligibility was measured by presenting the test material with background noise (speechspectrum shaped random noise) at signal-to-noise ratios of 12 and 4 dB. The measurements were made for diotic and digotic listening conditions. From the arctangent transformed scores for these two SNR's,the SNR value corresponding to 50% score was obtain by interpolation. The improvement in the percentage phoneme intelligibility score was averaged across the subjects. Improvements in the scores were not statistically significant, for all the three filter bandwidths, for both the normal as well the hearing-impaired subjects. Lyregaard suggested that the lack of significant improvement could be attributable to one of the three factors: unsuitable filtering, insufficient listening experience by the subjects, and non-feasibility of binaural fusion of dichotic signals.

Lunner *et al.* (1993) tested the use of an 8-channel digital filter bank in monaural, diotic, and dichotic modes. The filter bank was designed to give 8 parallel filtered outputs which are added together with individually adjustable weighting factors in order to obtain a proper fit of the gain frequency response of the hearing aid to the need of the individual hearing aid user. By combining alternate bands together, the filter bank was used for dichotic presentation, as shown in Fig. 3.2. All the filters in the filter bank had bandwidth of 700 Hz. The filter bank was realized using complementary interpolated linear phase FIR filters, in order to minimize the number of multiplication. The scheme was implemented in real time using TI/TMS320C25 digital signal processor based board and a PC. The signal sampling rate was 11.6 k samples/s and signal processing delay was about 4 ms. The processing system was later rebuilt using the processor TI/TMS320E25 and incorporated as part of a pocket type hearing aid. The listening tests were conducted under three conditions: (1) dichotic presentation of odd numbered bands to the left ear and even numbered bands to the other, (2) dichotic presentation of even numbered bands to the left ear and odd numbered bands to the other, (3) diotic presentation in which all the bands are presented to both the ears. Three subjects (age: 39-69 years) with bilateral, symmetrical, moderate sensorineural hearing loss with audiogram indicating hearing thresholds increasing with frequency for one subject and elevated thresholds for the frequencies in the range of 1-4 kHz for the other two, participated in the experiments. The gain of the filters was adjusted depending on the hearing loss configuration of an individual.

The experimental evaluation of speech recognition in presence of noise was done by finding the speech-to-noise ratio which satisfy the 50% correct word recognition (a method similar to that used by Lyregaard, 1982). The test material consisted of a list of ten low-redundancy five word sentences. The results indicated an overall improvement in speech-to-noise ratio of about 2 dB for the dichotic conditions over diotic. The speech-to-noise ratio obtained was almost same in case of both dichotic conditions. Further, they also studied the combined effect of temporal and spectral splitting of speech for one subject, by using symmetrical inter-aural switching of odd and even bands with a period of 20 ms. No additional improvement were observed due to temporal splitting.

In a preliminary study, Mithal (1996) conducted experiments involving splitting of speech into two signals on the basis of multiband filtering, using filters with 1/3 octave bandwidths as an approximation to critical band filtering. The scheme was implemented in off-line. The signal sampling rate was 10 k samples/s. The listening tests were conducted under two conditions: (1) diotic presentation in which all the bands are presented to both the ears and (2) dichotic presentation of even numbered bands to the left ear and odd numbered bands to the right, and vice versa. Five normal hearing subjects with simulated hearing loss participated in the experiments. The hearing loss was simulated using white noise, band limited to 4.6 kHz, as an additive masker with different S/N ratios. Listening test material consisted of 12 English consonants with the vowel /*a*/ in VCV context. The improvement in recognition scores due to processing for dichotic presentation was in the range of 1.3-3.7 % and 3-4.7 % for SNR of 3 and 6 dB respectively.

3.4 Proposed scheme

The studies, reviewed in the preceding section, on splitting speech for dichotic presentation on the basis of frequency have used absolute and relative bandwidths for filtering. One of the earliest reported studies of splitting speech on the basis of frequency (Lyregaard, 1982) did not find improvement in speech intelligibility with constant bandwidth comb filters implemented using time delay and analog circuits. He indicated that the lack of improvement was possibly due to unsuitable filtering, lack of subject's listening experience, and lack of the binaural fusion of dichotic signals. However, later studies (e.g., Franklin, 1981; Ho, 1987) have indicated perceptual fusion of dichotic signals. Further, Lunner *et al.* (1993) found a modest improvement in speech perception by splitting speech spectrally. They implemented the scheme using digital filtering with an efficient realization of constant bandwidth (700 Hz) filters.

The main objective of spectral splitting for dichotic presentation should be to enhance the perception of spectral contrast of resonance peaks without adversely affecting the perception of features cued by amplitude and duration. Although use of constant bandwidth filters can be exploited for very efficient realization, a single bandwidth value may not be optimal for reducing the effect of spectral masking over the entire frequency range of speech signal. Too wide a band will not resolve the lower formants for presentation to different ears. If the bandwidth is too narrow, the pitch harmonics under the same formant peak are likely to be presented to different ears, resulting in reduced perception of spectral contrast. Narrow bandwidths are also likely to smear timing related cues. Critical bands corresponding to auditory filter bandwidths as estimated from psychophysical tuning curves (for normal hearing subjects) may be appropriate for separating the adjacent formants to different ears, without separating the pitch harmonics under the same formant. Traditional estimates of critical bandwidths as given by Zwicker (1961) and more recently obtained estimates of auditory filter bandwidths (Moore, 1997) have been shown earlier in Fig. 2.6. In the 1-5 kHz range, both the estimates are in agreement and are about 15-17 % of the center frequency. Below 500 Hz, the bandwidth estimates given by Zwicker are nearly constant at about 100 Hz while the more recent estimates are in the range of 40-80 Hz. However, use of bandwidth values less than 100 Hz may result in reduced spectral contrast and smearing of timing related cues. Therefore we decided to use critical band estimates as given by Zwicker in our scheme for spectral splitting of speech signal, for dichotic presentation, so that frequency components in alternate bands are presented to the different ears. The scheme uses 18 bands over frequency range of 5 kHz, with the magnitude response of each band being an ideal filter approximation with bandwidth of the critical band.

Advantages of frequency dependent gain as well as multiband compression have been well established (as discussed in Section 2.6) and these can be combined with processing for dichotic presentation. However, in the present study, dynamic range compression has not been used. The spectral splitting for binaural dichotic presentation has been implemented in two ways. In the first implementation, all the bands have the same gains. In second implementation, the gains for the bands are adjustable in the range of -3 to +3 dB, as a way of partial matching of the filter response to the frequency characteristics of individual subject's hearing loss. The objective of evaluating the two implementations is to first study the effect of processing for binaural dichotic presentation with flat frequency response and then to study additional benefits of a frequency response selected in accordance with the individual's hearing loss characteristics. The scheme for setting the frequency response is given later in Section 5.2 (on p.5.2 and in Fig.5.2).

It will be desirable to preserve the relative phases of the frequency components in speech signal so that timing related cues are not affected. Consequently, the magnitude response based on critical band filtering should be coupled with a linear phase response, and hence the filter realization should make use of FIR filter with symmetric impulse response (Rabiner and Schafer, 1978; Kuc, 1982; Proakis and Manolakis, 1997).

The effectiveness of the scheme was studied by testing its off-line processing implementation and testing its real-time processing implementation, using normal hearing subjects and subjects with bilateral sensorineural hearing impairment. For off-line processing, the efficiency of filter realization is not critical. Efficient realization for real-time implementation can be obtained, if instead of using a bank of filters and combining their outputs, we use a single FIR filter for approximating the overall magnitude response coupled with linear phase for each channel.

The following section deals with the evaluation methods, to be used for the processing scheme. The implementation and test results are presented in the following two chapters.

3.5 Evaluation methods

For evaluating the performance of the speech processing scheme used for hearing aids, researchers have employed intelligibility test and judgment of perceived sound quality (Dirks and Carhart, 1962; Harris and Goldstein, 1979; Gabrielsson *et al*., 1988). In speech intelligibility test, subject listens a list of standard words (e.g. spondees, phonetically balanced words, Central Institute for the Deaf (CID) word list) and the number of correct responses is noted. For perceived sound quality judgments, clarity, loudness, etc. are taken into account. The speech intelligibility test is well established, but for comparing the hearing aids, judgment of sound quality has also been used (Gabrielson *et al*., 1988). DeSimio *et al.* (1996) compared binaural and monaural hearing aids for finding the perceptual robustness in noise. They plotted the recognition scores versus SNR (15, 20, 25, and ∞ dB) for both types of hearing aids and then compared their performance. Some researchers (Lyregaard, 1982; Hawkins and Yacullo, 1984; Gabrielsson *et al*., 1988; Lunner *et al*., 1993) have employed an evaluation scheme in which the processed speech is mixed with noise and listening tests are carried out to find the speech-to-noise ratios for 50% correct word recognition. The processing and presentation methods are compared on the basis of SNR required for 50 % scores. Lunner *et al*. (1993) used this method with a variation of SNR in 3 dB steps.

In using measure of speech quality; there is often difficulty in explaining the judgment parameters, particularly to the hearing impaired patients, and consequently it becomes difficult to compile and compare the test results across subjects. Hence it was decided to ask subjects only to make a comparative assessment of the overall quality of the processed and unprocessed speech, without specifying any judgment parameter.

Intelligibility scores are measured using various types of test material. List of sentences and words are preferred in audiological practice, because these do not require as much concentration on the part of the patient, as the nonsense syllables do. However, it becomes very difficult to analyze such intelligibility scores for gating an insight into the causes for success and failure of the hearing aids. Hence, it was decided to use a list of nonsense syllables that will minimize the contribution of linguistic factors and maximize the contribution of acoustic factors.

The intelligibility scores do not provide an insight into the features responsible for the relative improvement. Miller and Nicely (1955) conducted listening tests with 16 syllables involving different consonants in consonant-vowel context and scored subject responses in the form of stimulus-response confusion matrices. The cell entries were converted to stimulus-response confusion probabilities, and these probabilities were subjected to information transmission analysis to study the effect of various features. Recognition or articulation score is probability of correct responses, and can be obtained as the sum of probabilities in the diagonal cells. The recognition score might be influenced by the subject's response bias. Information transmission analysis used by Miller and Nicely (1955) furnishes a measure of covariance between stimuli and responses, employing mean logarithmic probability (MLP) measure of information. The information transmission analysis is described in Appendix E. This analysis has been used in a number of studies (Dowell *et al.*, 1982; Pandey, 1987; Christopher *et al.*, 1987; Tyler and Moore, 1992; Hou and Pavlovic, 1994; Tye-Murray *et al.*, 1995; Thomas, 1996; Skinner *et al.*, 1997; Dorman *et al.*, 1997; Fu *et al.*, 1998).

In addition to the information available from stimulus-response confusion matrices, the response time statistics could be used for comparing the processing and presentation schemes, in terms of load on perception process. Even if two schemes result in same recognition score or information transmitted, the scheme that requires less response time can be considered as superior. This method has been used by Pandey (1987) and Thomas *et al*. (1996) but is not an established technique.

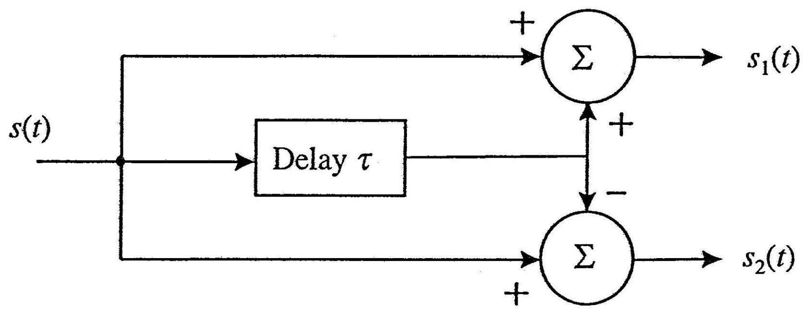
For the evaluation of benefits of speech processing for dichotic presentation, as discussed in the previous section, we will use stimulus-response confusion matrix for the close set of speech stimuli and also obtain the response time statistics. The matrices will be analyzed for obtaining recognition scores and information transmission. The cell entries in the matrix can be used to obtain matrices by grouping stimuli with the same feature, in order to study the contribution of various speech features. The scheme to be studied is aimed at reducing the effect of spectral smearing due to loss of spectral resolution. Hence it should result in improvement in the reception of "place" feature without adversely affecting the reception of other features. For this purpose the set of stimuli should be appropriately chosen. We have decided to use nonsense syllables stimuli with twelve consonants /p, b, t, d, k, g, m, n, s, z, f, v/. Two sets of stimuli where formed with these consonents and the vowel /*a*/ in vowel-consonant-vowel (VCV) context and consonant-vowel (CV) context. These can be used for studying the reception of consonantal features of voicing, place, manner, nasality, friction and duration (Miller and Nicely, 1955; Rabiner and Schafar, 1978; Ladefoged, 1982). The features groupings are given in Table E.1 in Appendix E dealing with information transmission analysis. A brief description of program used for computing recognition scores, information transmission analysis, and a sample test result analysis is given in Appendix F.

The listening tests for obtaining confusion matrix are very time consuming. Therefore a computerized test administration set-up has been developed. There are difficulties in having large number of hearing impaired subjects with bilateral hearing loss available and willing to participate in these experiments. And in any case the pattern of hearing loss varies across the subjects. Therefore, it was decided to conduct two set of experiments. In the first set, normal hearing impaired subjects participated and sensorineural hearing loss was simulated by a masking noise. In the second set of experiments, subjects with bilateral hearing loss participated.

Sensorineural loss is associated with hearing thresholds elevation, reduced dynamic range, and degraded temporal and frequency resolution. For studying the effectiveness of speech processing schemes, simulation of the effect of specific aspect of hearing impairment in normal hearing subjects helps in determining the effect of each of the above variables separately on speech intelligibility. In a simulation for the effect of reduced temporal resolution, developed by Hou and Pavlovic (1994), the signal is filtered by a bank of critical band (auditory) filters. Subsequently, each output signal intensity envelope is convolved with a resultant temporal window (a resultant temporal window is obtained by convolving smeared temporal window with temporal window of normal auditory system), and then the signals are added. The effect of reduced frequency resolution can be implemented by filtering the signal into bands and multiplying each band by a bandpass noise. The grading of the hearing loss is determined by the width of the bandpass noise (Villchur, 1977). In another study, it was implemented by taking the short-time fast Fourier transform of the signal segment followed by modifying the spectra and then getting the signal by applying inverse fast Fourier transform (Keur *et al.*, 1992).

The established practice for simulating sensorineural loss in the normal hearing subject is to present the stimuli in the presence of background noise. Leek *et al.* (1987) used broadband noise to simulate moderate hearing loss (about 72-75 dB of noise) in normal hearing listeners. Leek and Summers (1996) tested normal hearing listeners in the presence of low pass filtered Gaussian noise. Fletcher (1952) reported that the process of masking responsible for threshold elevation is of cochlear origin. Addition of broadband noise elevates the hearing threshold (DeGennaro *et al*., 1981; Jestead, 1997). These studies indicate that different methods can be used for simulating various aspects of sensorineural hearing loss in normal hearing subjects. We decided to use Gaussian noise band limited to the band of the speech signal as masking noise with different signal-to-noise ratios for varying the severity of simulated loss. The noise was added in such a way that signal-to-noise ratio was kept constant on the basis of short time (≈10 ms) energy of the signal. Thus during silence segments, there would not be any background noise.

The effect of presentation level on consonant identification for normal hearing subjects have been studied by several investigators (Simon, 1978; Dorman and Dougherty, 1981). They noticed that identification performance declines at very low (< 35 dB SPL) and very high (> 90 dB SPL) presentation levels. In one of the studies, the stimuli were presented at 75, 80 or 85 dB SPL (Leek and Summers, 1996; Skinner *et al*., 1997). In this study we decided to use presentation level as the most comfortable listening level of the individual subject. The level was selected by the subject for each listening condition and was kept constant for all the tests under a particular listening condition.



$$|H_1(f)| = 2\,|\cos(\pi f \tau)|$$
$$|H_2(f)| = 2\,|\sin(\pi f \tau)|$$

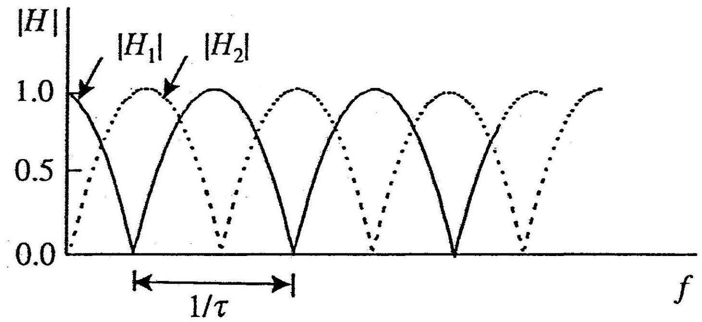


FIG. 3.1 A schematic of the comb-filtering scheme and magnitude response of the filter used by Lyregaard (1982). Waveform $s(t)$ is the input and $s_1(t)$, $s_2(t)$ are the output waveform for presenting to the two ears. $|H_1(f)|$ and $|H_2(f)|$ are the magnitude responses for the two channels.

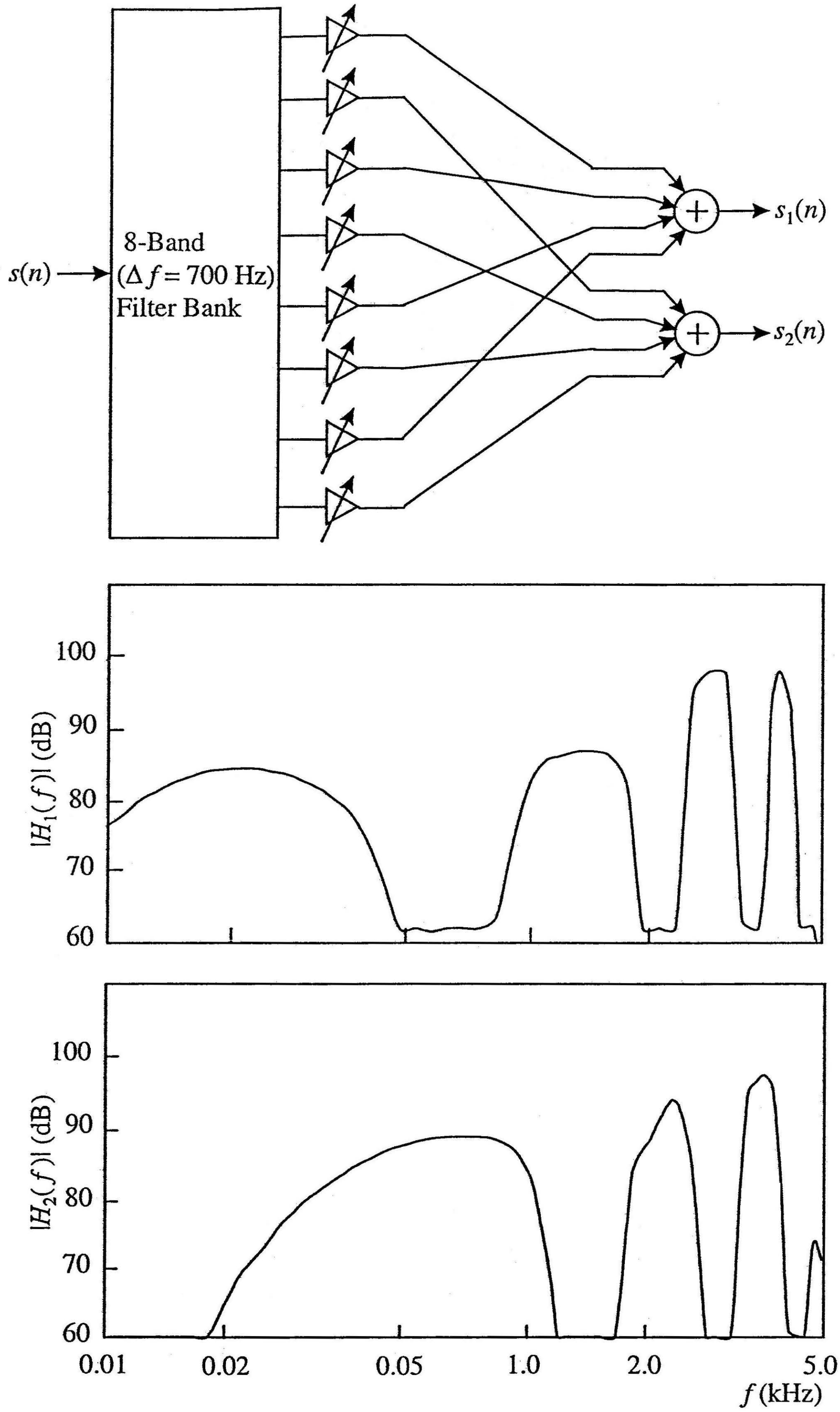


FIG. 3.2 A schematic diagram and magnitude response of the digital processor for dichotic presentation developed by Lunner *et al.* (1993). Schematic is based on the description given in Lunner *et al.* (1993) and Lunner and Hellgren (1991). Waveform $s(n)$ is the input and $s_1(n)$ and $s_2(n)$ are the output waveform for presenting to the two ears. $|H_1(f)|$ and the $|H_2(f)|$ are the magnitude responses for the two channels.

Chapter 4

EVALUATION WITH OFF-LINE PROCESSING

4.1 Introduction

Scheme for splitting speech signal for dichotic presentation on the basis of critical band filtering has been described in the previous chapter. This scheme was implemented for off-line processing of digitized speech signals. The implementation was tested using spectrographic analysis of the processed waveforms. Experiments were carried out with (a) normal hearing subjects with simulated hearing loss and (b) hearing impaired subjects; for evaluating the effectiveness of the scheme. In this chapter, we present implementation of the scheme, experimental procedure for the listening tests, and the test results, along with a discussion of these results.

4.2 Implementation

Speech processing for dichotic presentation can be implemented in three different ways. Fig. 4.1 shows a scheme with two filter banks, each with nine band pass filters. The two output signals are obtained by adding the outputs of the filters in the bank. In this approach using parallel combination of band pass filters, different bands can be added with different gains. However, there is a possibility of notches in the magnitude at crossover frequencies due to phase shifts in the adjacent filters. Fig. 4.2 shows a scheme in which the splitting of the input signal into two signals is realized by using cascade combination of band reject filters. The phase shifts of individual filters do not affect the magnitude response in this approach. In the scheme shown in Fig. 4.3, two filters with the desired comb filter response are designed. This may result in overall efficiency of realization.

For off-line processing, the cascade combination of band reject filters as shown in Fig. 4.2 was selected. All the filters are linear phase FIR filters with 255 coefficients. The filter coefficients were obtained from the sampled magnitude response using rectangular window (Rabiner and Schafer, 1978; Kuc, 1982) with the help of filter design package DSPLAY. Filter of such a high order was selected in order to obtain sharp cut-off, so that there is no significant overlap from the neighboring bands. Computational efficiency was not considered to be important since processing of signals was done in off-line mode. The program for implementing the scheme is described in Appendix C.

The log magnitude response of each of the two channels was obtained by taking 512-point FFT of their unit sample responses, and these are shown in Fig. 4.4. The pass band ripples are within 2 dB, and side lobe attenuation are more than 40 dB. The transition bands are less than 55 Hz. The filter response was also verified by obtaining the spectrograms of the processed outputs using a spectrographic analysis set-up (as described in Appendix A) and these are shown in the Figs. 4.5-4.8. Each figure shows the spectrograms of the unprocessed waveform and the two processed outputs. The waveform and gray scale plot are shown at the bottom and at the side of the spectrogram respectively. Fig. 4.5 shows wide band spectrogram for a swept sine wave. As the frequency is swept linearly with time over 1 s, we see the signal passes alternately between the two channels showing the widening of the critical bands with increasing frequency. Fig 4.6 shows narrowband spectrogram for random white noise, in this we see the complementary splitting of the spectra. Figs. 4.7 and 4.8 show wideband spectrograms for speech segment /*a*s*a*/ and /*a*g*a*/ respectively. These also show the complementary spectra for the two channels.

4.3 Listening tests

The purpose of the experimental investigation was to evaluate the effectiveness of the scheme in reducing the effect of spectral masking. Two sets of listening tests were conducted. In the first set, later referred as Experiment I, normal hearing subjects with simulated hearing impairment participated. The loss was simulated by adding broadband noise to the speech stimuli with five different SNR conditions (as described earlier in Section 3.5). In the second set of tests, later referred as Experiment II, were with hearing impaired subjects having bilateral sensorineural loss. These subjects were tested without adding any masking noise to the speech stimuli.

Listening tests were carried out for finding the confusions among the set of 12 English consonants. These tests happen to be repetitive and time consuming, and hence an automated computerized test administration system was used. The listening tests were administered for (a) unprocessed speech diotically presented and (b) processed speech dichotically presented. A description of test material, subjects participated, and experimental procedure is given in the following subsections.

4.3.1 Test material

In order to minimize the contribution of linguistic factors and maximize the contribution of acoustic factors, nonsense syllables were used for stimuli. Twelve consonants /p, b, t, d, k, g, m, n, s, z, f, v/ and the vowel /*a*/ as in 'father' were used to form two sets of stimuli with syllables in vowel-consonant-vowel (VCV) and consonant-vowel (CV) contexts, The number of stimuli was restricted to 12, so that they can be conveniently accommodated on subject's screen in the computerized test administration system.

The speech stimuli were acquired and analyzed by using the set-up consisting of a DSP board interfaced to a PC, as shown in Fig. 4.9. The signal from the microphone goes to an amplifier, attenuator, frequency-weighting filter, and buffer amplifier (microphone B&K 4176 along with sound level meter B&K 2235). This output is amplified, low pass filtered (f_p = 4.6 kHz, f_s = 5.0 kHz, pass band ripple < 0.3 dB, stop band attenuation > 40 dB), and digitized with 16-bit resolution at 10 k samples/s using the ADC of TI/TMS 320C25 based DSP board which is interfaced to PC. The acquired speech segment can be outputted using DAC of the DSP board. The set-up can be used for spectrographic analysis. The description of hardware and software is given in Appendix B & C respectively.

A male speaker spoke the syllables. Each syllable was recorded a number of times. These were analyzed spectrographically and played back, and from these recordings, a syllable considered as the most normal sounding was selected to be the stimulus in the experiments. The processing of syllables was done off-line. For simulating sensorineural hearing loss in normal hearing subjects, the stimuli were added with broadband noise at five SNR conditions of ∞ (no noise), 6, 3, 0, and -3 dB. The noise was obtained by digitizing the white Gaussian noise from a waveform generator (HP33120A). The noise was added in such a way that signal-to-noise ratio was kept constant on the basis of short-time (≈ 10 ms) energy of the signal. Thus during silence segments there would not be any background noise.

4.3.2 Subjects

In Experiment I, the scheme was tested on normal hearing subjects with simulated hearing loss and bilateral sensorineural hearing impaired subjects. The subjects who participated in the experiments were from different parts of India and they had no difficulty in clearly recognizing the test stimuli.

Seven normal hearing subjects (four male and three female) participated in the experiments. Three subjects (SAK: F 22, MSC: M 24, CKS: M 24) participated in both the VCV and CV tests. Two subjects (HBN: M 28, PK: F 22), participated in VCV test only, while two subjects (PCP: M 40, DSJ: F 35) participated in CV tests only. Thus we had five subjects for both the tests. All the subjects are right handed and well conversant with English language. These subjects had pure tone average (PTA) of better than 20 dB HL at audiometric test frequencies of 0.5, 1, and 2 kHz for both the ears.

In Experiment II, ten hearing impaired subjects (SG: M 27, SJH: F 19, KRN: M 35, DSD: M 19, LGR: M 27, SSN: M 31, KRV: M 49, BAS: M 58, SAV: M 46, and LDM: M 48) participated. The subjects were right handed and familiar with English. The hearing impaired subjects had 'mild'-to-'very severe' bilateral sensorineural hearing loss and their PTA difference between right and left ear was from 0 to 30 dB. Their pure tone hearing thresholds are given in Appendix G. The format of the forms for subject's background and willingness are given in Appendix H.

4.3.3 Experimental procedure and set-up

The speech stimuli were presented to the subjects through a pair of headphones (Telephonics TDH-39P). The headphones were calibrated before using for presentations, the procedure for callibration of headphones is given in Appendix D. Presentations were done at the most comfortable listening level for the subject, which ranged from 75 to 85 dB SPL. Once th e presentation level is decided, care was taken to keep it constant for the particular subject throughout the test, since the clarity of speech depends on loudness also. The presentation level for a particular setting of amplifier's volume control was checked at the beginning of each test session, using sound pressure level meter along with artificial ear (B&K, 4153) coupled to the headphone. The A-weighted SPL reading was taken by continuously outputting the vowel segment of the test syllable.

The listening tests were performed using an automated computerized test administration system as shown in Fig. 4.10. It includes a PC, a PC based data acquisition card (PCL-208, from Dynalog MicroSystem Limited, Mumbai) having two output ports, and a subject terminal (placed in acoustically isolated chamber). For dichotic presentation, the stimuli were outputted at a rate of 10 k samples/s through two D/A ports of data acquisition card. The D/A outputs were passed through a pair of smoothing low pass filters (with same specifications as that of input antialiasing filter used in the set-up for signal acquisition) and a pair of audio amplifiers. The subject terminal was used for displaying the response choices on its screen and for obtaining subject responses from its keyboard. The description of hardware and software of the experimental set-up is given in Appendix B & C respectively.

Before conducting the listening tests, the subject was briefed about the experimental procedure. A test consisted of a total of 60 presentations, each of the stimuli presented five times, in a randomized order. For each presentation, the subject should respond by pressing a key. Response choices were displayed on subject screen for each presentation, each choice corresponding to a key on the subject terminal's keyboard. The location of response choices was also randomized. Each test used a different order for presentations. The set-up also recorded the response time. Before each test session, the subject could listen to the stimuli any number of time in any order, in order to become fully familiar with them. Each stimulus was presented after flashing "Listen…" on the screen. The subject responded by pressing an appropriate key. At the end of each test, the stimulus-response confusion matrix and response time statistics were stored. The tests can be conducted with feedback to the subject or without feedback. Only results from tests without feedback were considered for analysis. The subject performance can vary with exposure to stimuli and fatigue, hence the tests with stabilized scores with variation of

10 % were used. The test with feedback was conducted in between the tests without feedback on request of the subject, in order to refresh the subject's familiarity with the test stimuli.

In Experiment I involving normal hearing subjects, listening tests were carried out for 5 SNR conditions (∞, 6, 3, 0, and -3 dB) randomized across the test sessions. Subjects were also asked to provide a qualitative assessment of the test stimuli for ascertaining the improvement in speech quality. Each test took about 6-10 min. Thus, the five SNR conditions with VCV and CV contexts took about thirteen to fourteen hours for completion for each subject. The test sessions were spread over two months as per the availability and willingness of the subjects. For each experimental run and for each subject, the three stabilized (in terms of recognition scores) confusion matrices were combined out of the five confusion matrices, and considered for the final analysis. Thus, the probabilities in the combined confusion matrices were obtained on the basis of 15 presentations for each stimulus.

In Experiment II involving hearing impaired subjects, all the tests were conducted without any masking noise. Each test took about 8-14 min. Ten tests were conducted for unprocessed and processed speech presented diotically and dichotically respectively, in VCV and CV contexts. It took about eight-to-nine hours of testing time for each subject. As per the availability and willingness of the ten hearing impaired subjects, the test sessions were spread over six months. Out of the ten confusion matrices, eight confusion matrices with stabilized recognition scores were combined for final analysis. Thus, the probabilities in the combined confusion matrices were obtained on the basis of 40 presentations for each stimulus.

4.4 Results with simulated hearing loss subjects (Experiment I)

Results for the listening tests conducted with five normal hearing subjects, at 5 SNR conditions of masking noise (∞, 6, 3, 0, and -3 dB), in VCV and CV contexts are presented here. A compilation of subjects' qualitative assessments about the set of test stimuli under various listening conditions for ascertaining the improvement in the speech quality was carried out. Average response time was used for comparing the effectiveness of the processing scheme, in terms of the load on the perception process. The stimulus-response confusion matrices were used to obtain the recognition scores. The confusion matrices were subjected to information transmission analysis (as described earlier in Section 3.5, and in more detail later in Appendix E) in order to obtain a measure that is not affected by subject's response bias. The twelve stimuli were combined in groups and the resulting matrices were analyzed for reception of the consonantal features of duration, frication, nasality, manner, place, and voicing.

Compilation of subjects' qualitative assessment indicated that the speech quality was better with processing for binaural dichotic presentation. The response times under various listening conditions for the VCV and CV contexts are given in Tables 4.1 and 4.2 respectively. These tables provide values for the five subjects and also averaged values across the subjects. The response times, for a subject and averaged across the subjects are plotted for the VCV and CV contexts in Figs. 4.11 and 4.12 respectively. The subject chosen for plotting the scores happen to be the first participant in the listening test. The relative improvement in response time (RT) were calculated as

$$[(R.S.)_{\text{processed speech}} - (R.S.)_{\text{unprocessed speech}}] / (R.S.)_{\text{unprocessed speech}}$$

With decreasing SNR response time increases. For unprocessed speech, as compared to SNR= ∞ dB, averaged increasing the response time with SNR for 6, 3, 0, and -3db is 8, 10, 36, and 45% and 2, 8, 8, and 15% for the VCV and CV contexts respectively. This indicates that the load on perception process increases with decreasing SNR processing brought and improvement in response time. The averaged improvements for the SNR conditions of ∞, 6, 3, 0, and -3db were 19, 15, 17, 29, and 16% for VCV context and 14, 9, 9, 10, and 16% for CV context. The improvements in response time were tested for statistical significance using t-test (Snedecor and Cochran, 1980). All the subjects showed less response time for processed speech, indicating a reduction in the load on the perception process.

The recognition scores for individual subjects and averaged across the subjects for VCV and CV contexts, are given in Tables 4.3 and 4.4 respectively. Under no noise condition, all the subjects showed nearly perfect scores with both unprocessed and processed speech. For all the subjects, score generally decreased as the masking noise level increased. Further, the scores for processed speech are higher than those for unprocessed speech under the same condition of masking noise. It is to be noted that the improvements due to processing were more for higher levels of masking noise (i.e. higher level of simulated sensorineural loss). The recognition scores for a subject and averaged across the subjects are plotted for the VCV and CV contexts in Figs. 4.13 and 4.14 respectively. Relative improvements in recognition score (RS) were calculated as

$$[(\mathrm{R.S.})_{\text{processed speech}} - (\mathrm{R.S.})_{\text{unprocessed speech}}] / (\mathrm{R.S.})_{\text{unprocessed speech}}$$

and these are also given in Tables 4.3 and 4.4. For SNR conditions of 6, 3, 0, and -3 dB, the relative improvements in recognition scores range from 2.2 to 12.7, 2.9 to 17.1, 2.2 to 19.2, and 5.6 to 19.4 respectively in VCV context, and from 1.2 to 14.8, 2.1 to 33.9, 2.6 to 44.6, and 4.9 to 38.2 respectively in CV context. Averaged across the subjects, the percentage improvements at these SNR levels are 0.9, 5.5, 8.6, 11.0, and 11.1 respectively in VCV context, and 0.1, 6.4, 12.2, 15.5, and 15.0 respectively in CV context, indicating that processing of the speech and dichotic presentation improves recognition scores. It was observed that the improvements were higher under adverse listening conditions.

The recognition scores, for individual subjects obtained in VCV and CV contexts were subjected to t-test for testing the statistical significance of improvements in scores due to processing. For all the subjects, under low SNR (< 3 dB) conditions, the improvements due to processing are highly significant ($p < 0.05$). Paired t-test, across the subjects for testing the significance of improvement in recognition scores due to processing, was carried out and improvements are highly significant ($p < 0.05$) for almost all SNR conditions.

Information transmission analysis of the confusion matrices was carried out and relative information transmitted are given in Tables 4.5 and 4.6 for all the subjects The relative information transmitted for the two contexts of VCV and CV are plotted in Figs. 4.15 and 4.16 respectively. Under both the unprocessed and processed condition, the highest information transmission was observed for nasality, voicing, and duration, whereas frication and place were relatively poorly transmitted.

Under high SNR conditions, the information transmission was near perfect even with unprocessed speech and improved to 100 % with processed speech. With poor SNR, the information transmission with unprocessed speech decreased, and improvements were seen with the processed speech. For unprocessed speech, most of the decrease in the relative information trransmission could be attributed to decrease in the recxeption of placxe feature. This indicates that masking noise resulted in simulation of spectral masking. The overall improvements were contributed by better reception of almost all the six features of duration, frication, nasality, manner, place, and voicing. However, most of the subjects indicated the maximum improvement for the place feature among the features of place, voicing, and manner.

Relative improvements in the relative information transmission were calculated in the same way as for recognition scores. These improvements for the place feature are given in Table 4.7, and the values averaged across the subjects are plotted in Fig. 4.17. It was seen that relative improvements were very high at higher levels of masking noise. As the reception of the place feature is related to frequency resolving capacity of the auditory processing, it can be inferred that the processing scheme has reduced the effect of spectral masking.

4.5 Results with hearing impaired subjects: Experiment II

Results from listening tests conducted, without any masking noise, with ten hearing impaired subjects in VCV and CV contexts are presented here. A compilation of subjects' qualitative assessments about the test stimuli for ascertaining the improvement in speech quality was carried out. Average response time was used for comparing the effectiveness of the processing scheme, in terms of the load on the perception process. The stimulus-response confusion matrices were used to obtain the recognition scores. The confusion matrices were subjected to information transmission analysis in order to obtain a measure that is not affected by subject's response bias. The twelve stimuli were combined in groups and the resulting matrices were analyzed for reception of the consonantal features of duration, frication, nasality, manner, place, and voicing.

Compilation of subjects' qualitative assessment indicated that the speech quality was better with processing for binaural dichotic presentation. The response times for the VCV and CV contexts are given in Tables 4.8 and 4.9 respectively. The relative improvements in response time, were ranged from 3.3 to 17% and 1.1 to 15.1% in VCV and CV contexts respectively. All the subjects showed decrease in response time due to processing. For most of the subjects, the decrease was statistically significant (as seen by t-test) indicating an improvement in listening condition due to processing. The response time, for a subject and averaged across the subjects are plotted in Fig. 4.18, and these show a reduction in the response time for both the contexts. Paired t-test, across the subjects showed that the decreases in response time are highly significant ($p < 0.0005$) in both the contexts.

The recognition scores for individual subjects and averaged across the subjects for both the contexts, are given in Tables 4.10 and 4.11. The recognition scores for a subject and averaged across the subjects are plotted in Fig. 4.19. For all the subjects, the scores for processed speech are higher than those for unprocessed speech. The percentage relative improvements in recognition scores ranged from 4.6 to 26.6 in VCV context and 6.4 to 25.1 in CV context. The recognition scores were subjected to t-test, for testing the statistical significance of improve ments in scores due to processing. The significance levels for individual subjects in both the contexts were almost similar. Majority of the subjects showed highly significant improvement due to processing ($p < 0.01$). Paired t-test, across the subjects for testing the significance of improvement in recognition scores due to processing, was carried out and improvements are highly significant ($p < 0.0005$) in both the contexts.

The confusion matrices were subjected to information transmission analysis and overall information transmitted as well as information transmission for specific features is given in Tables 4.12 and 4.13, for all the subjects. The results for a subject and averaged across the subjects are plotted for the VCV and CV contexts in Figs. 4.20 and 4.21 respectively. The overall improvements were contributed by better reception of almost all the six features of duration, frication, nasality, manner, place, and voicing. It was observed that, among the features of place, voicing, and manner, largest number of subjects showed improvement in place feature (six subjects), followed by voicing (3 subjects) and manner (one subject) in VCV context. Almost similar pattern of improvement in the reception of feature was observed in CV context also.

The relative improvements in the relative transmission of information for different features are given in Table 4.14. All the six features of duration, frication, nasality, manner, place, and voicing contributed the overall relative improvements in transmission, with nearly maximum relative improvements for the place feature. Averaged across the ten subjects, the relative improvement for the place feature were 29 and 25 % in VCV and CV contexts respectively.

4.6 Discussion

Two experiments were conducted for evaluation of the off-line implementation of the processing scheme for dichotic presentation. Experiment I consisted of a set of tests conducted with five normal hearing subjects with hearing loss simulated by mixing broadband noise as a masker at different SNR conditions. In the listening tests of Experiment II, ten subjects with bilateral hearing loss participated.

A compilation of both types of subjects' qualitative assessment indicated that the subjects preferred the processed dichotic presentation to the unprocessed diotic presentation.

For all normal hearing subjects, the average response time increased with masking noise and significantly decreased for processed speech as compared to unprocessed one. Here it is to be noted that under no masking noise condition, the recognition scores were near perfect for both unprocessed and processed conditions. But there was a decrease in response time due to processing. This indicates an improvement in listening condition with processing, resulting in reduction in the load on the perception process. Most of the hearing impaired subjects showed significant decrease in response time for processed speech. Thus, response time can be considered as an independent indication of effectiveness of the processing.

It was clearly observed from the results for all the normal hearing subjects that the recognition score generally decreased with increased in the masking noise level. Further, it was noticed that for a particular level of masking noise, the score for processed speech was higher than that for the unprocessed one. The important finding is that the improvements due to processing were more for higher levels of masking noise, i.e. higher levels of simulated sensorineural loss. However, these improvements tend to level, at very high levels of simulated loss. Most of the hearing impaired subjects indicated highly significant improvement in recognition score.

Information transmission analysis of the stimulus-response confusion matrices with both set of subjects indicated that an overall improvement was contributed by better reception of all the six features (duration, frication, nasality, manner, place, and voicing). In case of normal hearing subjects, the relative improvement in the reception of place feature was seen to be higher under adverse listening condition. In case of hearing impaired subjects, the improvements were almost highest for the place feature. As the reception of place feature is related to frequency resolving capacity of the auditory process, one can say that the implemented scheme has reduced the effect of spectral masking.

For gaining an insight into relationship between the improvement due to processing and the extent and nature of hearing loss of individual subjects, the relative improvement in recognition score and transition of place features were averaged for the VCV and CV context and these are given in table 4.15. We see that the maximum improvement for subjects SG and SJH. This two happen to subjects with very severe hearing loss. Therefore we may provisionally conclude that more the extent of sensorineural loss, i.e. higher the extent of spectral masking, more the improvement due to processing. However for the final conclusion, psychophysical measurements of the spread of spectral masking for the individual subject need to be carried out.

Comparing the test result for experiment I and II, we see that recognition scores for hearing-impaired listeners are lesser than that of normal hearing under simulated loss even at SNR=-3dB. Hence, for realizable simulation of sensorineural loss, still less SNR is needed. The hearing impaired subjects have a much larger reaction time, which can be expected as they do not have earlier exposure to the experimental set-up and computer keyboard and screen. However, the average reaction in the response time in case of both types of subjects is comparable.

From the qualitative assessment of speech, response time statistics, recognition scores and information transmission analysis of confusion matrices, it can be concluded that the scheme of splitting the speech on the basis of critical band filtering is helpful in improving speech quality, reducing the load on perception process, and in improving the reception of spectrally coded place feature without adversely affecting the reception of the features cued by amplitude and duration.

TABLE 4.1 Experiment I. Response times in VCV context. US : unprocessed speech, PS : processed speech. t-test for individual subjects : *n* (number of tests) = 3, df = 4; paired t-test for scores averaged across the subjects : *n* (number of subjects) = 5, df = 4.

Subject	SNR (dB)	Response Time (s) US mean	US s.d.	PS mean	PS s.d.	Relative Improve-ment (%)	t-test (one-tailed) t	p
SAK	∞	1.21	0.08	1.03	0.05	14.9	3.30	0.025
	6	1.11	0.11	1.10	0.03	0.9	0.15	0.25
	3	1.32	0.04	1.17	0.06	11.4	3.60	0.0125
	0	1.30	0.09	1.21	0.19	6.9	0.74	0.25
	-3	1.49	0.10	1.30	0.12	12.8	2.11	0.1
MSC	∞	1.50	0.41	1.23	0.18	18.0	1.04	0.2
	6	2.08	0.53	1.44	0.35	30.8	1.75	0.1
	3	1.84	0.06	1.77	0.06	3.8	1.43	0.2
	0	1.86	0.37	1.75	0.03	5.9	0.51	0.25
	-3	2.24	0.17	1.80	0.04	19.6	4.36	0.0125
CKS	∞	2.20	0.31	1.66	0.54	24.6	1.5	0.1
	6	2.17	0.98	1.75	0.03	19.4	0.74	0.25
	3	2.19	0.81	1.99	0.30	9.1	0.40	0.25
	0	3.03	0.57	2.80	0.29	7.6	0.62	0.25
	-3	3.12	0.18	2.81	0.14	9.9	2.35	0.05
HBN	∞	2.33	0.90	1.75	1.02	24.9	0.74	0.25
	6	2.63	0.09	2.39	0.57	9.1	0.72	0.25
	3	2.68	0.09	1.69	0.54	36.9	3.13	0.025
	0	3.78	0.70	1.25	0.11	66.9	6.18	0.0025
	-3	3.83	0.04	3.15	0.04	17.8	20.82	0.0005
PK	∞	1.42	0.08	1.34	0.07	5.6	1.30	0.2
	6	1.34	0.15	1.31	0.24	2.2	0.18	0.25
	3	1.50	0.22	1.28	0.13	14.7	1.49	0.2
	0	1.82	0.50	1.39	0.32	23.6	1.25	0.2
	-3	1.85	0.14	1.48	0.80	20.0	0.79	0.25
Avg.	∞	1.73	0.50	1.40	0.30	19.1	3.34	0.025
	6	1.86	0.56	1.59	0.50	14.5	2.24	0.05
	3	1.90	0.55	1.58	0.34	16.8	1.94	0.1
	0	2.36	1.02	1.68	0.66	28.8	1.45	0.2
	-3	2.50	0.96	2.11	0.82	15.6	1.87	0.005

TABLE 4.2 Experiment I. Response times in CV context. US : unprocessed speech, PS: processed speech, NS = not significant. t-test for individual subjects: *n* (number of tests) = 4, df = 6; paired t-test for scores averaged across the subjects: *n* (number of subjects) = 5, df = 4.

Subject	**SNR (dB)**	**Response Time (s)**				**Relative Improvement (%)**	**t-test (one-tailed)**	
		US		**PS**				
		mean	**s.d.**	**mean**	**s.d.**		**t**	**p**
SAK	∞	1.09	0.08	1.05	0.13	3.7	0.52	NS
	6	1.18	0.04	1.18	0.06	0.0	0.00	NS
	3	1.36	0.13	1.16	0.12	14.7	2.26	0.05
	0	1.43	0.17	1.22	0.06	14.7	2.33	0.05
	-3	1.70	0.22	1.37	0.23	19.4	2.07	0.05
MSC	∞	1.52	0.03	1.31	0.03	13.8	9.90	0.0005
	6	1.55	0.08	1.39	0.14	10.3	1.98	0.05
	3	1.89	0.32	1.46	0.12	22.8	2.52	0.025
	0	1.62	0.22	1.60	0.21	1.2	0.13	NS
	-3	2.03	0.14	1.54	0.10	24.1	5.70	0.0025
CKS	∞	1.70	0.36	1.55	0.48	8.8	0.50	NS
	6	1.78	0.38	1.78	0.47	0.0	0.00	NS
	3	2.10	0.42	1.90	0.53	9.5	0.59	NS
	0	2.36	0.08	1.97	0.32	16.5	2.36	0.05
	-3	2.10	0.07	1.56	0.03	25.7	14.18	0.0005
DSJ	∞	2.35	0.70	1.53	0.09	34.9	2.32	0.05
	6	1.61	0.07	1.45	0.24	9.9	1.28	0.2
	3	1.61	0.13	1.56	0.26	3.1	0.34	NS
	0	1.79	0.19	1.76	0.23	1.7	0.20	NS
	-3	2.01	0.17	1.92	0.12	4.5	0.86	0.25
PCP	∞	2.84	0.24	2.76	0.46	2.8	0.31	NS
	6	3.54	0.46	2.96	0.11	16.4	2.45	0.025
	3	3.36	0.09	3.32	0.40	1.2	0.20	NS
	0	3.10	0.48	2.77	0.46	10.7	0.99	0.2
	-3	3.10	0.51	2.77	0.43	10.7	0.99	0.2
Avg.	∞	1.90	0.69	1.64	0.66	13.7	1.82	0.1
	6	1.93	0.92	1.75	0.71	9.3	1.69	0.1
	3	2.06	0.78	1.88	0.85	8.7	2.61	0.05
	0	2.06	0.68	1.86	0.57	9.7	2.59	0.05
	-3	2.19	0.53	1.83	0.56	16.4	4.53	0.0125

TABLE 4.3 Experiment I. Recognition scores in VCV context. US: unprocessed speech, PS : processed speech. t-test for individual subjects : *n* (number of tests) = 3, df = 4; paired t-test for scores averaged across the subjects : *n* (number of subjects) = 5, df = 4.

Subject	SNR (dB)	Recognition Score (%) US mean	US s.d.	PS mean	PS s.d.	Relative Improvement (%)	t-test (one-tailed) t	p
SAK	∞	98.3	0.0	100.0	0.0	1.7	∞	0.0005
	6	96.7	0.0	100.0	0.0	3.4	∞	0.0005
	3	92.8	2.5	98.3	0.2	5.9	3.79	0.0125
	0	85.0	5.0	95.6	2.5	12.5	3.28	0.025
	-3	83.3	0.0	90.0	0.0	8.0	∞	0.0005
MSC	∞	100.0	0.0	100.0	0.0	0.0	--	--
	6	97.8	0.9	100.0	0.0	2.2	4.23	0.0125
	3	96.7	0.0	100.0	0.0	3.4	∞	0.0005
	0	96.1	1.9	98.3	0.0	2.2	2.01	0.1
	-3	90.0	1.7	98.3	0.0	9.2	8.46	0.0025
CKS	∞	98.9	1.0	100.0	0.0	1.1	1.85	0.1
	6	97.8	0.9	100.0	0.0	2.2	4.23	0.0125
	3	96.1	1.9	98.9	1.0	2.9	2.26	0.05
	0	92.2	0.9	95.6	1.0	3.7	4.38	0.0125
	-3	90.0	1.7	95.0	0.0	5.6	5.09	0.005
HBN	∞	98.3	0.0	100.0	0.0	1.7	∞	0.0005
	6	87.2	1.0	98.3	0.0	12.7	19.22	0.0005
	3	81.1	1.7	95.0	0.0	17.1	14.16	0.0005
	0	78.8	1.9	93.9	1.0	19.2	12.18	0.0005
	-3	74.4	1.0	84.4	1.0	13.4	12.25	0.0005
PK	∞	100.0	0.0	100.0	0.0	0.0	--	--
	6	93.3	0.0	100.0	0.0	7.2	∞	0.0005
	3	87.8	1.9	100.0	0.0	13.9	14.35	0.0005
	0	82.8	3.5	97.2	0.9	17.4	8.91	0.0005
	-3	78.2	2.5	93.4	1.2	19.4	12.25	0.0005
Avg.	∞	99.1	0.9	100.0	0.0	0.9	2.35	0.05
	6	94.6	4.5	99.7	0.8	5.4	2.98	0.025
	3	90.9	6.5	98.4	2.1	8.3	3.26	0.05
	0	87.0	7.0	96.1	1.7	11.0	3.38	0.025
	-3	83.2	7.0	92.2	5.3	11.1	5.18	0.005

TABLE 4.4 Experiment I. Recognition scores in CV context. US : unprocessed speech, PS : processed speech. t-test for individual subjects : *n* (number of tests) = 4, df = 6; Paired t-test for scores averaged across the subjects : *n* (number of subjects) = 5, df = 4.

Subject	**SNR (dB)**	**Recognition Score (%)**				**Relative Score Improve-ment (%)**	**t-test (one-tailed)**	
		US		**PS**				
		mean	**s.d.**	**mean**	**s.d.**		**t**	**p**
SAK	∞	100.0	0.0	100.0	0.0	0.0	--	--
	6	96.2	0.8	100.0	0.0	4.0	9.50	0.0005
	3	89.6	3.4	96.2	3.2	7.4	2.83	0.025
	0	85.4	6.3	94.6	5.3	10.7	2.23	0.05
	-3	82.1	2.9	93.8	4.8	14.3	4.17	0.005
MSC	∞	100.0	0.0	100.0	0.0	0.0	--	--
	6	98.8	1.6	100.0	0.0	1.2	1.50	0.1
	3	97.9	0.9	100.0	0.0	2.1	4.67	0.0025
	0	96.2	0.8	100.0	0.0	4.0	9.50	0.0005
	-3	93.8	1.6	98.8	0.8	5.3	5.59	0.0025
CKS	∞	99.6	0.9	100.0	0.0	0.4	0.88	0.25
	6	97.5	1.0	99.6	0.9	2.1	3.12	0.0125
	3	97.5	2.1	99.2	1.0	1.7	1.46	0.1
	0	96.7	0.0	99.2	1.0	2.6	5.00	0.0025
	-3	94.2	1.0	98.8	0.8	4.9	7.18	0.0005
DSJ	∞	100.0	0.0	100.0	0.0	0.0	--	--
	6	90.8	1.0	99.6	0.9	9.7	13.08	0.0005
	3	85.4	2.5	99.2	1.7	16.1	9.13	0.0005
	0	81.2	0.8	93.8	2.1	15.5	11.21	0.0005
	-3	79.6	3.4	89.2	1.0	12.1	5.42	0.0025
PCP	∞	99.6	0.9	99.6	0.9	0.0	--	--
	6	84.2	3.9	96.7	2.7	14.8	5.27	0.0025
	3	68.8	0.8	92.1	2.1	33.9	20.74	0.0005
	0	62.5	4.4	90.4	0.8	44.6	12.48	0.0005
	-3	63.3	7.1	87.5	5.0	38.2	5.57	0.0025
Avg.	∞	99.8	0.2	99.9	0.2	0.1	1.00	0.2
	6	93.5	6.0	99.2	1.4	6.4	2.64	0.05
	3	87.8	11.9	97.3	3.3	12.2	2.33	0.05
	0	84.4	14.0	95.6	4.0	15.5	2.46	0.05
	-3	82.6	12.7	93.6	5.3	15.0	3.09	0.025

TABLE 4.5 Experiment I. Relative information transmitted in VCV context. US: unprocessed speech, PS: Processed speech.

Sub-ject	**Feature**	**Percentage Relative Information Transmitted**									
		SNR= ∞ dB		**SNR= 6 dB**		**SNR= 3 dB**		**SNR= 0 dB**		**SNR= -3dB**	
		US	**PS**	**US**	**PS**	**US**	**PS**	**US**	**PS**	**US**	**PS**
SAK	Overall	97	100	96	100	92	97	88	94	85	92
	Duration	100	100	93	100	83	93	52	93	68	86
	Frication	96	100	100	100	84	96	74	87	84	90
	Nasality	87	100	100	100	95	100	100	95	94	90
	Manner	92	100	100	100	88	97	84	92	90	96
	Voicing	96	100	92	100	95	91	100	96	96	100
	Place	100	100	90	100	75	94	57	86	52	71
MSC	Overall	100	100	98	100	95	99	92	99	89	97
	Duration	100	100	84	100	85	100	72	100	58	100
	Frication	100	100	100	100	100	100	100	100	86	100
	Nasality	100	100	100	100	100	100	100	100	100	100
	Manner	100	100	100	100	100	100	100	100	91	100
	Voicing	100	100	100	100	91	100	88	100	96	100
	Place	100	100	91	100	88	98	79	95	64	90
CKS	Overall	98	100	97	100	95	100	92	94	87	94
	Duration	100	100	83	100	78	100	82	83	55	96
	Frication	93	100	100	100	100	100	100	95	94	100
	Nasality	100	100	95	100	95	100	91	100	90	96
	Manner	95	100	98	100	98	100	96	97	92	98
	Voicing	100	100	100	100	100	100	96	100	85	94
	Place	96	100	92	100	86	100	77	82	68	84
HBN	Overall	98	100	91	97	88	96	82	94	76	90
	Duration	87	100	34	93	37	100	22	76	29	87
	Frication	100	100	54	100	100	100	41	100	45	88
	Nasality	95	100	100	95	100	100	100	100	100	87
	Manner	98	100	72	98	100	100	64	100	67	89
	Voicing	100	100	100	95	92	100	100	100	95	100
	Place	95	100	66	92	48	82	40	80	41	60
PK	Overall	100	100	93	100	88	100	87	97	86	96
	Duration	100	100	73	100	57	100	45	100	41	100
	Frication	100	100	91	100	87	100	96	100	95	100
	Nasality	100	100	100	100	100	100	100	88	100	100
	Manner	100	100	95	100	92	100	97	94	96	98
	Voicing	100	100	100	100	95	100	92	100	90	98
	Place	100	100	77	100	60	100	50	95	52	94
Avg.	Overall	99	100	95	99	92	98	88	96	85	94
	Duration	97	100	73	99	68	99	55	90	50	94
	Frication	98	100	89	100	94	99	82	96	81	96
	Nasality	96	100	99	99	98	100	98	97	97	95
	Manner	97	100	93	100	96	99	88	97	87	96
	Voicing	99	100	98	99	95	98	95	99	92	98
	Place	98	100	83	98	71	95	61	88	55	80

TABLE 4.6 Relative information transmitted in CV context. US: unprocessed speech, PS: processed speech.

Sub-ject	Feature	Percentage Relative Information Transmitted									
		SNR= ∞ dB		**SNR= 6 dB**		**SNR= 3 dB**		**SNR= 0dB**		**SNR= -3dB**	
		US	**PS**	**US**	**PS**	**US**	**PS**	**US**	**PS**	**US**	**PS**
SAK	Overall	100	100	95	100	88	95	86	92	81	92
	Duration	100	100	100	100	96	100	100	100	96	100
	Frication	100	100	90	100	86	94	73	84	73	96
	Nasality	100	100	96	100	87	100	100	90	100	90
	Manner	100	100	94	100	85	96	83	88	83	96
	Voicing	100	100	94	100	97	100	100	94	97	96
	Place	100	100	89	100	68	86	59	86	50	78
MSC	Overall	100	100	98	100	97	100	95	100	92	98
	Duration	100	100	100	100	100	100	84	100	94	100
	Frication	100	100	96	100	100	100	88	100	90	100
	Nasality	100	100	100	100	100	100	100	100	100	100
	Manner	100	100	98	100	100	100	92	100	94	100
	Voicing	100	100	100	100	100	100	88	100	93	97
	Place	100	100	97	100	91	100	89	100	83	96
CKS	Overall	99	100	97	99	96	99	95	99	92	98
	Duration	100	100	100	100	93	100	100	100	94	100
	Frication	100	100	96	100	92	100	87	100	90	100
	Nasality	100	100	100	100	100	100	100	100	100	100
	Manner	100	100	98	100	95	100	92	100	94	100
	Voicing	97	100	100	97	90	100	93	94	97	97
	Place	100	100	91	100	96	96	92	98	83	96
DSJ	Overall	100	100	94	99	90	99	86	93	81	87
	Duration	100	100	94	100	100	100	90	90	100	90
	Frication	100	100	88	100	89	96	74	81	60	71
	Nasality	100	100	100	100	100	100	100	100	90	100
	Manner	100	100	92	100	93	98	84	88	72	82
	Voicing	100	100	100	100	100	100	94	94	94	100
	Place	100	100	84	98	67	98	60	90	54	69
PCP	Overall	99	99	85	96	78	93	70	93	71	88
	Duration	94	94	41	82	34	69	48	45	39	50
	Frication	96	96	52	88	31	76	31	60	29	58
	Nasality	100	100	100	100	96	100	96	100	100	100
	Manner	98	98	72	93	57	86	57	77	59	75
	Voicing	100	100	90	96	100	94	94	100	90	97
	Place	98	98	61	87	28	76	16	79	18	66
Avg.	Overall	100	100	94	99	90	97	86	95	83	93
	Duration	99	99	87	96	85	94	84	87	85	88
	Frication	99	99	84	98	80	93	71	85	68	85
	Nasality	100	100	99	100	97	100	99	98	98	98
	Manner	100	100	91	99	86	96	82	91	80	91
	Voicing	99	100	97	99	97	99	94	96	94	97
	Place	100	100	84	97	70	91	63	91	58	81

TABLE 4.7 Experiment I. Relative improvement in relative information transmission of place feature.

Sub-ject	Relative Improvement (%)									
	Context: VCV					Context: CV				
	SNR (dB)					SNR (dB)				
	∞	**6**	**3**	**0**	**-3**	**∞**	**6**	**3**	**0**	**-3**
SAK	0	11	25	51	37	0	12	27	46	56
MSC	0	10	11	20	41	0	3	10	12	16
CSK	4	9	16	7	24	0	10	0	7	16
HBN	5	39	71	100	46	--	--	--	--	--
PK	0	30	67	90	81	--	--	--	--	--
DSJ	--	--	--	--	--	0	14	46	50	28
PCP	--	--	--	--	--	0	43	171	394	267
Avg.	2	20	38	54	46	0	17	51	102	76

TABLE 4.8 Experiment II. Response times in VCV context. US : unprocessed speech, PS : processed speech. NS = not significant. t-test for individual subjects: *n* (number of tests) = 8, df = 14; paired t-test for scores averaged across the subjects: *n* (number of subjects) = 10, df = 9.

Subject	Response Time (s)				t-test (one-tailed)	
	US		PS			
	mean	s.d.	mean	s.d.	t	p
SG	3.73	0.62	3.17	0.22	2.41	0.025
SJH	3.69	0.43	3.57	0.41	0.57	NS
KRN	4.33	0.96	3.94	0.66	0.95	0.2
DSD	4.81	0.91	4.37	0.54	1.18	0.2
LGR	3.78	0.28	3.45	0.47	1.71	0.1
SSN	3.00	0.34	2.60	0.24	2.72	0.0125
KRV	3.67	0.23	3.50	0.19	1.61	0.1
BAS	2.88	0.65	2.39	0.45	1.75	0.1
SAV	3.39	0.65	2.92	0.60	1.50	0.1
LDM	4.21	1.05	3.65	0.82	1.89	0.05
Avg.	3.75	0.59	3.36	0.60	8.31	0.0005

TABLE 4.9 Experiment II. Response times in CV context. US : unprocessed speech, PS : processed speech. NS = not significant. t-test for individual subjects: *n* (number of tests) = 8, df = 14; paired t-test for scores averaged across the subjects: *n* (number of subjects) = 10, df = 9.

Subject	Response Time (s)				t-test (one-tailed)	
	US		PS			
	mean	s.d.	mean	s.d.	t	p
SG	3.73	0.62	3.17	0.22	2.41	NS
SJH	4.39	0.32	4.20	0.30	1.27	0.2
KRN	3.80	0.27	3.44	0.20	3.03	0.005
DSD	4.49	0.52	4.44	0.67	0.17	NS
LGR	3.91	0.39	3.32	0.45	2.80	0.0125
SSN	3.05	0.65	2.70	0.52	1.19	0.2
KRV	3.70	0.31	3.45	0.26	1.75	0.1
BAS	2.46	0.32	2.25	0.21	1.55	0.1
SAV	3.65	0.77	3.35	0.90	0.72	0.25
LDM	3.31	0.84	2.95	0.67	0.95	0.2
Avg.	3.54	0.68	3.26	0.69	6.21	0.0005

TABLE 4.10 Experiment II. Recognition scores in VCV context. US: unprocessed speech, PS : processed speech. t-test for individual subjects : *n* (number of tests) = 8, df = 14; paired t-test for scores averaged across the subjects : *n* (number of subjects) = 10, df = 9.

Subject	**Recognition Score (%)**				**Relative Improvement (%)**	**t-test (one-tailed)**	
	US		**PS**				
	mean	**s.d.**	**mean**	**s.d.**		**t**	**p**
SG	48.2	4.2	61.0	6.1	26.6	4.89	0.0005
SJH	46.2	4.9	55.0	6.8	19.0	2.97	0.005
KRN	69.8	4.1	76.4	2.8	9.6	3.82	0.0025
DSD	89.7	12.5	93.8	8.9	4.6	0.76	0.25
LGR	82.7	2.7	89.4	3.3	8.1	4.44	0.0005
SSN	59.0	6.2	66.0	7.4	11.9	2.05	0.05
KRV	77.0	2.6	86.0	4.6	11.7	4.82	0.0005
BAS	78.8	4.0	84.4	3.3	7.1	3.05	0.005
SAV	52.2	2.9	61.0	4.6	16.9	4.58	0.0005
LDM	82.1	2.8	88.8	3.3	8.2	4.38	0.0005
Avg.	68.6	15.9	76.2	14.2	11.1	10.13	0.0005

TABLE 4.11 Experiment II. Recognition scores in CV context. US : unprocessed speech, PS: processed speech. t-test for individual subjects : *n* (number of tests) = 8, df = 14; paired t-test for scores averaged across the subjects : *n* (number of subjects) = 10, df = 9.

Subject	**Recognition Score (%)**				**Relative Score Improvement (%)**	**t-test (one-tailed)**	
	US		**PS**				
	mean	**s.d.**	**mean**	**s.d.**		**t**	**p**
SG	62.7	3.1	71.2	3.3	13.55	5.31	0.0005
SJH	39.8	17.6	49.8	13.9	25.1	1.26	0.2
KRN	56.9	3.3	62.8	2.3	10.4	4.15	0.0005
DSD	84.4	9.0	89.6	8.2	6.2	1.21	0.2
LGR	76.5	1.9	85.8	5.8	12.2	4.31	0.0005
SSN	76.7	9.6	86.2	8.9	12.4	2.05	0.05
KRV	84.4	4.1	90.2	2.9	6.9	3.27	0.005
BAS	78.0	5.1	83.0	3.7	6.4	2.24	0.025
SAV	72.5	4.5	80.8	5.3	11.4	3.38	0.0025
LDM	68.8	3.3	74.1	2.0	7.7	3.88	0.0025
Avg.	71.1	13.8	77.4	13.0	10.1	11.32	0.0005

TABLE 4.12 Experiment II Relative information transmitted in VCV context. US: unprocessed speech, PS: processed speech.

	Percentage Relative Information Transmitted													
Feature	Overall		Duration		Frication		Nasality		Manner		Voicing		Place	
Subject	US	PS	US	PS	US	PS	US	PS	US	PS	US	PS	US	PS
SG	58	68	29	40	39	68	59	59	47	64	49	54	17	33
SJH	57	55	12	15	11	14	43	58	26	33	43	53	23	34
KRN	71	75	18	27	31	41	89	90	56	63	66	80	44	48
DSD	85	90	82	93	80	82	89	90	83	85	76	81	77	86
LGR	81	86	94	82	60	62	75	96	66	76	95	86	65	75
SSN	58	64	25	29	28	32	30	28	34	34	57	68	29	37
KRV	82	84	97	97	83	79	69	81	75	79	88	96	46	57
BAS	81	85	89	85	58	56	100	89	74	68	86	87	64	77
SAV	49	56	24	25	24	29	31	27	33	31	46	55	27	32
LDM	79	86	94	90	63	64	79	96	69	76	83	90	65	77
Avg.	70	75	56	58	48	53	66	71	56	61	70	75	46	56

TABLE 4.13 Experiment II Relative information transmitted in CV context. US: unprocessed speech, PS: processed speech.

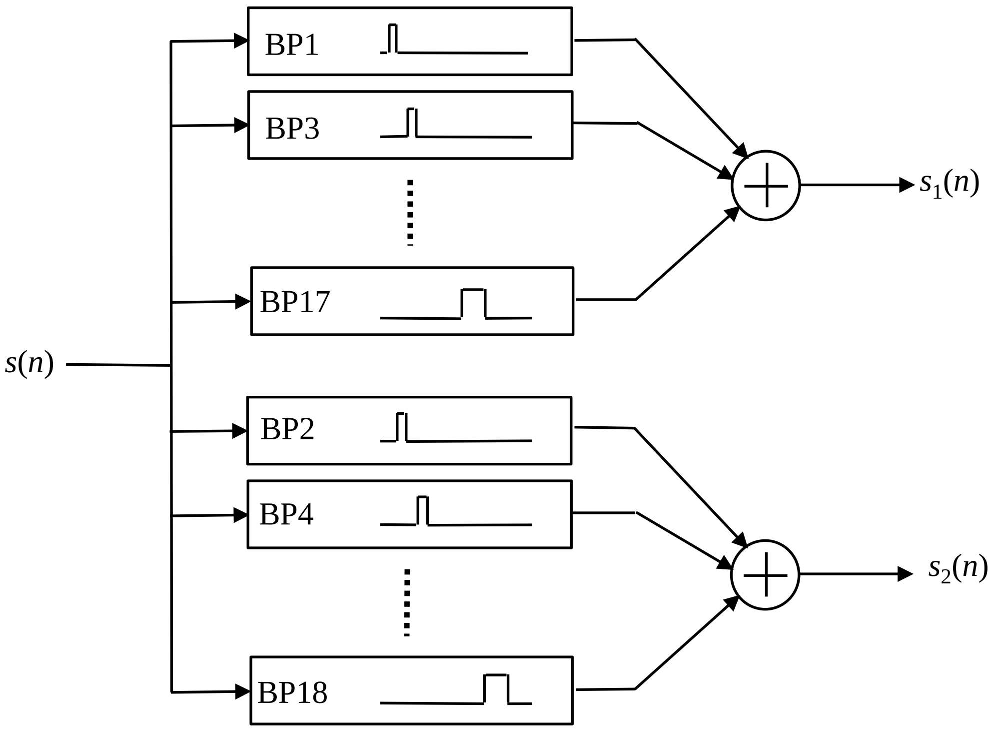
	Percentage Relative Information Transmitted													
Feature	Overall		Duration		Frication		Nasality		Manner		Voicing		Place	
Subject	US	PS	US	PS	US	PS	US	PS	US	PS	US	PS	US	PS
SG	74	73	01	09	14	22	100	100	50	55	100	96	29	39
SJH	38	42	17	24	16	24	26	31	19	26	40	46	09	12
KRN	66	72	20	28	14	21	33	34	29	37	36	46	38	45
DSD	81	86	86	79	71	77	68	80	70	78	62	74	84	86
LGR	73	83	97	93	68	88	84	96	74	91	92	94	39	57
SSN	71	81	71	80	57	73	80	63	67	69	55	71	57	81
KRV	82	89	94	100	86	85	65	92	79	87	86	95	69	76
BAS	76	82	87	96	54	65	82	97	68	77	63	73	61	72
SAV	68	75	76	63	60	64	65	90	61	74	63	72	47	60
LDM	67	72	17	17	24	29	84	100	49	59	77	80	43	51
Avg.	70	76	57	59	46	55	69	78	57	65	67	75	48	58

TABLE 4.14 Experiment II. Relative improvement in relative information transmission of different features. OV: overall, DU: duration, FR: friction, NA: nasality, MA: manner, VO: voicing, PL: place.

Subject	Relative Improvement (%)													
	Context: VCV							Context: CV						
	OV	DU	FR	NA	MA	VO	PL	OV	DU	FR	NA	MA	VO	PL
SG	17	38	74	0	36	10	94	-1	800	57	0	10	-4	34
SJH	-4	25	27	35	27	23	48	11	41	50	19	37	15	33
KRN	6	50	32	1	13	21	9	9	40	50	3	28	28	18
DSD	6	13	3	1	2	7	12	6	-8	8	18	11	19	2
LGR	6	-13	3	28	15	-9	15	14	-4	29	14	23	2	46
SSN	10	16	14	-7	0	19	28	14	13	28	-21	3	29	42
KRV	2	0	-5	17	5	9	24	9	6	-1	42	10	10	10
BAS	5	-4	-3	-11	-8	1	20	8	10	20	18	13	16	18
SAV	14	-4	21	-13	-6	20	19	10	-17	7	38	21	14	28
LDM	9	-4	2	22	10	8	18	7	0	21	19	20	4	19
Avg.	7	12	17	7	9	11	29	9	88	27	15	18	13	25

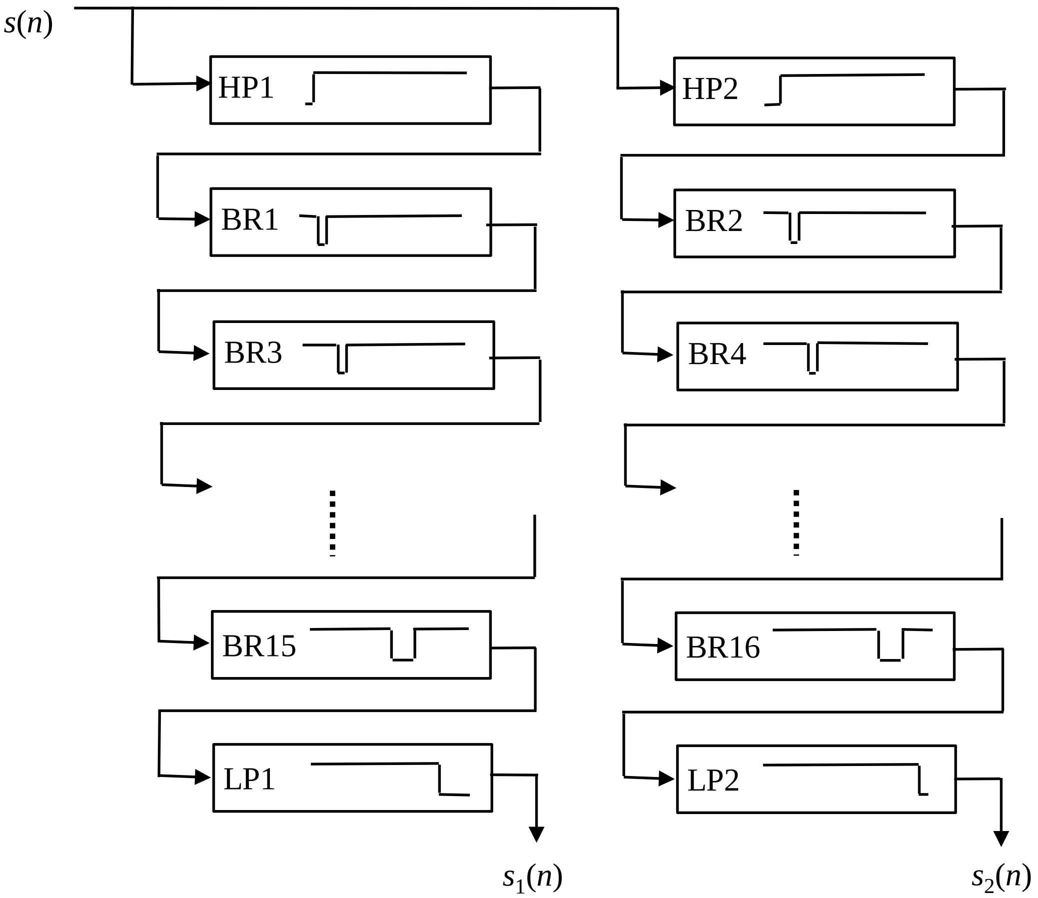
TABLE 4.15 Experiment II. Relative improvement in recognition score (RS) and transmission of place features (Place Tr.), averaged for VCV and CV contexts. Subjects' pure tone average (PTA) hearing threshold for the two ears are also give.

Subject	PTA hearing Threshold (dB)		Average Improvement In RS (%)	Average Improvement In place Tr. (%)
	Left	Right		
SG	73	77	20.1	64.0
SJH	98	88	22.1	40.5
KRN	42	58	10.0	13.5
DSD	85	73	5.4	7.0
LGR	68	68	10.2	30.5
SSN	75	65	12.2	35.0
KRV	60	52	9.3	17.0
BAS	33	38	6.8	19.0
SAV	45	67	14.2	23.5
LDM	52	82	8.0	18.5



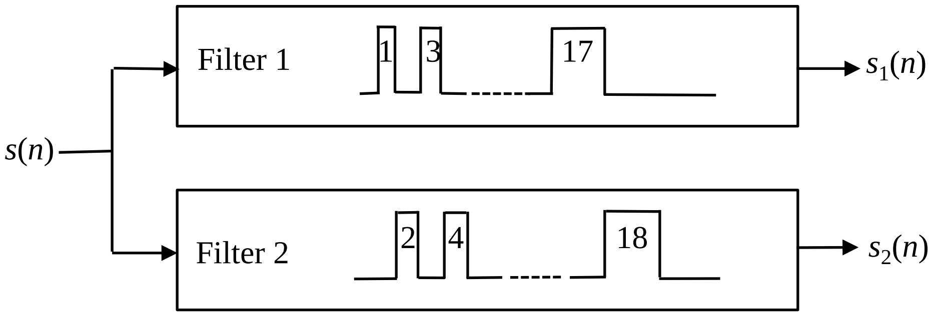
Filter	Freq. (kHz)	Filter	Freq. (kHz)
BP1	-- -0.20	BP2	0.20-0.30
BP3	0.30-0.40	BP4	0.40-0.51
BP5	0.51-0.63	BP6	0.63-0.77
BP7	0.77-0.92	BP8	0.92-1.08
BP9	1.08-1.27	BP10	1.27-1.48
BP11	1.48-1.72	BP12	1.72-2.00
BP13	2.00-2.32	BP14	2.32-2.70
BP15	2.70-3.15	BP16	3.15-3.70
BP17	3.70-4.40	BP18	4.40-5.00

FIG. 4.1 Splitting of speech signal using two banks of band pass filters. The filter magnitude response is shown in each block (table shows 3-dB cut-off frequencies).



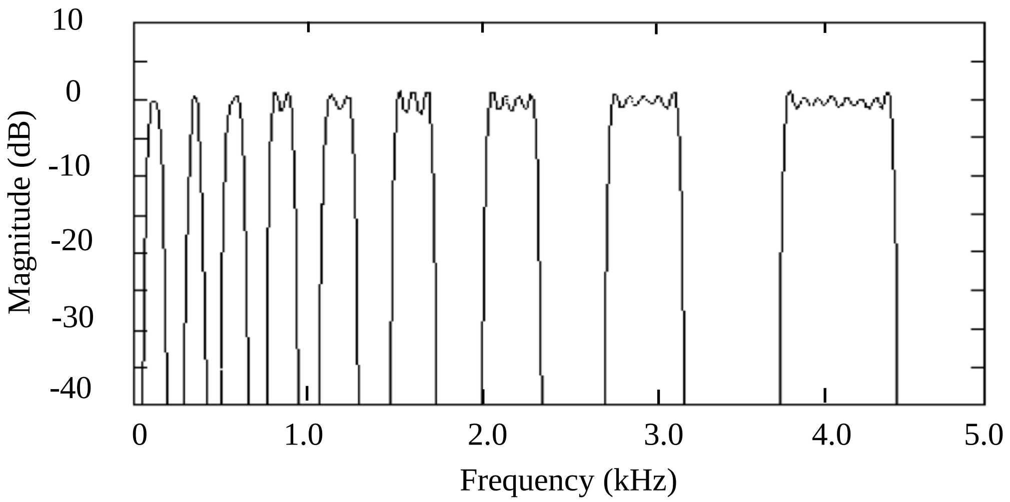
Filter	Freq.(kHz)	Filter	Freq. (kHz)
HP1	0.07	HP2	0.20
BR1	0.20-0.30	BR2	0.30-0.40
BR3	0.40-0.51	BR4	0.51-0.63
BR5	0.63-0.77	BR6	0.77-0.92
BR7	0.92-1.08	BR8	1.08-1.27
BR9	1.27-1.48	BR10	1.48-1.72
BR11	1.72-2.00	BR12	2.00-2.32
BR13	2.32-2.70	BR14	2.70-3.15
BR15	3.15-3.70	BR16	3.70-4.40
LP1	4.40	LP2	5.00

FIG. 4.2 Splitting of speech signal using cascade combination of band reject filters. The filter magnitude response is shown in each block (table shows 3-dB cut-off frequencies).

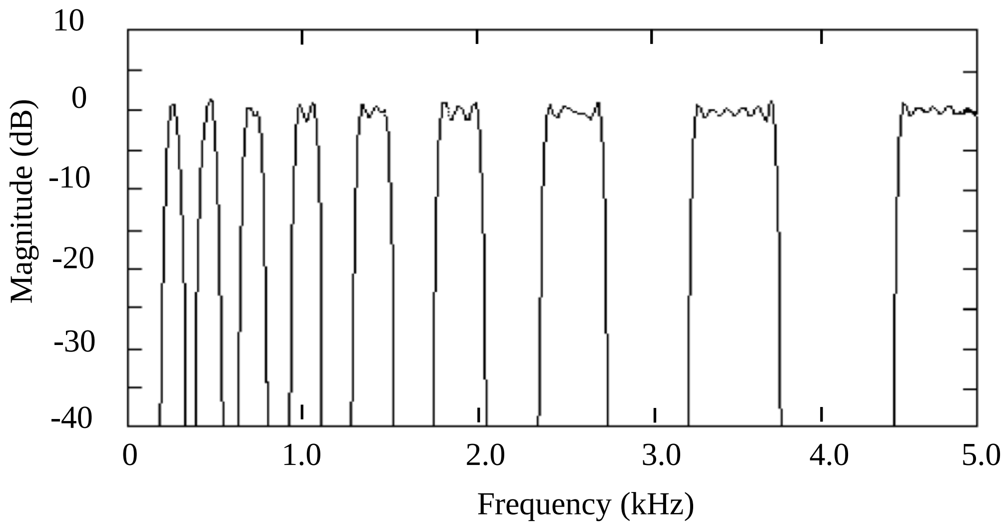


Band	Passband frequency	Band	Passband frequency
1	0.07 -0.20	2	0.20-0.30
3	0.30-0.40	4	0.40-0.51
5	0.51-0.63	6	0.63-0.77
7	0.77-0.92	8	0.92-1.08
9	1.08-1.27	10	1.27-1.48
11	1.48-1.72	12	1.72-2.00
13	2.00-2.32	14	2.32-2.70
15	2.70-3.15	16	3.15-3.70
17	3.70-4.40	18	4.40-5.00

FIG. 4.3 Schematic representation for splitting of speech signal using two comb filters. The filter magnitude response is shown in each block (table shows 3-dB cut-off frequencies).



(a)



(b)

FIG. 4.4 Magnitude response of the filters used in off-line processing: (a) left ear (b) right ear.

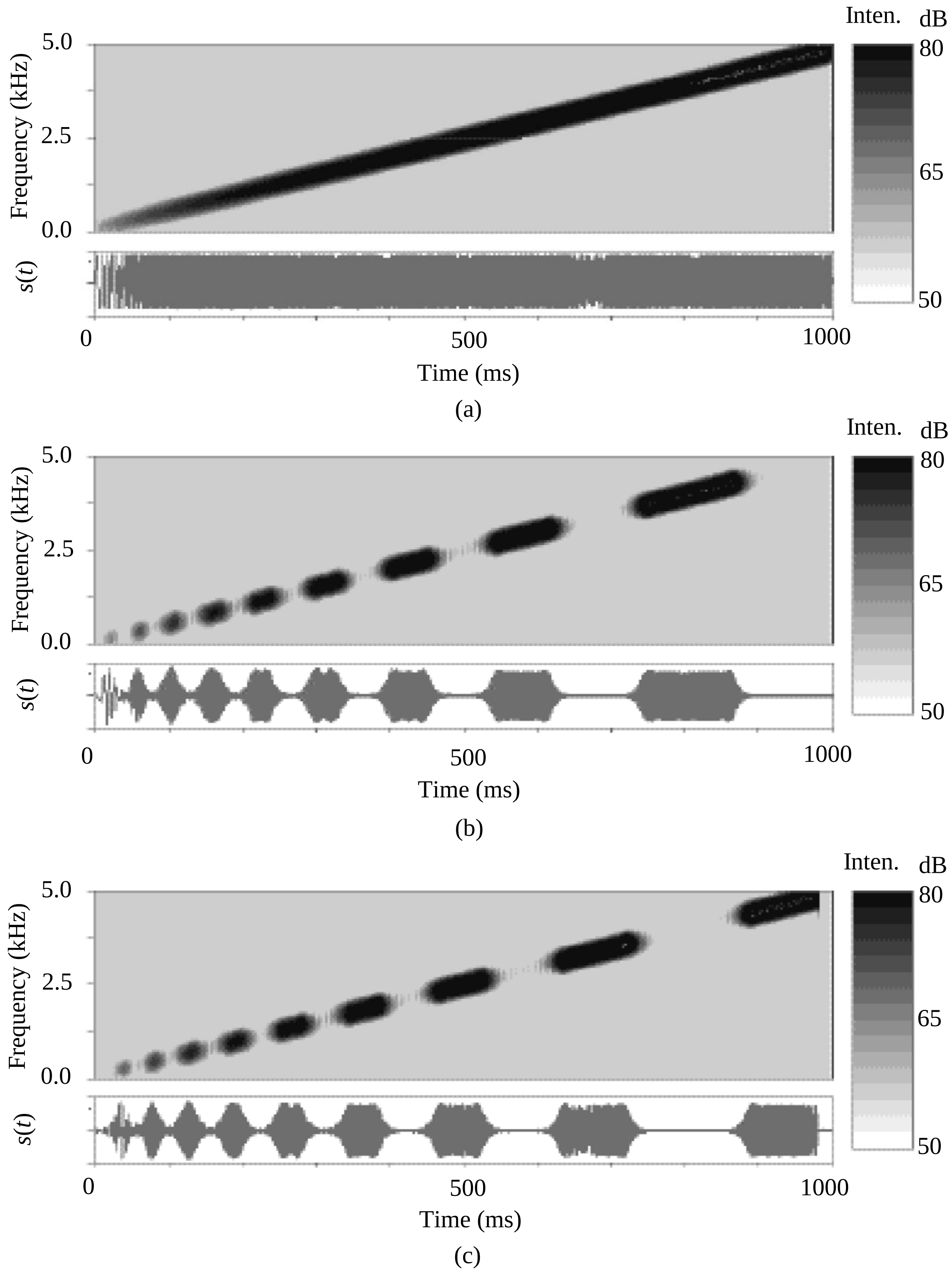


FIG. 4.5 Wideband spectrograms of swept sine wave (50 Hz to 5 kHz) $s(t)$: (a) unprocessed (b) processed (left ear) (c) processed (right ear). S. R. = 10 k samples/s and Δf = 300 Hz.

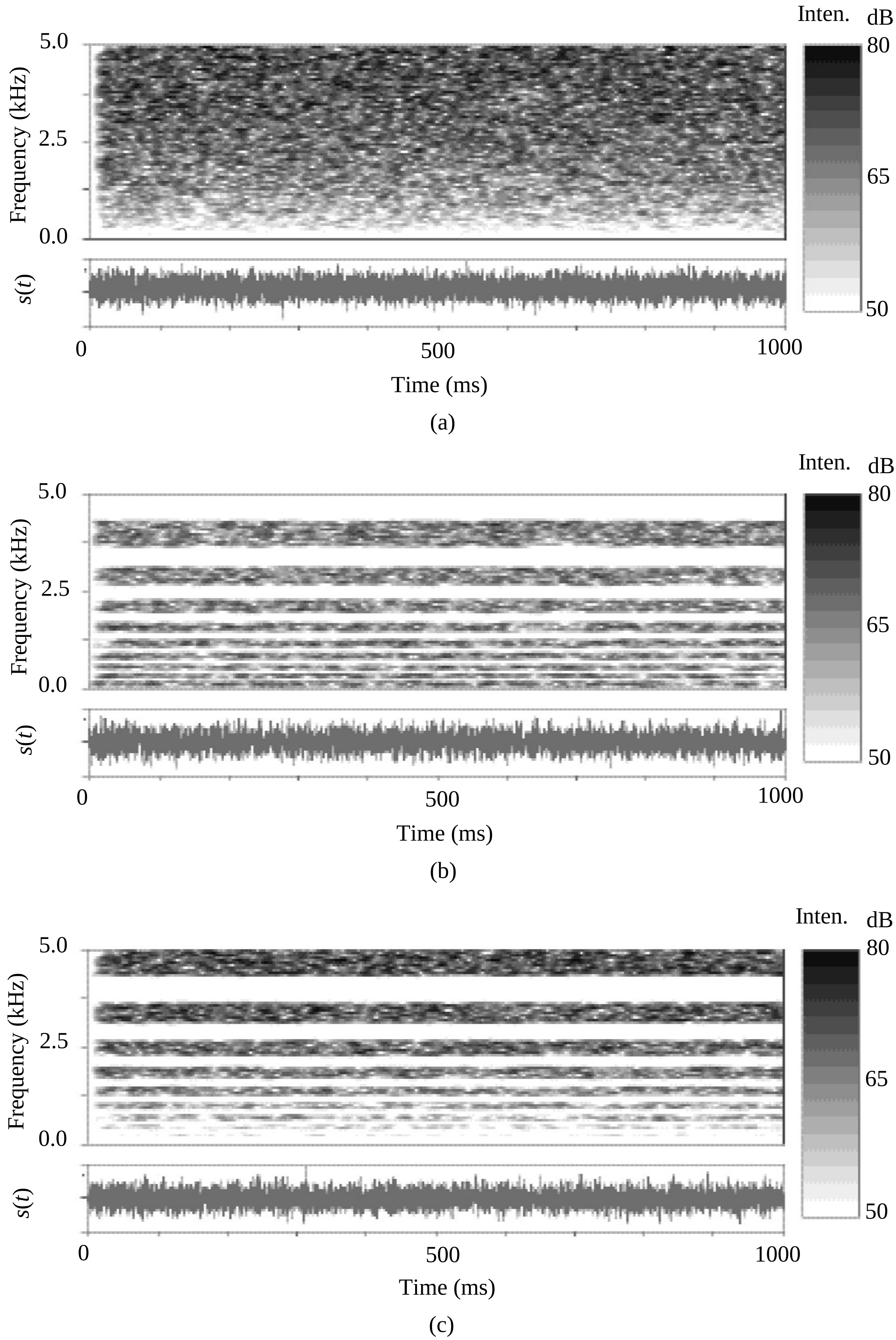


FIG. 4.6 Narrowband spectrograms of random noise $s(t)$: (a) unprocessed (b) processed (left ear) (c) processed (right ear). S. R. = 10 k samples/s and Δf = 53 Hz.

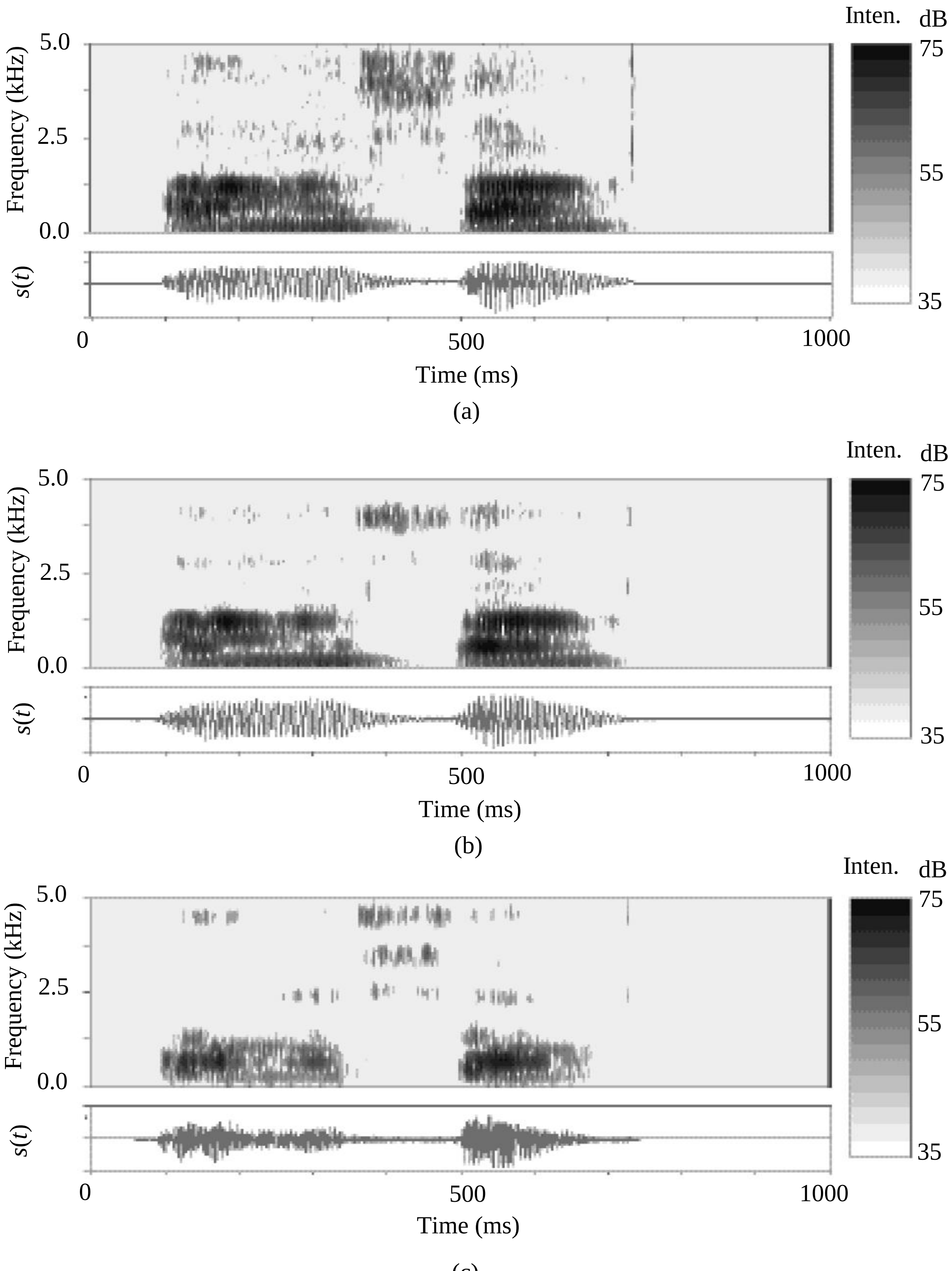


FIG. 4.7 Wideband spectrograms of speech waveforms $s(t)$, for utterance /*asa*/: (a) unprocessed (b) processed (left ear) (c) processed (right ear). S. R. = 10 k samples/s and Δf = 300 Hz.

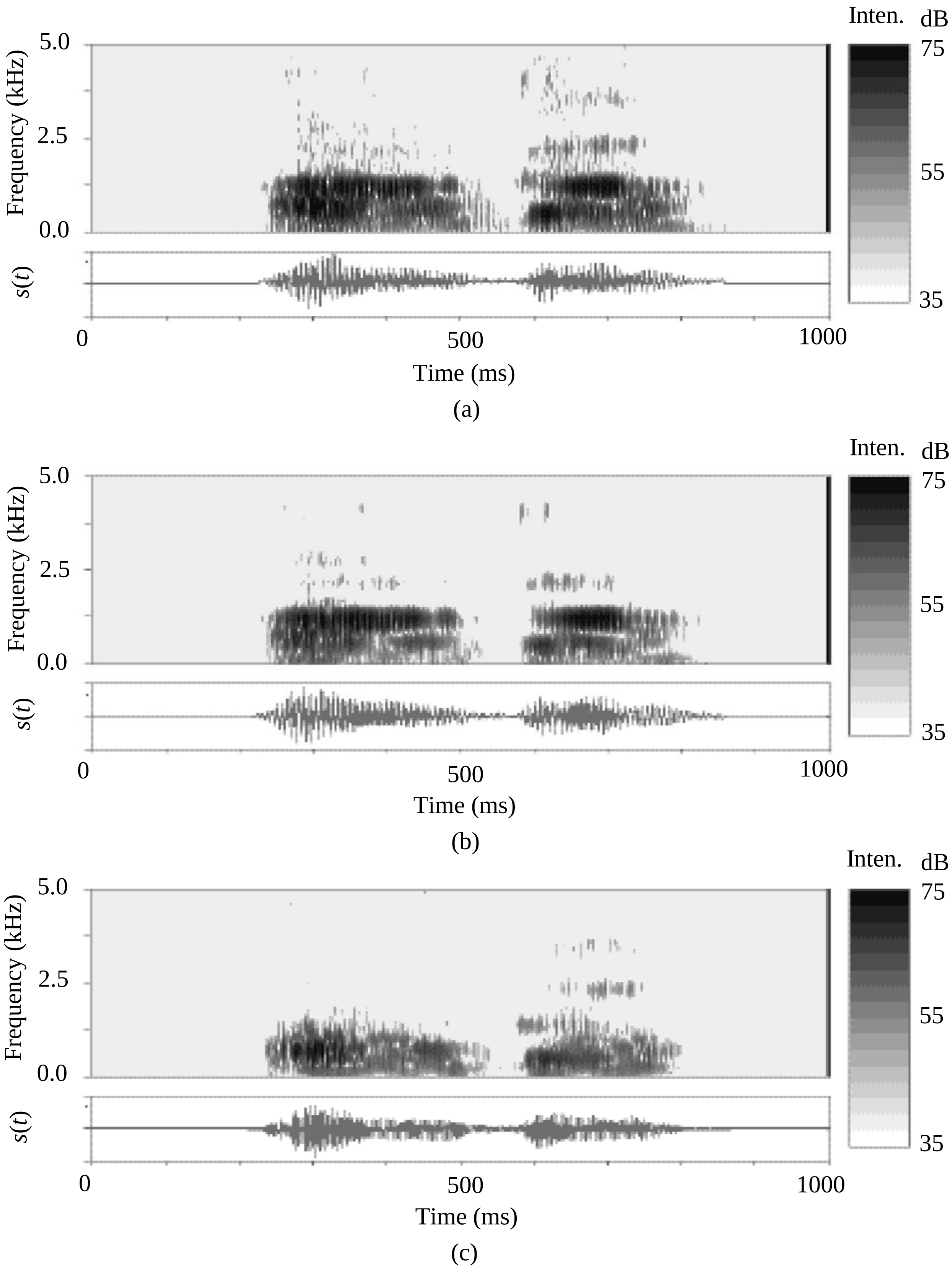


FIG. 4.8 Wideband spectrograms of speech waveforms $s(t)$, for utterance /*aga*/: (a) unprocessed (b) processed (left ear) (c) processed (right ear). S. R.=10 k samples/s and Δf =300 Hz.

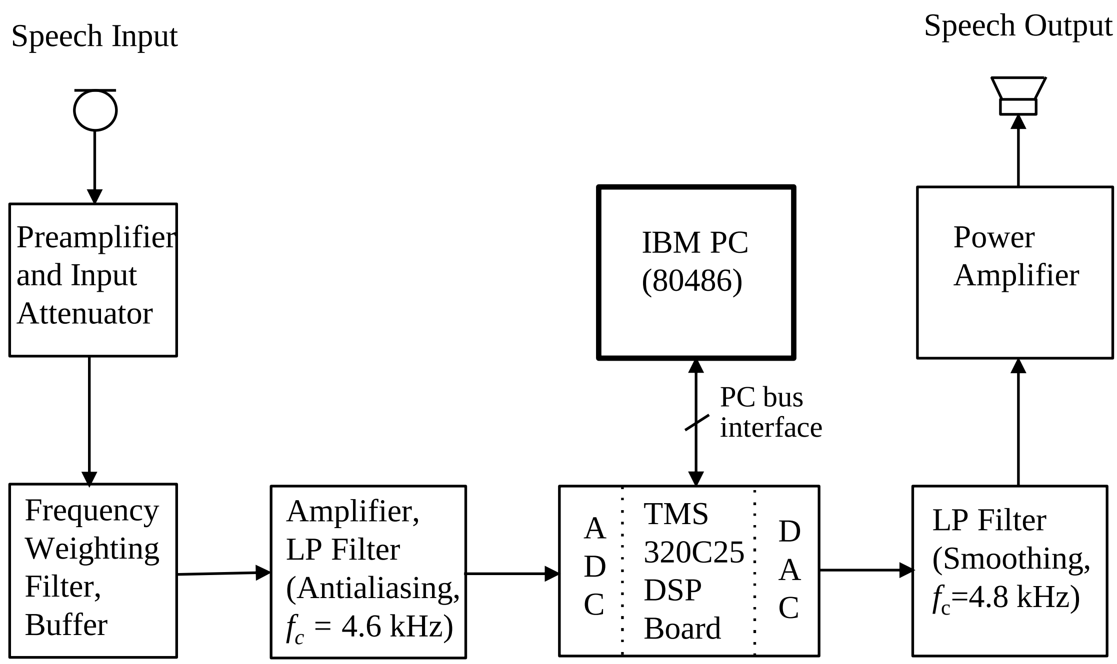


FIG. 4.9 Experimental setup for acquisition and analysis of speech segments.

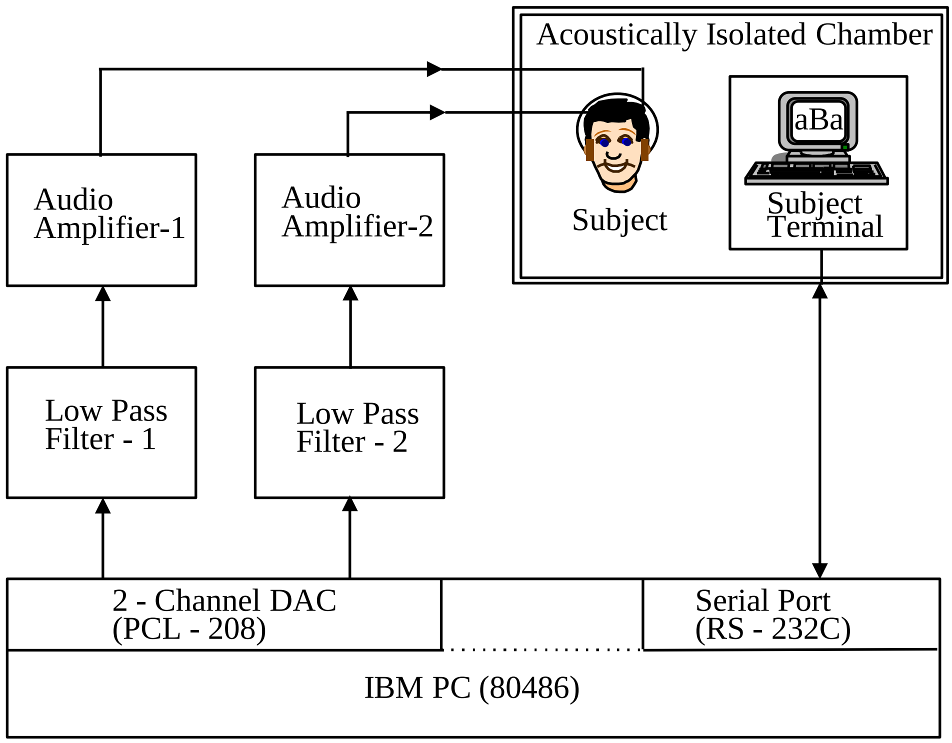
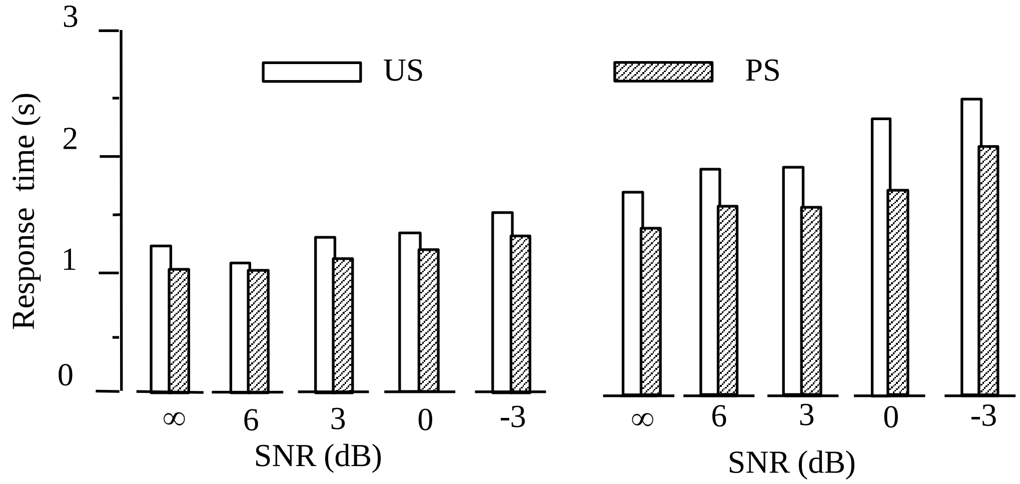
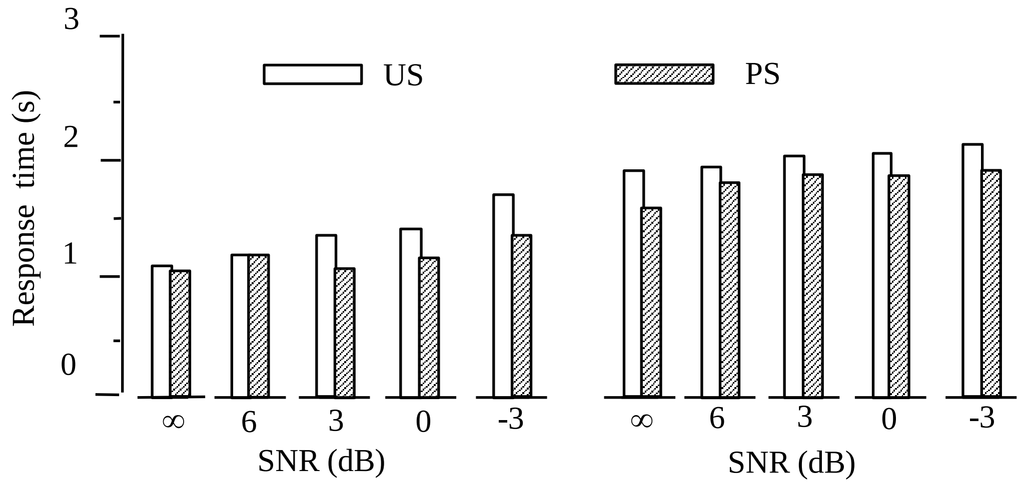


FIG. 4.10 Experimental setup used in the computerized test administration system for listening tests in off-line processing.



(a) (b)

FIG. 4.11 Experiment I. Response times at different SNRs, in VCV context: (a) for subject SAK (b) averaged for the five subjects. US: unprocessed speech, PS: processed speech.



(a) (b)

FIG. 4.12 Experiment I. Response times at different SNRs, in CV context: (a) for subject SAK (b) averaged for the five subjects. US: unprocessed speech, PS: processed speech.

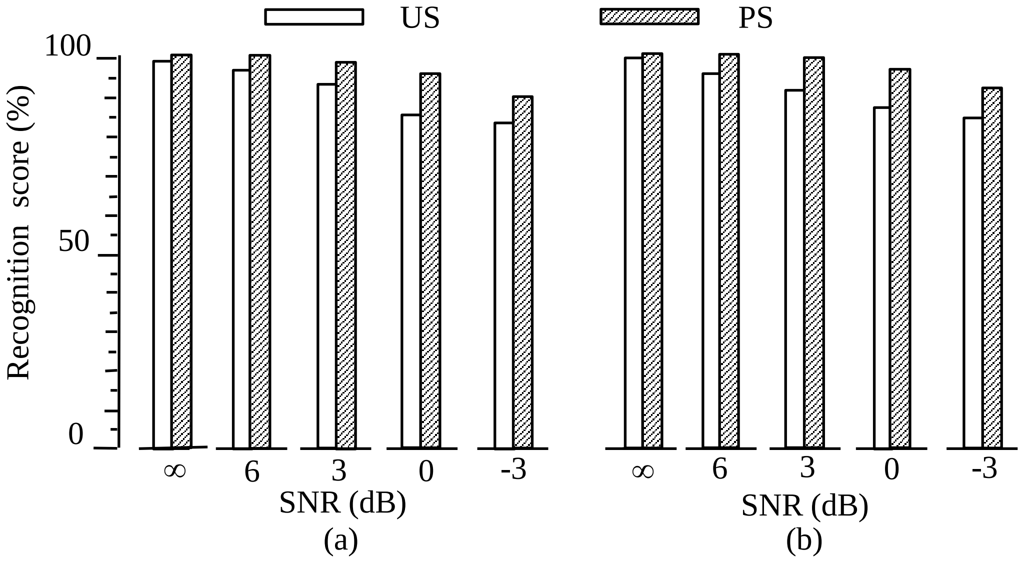


FIG. 4.13 Experiment I. Recognition scores at different SNRs, in VCV context: (a) for subject SAK (b) averaged for the five subjects. US: unprocessed speech PS: processed speech.

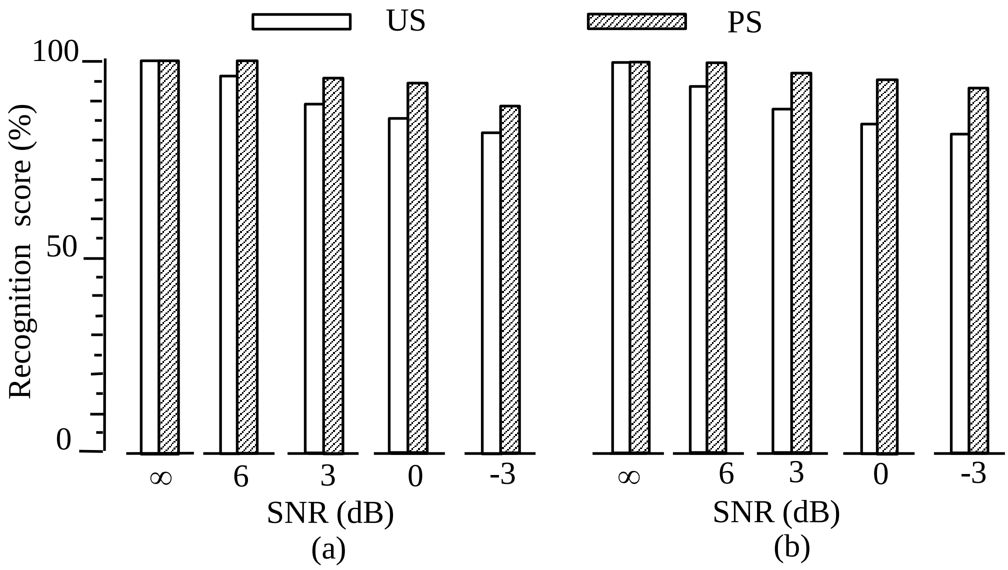
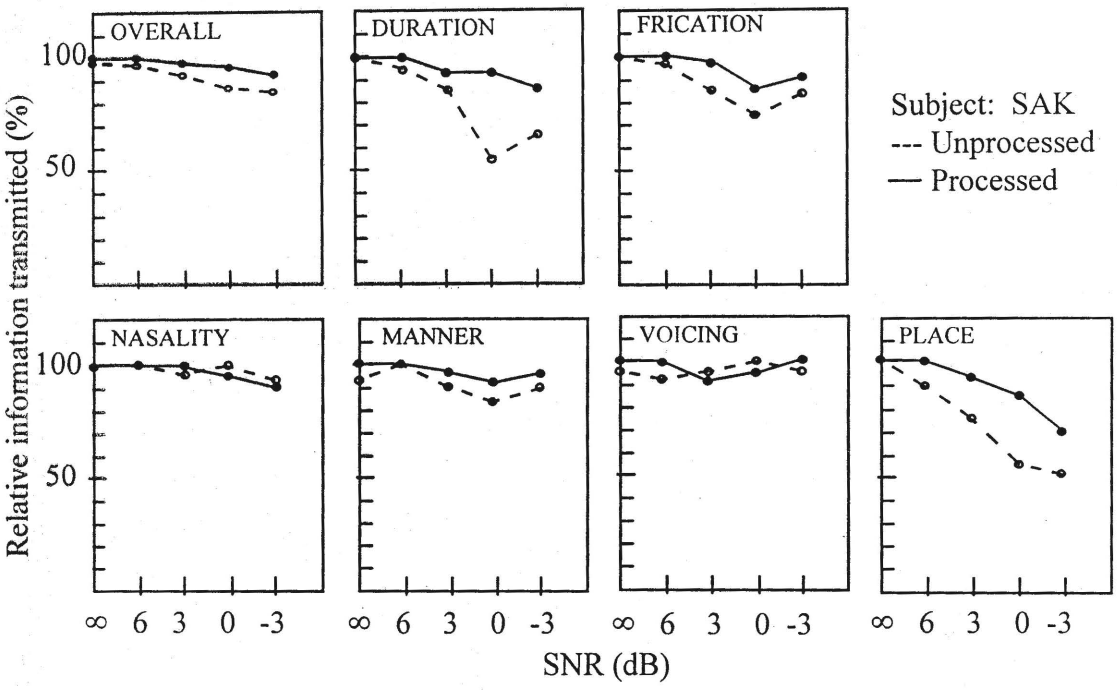
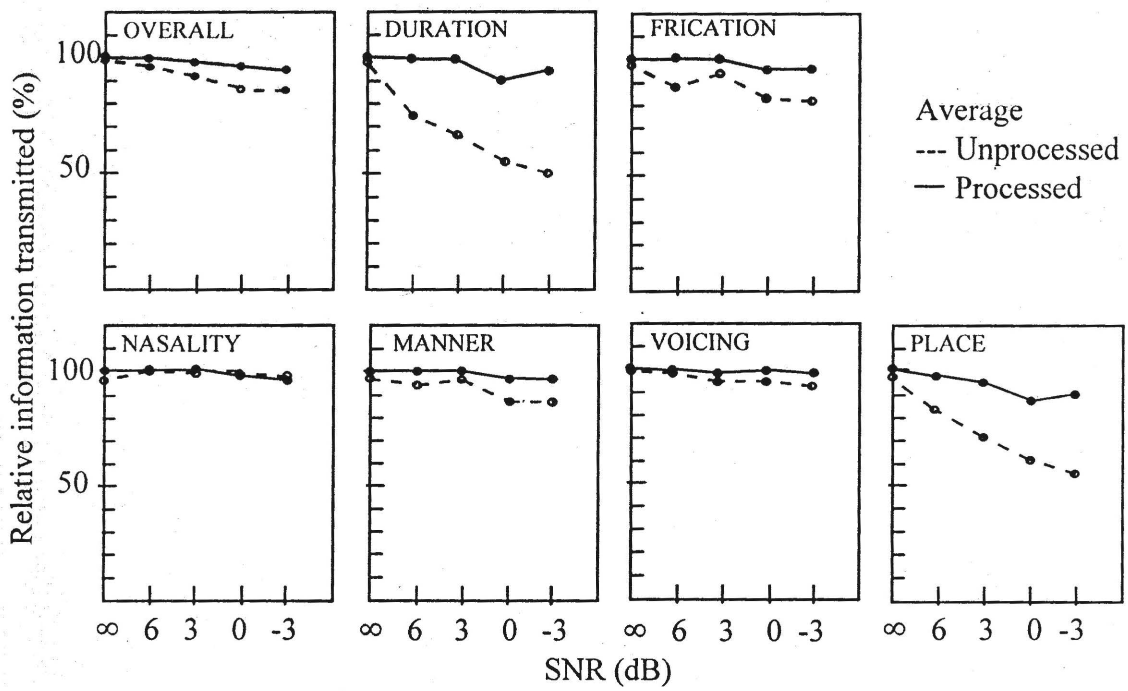


FIG. 4.14 Experiment I. Recognition scores at different SNRs, in CV context: (a) for subject SAK (b) averaged for the five subjects. US: unprocessed speech PS: processed speech.



(a)



(b)

FIG. 4.15 Experiment I. Percentage relative information transmitted, for unprocessed versus processed speech, in VCV context: (a) for subject SAK (b) averaged for the five subjects.

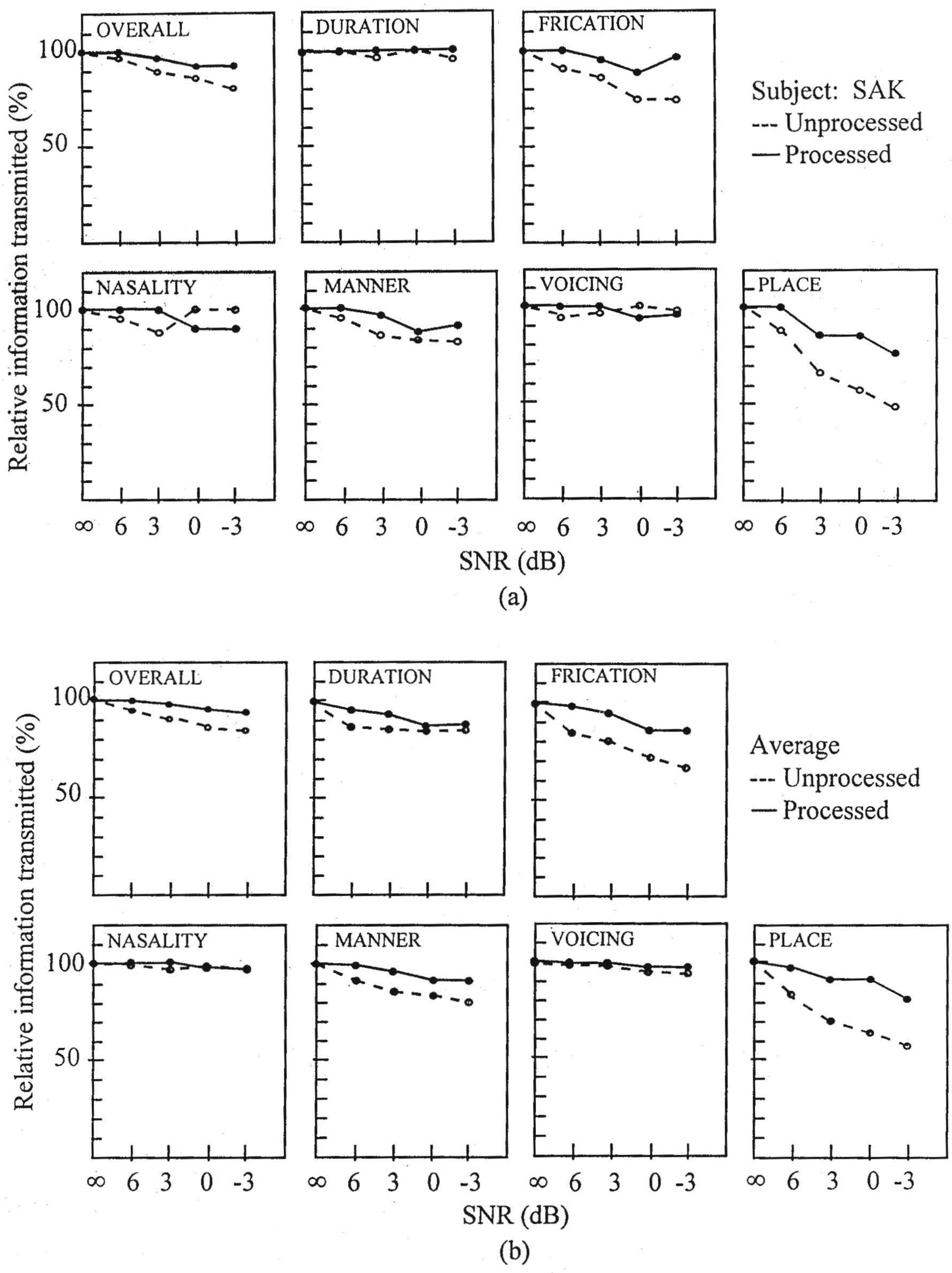


FIG. 4.16 Experiment I. Percentage relative information transmitted, for unprocessed versus processed speech, in CV context: (a) for subject SAK (b) averaged for the five subjects.

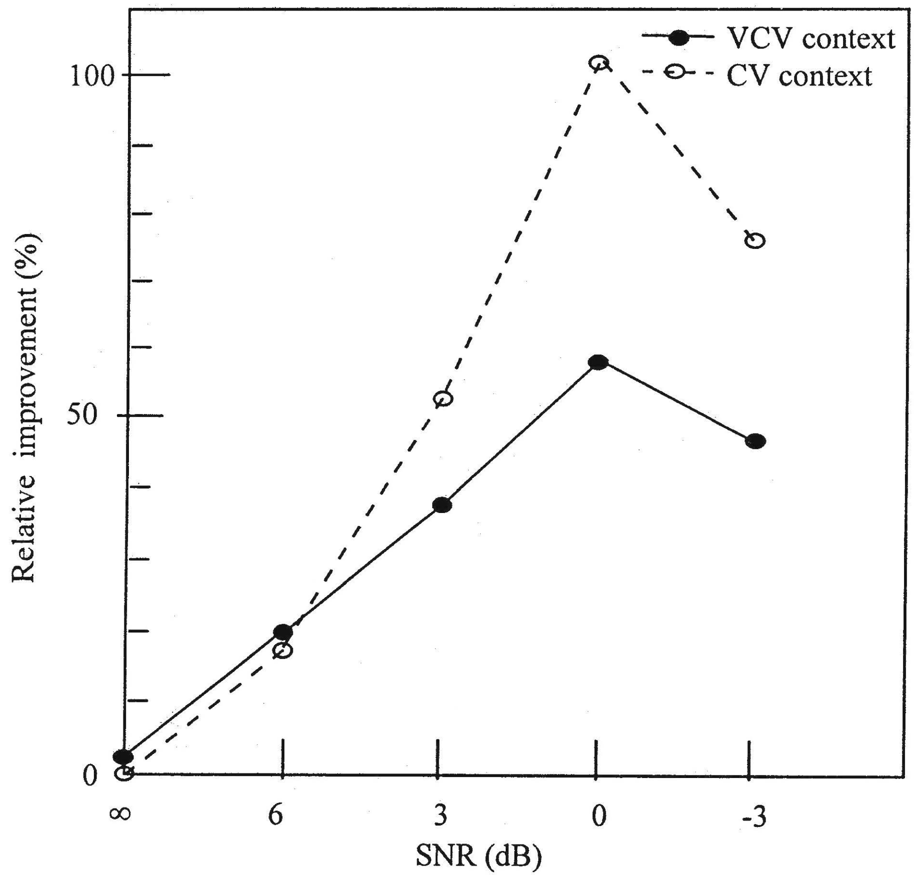


FIG. 4.17 Relative improvement in the relatve transmission of place feature, averaged for the five subjects, in VCV and CV contetxs.

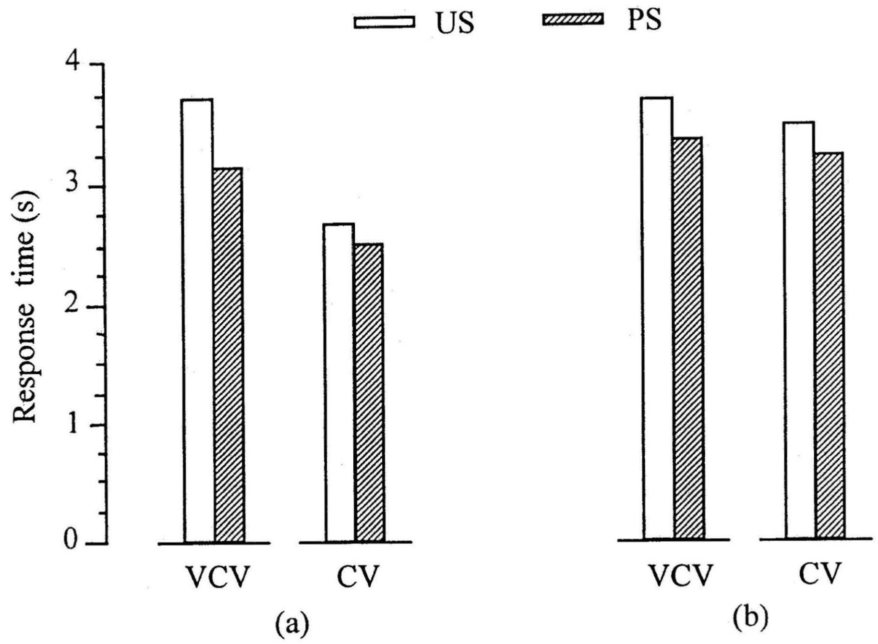


FIG. 4.18 Experiment II. Response times in VCV and CV contexts: (a) for subject SG (b) averaged for the ten subjects. US: unprocessed speech, PS: processed speech.

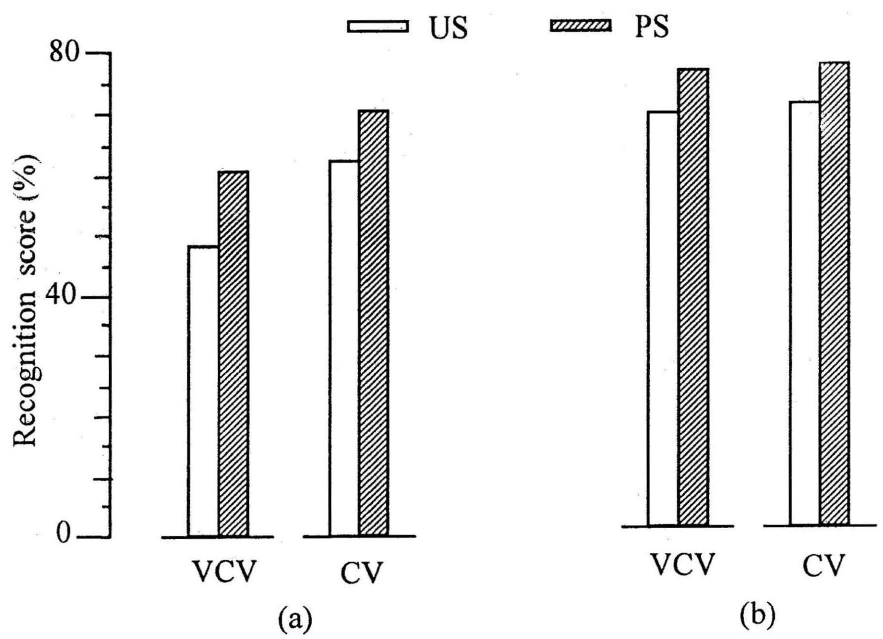


FIG. 4.19 Experiment II. Recognition scores in VCV and CV contexts: (a) for subject SG (b) averaged for the ten subjects. US: unprocessed speech PS: processed speech.

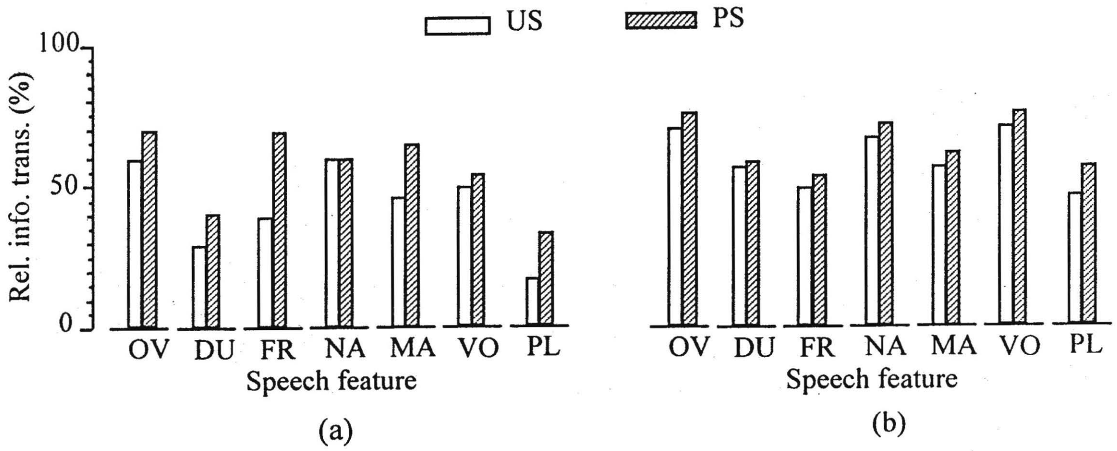


FIG. 4.20 Experiment II. Percentage relative information transmitted, in VCV context (a) for subject SG (b) averaged for the ten subjetcs. OV: overall, DU: duration, FR: frication, NA: nasality, MA: manner, VO: voicing, PL: place. US: unprocessed speech, PS: processed speech.

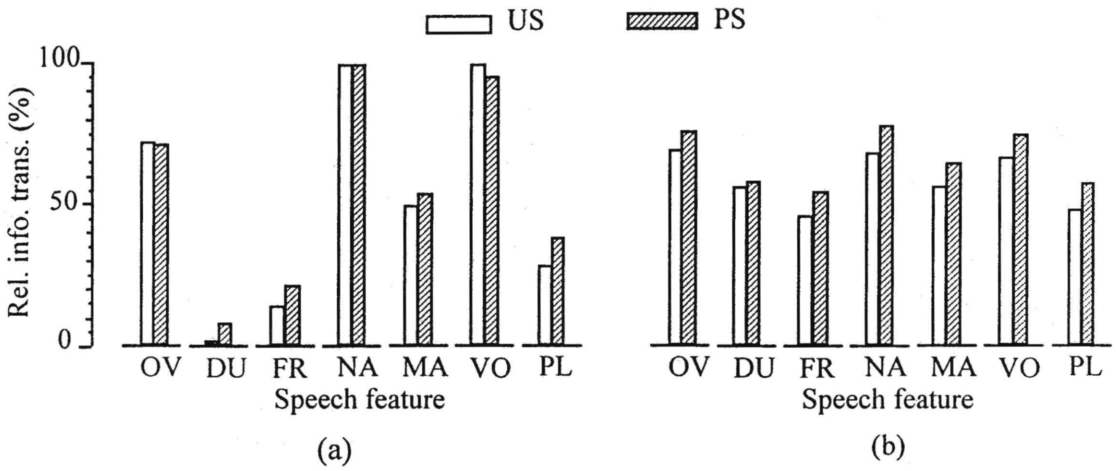


FIG. 4.21 Experiment II. Percentage relative information transmitted, in CV context (a) for subject SG (b) averaged for the ten subjetcs. OV: overall, DU: duration, FR: frication, NA: nasality, MA: manner, VO: voicing, PL: place. US: unprocessed speech, PS: processed speech.

Chapter 5

EVALUATION WITH REAL-TIME PROCESSING

5.1 Introduction

A scheme for splitting speech signal into two signals with complementary spectra, on the basis of critical band filtering for binaural dichotic presentation was presented in Chapter 3. An off-line implementation of the scheme and results from listening tests for its evaluation have been presented and discussed in chapter 4. The scheme was found helpful in improving speech quality, response time, recognition scores, and transmission of features, particularly the place feature, indicating the usefulness of the scheme for better reception of the spectral characteristics. On the basis of these results obtained from off-line processing, the scheme was implemented in real-time processing for use as a binaural hearing aid. In this chapter, we present real-time implementation of the scheme, listening tests involving subjects with bilateral hearing impairment, and the results, along with a discussion of these results.

5.2 Implementation

The real-time processing was done using two DSP boards based on 16-bit fixed point processor, TI/TMS320C50 (TMS320C5X, user's guide, 1993; TMS320C5X, starter kit user's guide, 1994). Each board consists of a processor along with an analog interface circuit (AIC) with 14-bit ADC and DAC, and a programmable timer which can be used for setting the sampling rate. The AIC also has a low pass filter (using switch capacitor circuit) at the input of ADC and at the output of DAC. The processing set-up, as shown in Fig. 5.1, consists of an input low pass filter, two DSP boards operating with sampling rate of 10 k samples/s, and two audio amplifiers.

The processing scheme was implemented in two ways. In the first, the gain of all the filter bands was kept constant (same as in case of off-line processing) This implementation is later referred as PS-CG. The off-line processing implementation had been done using cascade combination of band reject filters, as shown earlier in Fig. 4.2 and is not suitable for real-time implementation due to very high order calculations required. The second implementation provided adjustable filter gains, in the range of -3 to +3 dB, as a way of partial matching of the filter response to the frequency characteristics of the individual subject's hearing loss. The pure tone audiogram of the subject was interpolated to obtain the hearing loss up to 5 kHz. The adjustable filter magnitude response in dB, as a function of frequency f, is given as

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

where $A_c(f)$ is the gain in dB for the constant gain implementation, $\alpha(f)$ is the interpolated value of hearing loss in dB, and $\alpha_{\min}$ and $\alpha_{\max}$ are the minimum and maximum values over the 125 Hz to 5 kHz frequency range. The relationship between filter gain and hearing loss is shown in Fig. 5.2. Since 16-bit fixed-point processors are being used and amplitude compression is not being implemented, it was decided to keep the frequency compensation within ± 3 dB. The second implementation is later referred as PS-AG.

For real-time implementation, the scheme makes use of a FIR filter with comb filter magnitude response as shown earlier in Fig. 4.3 was selected. For both the desired comb filters, the magnitude response was approximated with 128 coefficients using frequency sampling technique of linear phase FIR filter design. The filter program and coefficients can be loaded into the program RAM on the DSP chip using serial port interface. Once loading is over the serial port can be disconnected, keeping DSP boards 'on.' It is to be noted that no data transfer takes place between the two boards. The program for implementing the scheme is described in Appendix C.

The log magnitude response of the two banks of filters for the PS-CG implementtation was obtained by applying sine waves of constant amplitude from 100 Hz to 5 kHz (Δf = 20 Hz) and plotted as shown in Fig. 5.3 (a), (b). The pass band ripples are within 2 dB, and side band attenuation are more than 28 dB. The transition bands are less than 90 Hz. The filter responses were also verified by obtaining spectrograms using a spectrographic analysis set-up (described in Appendix A), and three spectrograms are shown in Figs. 5.4-5.6. Fig. 5.4 shows narrowband spectrogram for random white noise, in this the complementary splitting of spectra is clearly seen. The wideband spectrograms for speech segment /*asa*/ and /*aga*/ are shown in Figs. 5.5 and 5.6 respectively. These also indicate the complementary spectra for the two channels.

In the PS-AG implementation, the frequency response of the comb filter was adjusted for individual subjects. For subject SG, the log-magnitude responses for the two ears and subject's audiogram are shown in Fig. 5.7. The filter responses were also verified spectrographically.

5.3 Listening tests

As described in the preceding section, the scheme was implemented in two ways. In the first implementation, all the filter bands had constant gain and this implementation is referred as PS-CG. In the second implementation, all the filter bands had adjustable gain as per the hearing loss in the corresponding band of an individual subject and this is referred as PS-AG. The listening tests were conducted without any masking noise on bilateral sensorineural hearing-impaired subjects for evaluation of both the implementations. The listening tests for implementation PS-CG and PS-AG are later referred as Experiment III and Experiment IV respecttively.

In Experiment III, listening tests were conducted to evaluate advantages of the processed signal presented dichotically over the unprocessed signal diotically presented for the evaluation of PS-CG implementation.

In Experiment IV, for the evaluation of PS-AG implementation, listening tests were conducted for evaluating the advantages of the PS-AG implementation over PS-CG implementation, with dichotic signal presentation in both the cases. We could have saved on total number of listening tests, by carrying out listening tests for PS-AG implementation and comparing the results obtained earlier for PS-CG implementation. However, the subjects who participated in listening tests for the evaluation of the PS-CG implementation were not going to necessarily participate in the second set of implementation. Therefore, it was decided to conduct the listening tests for PS-CG and PS-AG implementations.

5.3.1 Test material

In order to minimize the contribution of linguistic factors and maximize the contribution of acoustic factors, nonsense syllables were used for stimuli. Twelve consonants /p, b, t, d, k, g, m, n, s, z, f, v/ and the vowel /a/ as in 'father' were used in vowel-consonant-vowel (VCV) and consonant-vowel (CV) contexts, to form two setsa of stimuli with syllables. The number of stimuli was restricted to 12, so that they can be conveniently accommodated on subject's screen in the computerized test administration system. This test material was the same as used for the off-line implementation in Experiment I and II.

The acquisition and analysis of the test material has been discussed earlier in Section 4.3.1. The processing of syllables was done in real-time, by using the implementation set-up as described in 5.2.

5.3.2 Subjects

In the Experiment III and IV, the scheme was tested on bilateral sensorineural hearing impaired subjects. The subjects who participated were from different parts of India and they had no difficulty in clearly recognizing the test stimuli.

Seven hearing impaired subjects participated in the experiments. Five subjects (SG: M 27, SSN: M 31, KRV: M 49, BAS: M 58, SAV: M 46) participated in both the experiments. Subject (LDM: M 48) and subject (KIT: M 48) participated only in Experiment III and IV respectively. Thus, we had six subjects for both the experiments. The subjects were right handed and familiar with the English. The subjects had 'mild'-to-'very severe' bilateral sensorineural hearing loss. The pure tone threshold averages (PTAs) are given in Appendix E. Subjects' PTA difference between right and left ear was from 4 to 30 dB. The format of forms for subject's background and willingness are given in Appendix H.

5.3.3 Experimental procedure and set-up

The computerized test administration system used for the evaluation of off-line implementation (as described earlier in Section 4.3.3) was modified for use in the listening tests for evaluation of real-time implementation, and this set-up is shown in Fig. 5.8. It includes a PC, a PC based data acquisition card (PCL-208, from Dynalog Microsystems Limited, Mumbai) having two DAC outputs, and a subject terminal (placed in acoustically isolated chamber). In this set-up only one DAC output was used. The splitting into two signals was done by the TI/TMS 320C50 based DSP boards. The stimuli stored in computer memory were outputted at a rate of 10 k samples/s through a D/A port of data acquisition card. For diotic presentation of unprocessed speech (US) the D/A outputs were passed through antialiasing filter (f_c = 4.6 kHz), and a pair of audio amplifiers. In dichotic presentation of processed speech (PS) the D/A outputs were passed through antialiasing filter, a pair of DSP boards, and a pair of audio amplifiers. In both the presentations, the outputs of the audio amplifier drive the headphones. One serial port was used for loading the filter program to the DSP boards using a switch, and the other serial port was used for communicating with the subject terminal. The subject terminal was used for displaying the response choices on its screen and for obtaining subject responses from its keyboard.

The listening test procedure is similar to that described earlier in Section 4.3.3. The subject was seated in an acoustically isolated chamber during the testing. The presentation level was kept to the subject's most comfortable listening level and presentation were made using a pair of headphones (Telephonics TDH-39P). The telephones were calibrated before using for presentation, the procedure for calibrationn of headphones is given in Appendix D.A test consisted of 5 presentation of each stimulus, i.e. a total of 60 presentations. Each test took about 8-14 min. Two to four tests were conducted in a typical test session. Five confusion matrices with stabilized recognition scores were combined and considered for final analysis. Thus, the confusion probabilities were obtained on the basis of 25 presentations of each stimulus.

Each subject took about 4-6 hours for completion of the listening tests in VCV and CV contexts. As per the availability and willingness of the six hearing impaired subjects, the test sessions were spread over four months. There was a time gap of about two months between Experiment III and IV.

5.4 Results with PS-CG implementation: Experiment III

Results from listening tests conducted with six hearing impaired subjects in VCV and CV contexts for comparing the diotic presentation with unprocessed speech (US) and the dichotic presentation with constant gain filter implementation (PS-CG) are presented here. A compilation of subjects' qualitative assessments about the test stimuli for ascertaining the improvements in speech quality was carried out. Average response time was used for comparing the effectiveness of the processing scheme, in terms of load on perception process. The stimulus-response confusion matrices were used to obtain the recognition scores. The confusion matrices were subjected to information transmission analysis in order to obtain a measure that is not affected by subject's response bias. The twelve stimuli were combined in groups and the resulting matrices were analyzed for reception of the consonantal features of duration, frication, nasality, manner, place, and voicing.

Compilation of subjects' qualitative assessment indicated that the speech quality was better with processing for binaural dichotic presentation. The response times for both the contexts are given in Tables 5.1 and 5.2. All the subjects showed decrease in response time due to processing. The relative improvement in response time ranged from 8.8 to 13.7% and 7.7 to 22.2%, with an average of 11.5% and 12.6% across the subjects in VCV and CV contexts respectively. For most of the subjects the decrease was statistically significant, as seen by t-test, showing an improvement in listening condition due to processing. The response times, for a subject and averaged across the subjects, are plotted in Fig. 5.9, and these show a reduction for both the contexts. The subject chosen for plotting the scores happen to be the first participant in the listening tests.Paired t-test, across the subjects, showed that the decreases are highly significant ($p < 0.005$) in both the contexts.

The recognition scores for individual subjects and averaged across the subjects for both the contexts are given in Tables 5.3 and 5.4. The recognition scores for a subject and averaged across the subjects are plotted in Fig. 5.10. For all the subjects, the PS-CG scores are higher than the US scores. The percentage relative improvement in the scores (calculated as in Section 4.4) range from 9.2 to 23.6 and 14.4 to 19.2 in VCV and CV contexts respectively. Averaged across the subjects, the percentage improvements in the scores were 14 and 16.3 in VCV and CV contexts respectively. The recognition scores were subjected to t-test, for testing the statistical significance of improvements in scores due to processing. Almost all subjects showed highly significant ($p < 0.005$) improvement in both contexts. Paired t-test, across the subjects for testing the significance of the improvement due to processing, was also carried out and the improvements are highly significant ($p < 0.0005$) for both the contexts.

The confusion matrices were subjected to information transmission analysis. The overall information transmitted as well as information transmitted for specific features are given in Tables 5.5 and 5.6 for all the subjects. These are plotted for a subject and averaged across the six subjects for VCV and CV contexts in Figs. 5.11 and 5.12 respectively. The overall improvements in speech reception are contributed by better reception of almost all the six features of duration, frication, nasality, manner, place, and voicing. It is observed that, in both the contexts, these improvements are higher for the features of manner, voicing, and place. Four sub jects showed maximum improvement in place feature, while two subjects showed maximum improvements in manner and voicing features.

The relative improvements in relative transmission of information for different features are given in Table 5.7. The overall relative improvement in transmission was contributed by all the six features of duration, frication, nasality, manner, place, and voicing, with nearly maximum improvement for the place feature. Averaged across the six subjects, the relative improvement for place feature were 34 and 41 % in VCV and CV contexts respectively.

5.5 Results with PS-AG gain implementation: Experiment IV

Results from listening tests, conducted with six hearing impaired subjects in VCV and CV contexts for comparing dichotic presentation with constant filter gain implementation (PS-CG) and adjustable filter gain (PS-AG) are presented here. A compilation of subjects' qualitative assessments for ascertaining the improvement in speech quality was carried out. Average response time can be used for comparing the effectiveness of the processing scheme, in terms of the load on the perception process. The stimulus-response confusion matrices were used to obtain the recognition scores. The confusion matrices were subjected to information transmission analysis in order to obtain a measure that is not affected by subject's response bias. The twelve stimuli were combined in groups and the resulting matrices were analyzed for reception of the consonantal features of duration, frication, nasality, manner, place, and voicing.

Compilation of subjects' qualitative assessment with two kinds of processing did not show a clear preference for either of the two implementations. The response times for both the contexts are given in Tables 5.8 and 5.9. The relative improvement in response time range from 3.3 to 13.7 %. And 4.1 to 9.5%, with an average of 6.7% and 6.8% across the subjects in VCV and CV contexts respectively. The response time, for a subject and averaged across the subjects is plotted in Fig. 5.13. These show the reduction in the response time with PS-AG over PS-CG. The response times were subjected to t-test for testing statistical significance of decrease in response time. For individual subjects, the decreases in the response time were, in general, not very significant. However, paired t-test across the subjects, showed that the decreases in response time were highly significant ($p < 0.005$) in both the contexts.

The recognition scores for individual subjects and averaged across the subjects for both the contexts are given in Tables 5.10 and 5.11. The scores, for a subject and averaged across the subjects, are plotted in Fig. 5.14. Out of the six subjects, five had participated in the first implementation i.e. Experiment III also. For these subjects the scores for PS-CG in Tables 5.10 and 5.11 are about the same as the corresponding scores for PS in Tables 5.3 and 5.4. All the subjects showed higher scores under PS-AG as compared to the scores for PS-CG. The percentage relative improvement in the scores (as calculated in Section 4.4) range from 2.2 to 6.4 and 1.6 to 7.8 in VCV and CV contexts respectively. Averaged across the subjects, the percentage improvement in the scores are 5 and 3.9 in VCV and CV contexts respectively. The recognition scores were subjected to t-test, for testing the statistical significance of improvements in scores with PS-AG over PS-CG. In both contexts, three subjects showed highly significant improvement ($p < 0.025$). Paired t-test, across the subjects for testing the significance of the processing was also carried out and the improvements are highly significant ($p < 0.005$) for both the contexts.

The results of relative information transmitted are given in Tables 5.12 and 5.13 and are plotted, for a subject and averaged across the six subjects, for VCV and CV contexts in Figs. 5.15 and 5.16 respectively. In these results, the overall improvements were contributed by almost all the six features for some subjects, whereas some subjects showed improvements for few features only. In case of individual subjects, improvements in some features are accompanied by degradation for some other feature(s). Almost all the subjects showed modestly higher transmission for place and manner features, in VCV context. In CV context, one subject showed highest improvement for place feature, one subject for place and voicing features, and the others showed very little or no improvement for place feature.

The relative improvements in the transmission of different features are given in Ta ble 5.14. The relative improvement for individual subject is modest, in the range of 2 to 7 % and 1 to 11 % for VCV and CV context respectively. Improvements for some features are seen to be accompanied by decrease for some other features, and different subjects have showed improvements for different features. For most of the subjects higher relative improvements in transmission are observed for the place feature. Averaged across the six subjects, for the place features, these are 17 and 10 % for VCV and CV contexts respectively.

5.6 Discussion

For real-time implementation of the scheme for dichotic presentation, listening tests for two types of implementation were conducted i.e. Experiment III and IV. In the implement-tation PS-CG, the gain for all the bands was constant. In the implementation PS-AG, the frequency response of the filter was altered within ± 3 dB as a partial compensation for the frequency dependence of the hearing loss of the individual subjects. In Experiment III, listening tests, involving six subjects with bilateral hearing impairment, were conducted to evaluate the advantages of PS-CG over unprocessed speech (US). Listening tests in Experiment IV, on six hearing impaired subjects, were carried out to evaluate the advantages of PS-AG over PS-CG. The evaluation was done by comparing (a) qualitative assessment of the stimuli, (b) response times, (c) recognition scores as obtained from confusion matrices, and (d) information transmission for various features.

In Experiment III, the qualitative assessment indicated a definite preference for PS-CG over unprocessed speech. In the response time analysis, most of the hearing impaired subjects showed significant decrease in response time for processed speech. This indicates an improvement in listening condition with processing. All the subjects showed highly significant improvement in recognition score due to processing, indicating the usefulness of the implemented scheme. Information transmission analysis of the stimulus-response confusion matrices indicated that the overall improvement is contributed by better reception of all the six features of duration, frication, nasality, manner, place, and voicing, with nearly maximum improvement for place feature for almost all the subjects.

In Experiment IV, the qualitative assessment of stimuli did not show a clear preference for either PS-CG or PS-AG. There is a decrease in response time due to PS-AG over PS-CG. However, this is statistically not very significant. In recognition scores, three subjects (out of six) showed highly significant advantage of PS-AG over PS-CG. In information transmission analysis, the overall improvement is contributed by different features. The relative improvements are highest for the place feature. But these are in case of some subjects accompanied by relative deterioration for some other feature(s). The relative improvements as well as deterioration are of modest value.

For gaining an insight into relationship between improvements in processing and the extent and nature of hearing loss of individual subjects, the relative improvements in recognition scores and transmission of place feature were averaged for the VCV and CV contexts and these are given in Table 5.15 for both the experiments. In Experiment III, we see that the improvements in recognition scores are highly related to improvements in transmission of place feature. The improvements are maximum for subjects SG, the subjects with highest level of hearing loss. Therefore we may provisionally conclude that more the extent of sensorineural loss, i.e. higher the extent of spectral masking, more the improvement due to processing. However, for a final conclusion, psychophysical measurements of the spread of spectral masking for the individual subject need to be carried out. In case of Experiment IV, the additional improvements due to PS-AG do not show any clear relationship between the improvements in the recognition scores and the improvements in the transmission of place feature. The improvements are modest in value and do not appear to be related to the extent of the hearing loss of individual subjects.

On the basis of Experiment III for evaluating the advantages of dichotic presentation using PS-CG over diotic presentation using unprocessed speech, it may be concluded that PS-CG improves the quality of speech and results in a decrease in response time indicating reduction in the load on the perception process. The processing results in significant improvement in recognition scores and information transmission. It is seen that a large part of improvement is contributed by place feature. As the place information is related to frequency resolving capacity of the auditory process, one can say that the implemented scheme has reduced the effect of spectral masking without adversely affecting the reception of the features cued by amplitude and duration.

The Experiment IV was carried out to study additional advantages obtained by adjusting the magnitude response of filter bands. In this study, we have not employed any dynamic range compression, and the frequency response adjustment was within ± 3 dB. Averaged across the subjects, the test results show a mild advantage of PS-AG over PS-CG, but these advantages are significant for some subjects only. Out of the six subjects, three (SSN, KRV, BAS) showed highly significant improvement in recognition scores. Inspection of audiogram of these subjects indicates less than 25 dB variation of hearing loss with frequency. For other three subjects, the improvements are not as significant. These subjects have relatively larger variation in hearing loss (85, 40, 32 dB for SG, SAV, and KIT respectively).

Information transmission analysis for various speech features. It is seen that the place feature does not make a distinct contribution to advantages of PS-AG over PS-CG. This indicates that PS-CG is effective in reducing the effect of spectral masking, because of separation of speech spectra in accordance with critical band filtering. Further shaping of filter bands in accordance with hearing loss characteristics of the individual subject does not contribute to a further significant reduction in the effect of spectral masking. However, it does help in overall better reception. For subjects with high degree of frequency dependence of hearing loss, advantages of gain adjustments over a range larger than ± 3 dB along with multiband amplitude compression needs to be investigated.

TABLE 5.1 Experiment III. Response time in VCV context. US: unprocessed speech, PS: processed speech with constant gain filter (PS-CG). t-test for individual subjects: *n* (number of tests) = 5, df = 8; paired t-test for scores averaged across the subjects: *n* (number of subjects) = 6, df =5.

Subject	Response Time (s)				Relative Improvement (%)	t-test (one-tailed)	
	US		PS				
	mean	s.d.	mean	s.d.		t	p
SG	3.82	0.09	3.35	0.30	12.3	3.36	0.005
SSN	3.91	0.73	3.41	0.61	12.8	1.18	0.2
KRV	3.93	0.42	3.54	0.38	9.9	1.54	0.1
BAS	3.18	0.38	2.79	0.38	12.3	1.62	0.1
SAV	4.21	0.19	3.84	0.33	8.8	2.17	0.05
LDM	4.01	0.42	3.46	0.31	13.4	2.36	0.025
Avg.	3.84	0.35	3.40	0.34	11.5	15.05	0.0005

TABLE 5.2 Experiment III. Response time in CV context. US: unprocessed speech, PS: processed speech with constant gain filter (PS-CG). t-test for individual subjects: *n* (number of tests) = 5, df = 8; paired t-test for scores averaged across the subjects: *n* (number of subjects) = 6, df =5.

Subject	Response Time (s)				Relative Improvement (%)	t-test (one-tailed)	
	US		PS				
	mean	s.d.	mean	s.d.		t	p
SG	3.26	0.09	3.01	0.22	7.7	2.35	0.025
SSN	3.97	0.69	3.19	0.08	19.7	2.51	0.025
KRV	3.66	0.23	3.37	0.19	7.9	2.17	0.05
BAS	3.34	0.21	2.95	0.34	11.7	2.18	0.05
SAV	4.23	0.29	3.29	0.33	22.2	4.78	0.0025
LDM	3.92	0.28	3.16	0.49	19.4	3.01	0.0125
Avg.	3.73	0.38	3.26	0.16	12.6	4.75	0.005

TABLE 5.3 Experiment III. Recognition Scores in VCV context. US: unprocessed speech, PS: processed speech. t-test for individual subjects: *n* (number of tests) = 5, df = 8; paired t-test for scores averaged across the subjects: *n* (number of subjects) = 6, df =5.

Subject	**Recognition Score (%)**				**Relative Improvement (%)**	**t-test (one-tailed)**	
	US		**PS**				
	mean	**s.d.**	**mean**	**s.d.**		**t**	**p**
SG	60.7	3.8	75.0	1.7	23.6	7.68	0.0005
SSN	58.7	2.2	68.3	3.9	16.4	4.79	0.0025
KRV	84.0	3.4	92.3	3.5	9.9	3.80	0.005
BAS	76.7	2.7	87.7	0.9	14.3	8.64	0.0005
SAV	81.3	1.4	90.0	1.7	10.7	8.83	0.0005
LDM	84.0	2.5	91.7	2.0	9.2	5.38	0.0005
Avg.	74.23	11.58	84.17	10.05	14.0	10.02	0.0005

TABLE 5.4 Experiment III. Recognition Scores in CV context. US: unprocessed speech, PS: processed speech. t-test for individual subjects: *n* (number of tests) = 5, df = 8; paired t-test for scores averaged across the subjects: *n* (number of subjects) = 6, df =5.

Subject	**Recognition Score (%)**				**Relative Improvement (%)**	**t-test (one-tailed)**	
	US		**PS**				
	mean	**s.d.**	**mean**	**s.d.**		**t**	**p**
SG	74.3	3.0	86.7	3.4	16.7	6.11	0.0005
SSN	72.3	2.5	82.7	1.9	14.4	7.41	0.0005
KRV	74.3	1.9	86.0	2.5	15.8	8.33	0.0005
BAS	69.3	1.9	81.0	2.0	16.9	9.48	0.0005
SAV	79.7	2.6	95.0	1.2	19.2	11.94	0.0005
LDM	76.0	1.4	87.0	2.5	14.5	8.58	0.0005
Avg.	74.3	3.5	86.4	4.8	16.2	16.68	0.0005

TABLE 5.5. Experiment III. Relative information transmitted in VCV context. US : unprocessed speech, PS : processed speech with constant filter gain (PS-CG)

	Percentage Relative Information Transmitted													
Feature	**Overall**		**Duration**		**Frication**		**Nasality**		**Manner**		**Voicing**		**Place**	
Subject	**US**	**PS**	**US**	**PS**	**US**	**PS**	**US**	**PS**	**US**	**PS**	**US**	**PS**	**US**	**PS**
SG	65	72	34	50	49	59	58	75	52	64	60	78	25	48
SSN	60	68	23	22	38	33	29	40	31	40	48	82	33	39
KRV	85	90	100	100	91	100	73	72	82	88	93	97	58	76
BAS	80	86	91	92	66	62	91	100	76	77	71	82	61	80
SAV	84	91	53	65	37	58	91	96	58	73	87	95	78	86
LDM	83	91	100	100	65	71	91	94	75	80	84	97	68	84
Avg.	76	83	67	72	58	64	72	80	62	70	74	89	54	69

TABLE 5.6. Experiment III. Relative information transmitted in CV context. US : unprocessed speech, PS : processed speech with constant filter gain (PS-CG)

	Percentage Relative Information Transmitted													
Feature	**Overall**		**Duration**		**Frication**		**Nasality**		**Manner**		**Voicing**		**Place**	
Subject	**US**	**PS**	**US**	**PS**	**US**	**PS**	**US**	**PS**	**US**	**PS**	**US**	**PS**	**US**	**PS**
SG	78	84	48	68	44	66	100	100	66	79	100	100	43	62
SSN	64	80	68	85	54	72	46	69	53	71	52	64	52	77
KRV	74	85	94	100	68	81	49	78	63	80	84	85	51	67
BAS	72	81	79	94	65	79	94	89	78	83	64	81	46	58
SAV	76	94	87	91	56	94	85	82	69	89	65	84	65	92
LDM	72	83	76	95	66	75	76	81	70	77	79	77	46	72
Avg.	73	85	75	89	59	78	75	83	67	80	74	82	51	71

TABLE 5.7 Experiment III. Relative improvement in relative information transmission of different features. OV: overall, DU: duration, FR: friction, NA: nasality, MA: manner, VO: voicing, PL: place.

Subject	Relative Improvement (%)													
	Context: VCV							Context: CV						
	OV	**DU**	**FR**	**NA**	**MA**	**VO**	**PL**	**OV**	**DU**	**FR**	**NA**	**MA**	**VO**	**PL**
SG	11	47	20	29	23	30	92	8	42	50	0	20	0	44
SSN	13	-4	-13	38	29	71	18	25	25	33	50	34	23	48
KRV	6	0	10	-1	7	4	31	15	6	19	59	27	1	31
BAS	8	1	-6	10	1	15	31	13	19	22	-5	6	27	26
SAV	8	23	57	5	26	9	10	24	5	68	-4	29	29	42
LDM	10	0	9	3	7	15	24	15	25	14	7	10	-3	57
Avg.	9	11	13	14	16	24	34	17	20	34	18	21	13	41

TABLE 5.8. Experiment IV. Response times in VCV context. PS-CG : processed speech with constant filter gain, PS-AG : processed speech with adjustable filter gain. NS = not significant. t-test for individual subjects : *n* (number of tests) = 5, df = 8; paired t-test for scores averaged across the subjects : *n* (number of subjects) = 6, df =5.

Subject	Response Time (s)				Relative Improve-ment (%)	t-test (one-tailed)	
	PS-CG		PS-AG				
	mean	s.d.	mean	s.d.		t	p
SG	3.43	0.48	3.31	0.41	3.5	0.43	NS
SSN	3.60	0.28	3.48	0.26	3.3	0.70	0.25
KRV	4.39	0.26	3.98	0.23	9.3	2.64	0.025
BAS	2.92	0.63	2.71	0.43	7.2	0.62	NS
SAV	3.50	0.57	3.02	0.44	13.7	1.49	0.1
KIT	4.61	1.60	4.41	1.58	4.3	0.20	NS
Avg.	3.74	0.64	3.49	0.62	6.7	4.13	0.005

TABLE 5.9. Experiment IV. Response times in CV context. PS-CG : processed speech with constant filter gain, PS-AG : processed speech with adjustable filter gain. NS = not significant. t-test for individual subjects : *n* (number of tests) = 5, df = 8; paired t-test for scores averaged across the subjects : *n* (number of subjects) = 6, df = 5.

Subject	Response Time (s)				Relative Improve-ment (%)	t-test (one-tailed)	
	PS-CG		PS-AG				
	mean	s.d.	mean	s.d.		t	p
SG	3.16	0.11	2.86	0.29	9.5	2.16	0.05
SSN	3.17	0.82	2.99	0.73	5.7	0.36	NS
KRV	3.95	0.39	3.64	0.23	7.9	1.53	0.1
BAS	2.93	0.13	2.81	0.16	4.1	1.30	0.2
SAV	2.73	0.27	2.49	0.08	8.8	1.91	0.05
KIT	5.21	0.59	4.93	0.55	5.4	0.78	0.25
Avg.	3.53	0.92	3.29	0.89	6.8	7.79	0.0005

TABLE 5.10 Experiment IV. Recognition scores in VCV context. PS-CG : processed speech with constant filter gain, PS-AG : processed speech with adjustable filter gain. t-test for individual subjects : *n* (number of tests) = 5, df = 8; paired t-test for scores averaged across the subjects : *n* (number of subjects) = 6, df = 5.

Subject	**Recognition Score (%)**				**Relative Improve-ment (%)**	**t-test (one-tailed)**	
	PS-CG		**PS-AG**				
	mean	**s.d.**	**mean**	**s.d.**		**t**	**p**
SG	75.3	1.4	79.3	6.1	5.3	1.43	0.1
SSN	67.3	2.5	71.3	2.2	5.9	2.69	0.025
KRV	87.7	1.9	92.0	2.7	4.9	2.91	0.0125
BAS	89.0	2.5	94.7	2.2	6.4	3.83	0.0025
SAV	91.3	3.5	93.3	3.1	2.2	0.96	0.2
KIT	64.0	2.5	67.3	2.5	5.2	2.09	0.05
Avg.	79.1	11.9	83.09	12.0	5.0	7.83	0.0005

TABLE 5.11 Experiment IV. Recognition scores in CV context. PS-CG : processed speech with constant filter gain, PS-AG : processed speech with adjustable filter gain. t-test for individual subjects : *n* (number of tests) = 5, df = 8; paired t-test for scores averaged across the subjects : *n* (number of subjects) = 6, df = 5.

Subject	**Recognition Score (%)**				**Relative Improve-ment (%)**	**t-test (one-tailed)**	
	PS-CG		**PS-AG**				
	mean	**s.d.**	**mean**	**s.d.**		**t**	**p**
SG	77.3	3.0	83.3	5.7	7.8	2.08	0.05
SSN	81.0	1.9	82.3	2.2	1.6	1.00	0.2
KRV	91.0	1.9	94.0	1.9	3.3	2.49	0.025
BAS	84.7	2.7	89.7	3.0	5.9	2.77	0.0125
SAV	95.3	1.4	97.3	1.9	2.1	1.89	0.05
KIT	76.7	1.2	78.7	2.2	2.6	1.78	0.1
Avg.	84.4	7.5	87.6	7.3	3.9	4.2	0.005

TABLE 5.12 Experiment IV Relative information transmitted in VCV context. CG : processed speech with constant gain filter (PS-CG), AG: processed speech with adjustable gain filter (PS-AG).

	Percentage Relative Information Transmitted													
Feature	Overall		Duration		Frication		Nasality		Manner		Voicing		Place	
Subject	CG	AG	CG	AG	CG	AG	CG	AG	CG	AG	CG	AG	CG	AG
SG	72	75	40	48	42	53	79	83	57	65	89	84	45	53
SSN	67	69	32	17	42	33	53	66	51	54	74	65	32	35
KRV	86	91	95	100	97	91	66	74	83	84	95	90	60	76
BAS	88	94	95	100	67	84	94	91	77	86	93	93	84	93
SAV	92	94	64	83	61	70	100	100	76	82	95	100	82	91
KIT	65	66	43	41	50	49	56	53	55	52	73	73	29	37
Avg.	78	82	62	65	59	63	75	78	67	71	87	84	55	64

TABLE 5.13 Experiment IV Relative information transmitted in CV context. CG : processed speech with constant gain filter (PS-CG), AG : processed speech with adjustable gain filter (PS-AG).

	Percentage Relative Information Transmitted													
Feature	Overall		Duration		Frication		Nasality		Manner		Voicing		Place	
Subject	CG	AG	CG	AG	CG	AG	CG	AG	CG	AG	CG	AG	CG	AG
SG	74	82	80	92	43	54	95	100	65	72	86	92	47	53
SSN	77	78	76	76	72	72	71	71	74	74	55	56	69	70
KRV	89	93	91	100	83	91	72	85	80	94	84	90	75	88
BAS	84	89	100	100	82	82	100	100	89	89	91	97	63	76
SAV	94	96	81	100	87	95	82	100	86	97	92	100	90	93
KIT	76	78	86	86	91	91	85	85	89	89	54	58	49	51
Avg.	82	86	86	92	76	81	84	90	81	86	77	82	66	72

TABLE 5.14 Experiment IV. Relative improvement in relative transmission of different features. OV: overall, DU: duration, FR: friction, NA: nasality, MA: manner, VO: voicing, PL: place.

Sub-ject	Relative Improvement (%)													
	Context: VCV							Context: CV						
	OV	**DU**	**FR**	**NA**	**MA**	**VO**	**PL**	**OV**	**DU**	**FR**	**NA**	**MA**	**VO**	**PL**
SG	4	20	26	5	14	-6	18	11	15	26	5	11	7	13
SSN	3	-47	-	25	6	-12	9	1	0	0	0	0	2	1
KRV	6	5	-6	12	1	-5	27	4	10	10	18	18	7	17
BAS	7	5	25	-3	12	0	11	6	0	0	0	0	7	21
SAV	2	30	15	0	8	5	11	2	23	9	22	13	9	3
KIT	2	-5	-2	-5	-5	0	28	3	0	0	0	0	7	4
Avg.	4	1	6	6	6	-3	17	5	8	7	8	7	6	10

TABLE 5.15 Experiment III and IV. Relative improvement in recognition score (RS) and transmission of place features (Place Tr.), averaged for VCV and CV contexts. Subjects' pure tone average (PTA) hearing threshold for the two ears are also given.

Subject	**PTA hearing Threshold (dB)**		**Experiment III: Average Improvement (%)**		**Experiment IV: Average Improvement (%)**	
	Left	**Right**	**RS**	**Place Tr.**	**RS**	**Place Tr.**
SG	73	77	20.2	68.0	6.6	15.5
SSN	75	65	15.4	33.0	3.8	5.0
KRV	60	52	12.9	31.0	4.1	22.0
BAS	33	38	15.6	25.5	6.2	16.0
SAV	45	67	15.0	26.0	2.2	7.0
LDM	52	82	11.9	40.5	--	--
KIT	42	88	--	--	3.9	16.0

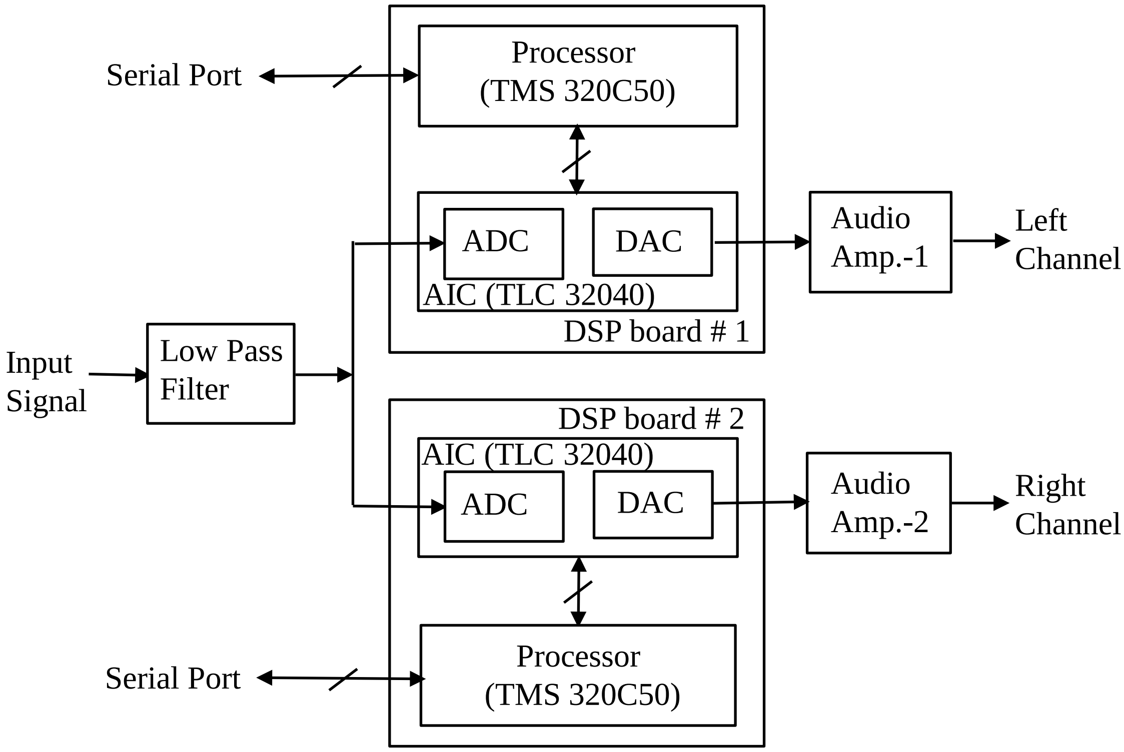


FIG. 5.1 Speech processing set-up using two TI/TMS320C50 based DSP boards for dichotic presentation.

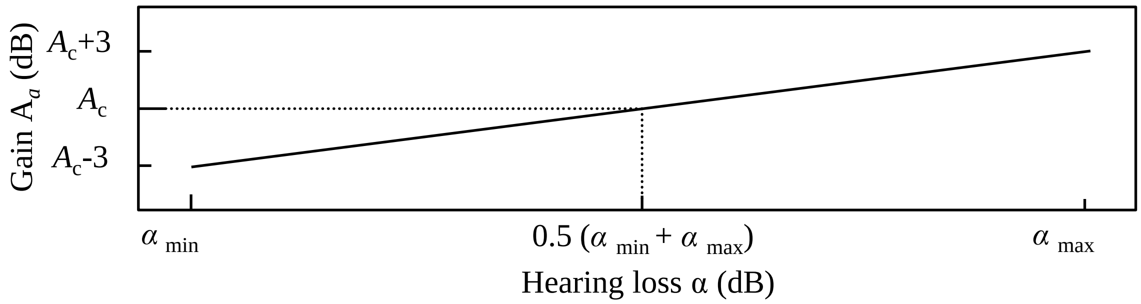
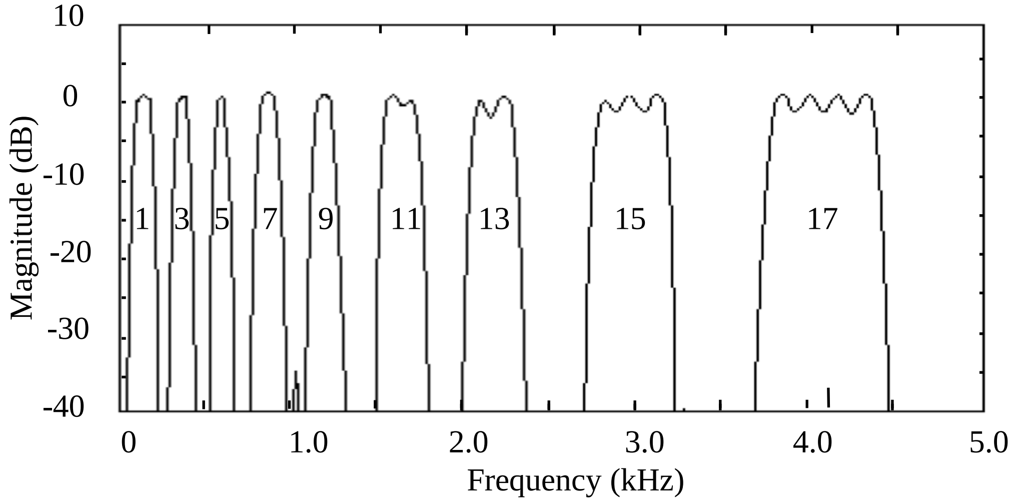
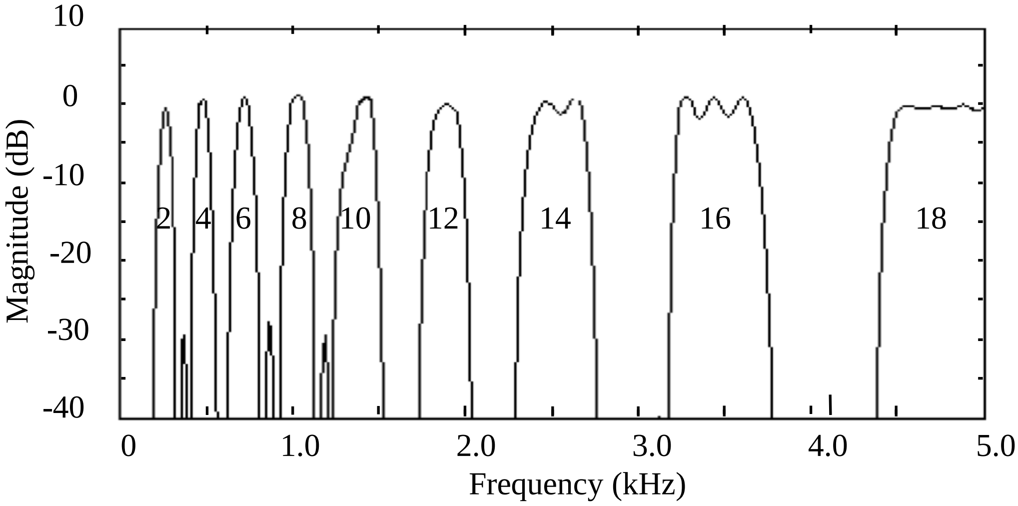


FIG. 5.2 Relationship between filter gain and hearing loss, used for frequency dependent gain compensation in PS-AG implementation.



(a)



(b)

FIG. 5.3 Magnitude response of the filters used in real-time processing implementation, PS-CG: (a) left ear (b) right ear.

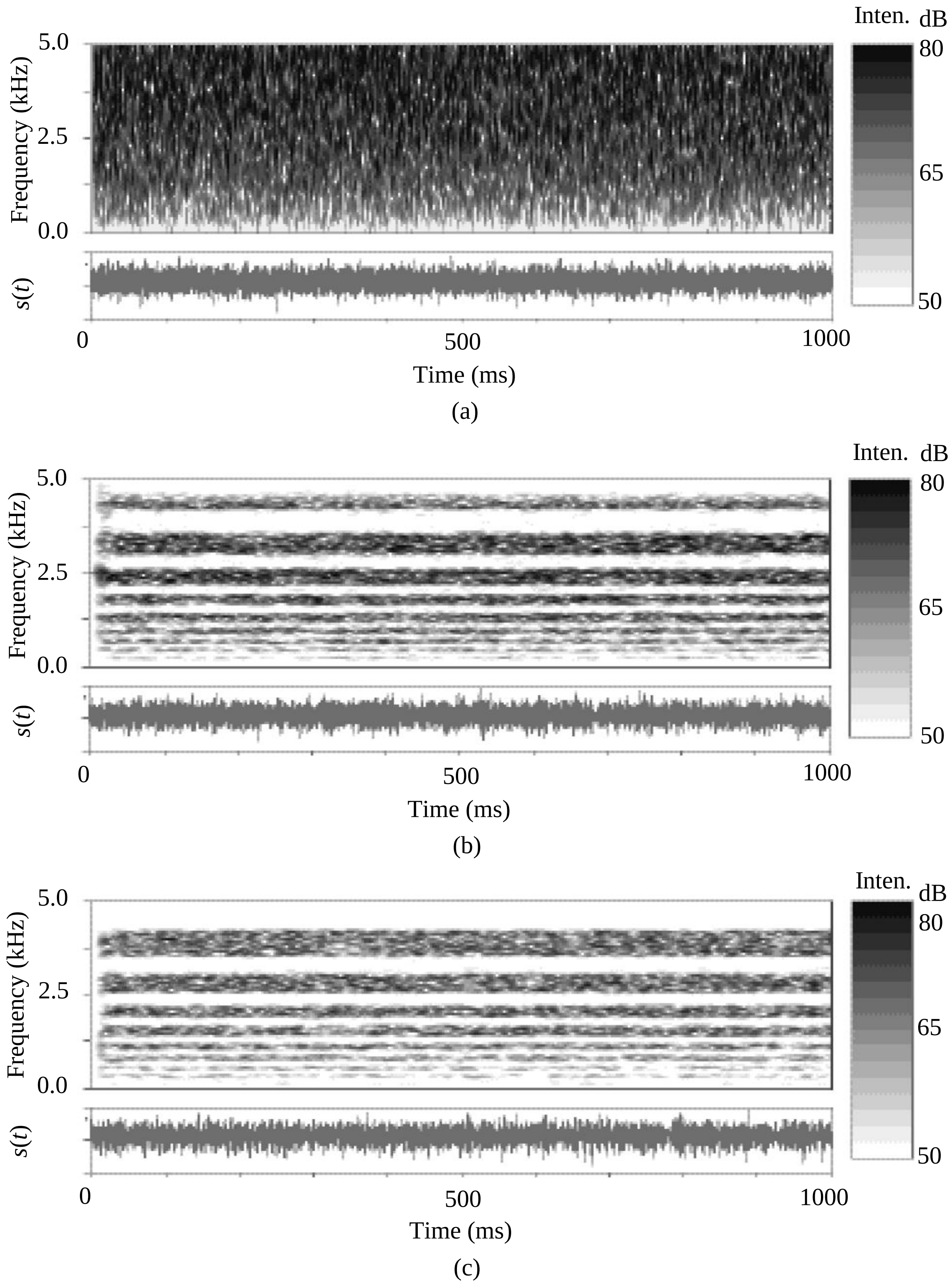
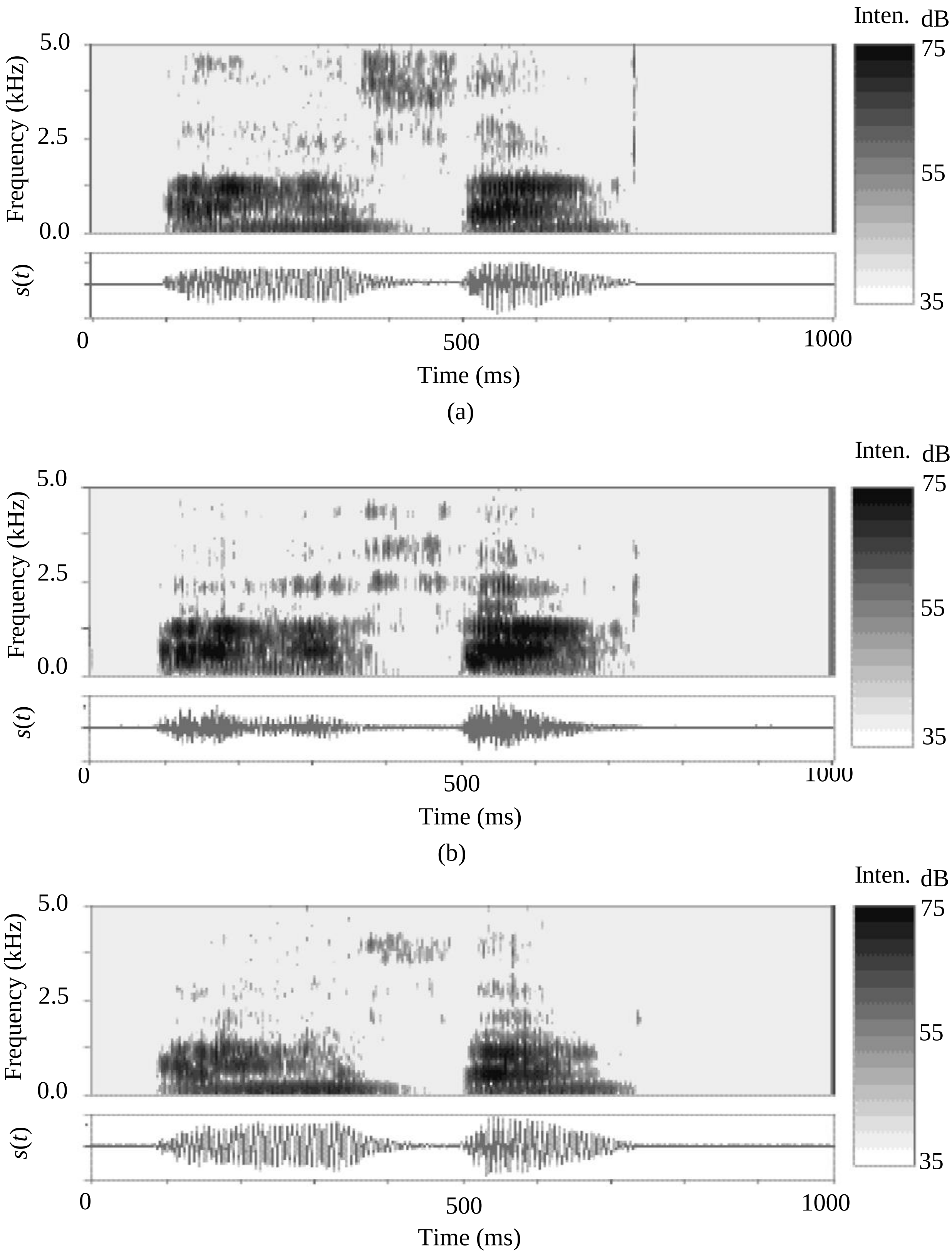


FIG. 5.4 Narrow-band spectrograms of random noise $s(t)$, processed using implementation PS-CG (a) unprocessed (b) processed (left ear) (c) processed (right ear). S. R. = 10 k samples/s and Δf = 53 Hz.



(a)

(b)

(c)

FIG. 5.5 Wide-band spectrograms of speech waveforms $s(t)$, for utterance /*asa*/, processed using implementation PS-CG: (a) unprocessed (b) processed (left ear) (c) processed (right ear). S. R. = 10 k samples/s and Δf =300 Hz.

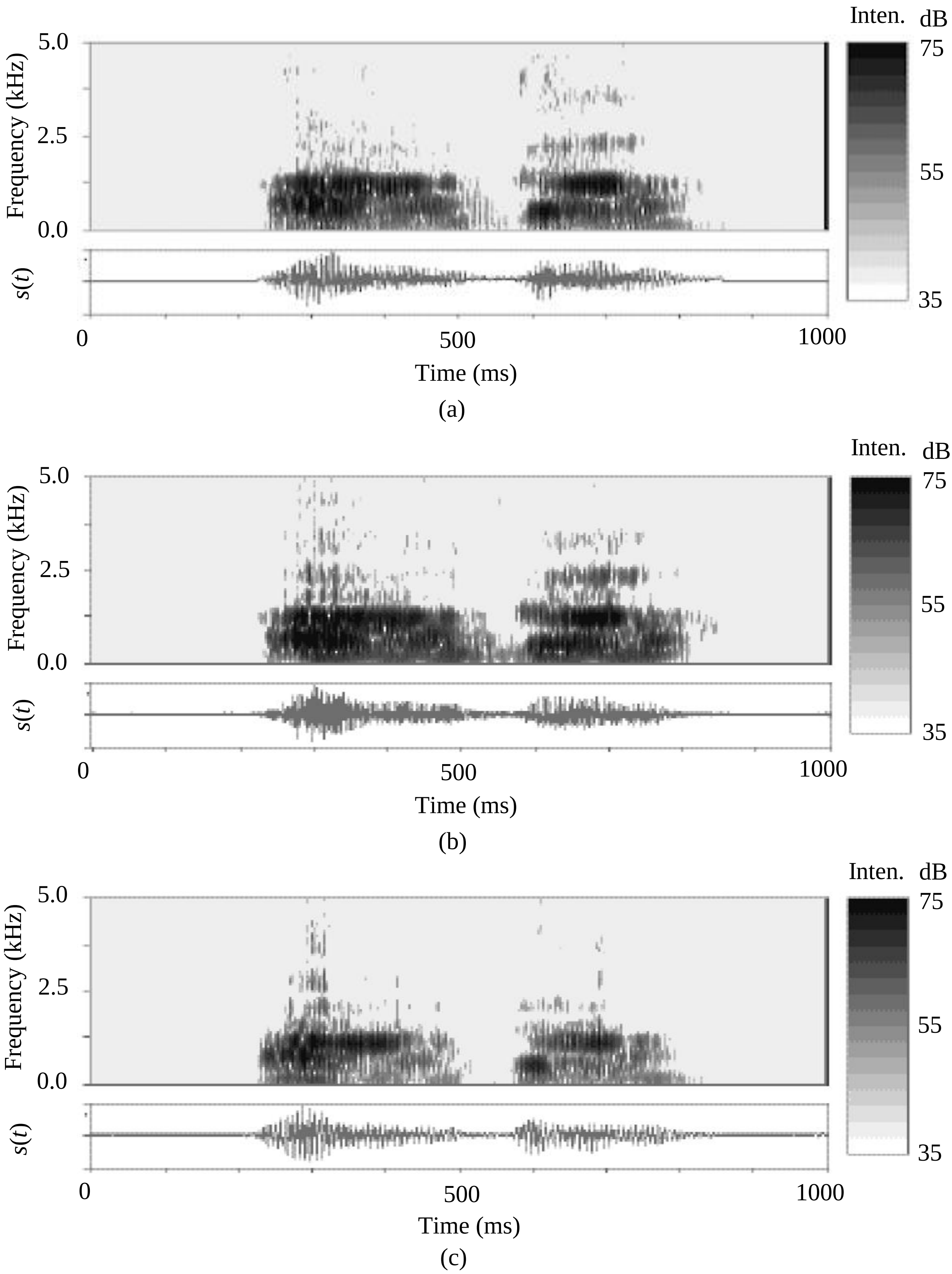


FIG. 5.6 Wide-band spectrograms of speech waveforms $s(t)$, for utterance /*aga*/, processed using implementation PS-CG : (a) unprocessed (b) processed (left ear) (c) processed (right ear). S. R.=10 k samples/s and Δf =300 Hz.

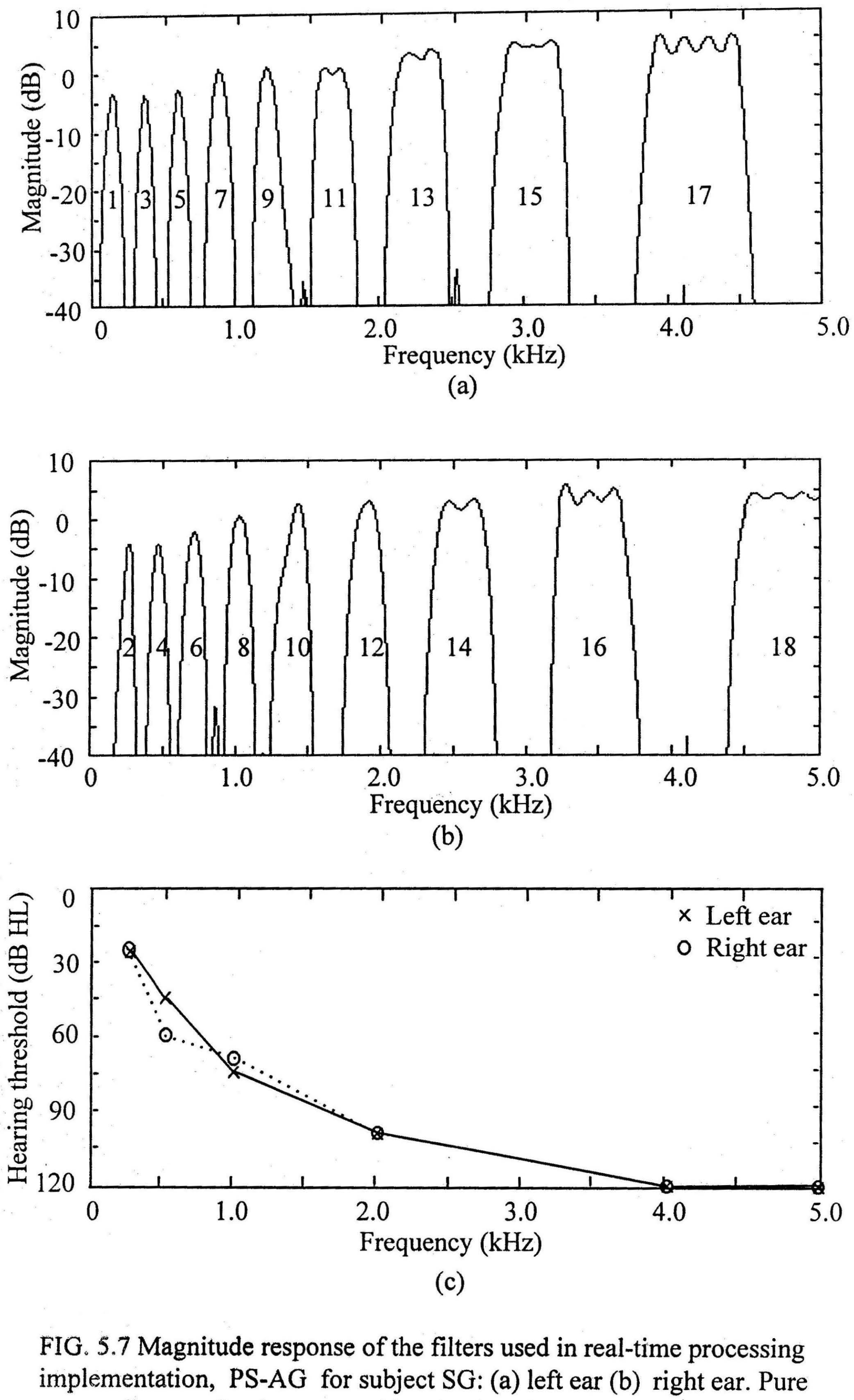


FIG. 5.7 Magnitude response of the filters used in real-time processing implementation, PS-AG for subject SG: (a) left ear (b) right ear. Pure tone audiogram for the same subject is given in (c).

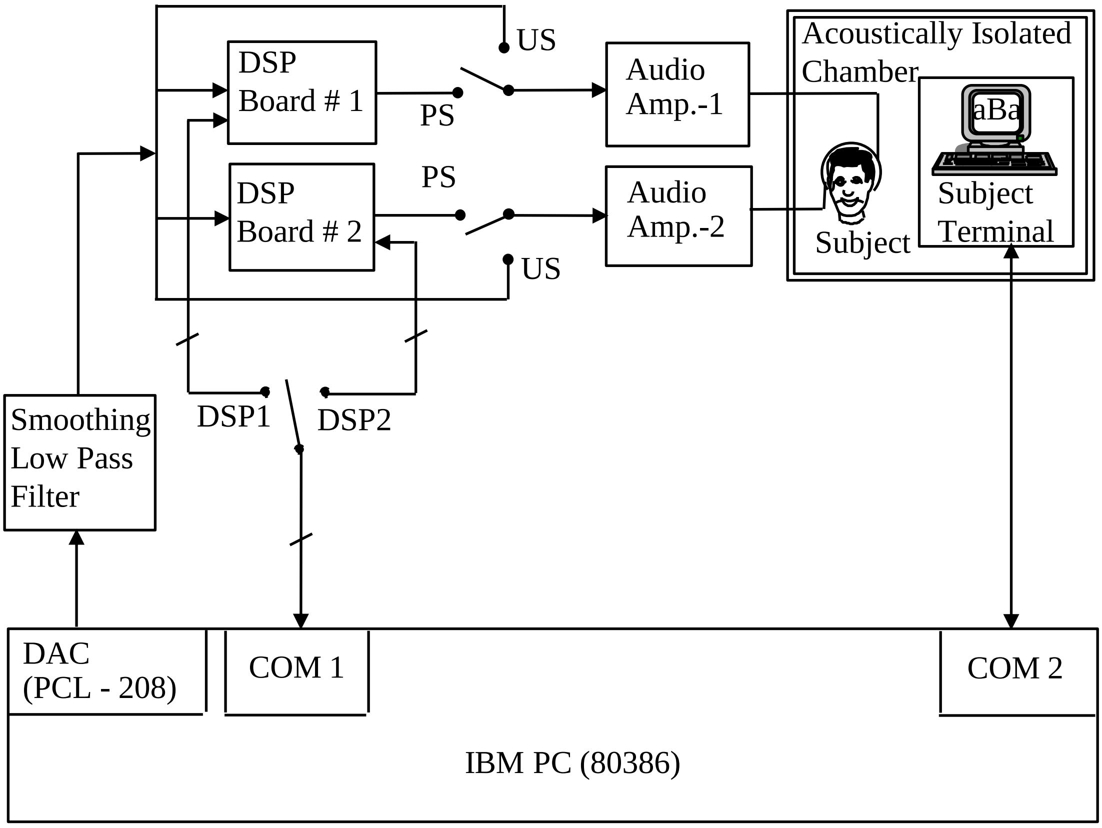


FIG. 5.8 Experimental set-up for the computerized test administration of listening tests for the evaluation of real-time processing implementation.

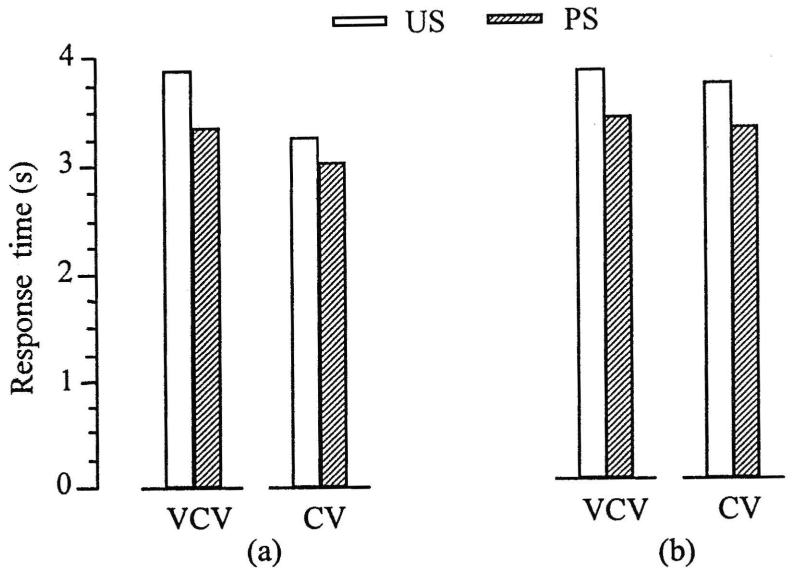


FIG. 5.9 Experiment III. Response times in VCV and CV contexts: (a) for subject SG (b) averaged for the six subjects. US: unprocessed speech, PS: processed speech.

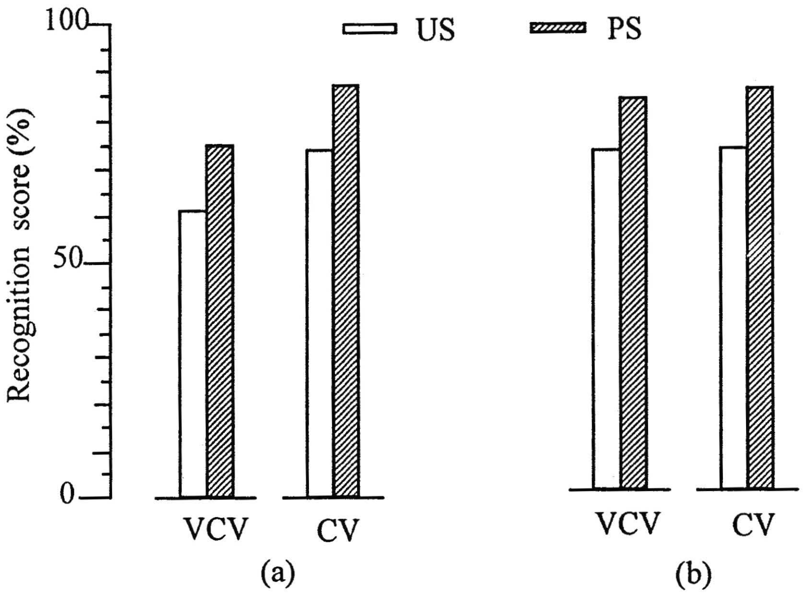


FIG. 5.10 Experiment III. Recognition scores in VCV and CV contexts: (a) for subject SG (b) averaged for the six subjects. US: unprocessed speech, PS: processed speech.

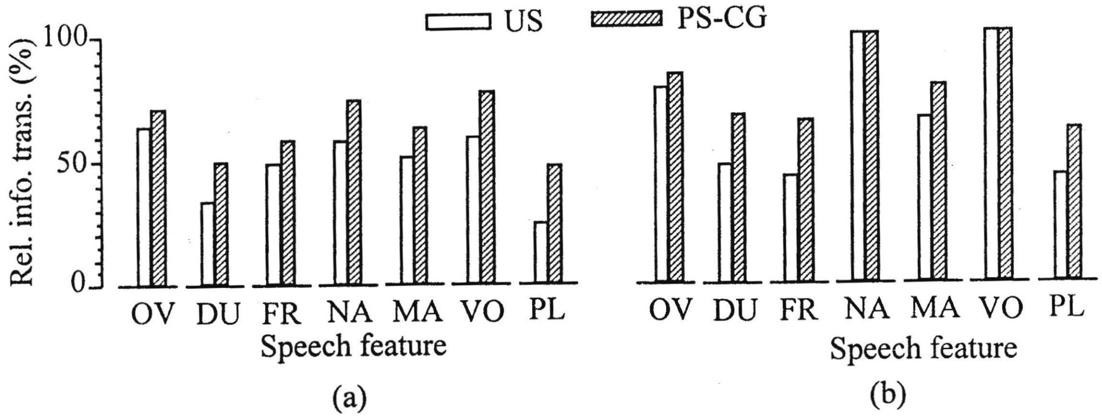


FIG. 5.11 Experiment III. Percentage relative information transmitted for subject SG: (a) VCV and (b) CV contexts. OV: overall, DU: duration, FR: frication, NA: nasality, MA: manner, VO: voicing, PL: place. US: unprocessed speech, PS-CG: processed speech with constant gain filter.

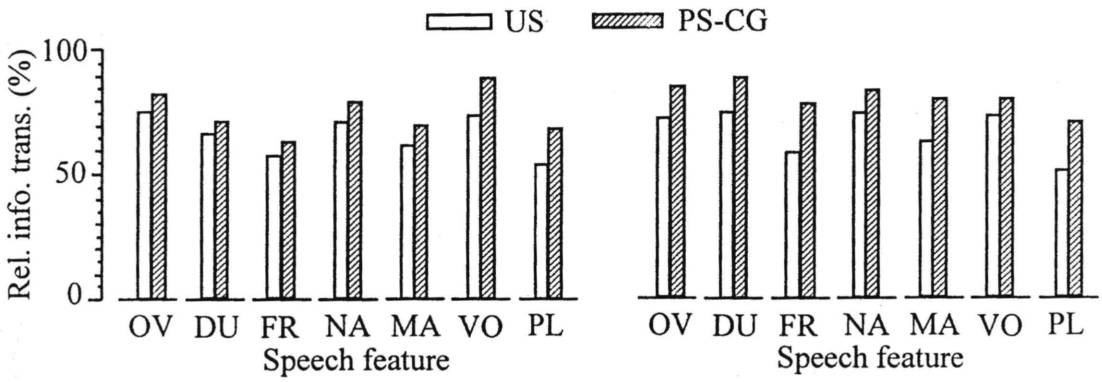


FIG. 5.12 Experiment III. Percentage relative information transmitted averaged for the six subjects: (a) VCV and (b) CV contexts. OV: overall, DU: duration, FR: frication, NA: nasality, MA: manner, VO: voicing, PL: place. US: unprocessed speech, PS-CG: processed speech with constant gain filter.

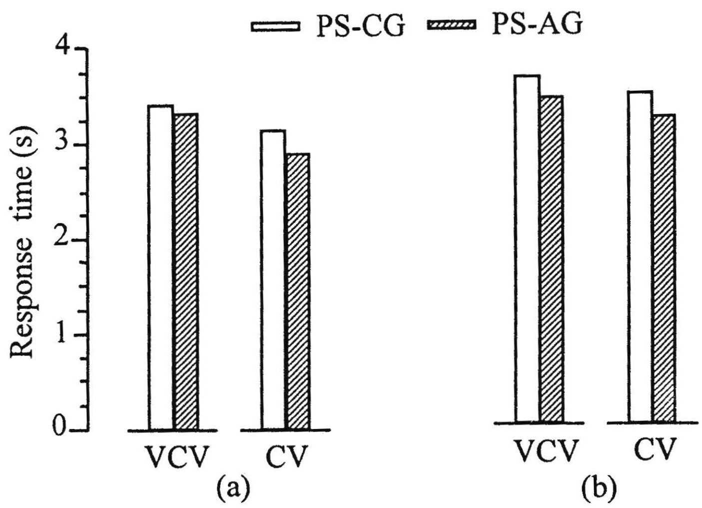


FIG. 5.13 Experiment IV. Response times in VCV and CV contexts: (a) for subject SG (b) averaged for the six subjects. PS-CG: processed speech with constant filter gain, PS-AG: processed speech with adjustable filter gain.

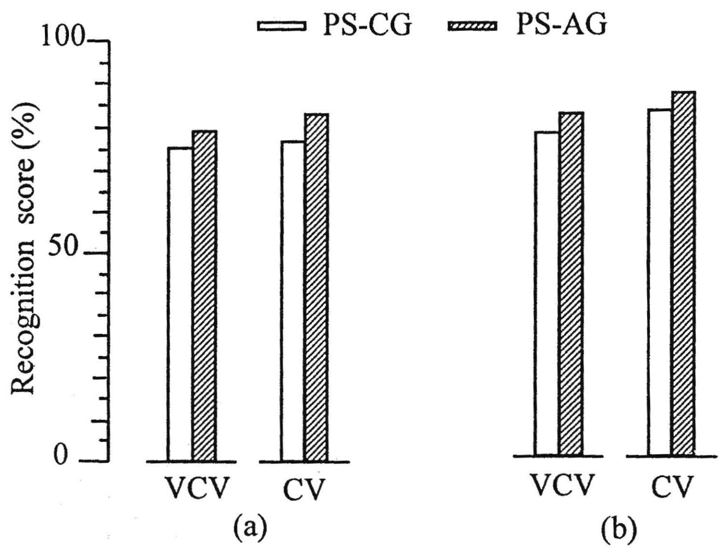


FIG. 5.14 Experiment IV. Recognition scores in VCV and CV contexts: (a) for subject SG (b) averaged for the six subjects. PS-CG: processed speech with constant filter gain, PS-AG: processed speech with adjustable filter gain.

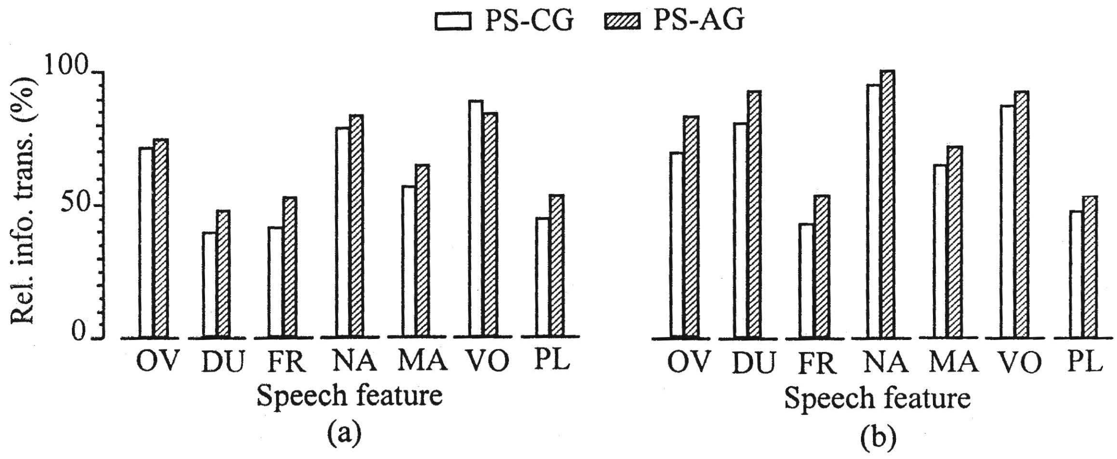


FIG. 5.15 Experiment IV. Percentage relative information transmitted for subject SG: (a) VCV and (b) CV contexts. OV: overall, DU: duration, FR: frication, NA: nasality, MA: manner, VO: voicing, PL: place. PS-CG: processed speech with constant gain filter, PS-AG: processed speech with adjustable gain filter.

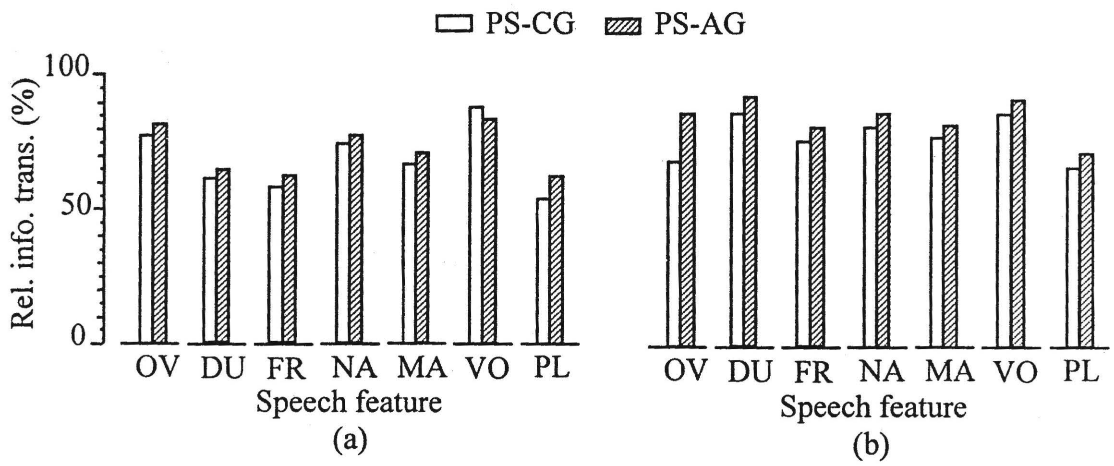


FIG. 5.16 Experiment IV. Percentage relative information transmitted averaged for the six subjects: (a) VCV and (b) CV contexts. OV: overall, DU: duration, FR: frication, NA: nasality, MA: manner, VO: voicing, PL: place. PS-CG: processed speech with constant gain filter, PS-AG: processed speech with adjustable gain filter.

Chapter 6

SUMMARY AND CONCLUSIONS

6.1 Introduction

Sensorineural impairment of the hearing mechanism is associated with decrease in frequency resolving capacity of the auditory system due to spread of spectral masking along the cochlear partition. As the consonantal place feature is cued by spectral differences, the hearing impaired persons may find difficulties particularly in identifying this feature. A possible solution for this problem is a binaural dichotic presentation by splitting the speech signal, using a filter bank and adding signals from alternate bands for presenting to the two ears. Thus, two adjacent bands which are likely to mask each other get presented to different ears.

Lyregaard (1982) used constant bandwidth comb filters (implemented using analog time delay) for dichotic presentation. Improvements in the scores were not statistically significant. He suggested that the lack of significant improvement could be attributable to one of the three factors: unsuitable filtering, insufficient listening experience by the subjects, and non-feasibility of binaural fusion of dichotic signals. Lunner *et al.* (1993) tested the use of an 8-channel digital filter bank in monaural, diotic, and dichotic modes. The filter bank was designed to give 8 parallel filtered outputs which are added together with individually adjustable weighting factors in order to obtain a proper fit of the gain frequency response of the hearing aid as per the need of the individual hearing aid user. By combining alternate bands together, the filter bank was used for dichotic presentation. All the filters in the filter bank had bandwidth of 700 Hz. The experimental evaluation of speech recognition in noise was done by finding the speech-to-noise ratio which satisfy the 50% correct word recognition. The results indicated an overall improvement in speech-to-noise ratio of about 2 dB for the dichotic conditions over diotic.

The research reported here involves implementation and experimental evaluation of a scheme for splitting speech for binaural dichotic presentation. This scheme uses critical bands corresponding to auditory filters based on psychophysical tuning curves as described by Zwicker (1961). We used 18 critical band filters over 5 kHz frequency range. The magnitude response for each band was an approximation of an ideal filter with bandwidth of the critical band. Initially we studied the effect of the scheme using constant filter gains. Further investigation was done using adjustable filter gains (in the range of -3 to +3 dB) as a way of partial matching of the filter response to the frequency characteristics of the individual subject's hearing loss. In order to maintain the timing related cues, it will be desirable to preserve the relative phases of the frequency components in speech signal. To achieve this, the magnitude response of the filters was coupled with linear phase response.

For ascertaining the improvement in the speech quality due to processing, a compilation of qualitative assessment, by the subjects, about the test stimuli under various listening conditions was carried out. Average response was used for comparing the effectiveness of the processing scheme, in terms of load on the perception process. The stimulus-response confusion matrices were analyzed for obtaining recognition scores and relative transmission of information . The cell entries in the confusion matrix were used to obtain confusion matrices by grouping stimuli with the same features, for studying the contribution of various speech features.

The reported scheme was aimed at reducing the effect of spectral masking due to loss of spectral resolution. Hence it should result in improvement in the reception of "place" feature without adversely affecting the reception of other features. The test stimuli consisted of nonsense syllables, with twelve English consonants /p, b, t, d, k, g, m, n, s, z, f, v/, in VCV and CV contexts, and the vowel being /*a*/. These were used for studying the reception of consonantal features of duration, frication, nasality, manner, place, and voicing. A PC-based set-up was developed for (i) signal acquisition and analysis (ii) off-line and real-time processing (iii) automated administration of listening tests.

The scheme was implemented for off-line processing of digitized speech signals. Two sets of listening tests were carried out: (a) Experiment I involving normal hearing subjects with simulated hearing loss, (b) Experiment II involving hearing-impaired subjects. The scheme was able to improve speech quality, response time, recognition scores, and transmission of features, particularly the place feature, indicating the usefulness of the scheme for better reception of the spectral characteristics. On the basis of these results obtained from off-line processing, the scheme was implemented for real-time processing for use as a binaural hearing aid. Two sets of listening tests, involving subjects with bilateral hearing impairment, were conducted (a) Experiment III to evaluate the advantages of processed speech with constant filter gain for all bands (PS-CG) over unprocessed speech, (b) Experiment IV to evaluate the advantages of processed speech with adjustable filter band gains (PS-AG) over PS-CG.

The conclusions drawn on the basis of results obtained from the experiments of splitting speech signal to reduce the effect of spectral masking, in off-line and real-time processing are given in the following sections, followed by some suggestions for further studies and development.

6.2 Experiments using off-line processing

Two sets of listening test, Experiment I and II, were conducted for evaluation of the off-line implementation of the processing scheme for dichotic presentation. In Experiment I, listening tests were carried out on five normal hearing subjects with hearing loss simulated by mixing broadband noise as a masker at different SNR conditions. The listening tests in Experiment II were conducted on ten subjects with bilateral sensorineural hearing loss.

Compilation of qualitative assessment for normal hearing subjects indicated that they preferred the processed dichotic presentation over unprocessed diotic presentation. For all the subjects averaged response time increased with increase in the level of masking noise (i.e. decrease in SNR). This indicate that the masking noise resulted in an increased load on the perception process. The recognition scores and information transmitted decreased with decrease in SNR. For unprocessed speech, most of the decrease in the relative information transmission could be attributed to decrease in the reception of place feature. This indicates that masking noise resulted in a simulation of spectral masking. It was observed that response time for processed speech was significantly lower than that for unprocessed one. Further, it was observed that for a particular level of masking noise, the score for processed speech was significantly higher than that for the unprocessed one ($p < 0.05$). For different subjects, the percentage relative improvements in scores for -3 dB SNR condition ranged from 5.6 to 19.4 and 4.9 to 38.2 in VCV and CV contexts respectively. The important finding was that the improvements due to processing were more for higher levels of masking noise, i.e. higher levels of simulated sensorineural loss. However, these improvements tend to level at very high levels of simulated loss. Paired t-test, across the subjects for testing the significance of improvement in recognition scores due to processing, was carried out and improvements are highly significant ($p < 0.05$) for all SNR conditions.

Information transmission analysis of the stimulus-response confusion matrices indicated that a better reception of all the six features (duration, frication, nasality, manner, place, and voicing) contributed to an overall improvement in relative information transmitted. Nearly maximum improvement was observed for place, among the features of place, voicing, and manner, in almost all normal hearing subjects. For -3 dB SNR condition, the relative improvement in relative information transmission of place features ranged from 24 to 81 % and 16 to 267 % in VCV and CV contexts respectively. As the place information is related to frequency resolving capacity of the auditory process, one can say that the implemented scheme has reduced the effect of spectral masking.

In the Experiment II listening tests were conducted on ten subjects with bilateral hearing loss. Compilation of qualitative assessment for the subjects indicated that they preferred the processed dichotic presentation to unprocessed diotic presentation. For most of the subjects, the averaged response time significantly decreased for processed speech as compared to unprocessed one. This indicates an improvement in listening condition and reduction in load on perception process due to processing. Paired t-test, across the subjects showed that the decrease in response time are highly significant ($p < 0.005$) in both the contexts. Most of the subjects indicated highly significant improvement ($p < 0.01$) in recognition score. For different subjects, the percentage relative improvements ranged from 4.6 to 26.6 and 6.4 to 25.1 in VCV and CV contexts respectively. Paired t-test, across the subjects for testing the significance of improvement in recognition scores due to processing, was carried out and improvements are highly significant ($p < 0.0005$) in both the contexts.

Information transmission analysis of the stimulus-response confusion matrices indicated that an overall improvement is contributed by reception of all the six features (duration, frication, nasality, manner, place, and voicing) with nearly maximum improvement for place, among the features of place, voicing, and manner, in almost all hearing-impaired subjects. Nearly maximum relative improvement in relative information transmission was observed for the place feature. These improvements for the place feature ranged from 9 to 94 % and 2 to 46 % in VCV and CV contexts respectively. Averaged across the ten subjects, the relative improvements were 29 and 25 % in VCV and CV contexts respectively. As the place information is related to frequency resolving capacity of the auditory process, one can say that the implemented scheme has reduced the effect of spectral masking.

From the qualitative assessment of speech, response time statistics, recognition scores and information transmission analysis of confusion matrices, it can be concluded that the scheme of splitting of speech on the basis of critical band filtering, for dichotic presentation is helpful in improving speech quality, reducing the load on perception process, and in improving the reception of spectrally coded place feature without adversely affecting the reception of the features cued by amplitude and duration.

6.3 Experiments using real-time processing

For real-time implementation of the scheme for dichotic presentation, tests for two types of implementation, PS-CG and PS-AG, were conducted. In implementation PS-CG, the gain for all the filter bands was constant. In implementation PS-AG, the magnitude response of the filter was altered to vary the gain of the bands within ± 3 dB as a partial compensation for the frequency dependence of the hearing loss of the individual subjects. Two sets of listening tests, both Experiment III and IV, involving subjects with bilateral hearing impairment, were conducted to evaluate the real-time implementation: (a) Experiment III: for evaluating the advantages of PS-CG over unprocessed speech (US) and (b) Experiment IV for evaluating the advantages of PS-AG over PS-CG. The evaluation was done by comparing (i) qualitative assessment of the stimuli, (ii) response times, (iii) recognition scores as obtained from confusion matrices, and (iv) information transmission for various features as obtained from analysis of confusion matrices.

In Experiment III, the qualitative assessment by the six subjects indicated a definite preference for PS-CG over unprocessed speech. Most of the hearing impaired subjects showed statistically significant decrease in response time for processed speech. The percentage relative improvement in the recognition scores was highly significant ($p < 0.005$) and ranged from 9.2 to 23.5 and 14.4 to 19.2 in VCV and CV contexts respectively. Information transmission analysis of the stimulus-response confusion matrices indicated that the overall improvement is contributed by reception of all the six features of duration, frication, nasality, manner, place, and voicing. Nearly maximum improvement was observed for place, among the features of place, voicing, and manner, for almost all the subjects. These improvements for the place feature ranged from 10 to 92 % and 26 to 57 %, and averaged across the six subjects the relative improvements were 34 and 41 %, in VCV and CV contexts respectively.

In Experiment IV, the qualitative assessment of stimuli by the six subjects did not show a clear preference for either PS-CG or PS-AG. There was a decrease in response time due to PS-AG over PS-CG. However, this was statistically not very significant. Paired t-test, across the subjects, showed that improvements are highly significant (p < 0.005) for both the contexts. In recognition scores, three subjects (out of six) showed highly significant advantage of PS-AG over PS-CG. For different subjects, the percentage relative improvement in the scores ranged from 2.2 to 6.4 and 1.6 to 7.8 in VCV and CV contexts respectively. Paired t-test, across the subjects, showed that improvements are highly significant (p < 0.005) for both the contexts. In information transmission analysis, the overall improvement was contributed by different features, but there was no distinct contribution by the place feature.

On the basis of Experiment III, for evaluating the advantages of dichotic cy reponse slope in or presentation using PS-CG over unprocessed speech it may be concluded that PS-CG improved the quality of speech and resulted in a decrease in response time indicating reduction in the load on the perception process. The processing resulted in significant improvement in recognition scores and information transmission. It was seen that a large part of improvement was contributed by place feature. As the place information is related to frequency resolving capacity of the auditory processing, one can say that the implemented scheme reduced the effect of spectral masking.

The Experiment IV was conducted to study additional advantages obtained by adjusting the magnitude response of filter bands, in accordance with frequency dependence of hearing loss of the individual subject. The frequency dependent gain variation was restricted to ± 3 dB (due to restrictions post by signal processing implementation). It is to be noted that adjustable frequency response is the feature of the most of the advance signal processing hearing aids, and the frequency response slope in our implementation is considerably less that implemented in such hearing aids. Averaged across the subjects, the test results showed a mild advantage of PS-AG over PS-CG. These advantages were significant for subjects having relatively less variation (< 25 dB) in the hearing threshold over the 5 kHz frequency range. It was seen that the place feature did not make a distinct contribution to advantages of PS-AG over PS-CG. This indicates that PS-CG was effective in reducing the effect of spectral masking, because of separation of speech spectra in accordance with critical band filtering. Further shaping of filter bands in accordance with hearing loss characteristics of the individual subject did not contribute to the reduction in the effect of spectral masking due to dichotic presentation.

6.4 Conclusions

In summary, the present study established that the binaural dichotic presentation of speech signal, by splitting it into two signals with complementary short-time spectra by using comb filters with magnitude response based on critical bands (corresponding to auditory filters) and linear phase response results in

- Improvement in overall speech quality
- Reduction in response time, indicating an improvement in listening condition and less load on perception process
- Improvements in recognition scores and overall relative information transmission for consonants in nonsense syllable test

Maximum improvement in the consonantal information transmission is contributed by place feature. This indicates that the processing scheme reduces the effect of spectral masking at the cochlear level. The scheme does not adversely affecting the reception of other consonantal features.

Further, the dichotic binaural presentation scheme can be combined with shaping of comb filter magnitude response with accordance with individual subjects audiogram for additional improvements in speech reception.

6.5 Suggestion of future work

It is possible that dichotic presentation with comb filter designed using bandwidths somewhat different from the Zwicker's estimation of auditory filter bandwidths give better speech reception. Further experiments are needed to establish the optimal values of filter bandwidths.In our filter design, the transmission were kept the sharpest possible, which resulted in pass-band ripple up to 2 dB and sidelobes of up to 28 dB. The filter can be redesigned with a trade-off between transition widths and ripples. Further experiments are needed to study the effects of variations in transition widths in order to establish filter design with optimal filter parameters (Rabiner el al, 1970; Ratanpal, 2000).

Effect of gain adjustment of the filter shapes to compensate for individual subject's hearing loss needs to be investigated in more details. This will need an implementation in which larger gain variations are possible along with the flexibility of using dynamic range compression for each band.

In this study, only the effect of spectral masking was considered; the scheme can further be extended to also reduce the effect of temporal masking. The previous studies have reported that the temporal masking effect can be reduced by using the properties of "clear" speech for consonantal enhancement by modification in intensity and duration (Picheny *el al*., 1985, 1986; Thomas *el al*., 1996). This requires the analysis/synthesis of the speech signal which needs a processing time delay extending over several sub-segment duration. It has been earlier reported that a delay of up to 120 ms in speech processing and stimulus encoding shouldn't interferes with the benefits of auditory signal in audiovisual comprehension of connected speech (Pandey *el al*., 1986). Therefore, speech processors could be built for reducing the effect of both the temporal and spectral masking. The scheme for temporal splitting is also likely to help in reducing the effect of temporal masking (Jangamshetty and Pandey, 2000). Studies need to be carried out for establishing appropriate combination of temporal and spectral splitting.

On the basis of favourible results obtained with normal hearing subjects at different SNR conditions of masking noise, it can be suggested that the processing scheme for dichotic presentation can be used for improving speech perception under adverse listening conditions. The processing can be integrated as a part of the multimedia systems to be used by person with sensorineural hearing impairment (Chaudhari and Pandey, 1999).

Spectral shape of auditory signal and relationship between binaurally received signal provide the necessary cues for sound source localization (Moore 1997; Hartmann, 1998). Investigations are required for assessing the effect of dichotic processing scheme on source localization. In case of tones with several harmonics, localization of the sound may not be affected, since the spectral splitting does not involve constant bandwidth filter and therefore harmonics are not likely to be presented to the same ear. However, narrow-band stimuli are presented to one ear only. The processing scheme may be modified to detect such stimuli changeover to diotic presentation in such a case.

For implementing the scheme as part of a binaural hearing aid, the factors like power consumption, circuit complexity, etc. are to be taken into consideration. Finally, real-life studies need to be carried out with patients wearing these binaural hearing aids. These studies need to be carried out with different types of test material.

Appendix A

SPECTROGRAPHIC ANALYSIS SET-UP

Short-time spectral characteristics of speech signal are helpful in correlating acoustic to phonetic characteristics. A spectrogram is a visual representation of temporal variations in spectral magnitudes at various frequencies of a dynamic signal. In a spectrogram, time varying spectral characteristics are displayed as a two-dimensional plot, with time and frequency along x and y axes respectively. The spectral magnitudes as a function of time and frequency, are viewed as intensity (gray level/color) variations (Koenig *et al.*, 1946; Kersta, 1948; Potter *et al.*, 1966; Oppenheim, 1970; Rabiner & Schafer, 1978; O'Shaughnessy, 1987). Spectrographic analysis also find applications in the study of music, analysis of Doppler ultrasound signals, vibration analysis, study of biomedical phenomena, etc. The analog spectrographic analyzer posed restriction on the size of the segment to be analyzed, and provided very limited option for changing the frequency resolution in the spectral analysis. Further, the display offered a limited dynamic range and the analysis was time consuming. Spectrograms with much greater dynamic range and adjustable time and frequency resolutions can be digitally generated and displayed on a monitor for readouts (Oppenheim, 1970; Thomas *et al.*, 1994). Spectral analysis of digitized waveform can be carried out by either using a bank of digital filters or computing discrete Fourier transform (DFT). The DSP chips have helped in cost-effective implementation of spectrographic analyzers, e.g. Kay Elemetric Sona-Graph 5500 using TMS320C20 DSP chip (Morris, 1985; Morris, 1988).

Making use of the earlier work at IIT Bombay (Thomas *et al.*, 1994; Thomas, 1996; Prasad, 1996; Baragi, 1996) a spectrographic analyzer is developed using a PC with VGA card and a DSP board, for signal acquisition, editing, and spectrographic analysis. The speed of analysis and display is improved by partitioning the tasks appropriately between the PC and the DSP board. The details of spectrographic analyzer are as the following.

A.1 Digital spectrographic analysis

Spectrograms can be generated by obtaining magnitude spectrum of digitized waveforms by using either a digital filter bank or short-time Fourier transform (Rabiner & Schafer, 1978; O'Shaughnessy, 1987), and displaying time-frequency plots. The short-time Fourier transform of a sampled waveform is

$$X(n,k) = \sum_{m=0}^{L-1} w(m)\, x(n-m)\, \mathrm{e}^{-\mathrm{j}2\pi km/N}; \quad 0 \le k \le N-1 \qquad (1)$$

where n is the number of discrete time samples, k is the discrete frequency and N is the DFT size. The window *w(m)* is an L-point ($L<N$) Hamming window given by

$$w(m) = 0.54 - 0.46\cos(2\pi m/(L-1)); \qquad 0 \le m \le L-1 \qquad (2)$$

Frequency spectrum is computed using fast Fourier transform (FFT) for each slice of sliding windowed data, across the signal, the magnitude spectrum is calculated, converted to dB scale, and displayed as a function of time along the x-axis, and frequency along the y-axis.

The frequency resolution of the spectrographic analysis with a particular window is its equivalent noise bandwidth (Harris, 1978) and for Hamming window it is

$$\Delta f \approx 1.36\ \frac{f_s}{L} \qquad (3)$$

where f_s is the sampling rate. Thus, the choice of window duration decides the time and frequency resolutions (Harris 1978; Morris, 1988). For speech analysis, wide-band spectrogram with spectral resolution of 300 Hz is useful in observing pitch period as vertical striations and for seeing formant transitions. Narrow-band spectrogram with spectral resolution of 45 Hz, on the other hand, is useful for observing the pitch harmonics and formant frequencies during vowel segments (O'Shaughnessy, 1987). Cheung and Lim (1992) have proposed the use of geometric mean of the narrow-band and wide-band spectra for displaying a combined spectrogram.

A.2 Implementation of the spectrographic analyzer

The main tasks in a spectrographic analyzer are: digitization of the input waveform, selection of segment for analysis, computation of magnitude spectra of windowed block of data for sliding window positioning, displaying the spectrogram and the waveform, and user interfacing for selecting the analysis parameters and making measurements. An optional facility for outputting the selected waveform segment is available. Several systems using a PC with graphic display card and a data acquisition card have been reported (Thomas *et al.*, 1994). The data acquisition card is used for digitizing the input waveform as well as for outputting the selected segment. Spectral computations, spectrogram display, and user interfacing are handled by the PC. The generation of the spectrogram can be speeded up by using a DSP board and appropriately partitioning the tasks between the PC and DSP board. By using a DSP board with on board analog-to-digital and digital-to-analog converters, the data acquisition tasks can also be handled. The transfer rate between the PC and DSP board should be high enough to fully utilize the advantages of task partitioning. A spectrographic analyzer has been implemented using the PC and the DSP board; its hardware and software aspects are described in the following subsections.

A.2.1 Hardware set-up

A display area of 500×200 pixels is required for displaying the waveform and the spectrogram (Morris, 1988). Thus 640×480 resolution monochrome VGA with 16 gray levels of pixel intensity, will be adequate for spectrogram display and the display of gray level scale, cursor readouts, magnitude spectrum, and prompts for user interaction.

The hardware set-up of the analyzer is shown in Fig. 4.9. It consists of a PC with a VGA card interfaced through the PC expansion bus to a DSP board based on 16-bit fixed point processor TMS320C25 from Texas Instruments (PCL-DSP25 user's manual, 1989; TI-TMS320C2X user's guide, 1993). The board has 64 K word program memory, 64-K word data memory, a programmable timer, and an analog-to-digital converter (resolution: 16-bit, conversion time: 17μs), and a digital-to-analog converter. The board is interfaced to the PC through the input/output ports of the DSP chip, and these are mapped into the PC memory address space. The analog signal conditioning circuit consists of an anti-aliasing low-pass filter for the input and smoothing low-pass filter at the output. A 600 dpi laser printer can be connected to the PC for hard copy records of the spectrograms.

A.2.2 Software

Software consists of program running on the PC and the DSP board, with appropriate task allocation between the two for quick generation of spectrogram. The program on the PC loads the program module to be executed on the DSP board and then onwards the two programs run with appropriate handshaking and data transfer.

Out of the 640×480 pixel display area, 500×128 pixels are used for spectrogram, permitting 128 spectral values and 500 time position of the analysis window. Below the spectrogram, 500×45 pixels are used for displaying the waveform. The rest of the area is used for gray scale, cursor readouts, plot of magnitude spectrum, and user interactions.

The operation of the two programs is shown as a flowchart in Fig. A.1. The signal can be acquired with the specified sampling rate, with a record length of up to 32 k samples, using the ADC of the DSP board, and this record can be stored. Alternatively, previously stored record can be read. For speech signal, with sampling rate = 10 k Sa/s, we can acquire and analyze record length of up to 3.2s. Initially, signal waveform and gray-level intensity scale are displayed. The selected segment is displayed using the x-axis (500 pixels). The desired signal can be outputted for listening or recording. Analysis is performed by partitioning the selected segment into 256-point data blocks with a spacing of M samples, where

$$M = \frac{\text{number of samples in the selected segment}}{\text{number of pixels along x - axis}}$$

The 256-point data block is windowed and pre-emphasized before calculating 256-point DFT via FFT. The magnitude of data is scaled down to avoid overflow and down-loaded to the DSP board. The pre-emphasis (boosting high frequency components) is done to visualize the high frequency components in voiced speech and reduce the dynamic range requirement of magnitude scale. The segment is Hamming win-dowed with user selected window length L (< 256), to form a data block. The data block is extended to 256 by padding it with N-L zeros and is uded to obtain 128-point inter-polated spectrum. A 256-point FFT is computed on the downloaded data block and the PC uploads 128 samples of the computed FFT. While FFT of one block is be-ing calculated on the DSP board, the program running on the PC calculates the log magnitude of the previous block. The log magnitude of FFT is mapped to 16 gray levels linearly and displayed by 128 vertical pixels above the last sample of the time window on the monitor. The mapping shows the highest spectral magnitude as black ('0' gray level), intermediate magnitude levels in shades of gray, and ab-sence of significant magnitude as white ('15' gray level). This parallel operation of the DSP board and the PC continues till the end of the selected segment. The spectral resolution, and displayed intensity level of acquired data segment are user selectable. It produces high quality spectrograms of speech input within a few sec-onds of signal acquisition. With data block length of 256 samples, and sampling rate of 10 k samples/s, we get the resolution Δf = 53 Hz.

After generation of a spectrogram, cursors can be used for reading out the spectral magnitude as a function of time (n) and frequency (k). The displayed waveform and spectrogram can be stored in the Postscript (Adobe systems, 1988) format. While storing the spectrogram, the resolution has been increased from 16 to 256 gray levels.

The spectrographic analysis results, obtained from implementation of the described hardware set-up and the program, are given in Chapters 4 and 5. The hard copies were taken with 600 dpi laser printer. The waveform and gray scale plot are shown at the bottom and at the right side of the spectrogram, respectively. The spectrograms for the unprocessed and processed stimuli in off-line and real-time processing are given in supplement to the thesis.

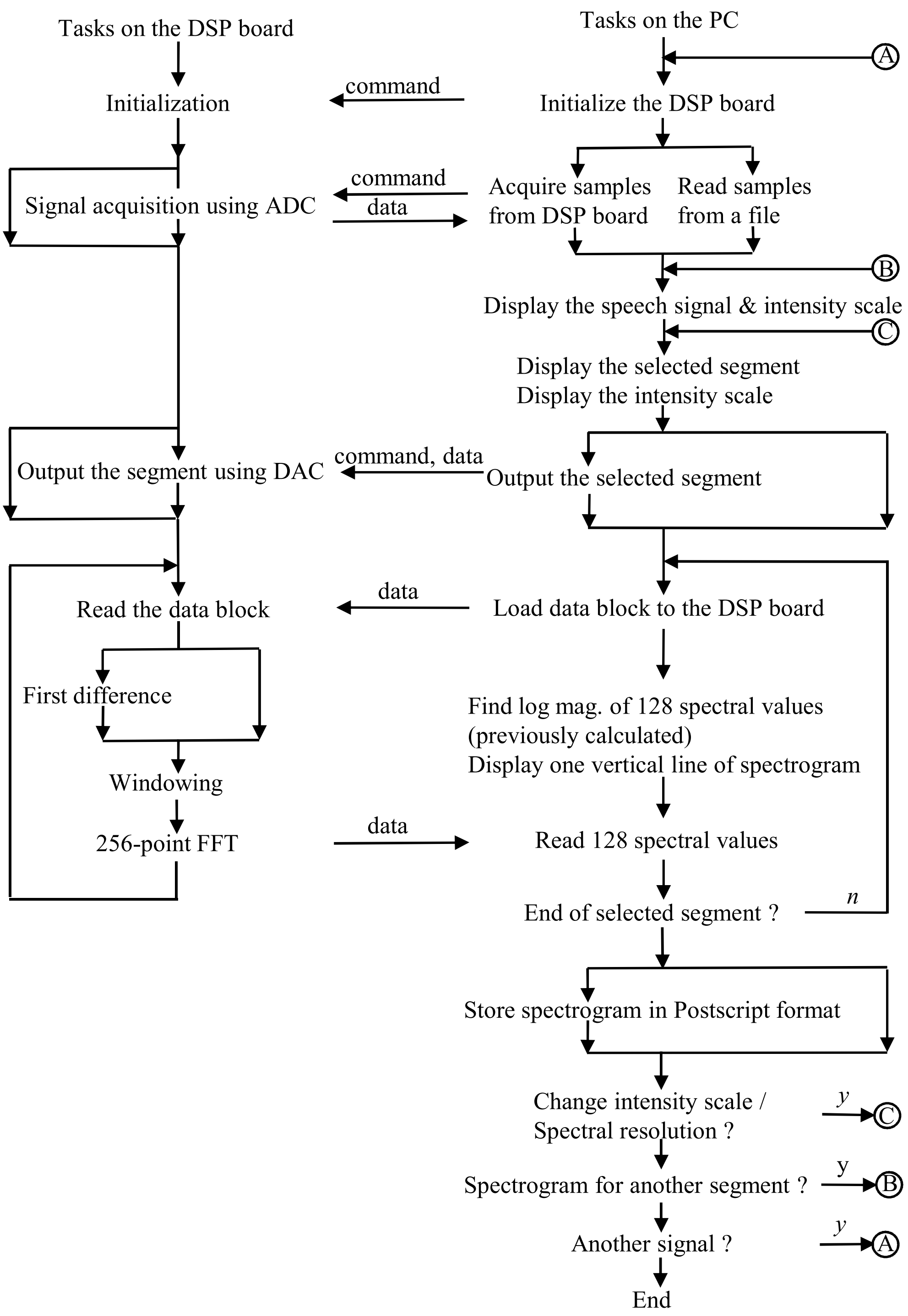


FIG. A. 1 Flowchart of task allocation, between the DSP board and the PC for spectrographic analyzer

Appendix B

HARDWARE FOR EXPERIMENTAL CONTROL

The set-up for speech signal acquisition and analysis is shown in Fig. 4.9. The speech signal was acquired using microphone, pre-amplifier, filter, and TI/TMS320C25 based DSP board interfaced to PC. The experimental set-up used for listening tests in off-line and real time processing are shown in Figs. 4.10 and 5.8 respectively. In off-line processing the signals were outputted using PCL 208 data acquisition card. This is followed by a pair of filters, power amplifiers for driving the headphones. In the real-time processing, the output signal from the PCL-208 board was processed through a pair of TI/TMS320C50 based DSP boards followed by a pair of power amplifier for dichotic presentation over headphones. This appendix provides a description of the hardware blocks used in the set-ups for the speech signal acquisition and processing and listening tests.

B.1 Microphone and input amplifier

The B&K 4176 microphone was used for acquiring the speech signal. The microphone has capacitance of 13 pF and sensitivity of 50 mV/Pa. The microphone is connected to sound level meter B&K 2235 and its AC output is taken as the signal for acquisition. The sound level meter produces 1V rms for 90 dB SPL (for the range 20–90dB) at the microphone input. The frequency weighting filter has A, C, and Lin. choices. For our purpose we used ‘C’ weighting (bandwidth 40 Hz–6 kHz). The signal is given to ADC of DSP25 board through amplifier (gain = 4) and low pass filter (f_p = 4.6 kHz).

B.2 TI/TMS320C25 based DSP board

PCL/DSP25 from M/S Dynalog Microsystem (user’s manual: PCL-DSP25, 1993) was used for data acquisition and spectrographic analysis. The DSP board is useful for numeric intensive operations like FFT computation. The DSP board fits in one of the expansion slot of the PC mother board. DSP25 board is based on the DSP chip TI/TMS320C25 which is a 16-bit fixed point processor operating at 40 MHz. It has 544-words of on chip programmable RAM. Its 32-bit ALU and single machine cycle multiplication makes it suitable for signal processing applications. Some of the important features of this board include

- Program memory, 64 K words and data memory 64 K words
- On board ADC has 35μs conversion time and resolution of 16-bit/50 kHz or 12-bit/100 kHz
- 16-bit programmable timer clocked at 5 MHz

The programmable timer can be programmed to provide start of conversion pulse to ADC at required sampling interval. The end of conversion pulse from ADC interrupts the processor for reading the digitized sample.

B.3 TI/TMS320C50 based DSP board

In real-time processing we were in need of serial port interface, efficient processor for numeric intensive operations, and input output interface. The TI/TMS320C50 based DSP boards (TI/DSP50 starter kit) were used for this purpose. The TMS320C50 is a 32-bit fixed point accumulator based microprocessor (TMS320C5x, users guide, 1993). It has full duplex synchronous serial port for direct communication with another serial device. The DSP50 board requires a 9 V ac, 250 mA power supply (TMS320C5x, starter kit user's guide, 1994). It consists of TLC32040 analog interface circuit (AIC) which interfaces to the TMS320C50 serial port.

The AIC provides a single channel, input/output voice quality analog interface. It incorporates a bandpass switched capacitor antialiasing input filter which can be bypassed, and a low pass switched capacitor output reconstruction filter, 14-bit resolution ADC, DAC and four microprocessor compatible serial port modes. It offers numerous combinations of master clock input frequencies and conversion/sampling rates, which can be changed via digital processor control (up to 19,200 samples per second). This is done by reloading RX counter A, and RX counter B for every analog-to-digital conversion period and TX counter A and TX counter B for digital-to-analog conversion.

The starter kit uses external flag output, branch control input, reset input, and ground pins for communicating with the RS232. During the serial transmission, the branch control input signal must be active (low) initially.

B.4 Data acquisition card PCL-208

In the listening test experiments for dichotic presentation involving off-line processing, two channels were required for outputting the processed signals which are not available on DSP25 board, therefore PCL-208 data acquisition card was used. Some of its important features include (PCL-208 user's manual, 1989).

- Switch selectable 16 single ended or 8 differential 12-bit ADC channels, with sampling rate up to 100 k samples/s, in DMA mode
- Two 12-bit digital-to-analog output channels with 0 to 5 V output range or adjustable output by applying external AC or DC reference.

The same set up was used for listening tests with real time processing also.

B.5 Low pass filter and audio amplifier

For analog-to-digital conversion, the input speech signal should be band-limited to a frequency less than 5 kHz, as we are using a sampling rate 10 k samples/s. Further, for getting the smooth waveform from the staircase waveform obtained at the output of the DAC, a low pass filter with f_c <5 kHz is needed. A seventh-order elliptic low-pass filter, called as antialiasing at input side and smoothing at output side, was used for this purpose. Its transition band is 4.6-5.0 kHz. Pass band ripples are less than 0.3 dB and stop band attenuation is greater than 40 dB. The magnitude response of the filter is as shown in Figure B.1. Details of the low pass filter design are as given in Pandey (1987).

The speech signal presented to the subject over TDH-39P headphones in the listening tests, is passed through the audio amplifier. A class B push-pull amplifier was used to drive headphones. The transistors are used in emitter follower (unity gain) mode and provide current busting. The input to the amplifier is fed through an attenuator using 10 kΩ logarithmic potentiometer. The input voltage range for the amplifier is 0 to 7.5 V. The audio amplifier board also has a facility to provide outputs of ± 300 mV for feeding to the tape recorder, ± 25mV for the microphone, and ±5 V as test signal. The audio amplifier boards and low pass filters were part of the experimental set-up developed earlier by Gavankar (1995) and Shah (1995) in the laboratory.

B.6 Subject terminal for listening tests

While conducting the listening tests, for displaying test instructions to the subject and stimuli response choices, and for obtaining subject's response to the stimuli, a VT-220 compatible terminal was used. The listening test program on the PC communicates with the terminal using RS-232 serial port. Only three lines of the serial port (transmit, receive, ground) were used for the link. In the serial port connector on the PC side, RTS and CTS lines are shorted together, and DSR, CD, DTR are shorted together. The communication was done using COM2 port of the PC at 9600 baud, and the terminal should be set at the same speed. The terminal along with the headphone for stimuli presentation were located in the acoustically isolated chamber. The PC controlling the experimental set-up, the signal processing and presentation set up, and amplifier are located outside the acoustically isolated chamber. This was done to keep the power dissipation and the noise from the equipment in the subject's chamber to a minimum.

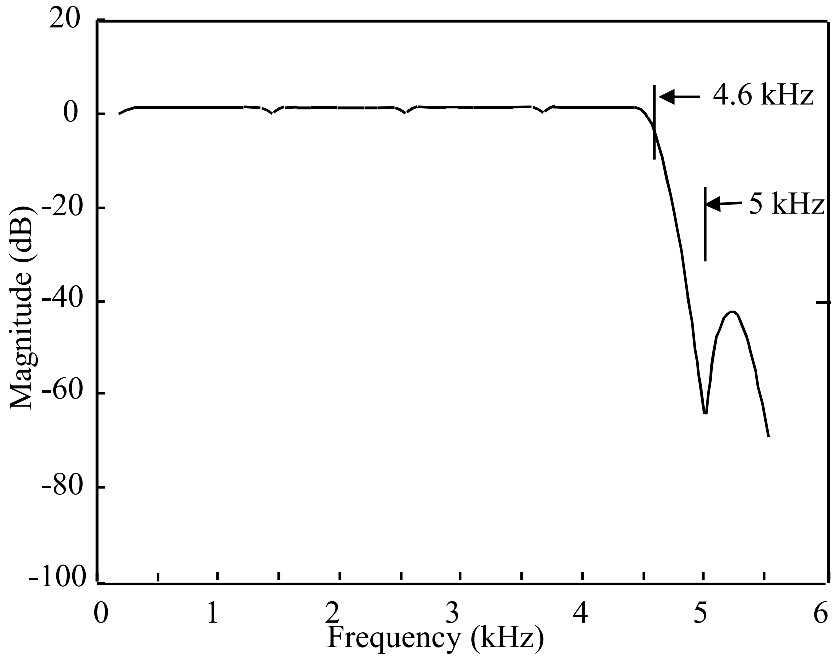


FIG. B.1 The measured magnitude response of the 7th order active elliptic low pass filters. The pass band ripples < 0.3 dB. The stop band attenuation > 40 dB.

Appendix C

SOFTWARE FOR SPEECH PROCESSING AND LISTENING TESTS

The software developed for speech processing and listening tests is described in this appendix.

C.1 Off-line processing of speech

The implementation for off-line processing is discussed in Chapter 4. The speech signal processing involves spectrally splitting it on the basis of critical bands corresponding to auditory filters. The speech $s(t)$, was filtered and divided into two parts in such a way that frequency components lying within a critical band are in one part, components lying in next non-overlapping critical band are in second part, component of third non-overlapping critical band in first part and so on, as shown in Fig. 4.2. The speech signal, $s_1(n)$ and $s_2(n)$ corresponding to odd and even numbered filter outputs were fed to one and other ear respectively. The processing was done by digitizing the waveforms. Final impulse response (FIR) filters were used so that degradation in speech quality due to non-linear phase response can be avoided.

The "filter" program was used for off-line processing of speech. This program was developed by Mithal (1996). While processing, the speech signal was padded with zeros (ncoef-2) to avoid abrupt ending of speech. This program passes the input signal data through two different sets of filters and store it in two different files by name 'tempxx.' These two files are again used as input to the processing loop. Finally, the 'temp19' and 'temp20' files are taken as /*a*sl/ (for one ear) and /*a*sr/ (for other ear). The input and final output files contain number of samples followed by the sample data in the binary format.

This program accepts the various parameters in the following way

After displaying information to user,

* Give the number of coefficients of the filter:<255>
* Give name of the BINARY speech file to be processed <*a*s*a*>
* Give the name of the file containing filter coefficient file names: <name2>

The 'name2' contains 'coef1, coef2 ...coef20' coefficient file names.

C.2 Real-time processing of speech

The real-time speech signal processing for our scheme was designed using frequency sampling technique of FIR filter design (Proakis and Manolakis, 1997). In this technique, the desired magnitude and phase responses of the are specified at uniformly spaced *M* frequency samples for filter order *M*. The impulse response is obtained by taking IDFT of the frequency response. The frequency response is recalculated with a finer resolution and is compared with the desired response. Add the specified frequency samples the responses exactly match, but there may be large variations from the desired responses at other frequencies. Therefore, there may be a need to modify the given desired response in order to reduce the errors at critical frequencies (Rabiner et al,1970). We have used linear phase FIR filter frequency sampling technique.

The filter coefficients are determined from the user specified magnitude values $|H(e^{j\omega k})|$ at *U* equally spaced frequencies samples i.e. ω_k around half the unit circle.

$\omega_k = 2\pi k/M$ for $0 \leq k \leq U$

where $U = (M-1)/2$ for *M* odd

$= (M/2)-1$ for *M* even

We define a set of real frequency samples

$G(k) = (-1)^k \, |H(e^{j\omega k})|$ for $0 \leq k \leq U$

The impulse response is calculated as

$$h(n) = \frac{1}{M}\left[G(0) + 2\sum_{k=1}^{U} G(k)\cos\frac{2\pi k}{M}(n+\frac{1}{2})\right]$$

This result in a filter with symmetric unit sample response and linear phase (Proakis and Manolakis, 1997). After *h(n)*, the *M*-point FIR impulse response, has been calculated, corresponding frequency response $|H(e^{j\omega k})|$ is calculated at *L* user specifies number of frequency sample using

$$H(e^{j\omega k}) = \sum_{k=0}^{M-1} H(e^{j\omega k})e^{-j\pi n/L}$$

where $\omega_k = 2\pi k/L$

The frequency response thus calculated is graphically superimposed over the desired response. This response can be inspected for ripple in the pass-band, transition overshoots, and sidelobes in the stop-band. If necessary, the desired frequency response, particularly the transition values can again be specified and the exercise can be repeated; to optimize the response in an interactive manner. The program can be used for $M \leq 400$.

The implementation for real-time processing is discussed in Chapter 5. For downloading and running of filter program on the DSP board using serial port interface and computation of filter coefficient of the FIR filter with graphically specified frequency response and online downloading of these coefficients a program was developed by Kasthuri (1997). This program was modified in two programs (1) dspfilt: for storing desired frequency response in the computer memory and (2) dspfilt1: for downloading the filter program with the filter coefficient created by the previous program on the DSP50 board.

Once the filter coefficients with the filter program are downloaded to DSP board through serial port, the serial port connection can be disconnected, leaving the DSP board ‘on.’ As the filter program with coefficients are made available on two DSP boards, the signal can be passed through the ADC, filtering program, and DAC. In this way, the desired filtered output was made available at the output of DAC. The first program reads the parameters in the following format

Sampling rate set to 10.28 k samples/s

* Filter specifications from data file (y/enter) : <CR>
* Length of filter (<= 180): <128>
* Chose the desired response: <M>
  Magnitude response ‘M’ (assumes linear phase)
  Magnitude and phase ‘B’ (both can be given)

Once ‘M’ is pressed the graphical display appears on the screen for inputting the desired frequency response. This can be reedited to get finer frequency response. After editing is over

* Save edited response (y/enter): <y>
* File name: <xxx.dat>

The difference in the “dspfilt” and “dspfilt1” is that, the later invokes the debugger and the user interface program (DSK5D). The later reads the parameter in the following way

After invoking the debugger

Default sampling rate: 10 k samples/s

* Sampling rate (4-20 k samples/s)/enter: <CR>

Sampling rate is set to 10.28 k samples/s

* Filter specifications from data file (y/enter): <y>
* Filter coefficient file name: <xxx.dat>
  Coefficients from the ‘xxx.dat’ are loaded to DSP board
* Edit filter response (y/enter) :<y>

Answering <CR> will come out of the program otherwise, the frequency response can be reedited to get finer frequency response and filter coefficients corresponding to this can be downloaded to DSP board. After downloading

* Save edited response (y/enter): <y>
* Filename: <xxx’.dat>

The DSK5D board from TI/TMS320 comes with a boot loder program. Once the debugger is invoked, it modifies the kernel program, and initializes the AIC. The program control is transferred to the starting location of kernel program (0x0800). Invoking DSK5D is a must for downloading and getting the filter program started on the DSP board through HLL program, as it makes kernel program running. It is also possible to load and execute assembly program while kernel program is running.

C.3 Simulation of hearing impairment

As mentioned in Chapter 4, the listening tests in first set of experiments were conducted on normal hearing subjects with simulated hearing impairment. As discussed in Chapter 3 the sensorineural impairment was simulated by adding broadband noise to speech signal at 5 SNR conditions of ∞, 6, 3, 0, and -3 dB. The noise was added in such a way that signal-to-noise ratio was kept constant on the basis of short-time (≈10 ms) energy of the signal. Thus during silence segments, there would not be any background noise.

The processing of speech signal and noise addition was done off-line using a program ‘snr2.’ developed by Prasad (1996) and modified by the author. The program assumes the noise samples in binary file named “noise.bin,” and it asks for the names for the binary files for input and output signals, and SNR (dB) value ρ. The input signal samples $s(n)$ and noise $d(n)$ are read in block of 100 samples (equal to 10 ms for SR = 10 k samples/s), and output samples $x(n)$ are calculated and written to output file. The output $x(n)$ is obtained by adding signal $s(n)$ with new $d(n)$, scaled by a factor such that SNR= ρ dB, on the basis of signal and noise energy in the block.

$$x(n) = s(n) + \alpha\, d(n)$$

$$\text{where } \alpha^2 = \sigma_s^2 \,/\, (\sigma_s^2 \times 10^{0.1\rho})$$

$$\sigma_s^2 = \Sigma\, s^2(n),\ \sigma_d^2 = \Sigma\, d^2(n),$$

since the noise is stationary, σ_d does not vary much from block-to-block, and therefore the weighting factor α depends on the short-time energy of the signal, i.e. σ_s

C.4 Program for listening tests

For qualitative analysis of listening test a program, ‘daap’ (data acquisition and presentation, using PCL-208) was developed by Kulkarni (1992) and used for presenting speech output on two channels. A computerized test administration system was used in order to automate process. For this purpose a program ‘test’ (listening test) was developed by Thomas (1995). It has been modified to present output on two channels, one for left and other for right. The output of this program was used for analysis. The program uses com2 port for communication to the subject terminal, and PCL-208 data acquisition card for outputting signal on digital-to-analog channel. This program accepts the parameters in the following format.

* Subject identification: <subject name>
* Test on specified sounds (to be presented) (y/n): <y>
  (For subject’s familiarization with the stimuli presented)
* Test no. (1-99): <1>
* List file number (1-12):<1>
  (Here 12 slist files are stored in working directory in each file 720 (12 x 60) numbers are stored for random presentation, this leads to automated process)

* Test using DAC channel 1/channel 1&2 software (1/0): <0>
* Response feedback (while hearing) (y/n):<y>
  (For getting feedback in response for presented stimuli)
* Signal information file: <sp2.dat >

In signal information file (e.g., ‘sp2.dat’) stimuli and calibration sounds file names are stored. The programs ‘cummat3’ (for combining confusion matrices) and ‘info1’ (for information transmission analysis) are discussed in the following Appendix F.

Appendix D

CALIBRATION OF HEADPHONES

In acoustics, the sound pressure is often expressed in the decibel unit of sound pressure level (SPL) which is defined as

$$SPL= 20 \log (P_{RMS}/P_{Ref})$$

Where P_{RMS} is the RMS value of the pressure and P_{Ref} = 20μPa. SPL measurements are carried out by using an SPL meter with calibrated microphone. Different type of filter responses (e.g. A, B, C, etc.) are used for weighting the frequency components present in the input sound for SPL measurements and this should be indicated (Peterson, 1980, Brül & Kjær, 1985; Hartmann, 1997).

Listening tests were performed using a pair of TDH-39P headphones. These type of headphones are commonly used in the audiological testing and speech perception experiments. The electroacoustical response of both the headphones in a pair was measured by using the set up shown in Fig D.1. The set-up was used to measure the input voltage required for driving the earpiece for a certain level of SPL to be developed by the earpiece when it is coupled to a calibrated acoustic cavity.

The sinusoidal waveform for driving earpiece was obtained from function/arbitrary waveform generator HP33120A (distortion < 0.04%) and an audio amplifier. The input voltage was measured in dBm using multimeter HP34401A. The earpiece was coupled to the calibrated acoustic cavity of Artificial Ear B&K4153, fitted with microphone B&K4176. The coupler has three acoustic cavities of 2.5 cm^3, 1.8 cm^3, 7.5 cm^3, connected in parallel (IEC R 318). The signal picked up by the microphone was connected to sound level meter B&K2235 for reading the SPL (with C weighting) During the measurements, a pressure of 500 g was applied on the earpiece, using spring tension mechanism of the artificial ear.

The elctroacoustic characteristics for the TDH-39P were obtained and a plot of the earpiece input voltage, in dBm, versus frequency for a 100 dB SPL at the earpiece output is shown for the two headphones in Fig. D.2. We see that the earpieces have almost similar response. The response varies by 10 dB over the 125 Hz-5 kHz range.

During listening tests, it is necessary to maintain the presentation level across various test sessions. The acoustic level selected by the subject can be found from the input voltage being provided by the audio amplifier, using a relationship between the electric input and acoustic output for speech signal. For this purpose, continuous presentation of synthesized vowel /*a*/ was used to obtain the SPL output in the artificial ear for different dBm levels of rms value of the input voltage to the earpiece. The most comfortable level as selected by normal hearing subject was about 75 dB SPL and this corresponds to -39 dBm electric input to the earpiece.

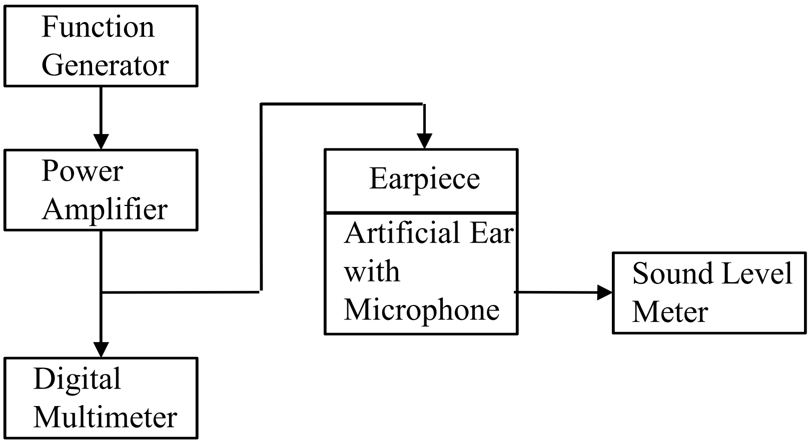


FIG. D.1 Experimental set-up for measuring the electroacoustic response of headphone TDH-39P

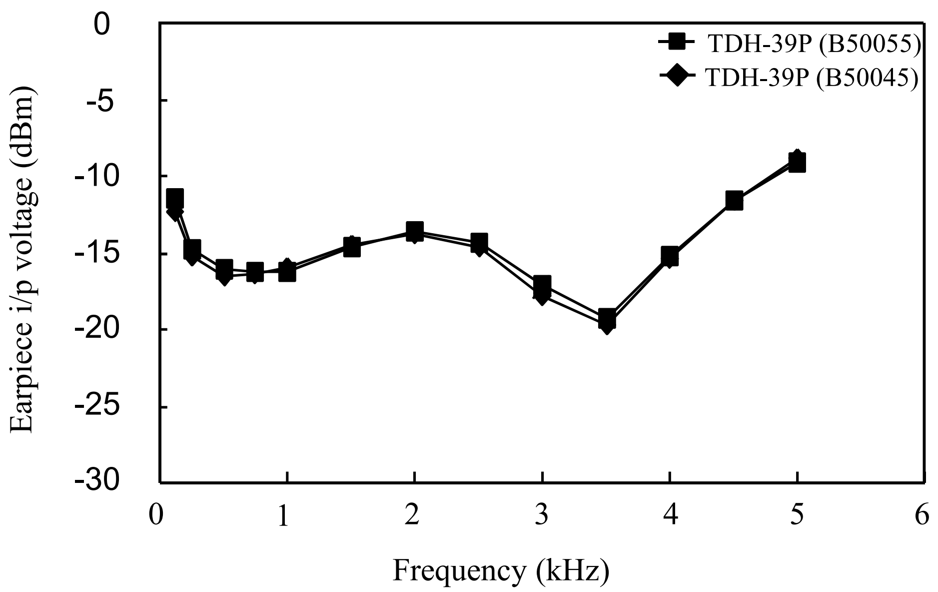


FIG. D.2 Earpiece input voltage versus frequency for 100 dB SPL at earpiece output. The earpiece input voltage was measured in dBm with V_{ref} = 224 mV (corresponding to 1 mW power in 50 ohm load resistance)

Appendix E

INFORMATION TRANSMISSION ANALYSIS

The number in each cell of a stimulus-response confusion matrix, with stimuli along the rows and responses along the columns represents either the frequency or probability of stimulus-response pair (Miller & Nicely, 1955).

Let ***s*** be the set of *n* stimuli ($s_1, s_2, \ldots, s_n$) and ***r*** be the set of *n* responses ($r_1, r_2, \ldots, r_n$) and $N(s_i)$, $N(r_j)$, $N(s_i; r_j)$ be the frequencies of stimulus s_i, response r_j, and the stimulus response pair (s_i; r_j), respectively, in a sample of *N* observations. The probabilities can be estimated as

$$p(s_i;r_j) = \frac{N(s_i;r_j)}{N}$$

$$p(s_i) = \frac{N(s_i)}{N} = \sum_{j=1}^{n} p(s_i;r_j)$$

$$p(r_i) = \frac{N(r_i)}{N} = \sum_{i=1}^{n} p(s_i;r_j) \tag{E.1}$$

In the confusion matrix the diagonal cell entries ($i = j$) indicate correct responses and off-diagonal entries ($i \neq j$) indicate incorrect one.

Recognition or articulation **s**core, R_s is ratio of sum of diagonal entries in a confusion matrix which gives the probability of correct responses

$$R_s = \sum_{i=1}^{n} p(s_i;r_i) \tag{E.2}$$

Even though the recognition score is useful, it doesn't provide any information about distribution of incorrect responses in confusion matrices. One way to generalize the recognition score is to combine stimuli and their responses in smaller groups so that confusions within the groups are more likely than those among the groups (Miller and Nicely, 1955). The new confusion matrix can be formed by combining stimuli with the common features that will give the recognition score for the transmission of this feature. In this way, several recognition scores can be used for specifying the transmission of different features.

The recognition score might be influenced by the subjects' response bias. For instance, if the subject adopted a scheme of giving the same phoneme response for all the stimuli presentations, the recognition score would be artificially high for that particular phoneme (chance scoring). Such a problem can be overcome if we express the results in terms of relative information transmitted. Information transmission analysis used by Miller and Nicely (1955) furnishes a measure of covariance between stimuli and responses, employing mean logarithmic probability (MLP) measure of information.

Information measure for input stimulus and output response, $I(s)$ and $I(r)$ respectively, are given, in bits, by

$$I(s) = \mathrm{MLP}(s) = -\sum_i p(s_i)\log_2[p(s_i)]$$
$$I(r) = \mathrm{MLP}(r) = -\sum_j p(r_j)\log_2[p(r_j)] \qquad \text{(E.3)}$$

The information measure of covariance of stimulus-response is given by

$$I(s;r) = \mathrm{MLP}(s) + \mathrm{MLP}(r) - \mathrm{MLP}(sr)$$
$$= -\sum_{i,j} p(s_i;r_j).\log_2 \frac{p(s_i)p(r_j)}{p(s_i;r_j)} \qquad \text{(E.4)}$$

The relative information transmission from s to r is given by

$$I_{trel}(s;r) = \frac{I(s;r)}{I(s)} \qquad \text{(E.5)}$$

Since $I(s) \geq I(s;\ r) \geq 0$; $0 \leq$ the $I_{trel}(s;r) \leq 1$.

In experiments usually the probabilities are estimated from relative frequencies obtained in a finite set of samples. Miller and Nicely (1955) observed that, like most maximum likelihood estimates, this estimate will be biased to overestimate $I(s;\ r)$ for small samples. If the sample size is sufficiently large this bias can be easily ignored.

The score by chance alone and patterning of incorrect response is being considered in above measure of information transmission. A relationship between recognition score R_s and relative information transmission I_{trel}; considering a special case in which the stimuli have equal probabilities, distribution of correct responses are equal among the diagonal cells, and incorrect responses are equally distributed among the off-diagonal cell, as obtained by Pandey (1987) is given in Fig. E.1. Information transmitted score of 0 % means chance scoring and score of 100 % means perfect identification.

In this study, nonsense syllables by using twelve consonants /p, b, t, d, k, g, m, n, s, z, f, v/ along with vowel /*a*/ have been used. These two sets of stimuli:

(a) syllables in vowel-consonant-vowel (VCV) context i.e. /*a*p*a*, *a*b*a*, *a*t*a*, *a*d*a*, *a*k*a*, *a*g*a*, *a*m*a*, *a*n*a*, *a*s*a*, *a*z*a*, *a*f*a*, *a*v*a*/

(b) syllables in consonant-vowel (CV) context, i.e. /p*a*, b*a*, t*a*, d*a*, k*a*, g*a*, m*a*, n*a*, s*a*, z*a*, f*a*, v*a*/.

In case of each set, the subjects have to identify the syllable heard from within the same set, and hence each set of stimuli is known as “closed set.” As the vowel remaining the same in all the syllables, the task of responding is that of identifying the consonants from the given set of consonants.

Information transmission analysis can also be applied to the confusion matrices derived by merging stimuli and responses in accordance with the different features, for evaluating the reception of specific feature. The twelve consonants can be grouped differently on the basis of six features of voicing, place, manner, nasality, frication, and duration (Miller and Nicely, 1955; Rabiner and Schefer, 1978; Ladefoged, 1982)

Each feature has two or more types and the consonants groups which can be numbered as 0, 1, 2. The following feature groupings were used

Voicing:	0- unvoiced (/p, t, k, s, f/)
	1- voiced (/b, d, g, m, n, z, v/)
Place:	0- front: bilabial /p, b, m)/ and labio-dental (/f, v/)
	1- mid (alveolar /t, d, n, s, z/)
	2- back (velar /k, g/)
Manner:	0- oral stop (/p, b, t, d, k, g/)
	1- frication (/s, z, f, v/)
	2- nasality (/m, n/)
Nasality:	0- oral (/p, b, t, d, k, g, s, z, f, v/)
	1- nasal (/m, n/)
Frication:	0- non-fricative (/p, b, t, d, k, g, m, n/)
	1- frication (/s, z, f, v/)
Duration:	0- short (/p, b, t, d, k, g, m, n, f, v/)
	1- long (/s, z)

Table E.1 Feature classification of 12 consonants used in listening tests

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

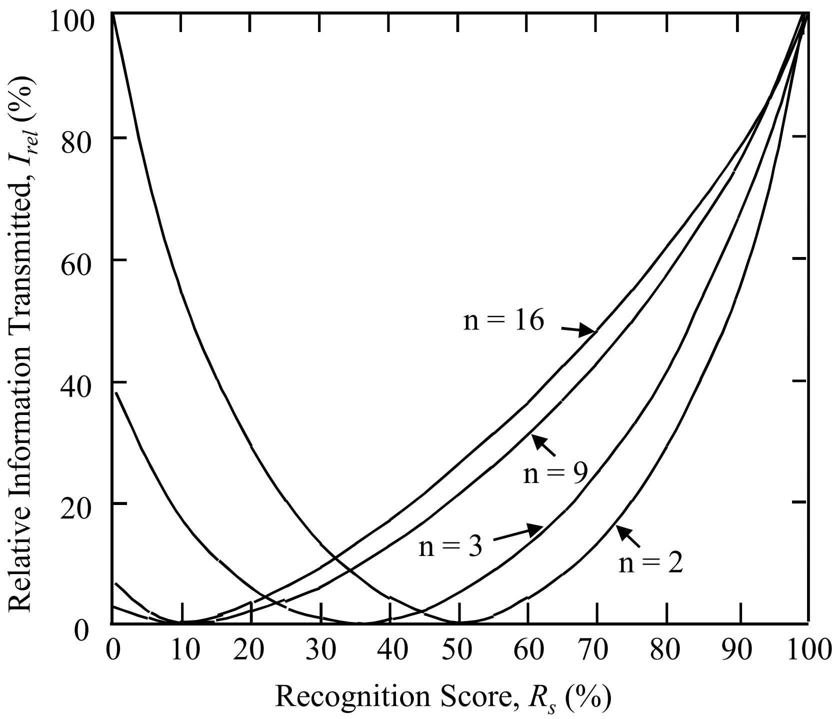


FIG. E.1 Recognition score versus relative information transmitted for a special case when the correct responses are equally distributed among the diagonal entries and the errors are equally distributed among the off diagonal entries in the stimulus-response confusion matrix. n = number of items. Adapted from Pandey (1987), Fig. 4.1.

Appendix F

ANALYSIS OF LISTENING TEST RESULTS

A brief description of the program used for the analysis of test results obtained from listening tests is given. A sample analysis for the test results for one listening condition is also given.

F.1 Analysis programs

The 'test' program is used for conducting the listening tests, at the end of this program, we get stimulus-response confusion matrix and response time statistics. For a particular listening condition, the confusion matrix and response time from a number of tests can be combined using program 'cummat3.'

The format for the input to this program (as generated by 'test') is

- N: number if stimuli, T: number of trials
- "S/R," stimulus names (0 to 80 characters, each name can be two or three characters long separated by one or more spaces), "+"
- N lines of confusion matrix row data, each line having
    - two or more character stimulus name, N cell entries separated by one or more spaces, sum of the cell entries in the row
- "+," sum of the N row entries for each column separated by one or more spaces.
- Test number, subject identification, feedback status, and presentation (1: monaural, 2: binaural)
- Date and time
- Total number of presentations, number of valid subject responses to the stimulus presentation, recognition score
- Minimum, maximum, mean, standard deviation of response time in seconds, and total test duration in minute

The output file has an almost similar format and an example is given later. The output of the above program is given as input to the program 'info1' for information transmission analysis. This program reads the stimulus grouping from file 'infogr.dat' in the following format

- N: number of stimuli, F: number of features
- Stimulus names (0-80 characters): These names can be two or three characters long separated by one or more spaces, and must be in the same order as for the input confusion matrix.

- Information about feature classification, feature groups are labeled as integer numbers (0, 1, 2)

Each row entry should be as follows

- The group numbers for various stimuli
- Feature name (limit 15 characters)
- Labels for the groups (these are separated from feature name and from each other by one or two spaces) up to 20 characters

This program outputs three files: 'infosc.dat' (information transmission scores), 'infotr.dat' (information transmission), 'infots.dat' (information transmission total summary).

F.2 Example of analysis of listening test results

The test results from 8 tests for a subject under a particular listening condition are given here. Subject: BAS (with bilateral sensorineural hearing loss), testing condition: no noise, unprocessed speech, off-line processing, VCV context.

1) Output of cummat3 (combination of results from 8 tests)

```
 12  480

 S\R  aPa aBa aTa aDa aKa aGa aMa aNa aSa aZa aFa aVa  +
 aPa   1   0   0   0   0   0   0   0   0   0  39   0   40
 aBa   0  23   0   0   0  13   0   0   0   0   0   4   40
 aTa   0   0  31   0   7   0   0   0   0   1   0   1   40
 aDa   0   0   0  36   0   4   0   0   0   0   0   0   40
 aKa   0   0   0   2  38   0   0   0   0   0   0   0   40
 aGa   0   0   0   5   0  35   0   0   0   0   0   0   40
 aMa   0   0   0   0   0   0  40   0   0   0   0   0   40
 aNa   0   0   0   0   0   0  16  24   0   0   0   0   40
 aSa   0   0   0   0   1   0   0   0  39   0   0   0   40
 aZa   0   0   0   0   0   0   0   0   3  37   0   0   40
 aFa   2   0   0   0   0   0   0   0   2   1  35   0   40
 aVa   1   0   0   0   0   0   0   0   0   0   0  39   40
  +    4  23  31  43  46  52  56  24  44  39  74  44  480

 No. of files:  8
--- 08-12-1997   16:04:44
--- 08-12-1997   16:22:35
--- 08-12-1997   17:41:19
--- 15-12-1997   15:39:16
--- 08-12-1997   17:41:19
--- 15-12-1997   16:11:05
--- 15-12-1997   17:01:05
--- 15-12-1997   18:00:05

73.3  83.3  78.8  4.0
2.14  4.25  2.88  0.65  1.83  6.0
```

2) File: infogr.dat (feature grouping 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   FRICATION ST FR
 0   0   0   0   0   0   1   1   0   0   0   0   NASALITY OR NA
 0   0   0   0   0   0   2   2   1   1   1   1   MANNER OS FR NA
 0   0   1   1   2   2   0   1   1   1   0   0   PLACE  FN MD BK
 0   1   0   1   0   1   1   1   0   1   0   1   VOICING  UV VO
```

3) Analysis results: Percentage scores

```
NO. OF STIMULI: 12

** PERCENTAGE SCORES **

* (12): OVERALL
 S/R  aPa  aBa  aTa  aDa  aKa  aGa  aMa  aNa  aSa  aZa  aFa  aVa
 aPa    3    0    0    0    0    0    0    0    0    0   97    0
 aBa    0   57    0    0    0   33    0    0    0    0    0   10
 aTa    0    0   77    0   18    0    0    0    0    3    0    3
 aDa    0    0    0   89    0   10    0    0    0    0    0    0
 aKa    0    0    0    5   94    0    0    0    0    0    0    0
 aGa    0    0    0   13    0   87    0    0    0    0    0    0
 aMa    0    0    0    0    0    0   99    0    0    0    0    0
 aNa    0    0    0    0    0    0   40   60    0    0    0    0
 aSa    0    0    0    0    3    0    0    0   97    0    0    0
 aZa    0    0    0    0    0    0    0    0    8   92    0    0
 aFa    5    0    0    0    0    0    0    0    5    3   87    0
 aVa    3    0    0    0    0    0    0    0    0    0    0   97

Correct:  78.8
```

* (2): DURATION

S/R	SH	LO
SH	100	2
LO	2	99

Correct: 99.0

* (2): FRICATION

S/R	ST	FR
ST	86	15
FR	3	98

Correct: 89.8

* (2): NASALITY

S/R	OR	NA
OR	100	0
NA	0	100

Correct: 100.0

* (3): MANNER

S/R	OS	FR	NA
OS	82	19	0
FR	3	98	0
NA	0	0	100

Correct: 89.8

* (3): PLACE

S/R	FN	MD	BK
FN	92	2	7
MD	9	86	6
BK	0	9	92

Correct: 89.2

* (2): VOICING

S/R	UV	VO
UV	98	3
VO	2	99

Correct: 98.1

4) *Analysis results: information transmission*

```
** INFORMATION TRANSMISSION **

* (12): OVERALL
Stimulus info =  3.5832
Response info =  3.4244
Transn info =  2.9079
Perc transn =  81.2

* (2): DURATION
Stimulus info =  0.6497
Response info =  0.6640
Transn info =  0.5806
Perc transn =  89.4

* (2): FRICATION
Stimulus info =  0.9180
Response info =  0.9806
Transn info =  0.5338
Perc transn =  58.2

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

* (3): MANNER
Stimulus info =  1.4587
Response info =  1.4829
Transn info =  1.0786
Perc transn =  73.9

* (3): PLACE
Stimulus info =  1.4829
Response info =  1.5240
Transn info =  0.9540
Perc transn =  64.3

* (2): VOICING
Stimulus info =  0.9796
Response info =  0.9786
Transn info =  0.8453
Perc transn =  86.3
```

5) *Analysis of results: summary of information transmission*

NO. OF STIMULI: 12

COR	IS	IR	IT	RTR	FEATURE	N
79	3.58	3.42	2.91	81	OVERALL	12
99	0.65	0.66	0.58	89	DURATION	2
90	0.92	0.98	0.53	58	FRICATION	2
100	0.65	0.65	0.65	100	NASALITY	2
90	1.46	1.48	1.08	74	MANNER	3
89	1.48	1.52	0.95	64	PLACE	3
98	0.98	0.98	0.85	86	VOICING	2

Appendix G

SUBJECT DATA

TABLE G-I: Hearing thresholds (dBHL) for the hearing impaired subjects. PTA: pure tone average hearing threshold levels for test frequencies of 0.5, 1, and 2 kHz.

Subject Code (Sex, Age)	**Ear L: left R:right**	**Hearing thresholds (dB HL)**						
		Frequency (kHz)						**PTA**
		0.25	**0.50**	**1.0**	**2.0**	**4.0**	**6.0**	
SG (M, 27)	L	25	45	75	100	120	120	73
	R	25	60	70	100	120	120	77
SJH (F, 18)	L	80	90	100	105	100	100	98
	R	70	85	90	90	85	85	88
KRN (M, 35)	L	40	35	45	45	45	45	42
	R	40	50	60	65	65	65	58
DSD (M, 19)	L	50	65	95	95	100	100	85
	R	50	60	75	85	90	90	73
LGR (M, 27)	L	70	65	70	70	70	70	68
	R	65	65	70	70	70	65	68
SSN (M, 31)	L	80	70	80	75	75	75	75
	R	65	60	70	65	85	85	65
KRV (M, 49)	L	50	60	60	60	60	65	60
	R	40	45	50	60	65	75	52
BAS (M, 58)	L	50	40	30	30	40	40	33
	R	45	50	35	30	30	30	38
SAV (M, 46)	L	50	45	45	45	35	40	45
	R	60	70	65	65	85	95	67
LDM (M, 52)	L	65	65	50	40	40	75	52
	R	70	80	85	80	80	95	82
KIT (M, 48)	L	40	35	45	45	45	75	42
	R	40	50	60	65	65	95	58

Appendix H

TEST INSTRUCTIONS AND FORMS

H.1 Test instructions to the normal hearing and hearing impaired subjects in the listening experiments for the evaluation of speech processing scheme

The purpose of this experiment is to evaluate certain speech processing scheme for binaural dichotic presentation as a possible solution to problem of spectral masking.

Your task will be to listen and identify the test sound (stimulus) presented. The sound will be presented binaurally (in two ears) over the headphone. Generally, the sound intensity during the test will be 80–95 dB (SPL). Further, it will adjusted as per your comfortable listening level. You will be seated in front of a subject terminal (computer) and will use the keyboard to indicate your response after each sound presentation. The number of test sounds is 12. A single test will take typically 4–12 minutes. A test session will involve several tests and may take 1–2 hours; however, you may request for a break at an earlier time.

Instructions for a particular test will be displayed on your terminal screen at the start of the test session. In the beginning you may undergo a trial test run with correct feedback so that you become familiar with the test sounds set. You should listen the complete set of test sounds several times in order to establish association between the sounds presented and names used to identify them.

During the test, the presentation number and set of choices will be displayed (in a random order). A "listen" message will be displayed before each presentation. You will indicate your response by hitting the appropriate key on the keyboard. The next presentation will follow after a brief pause (2-5 seconds). A presentation will not be repeated. If you are not sure, you can guess. The test will not proceed if you do not respond. If you missed a presentation, you may indicate this by hitting a key other than valid choices. You will be provided with feedback about the correct response in the first few tests in a given session.

H.2 Test instructions as displayed on the subject terminal screen during the VCV test

---

*** CONSONANT IDENTIFICATION TEST ***

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

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

After listening to the sound, please hit the corresponding key as quick as possible.

A presentation will not be repeated. If you are not sure, you can guess. The test will not proceed if you do not respond. If you miss a presentation, and cannot even guess, you may hit a key other than valid choices.

PLEASE HIT ANY KEY WHEN YOU ARE READY FOR THE TEST

H.3 Form for recording background information on the normal hearing and hearing impaired subjects

SUBJECT BACKGROUND INFORMATION

Date __/__/19__

Name ________________________________ Code ____________

Address ____

____

Phone ( ) ___ _______Extension _____

____

Sex _______ Age _____________

Occupation: ____

Place of birth: ____

First language: ____

Other languages: ____

Handedness: Left / Right

History of noise exposure: _____________________________

History of hearing problems: _____________________________

____

____

Other remarks: ____

H.4 Form for subject willingness for evaluation of speech processing scheme

CONSENT FORM

I have carefully read the test instructions provided by Mr. D.S. Chaudhari (Ph.D. Scholar, IIT Bombay) for participation in listening experiments for evaluation of speech processing schemes. I am willing to participate in tests conducted by him.

Signature: ______________________________

Name: ______________________________

Address: ______________________________

____

Date: ______________________________

REFERENCES

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

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 sensori-neural hearing listeners,” Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP), 3625-3628.

Balfour, P. B., and Hawkins, D. B. (**1992**). “A comparison of sound quality judgments for monaural and binaural hearing aid processed stimuli,” Ear and Hearing **13**(**5**), 331-339.

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

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

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).

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. and Pandey, P.C. (**1999**). “Binaural audio in multimedia system to improve auditory perception for the hearing-impaired,” Fifth Int. Sym. Signal Processing Appl. (ISSPA), Brisbane, Australia.

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

Cox, R. M., and Bisset, J. D. (**1984**). “Relationship between two measures of aided binau-ral advantage,” J. Speech Hear. Disord. **49**, 399-408.

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

Cristopher, W., Turner, and Michael, P. R. (**1987**). “Audibility and recognition of stop consonants in normal and hearing-impaired subjects,” J. Acoust. Soc. Am. **81**(**5**), 1566-1573.

DeGennaro, S., Braida, L. D., and Durlach, N. I. (**1981**). “Study of multi-band syllabic compression with simulated sensorineural hearing loss,” J. Acoust. Soc. Am. **69**, S16.

DeSimio, M. P., Anderson, T. R., and Westerkamp, J. J. (**1996**). “Phoneme recognition with a model of binaural hearing,” IEEE Trans. Speech Audio proc. **4**(**3**), 157-165.

Dirks, D., and Carhart, R. (**1962**). “A survey of reactions from users of binaural and mo-naural hearing aids,” J. Speech Hear. Disord. **27**(**4**), 311-322.

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., Loizon, P. C., and Rainey, D. (**1997**). “Simulating the effect of cochlear-implant electrode insertion depth on speech understanding,”J. Acoust. Soc. Am. **102,** 2993-2996.

Dowell, R. C., Martin, L. F. A., Tong, Y. C., Clark, G. M., Seligman, P. M., and Patrick, J. F. (**1982**). “A 12-consonant confusion study on a multiple-channel cochlear implant patient,” J. Speech Hear. Res. **25**, 509-516.

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

Duda, R. O. (**1996**). “Auditory localization demonstrations,” Acta Acoustica. **82**, 346-355.

Dunn, H. K., and White, S. D. (**1940**). “Statistical measurements on conversational speech,” J. Acoust. Soc. Am. **11**, 278-288.

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).

Fairbanks, G., Everett, W. L., and Jaeger, R. P. (**1954**). “Method for time or frequency compression-expansion of speech,” IRE Trans. Audio. **AU-2**, 7.

Fitzgibbons, P. J., and Wightman, F. L. (**1982**). “Gap detection in normal and hearing-impaired listeners,” J. Acoust. Soc. Am. **72(3)**, 761-765.

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

Fletcher, H. (**1938**). “Loudness, masking and their relation to the hearing process and the problem of noise measurement,” J. Acoust. Soc. Am. **9**, 275-293.

Fletcher, H. (**1952**). “The perception of sound by deafened persons,” J. Acoust. Soc. Am. **24**, 490-497.

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

Franklin, B. (**1981**). “Split-band amplification: a hi/lo hearing aid fitting,” Ear and Hearing **2**(**5**), 230-233.

Freyman, R. L., and Nerbonne, G. P. (**1989**). “The performance of consonant-vowel intensity ratio in the intelligibility of voiceless consonants,” J. Speech Hear. Res. **32**, 524-535.

Freyman, R. L., Nerbonne, G. P., and Cote, H. A. (**1991**). “Effect of consonant-vowel ratio modification on amplitude envelope cues for consonant recognition,”J. Speech Hear. Res. **34**, 415-426.

Fu, Q., 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., and Schenman, B. N., and Hagerman, B. (**1988**). The effects of different frequency responses on sound quality judgments and speech intelligibility,” J. Speech Hear. Res. **31**, 166-177.

Gavankar, P. S. (**1995**). “Speech Training Aid for the Deaf,” M. Tech. dissertation, Department of Electrical Engineering, IIT Bombay, India.

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.

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.

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

Guyton, A. C. (**1996**). *Textbook of Medical Physiology* (Prism Books, Bangalore).
Hannley, N., and Dorman, M. F. (**1983**). “Susceptibility to intraspeech masking in listeners with sensorineural hearing loss,” J. Acoust. Soc. Am. **74,** 40-51.
Harris, F. J. (**1978**). “On the use of windows for harmonic analysis with the discrete Fourier transform,” Proc. IEEE, **66**, 51-66.
Harris, R. W., and Goldstein, D. P. (**1979**). “Effect of room reverberation upon hearing aid quality judgments,” Audiology **18**, 253-262.
Hartmann, W. M. (**1998**). *Signals, Sound, and Sensation* (Springer, New York).
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.
Ho, W. (**1987**). “Split speech information for binaural presentation,” B. A. Sc. thesis, Department of Electrical Engineering, University of Toronto, Canada.
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.
Jangamshetty, 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. Nat. Conf. Biomed. Engineering, Roorkee, India.
Jerger, J., Carhart, R., and Dirks, D. (**1961**). “Binaural hearing aids and speech intelligibility,” J. Speech Hear. Res. **4**, 137-148.
Jesteadt, W. (Ed.). (**1997**). *Modeling Sensorineural Hearing Loss* (Lawrence Erlbaum Associates, Mahwah, New Jersey).
Johannson, B. (**1966**). “The use of the transposer for the management of the deaf child,” in Levitt et al. pp. 195-204.
Johnsson, L. G. (**1985**). “Cochlear anatomy and histopathology,” in *Cochlear Implants*, edited by R. F. Gray (College-Hill, San Diego, CA). pp. 50-73.
Kasthuri, P. (**1997**). “Real-time Filter Design with Tunable Frequency Response using TMS320C50 Processor,” Project report, Signal Processing and Instrumentation Lab., Department of Electrical Engineering, IIT Bombay, India.
Kersta, L. G. (**1948**). “Amplitude cross-section representation with sound spectrogram,” J. Acoust. Soc. Am. **20**, 796-801.
Koenig, R., Dunn, H. K., and Lacey, L. Y. (**1946**). “The sound spectrograph,” J. Acoust. Soc. Am. **18**, 19-49.
Kollmeier, B., Hohmann, V., and Peissig, J. (**1992**). “Digital signal processing for binaural hearing aids,” Proc. Int. Cong. Acoustics (ICA), Beijing **H3-4**.
Kuc, R. (**1982**). *Introduction to Digital Signal Processing* (McGraw-Hill, New York).
Kulkarni, D. M. (**1996**). “Development of Cascade Pole-Zero Synthesiser,” M. Tech. Dissertation, Department of Electrical Engineering, IIT Bombay, India.
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.
Leek, M. R., and Summers, V. (**1996**). “Reduced frequency selectivity and the preservation of spectral contrast in noise,” J. Acoust. Soc. Am. **100**, 1796-1806.
Levitt, H. (**1994**). “Technologies for signal processing hearing aids,” Proc. Int. Conf. Spoken Language Processing (ICSLP), 1-6.
Levitt, H., Pickett, J. M., and Houde, R. A. (Eds.). (**1980**). *Sensory Aids for the Hearing Impaired* (IEEE Press, New York).

Lunner, T. (**1997**). "A Digital Filterbank Hearing Aid," Doctoral dissertation (No. 479), Department of Neuroscience and Locomotion, Division of Audiology, Linköping 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.

Lunner, T., and Hellgren, J. (**1991**). "A digital filterbank hearing aid - design, implementation and evaluation," Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP), 3661-3664.

Lyregaard, P. E. (**1982**). "Frequency selectivity and speech intelligibility in noise," Scand. Audiol. Suppl. **15**, 113-122.

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.

Mithal, S. (**1996**). "Splitting of speech signal for binaural presentation," B. Tech. project, Department of Electrical Engineering, IIT Bombay, India.

Moncur, J. P., and Dirks, D. (**1967**). "Binaural and monaural speech intelligibility in reverberation," J. Speech Hear. Res. **10**(**2**), 186-195.

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* (Academic, London).

Moore, B. C. J. (**1997**). *An Introduction to the Psychology of Hearing* (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.

Morris, L. R. (**1985**). "Structural considerations for large FFT programs on a TI-TMS32010 DSP microchip," Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICASSP), 1648-1651.

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

Nabelek, A. K., and Pickett, J. M. (**1974**). "Monaural and binaural speech perception through hearing aids under noise and reverberation with normal and hearing-impaired listeners," J. Speech Hear. Res. **17**(**4**), 724-739.

Nebelek, A. K. and Robinson, P. K. (**1982**). "Monaural and binaural speech perception in reverberation for listeners of various ages," J. Acoust. Soc. Am. **71**, 1242–1248.

Neuman, A. C., Bakke, M. H., Hellman, S., and Levitt, H. (**1994**). "Effect of compression ratio in a slow acting compression hearing aid: Paired comparison judgments of quality," J. Acoust. Soc. Am. **96**(**3**), 1471-1478.

Ochs, M. T., Humes, L. E., Ohde, R. N., and Grantham, D. W. (**1989**). "Frequency discrimination ability and stop consonants identification in normally hearing and hearing-impaired subjects," J. Speech Hear. Res. **32**, 133-142.

Oppenheim, A. V. (**1970**). "Speech spectrogram using fast Fourier transform," IEEE Spectrum **7**, 57-62.

O'Shaughnessy, D. (**1987**). *Speech communication: human and machine* (Addison-Wesley, Reading, Massachusetts).

Pandey, P. C. (**1987**). "Speech processing for cochlear prosthesis," Ph.D. dissertation, Department of Electrical Engineering, University of Toronto, Canada.

Pandey, P. C., Kunov, H., Abel, S. M. (**1986**). "Disruptive effect of auditory signal delay on speech perception with lip-reading," J. Auditory Research. **26**, 27-41.

Pandey, P. C., Kunov, H., 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 (ICASSP) **34.3**, Dallas, Texas.

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

Patterson, A. P. G. (**1980**). *Handbook of Noise Measurement* (Gen Rad Inc., Concord, Massachusetts.

Picheny, M. A. (**1981**). "Speaking clearly for the hard of hearing," Ph.D. thesis, MIT, Cambridge, Massachusetts).

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).

Potter, R. K., Kopp, G. A., and Kopp, H. G. (**1966**). *Visible Speech* (Dover, New York).

Prasad, V. V. R. (**1996**). "Speech processing for single channel aid," M. Tech. dissertation, Department of Electrical Engineering, IIT Bombay, India.

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

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 Dichotic Presentation," M. Tech. dissertation, Department of Electrical Engineering, IIT Bombay, India.

Robles, L., Ruggero, M. A., and Rich, N. C. (**1986**). "Basilar membrane mechanics at the base of the chinchilla cochlea. I. input/output functions, tuning curves, and response phases," J. Acoust. Soc. Am. **80**, 1364-1374.

Shah, N. (**1995**). "A Sensory Aid for the Deaf," M. Tech. dissertation, Department of Electrical Engineering, IIT Bombay, India.

Simon, C. (**1979**). "On the use of comfortable listening levels in speech experiments," J. Acoust. Soc. Am. **64**, 744-751.

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.

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

Tejero, J. C., Bernal, S., Hidaldo, J. A., Fernandez, J., Urquiza, R., Gago, A. (**1995**). "A digital hearing aid that compensates loudness for sensorineural hearing impairments," Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP), 2991-2994.

ter Keurs, M., Festen, J. M., and Plomp, R. (**1992**). "Effect of spectral envelop smearing on speech reception," J. Acoust. Soc. Am. **91**, 2872-2880.

Texas Instruments (**1993**). *User's Guide, TI-TMS320C2X Digital Signal Processor Products* (Texas Instruments, USA).
Texas Instruments (**1993**). *User's Guide, TI-TMS320C5X Digital Signal Processor Products* (Texas Instruments, USA).
Texas Instruments (**1994**). *Starter Kit User's Guide, TI-TMS320C2X Digital Signal Processor Products* (Texas Instruments, USA).
Thibodeau, G. A. (**1992**). *Structure and Function of the Body* (Mosby-Year Book, St. Louis, Missouri).
Thomas, T. G. (**1996**). "Experimental evaluation of improvement in speech perception with consonantal intensity and duration modification," Ph.D. thesis, Department of Electrical Engineering, IIT Bombay, India.
Thomas, T. G., Pandey, P. C., and Agashe, S. D. (**1994**). "A PC-based multiresolution spectrograph," J. IETE (India). **40**, 105-108.
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. Ind. **XXIV**, VI-4.1-4.5.
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.
Villchur, E. (**1973**). "Signal processing to improve speech intelligibility in perceptive deafness," J. Acoust. Soc. Am. **53**, 1646–1657.
Villchur, E. (**1977**). "Electronic models to simulate the effect of sensory distortions on speech perception by the deaf," J. Acoust. Soc. Am. **62**, 665-674.
Wall, L. G. (**1995**). *Hearing for the Speech Language Pathologist and Health Care Professional* (Butterworth Heinmann, Boston).
Webster, J. G. (Ed.) (**1988**). *Encyclopedia of Medical Devices and Instrumentation, Vol 1* (Wiley, New York) pp. 159-169.
Wilson, K. J. L. (**1990**). *Ross and Wilson Anatomy and Physiology in Health and Illness* (ELBS with Churchill Livingstone, Hong Kong).
Zelnick, E. (**1970**). "Comparison of speech perception utilizing monotic and dichotic modes of listening," J. Aud. Res. **10**, 87-97.
Zwicker, E. (**1961**). "Subdivision of audible frequency range into critical bands (Frequenzgruppen)," J. Acoust. Soc. Am. **33**, 248.
Zwicker, E., and Schorn, K. (**1978**). "Psychoacoustical tuning curves in audiology," Audiology **17**, 120-140.