SPEECH PROCESSING FOR REDUCING THE EFFECTS OF SPECTRAL MASKING IN SENSORINEURAL HEARING LOSS

Thesis
submitted in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy

by

Pandurangarao N. Kulkarni
(Roll No. 03407803)

under the supervision of

Prof. P. C. Pandey
and
Prof. D. S. Jangamashetti

Department of Electrical Engineering
Indian Institute of Technology Bombay

December 2010

Indian Institute of Technology Bombay

Thesis Approval

Thesis entitled **Speech processing for reducing the effects of spectral masking in sensorineural hearing loss**

by **Pandurangarao N. Kulkarni**

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

Supervisors: PCPandey (Prof. P. C. Pandey)

Dhetti (Prof. D. S. Jangamashetti)

Internal Examiner: Preeti Rao (Prof. Preeti Rao)

External Examiner: ________ (Prof. C. Chandra Sekhar)

Chairman: ________ (Prof. S. Mukherji)

Date: 23rd December, 2010

Declaration

I declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Pandurangarao N. Kulkarni

Roll No.: **03407803**

Date: **23rd December 2010**

Dedicated to the memory of

my grandmother Shrimati Ambabai

and my father Shri Narayanarao S. Kulkarni.

Indian Institute of Technology Bombay

CERTIFICATE OF COURSE WORK

This is to certify that Mr. Pandurangarao N. Kulkarni (Roll No. 03407803) was admitted to the candidacy of Ph.D. Degree in January 2004, after successfully completing all the courses required for the Ph.D. Degree Programme. The details of the course work done are given below.

Sr. No.	Course No.	Course Name	Credits
1	EE603	Digital Signal Processing and its Applications	6
2	EE679	Speech Processing	6
3	EE610	Image Processing	6
4	EES801	Seminar	4
5	HS699	Communication and Presentation Skills	
		Total Credits	22

I.I.T. Bombay Dy. Register (Academic)

Dated:

P. N. Kulkarni / Supervisors: Prof. P. C. Pandey and Prof. D. S. Jangamashetti, "Speech processing for reducing the effects of spectral masking in sensorineural hearing loss", *Ph.D. Thesis*, Department of Electrical Engineering, Indian Institute of Technology Bombay, December 2010.

Abstract

Sensorineural hearing loss is associated with widening of auditory filters leading to increased spectral masking and degraded speech perception. The research objective is to investigate two techniques for improving speech perception by persons with sensorineural loss: (i) spectral splitting scheme for binaural dichotic presentation, and (ii) multi-band frequency compression for monaural presentation.

For binaural dichotic presentation, comb filter based spectral splitting scheme was investigated with respect to magnitude responses for perceptual balance and bandwidth. The listening tests were conducted (i) to assess the effectiveness of the scheme in improving speech perception and (ii) to study the effect of the scheme on sound source localization. The tests for consonant recognition showed that dichotic presentation with comb filters based on auditory critical bandwidth and magnitude response selected for monaural-binaural loudness balance resulted in an SNR advantage of 12 dB for normal-hearing listeners with hearing loss simulated by broad-band masking noise, and an improvement of 14 – 31 % in recognition scores for the hearing-impaired listeners. There was a significant decrease in the response time. The tests for sound source localization showed that identification of the direction of broad-band sound sources was not significantly affected.

For monaural hearing, multi-band frequency compression, applied on complex spectrum using overlap-add method, was implemented and investigated with different frequency mappings, bandwidths, segmentations for analysis-synthesis, and compression factors. Listening tests assessing quality of the processed speech showed best results for critical bandwidth based compression using spectral segment mapping and pitch-synchronous analysis-synthesis. The tests for consonant recognition on normal-hearing listeners with hearing loss simulated by broad-band masking noise and on listeners with moderate-to-severe sensorineural loss showed maximum improvement in speech perception for a compression factor of 0.6: the SNR advantage was 6 dB for the normal-hearing listeners and 9 - 21 % improvement in the recognition scores for the hearing-impaired listeners.

Contents

List of Figures

List of Tables

List of Symbols and Abbreviations

Symbols

a	beginning of the spectral segment contributing to a sample on the compressed scale
b	end of the spectral segment contributing to a sample on the compressed scale
c	compression factor
$G_l(f)$	gain of the comb filter for the left ear
$G_r(f)$	gain of the comb filter for the right ear
k	spectral sample index on the unprocessed spectrum
k'	spectral sample index on the frequency compressed spectrum
k_{ic}	center frequency of the i^{th} analysis band
m	lowest integer higher than a
n	highest integer lower than b
N	size of the DFT
p	one-tailed significance level of paired t-test
X	spectrum of the unprocessed speech signal
Y	spectrum of the frequency compressed speech signal
α	amplitude scaling factor for the left ear
β	amplitude scaling factor for the right ear
ρ	power relating amplitude to the loudness
σ_s	square-root of short-time energy of the speech signal
σ_e	square-root of short-time energy of the broad-band noise signal
Φ	mean of left/center and right/center discrimination thresholds
Φ_A	mean of left/center and right/center discrimination thresholds under diotic presentation
Φ_B	mean of left/center and right/center discrimination thresholds under dichotic presentation

Abbreviations

ACB	auditory critical bandwidth
AGC	automatic gain control
ANOVA	analysis of variance
AVC	automatic volume control
BB	broad-band
BLDEL	binaural level difference for equal loudness
BTE	behind-the-ear
CB	constant bandwidth
CB*n*	constant bandwidth filters with *n* bands
CB18	constant bandwidth filter with 18 bands
CV	consonant-vowel
CVC	consonant-vowel-consonant
CVR	consonant-to-vowel ratio
dB	decibel
df	degree of freedom
DFT	discrete Fourier transform
DRT	diagnostic rhyme test
ERB	equivalent rectangular bandwidth
FFT	fast Fourier transform
FIR	finite impulse response
GCI	glottal closure instant
HL	hearing level
HRTF	head related transfer function
IDFT	inverse discrete Fourier transform
IFFT	inverse fast Fourier transform
ILD	inter-aural level difference
ITD	inter-aural time difference
ITE	in-the-ear
ITC	in-the-canal
LPC	linear prediction coefficient
MCL	most comfortable level
MCL-6	6 dB below the most comfortable level
MCL+6	6 dB above the most comfortable level

MRT	modified rhyme test
MOS	mean opinion score
PB	perceptual balance
PTA	pure tone average
rms	root mean square
s.d.	standard deviation
SFM	spectral flatness measure
SNR	signal to noise ratio
SPL	Sound pressure level
VC	vowel-consonant
VCV	vowel-consonant-vowel
VOT	voice onset time
3I-3AFC	Three- interval, three- alternative, forced choice

Chapter 1

INTRODUCTION

1.1 Problem overview

Sensorineural hearing loss occurs when the functioning of the cochlea is affected or when there is a degeneration of auditory nerve. It is characterized by elevated hearing thresholds, reduced dynamic range and loudness recruitment, poor temporal and frequency resolution, and increased temporal and spectral masking. Sensorineural loss is generally associated with widening of auditory filters (Pick et al., 1977; Zwicker and Schorn, 1978; Florentine et al., 1980; Pickles, 1982; Carney and Nelson, 1983; Glasberg and Moore, 1986; Nelson et al., 1990; Pickett, 1999). Widening of auditory filters leads to poor frequency selectivity and increased spectral masking, causing difficulty in speech perception. In such cases, frequency-selective amplification of the speech signal by hearing aids can make the speech audible but may not significantly help in improving the speech perception. This necessitates the development of speech signal processing techniques for reducing the effects of masking.

Several investigations have been reported on binaural dichotic presentation, by splitting the speech signal using a pair of comb filters with complementary magnitude responses, for improving speech perception by persons with moderate bilateral sensorineural loss (Lyregaard, 1982; Lunner et al., 1993; Lunner, 1997; Chaudhari and Pandey, 1998a; Cheeran and Pandey, 2004b; Murase et al., 2004). Masking takes place primarily at the peripheral level, and integration of binaural information takes place at higher levels in the auditory system. In spectral splitting scheme, the spectral components likely to mask each other are presented to different ears, for reducing the effect of spectral masking. For spectral splitting without introducing perceived distortion, the magnitude responses of the filters should not result in a variation in the loudness of different spectral components. The comb filters should have nearly flat response in pass bands and large attenuation in stop bands. The spectral components in the pass bands of each comb filter are presented to the corresponding ear. The spectral components in the transition bands are presented to both the ears. Therefore the two comb filters should have magnitude responses such that perceived loudness of the spectral components in the pass bands and transition bands remain balanced. The earlier studies on spectral splitting have reported mixed results: from no advantage to improvements in the recognition scores corresponding to an SNR advantage of 2 – 9 dB. All the filters used in these studies had linear phase responses but they used different bandwidths and filter realizations. The variations in the results reported may be due to differences in the magnitude responses of the comb filters used. Further investigations are needed to optimize the scheme with respect to bandwidth and magnitude response in the transition bands, in order to obtain maximum improvement in speech perception. Another important issue that has not been addressed in the previous investigations is the effect of spectral splitting on source localization.

For monaural hearing, several studies have investigated the usefulness of spectral contrast enhancement schemes for improving the intelligibility of speech in noise for normal-hearing subjects and for subjects with sensorineural loss (Bunnel, 1990; Stone and Moore, 1992; Baer et al., 1993; Miller et al., 1999; Yang et al., 2003; Cohen, 2006). The processing involved enhancement of the spectral prominences which are perceptually significant. There may be errors in identification of the spectral prominences, and increase in the dynamic range of the speech signal may adversely affect the speech perception due to the reduced dynamic range of hearing associated with the sensorineural loss. Another technique that can be used for reducing the effect of spectral masking in monaural hearing is multi-band frequency compression (Yasu et al., 2002; Arai et al., 2004). In this technique, the speech spectrum was divided into a number of analysis bands, and frequency components in each of these bands were compressed towards its center by a constant compression factor, and it resulted in a modest improvement in speech perception for subjects with sensorineural loss. Further investigations are needed for optimizing the technique.

1.2 Research objectives

The objective of the research is to investigate two speech processing techniques for reducing the effect of intraspeech spectral masking in sensorineural loss: (i) spectral splitting for binaural dichotic presentation, and (ii) multi-band frequency compression for monaural presentation.

Listening tests are conducted to establish the relation between the filter gains for perceptual balance of loudness in the transition bands. This information is used for designing comb filters with different types of bandwidths, nearly satisfying the condition for perceptual balance. Binaural dichotic presentation with optimal set of comb filter pairs is evaluated for speech intelligibility through listening tests conducted on normal-hearing subjects with sensorineural loss simulated by broad-band masking noise and on subjects with moderate bilateral sensorineural loss. To investigate the effect of spectral splitting on source localization, head related transfer functions (HRTFs) are used to generate spatial sounds in the frontal azimuth plane. Degradation in the localization due to spectral splitting is quantified through listening tests conducted on normal-hearing and hearing-impaired subjects.

For improving speech perception in monaural presentation, an analysis-synthesis technique for multi-band frequency compression applied on the complex spectrum using overlap-add method is implemented and investigated. Listening tests are conducted on normal-hearing and hearing-impaired subjects for selecting the best combination of (a) segmentation for analysis-synthesis, (b) bandwidth, (c) frequency mapping scheme, and (d) compression factor.

1.3 Thesis outline

Chapter 2 gives a brief description of sensorineural loss, its characteristics and effect on speech perception, and a review of the signal processing techniques for improving the speech perception for persons with sensorineural loss. Investigations on comb filters for binaural dichotic presentation are presented in Chapter 3. Investigations to study the effect of spectral splitting on source localization are presented in the next chapter. Chapter 5 deals with implementation and evaluation of multi-band frequency compression. The last chapter provides a summary of the investigations, conclusions, and some suggestions for further work. Supplementary investigations on perceptual balance in binaural presentation and lateralization in binaural dichotic presentation are described in Appendix A and Appendix B, respectively. The last three appendices provide supplementary information on test material, instructions to subjects and forms, and tables of statistical analysis.

Chapter 2

SPEECH PROCESSING FOR SENSORINEURAL HEARING LOSS

2.1 Introduction

This chapter describes characteristics of sensorineural hearing loss along with its effects on speech perception. This is followed by a review of the speech processing techniques for improving the speech perception by persons with sensorineural loss.

2.2 Hearing impairment

Depending on the location of the defects in the auditory system, hearing loss can be broadly classified into four types: conductive, sensorineural, central, and functional (Levitt et al., 1980; Pickles, 1982; Yost, 1994; Moore, 1997). Conductive loss is due to an abnormality within the outer or the middle ear, impairing the transmission of sound to the inner ear. It results in increased hearing thresholds, but after a suitable amplification of the sound, speech discrimination is relatively unimpaired. Sensorineural loss is caused by pathology in the cochlea or in the auditory pathway from the inner ear to the brainstem. The loss specifically due to cochlear disorders is termed as sensory loss, while the loss due to disorders in the auditory pathway is termed as neural or retrocochlear loss. Sensorineural impairment can be congenital or acquired. Congenital loss may result from hereditary factors, malformation of the cochlea, intrauterine viral infections, or birth trauma. Acquired loss may occur due to noise exposure, acoustic tumor, head injury, infection, toxic drug effects, vascular disease, or aging.

Central loss occurs when a lesion exists in the central auditory pathway beyond the 8^{th} cranial nerve. It is mainly caused by damage to the brain cortex due to cerebral meningitis, skull trauma, or congenital defects. People with central loss have reduced ability to interpret, integrate, or understand speech. In addition to the organic types of hearing loss, there is also a possibility of functional loss, caused by psychological or emotional factors (Deutsch and Richards, 1979; Wall, 1995).

2.3 Characteristics of sensorineural hearing loss

Hearing loss is commonly quantified by pure-tone audiogram, a graph showing hearing threshold levels in dB HL as a function of frequency (Guyton, 1986; Rintelmann, 1991; Sataloff and Sataloff, 1993; Moore, 1997; Troost and Waller, 1998). The shape of the audiograms for sensorineural loss may vary depending on the type of pathology. A flat audiogram generally indicates weak and damaged auditory nerve fibers caused by salicylate poisoning. High frequency loss is caused by loss of hair cells and stiffening of basilar membrane. Low frequency loss is caused due to Meniere's disease. Many persons with sensorineural losses experience a loss only in the high frequency region (4 – 8 kHz). They generally have no difficulty in understanding speech at normal intensities in a quiet environment, but they experience difficulty in understanding speech in a noisy environment. Mixed loss, with conductive and sensorineural components in the same ear, has attributes of both the conductive and the sensorineural disorders. With a mixed loss, both air and bone conduction thresholds are elevated but the bone conduction thresholds are better than the air conduction thresholds. The difference between the two thresholds is referred to as the air-bone gap and represents the amount of the conductive component of the loss.

The characteristics of sensorineural loss are elevated hearing thresholds, reduced dynamic range and loudness recruitment, poor temporal resolution and increased temporal masking, and poor frequency resolution and increased spectral masking (Levitt et al., 1980; Dorman and Hannley, 1985; Humes et al., 1988; Baer and Moore, 1993; Hou and Pavlovic, 1994; Moore, 1998; Pickett, 1999; Moore, 2002). Many of these psychophysical characteristics tend to be interrelated.

2.3.1 Elevated hearing thresholds

Audibility is the primary requirement for speech intelligibility. Persons with severe loss may not hear any speech sounds unless presented at high levels. Figure 2.1 shows a plot of the sound pressure level vs. frequency, for speech and music, along with the auditory range (hearing and pain thresholds) as a function of frequency. When part of the speech spectrum is below the hearing threshold or is masked by background sound, then that information becomes inaudible, and intelligibility suffers.

2.3.2 Loudness recruitment

Loudness recruitment is an abnormally rapid growth in the sensation of loudness with increase in the level of an acoustic signal (Fowler, 1936; Steinberg and Gardner, 1937; Evans, 1975; Pickles, 1982; Sandlin, 1988; Moore, 1997; Oxenham and Plank, 1997; Derleth et al., 2001; Moore, 2003). It is related to the loss of compressive nonlinearity feature of the basilar membrane, and is generally attributed to the loss of hair cells, particularly outer hair cells.

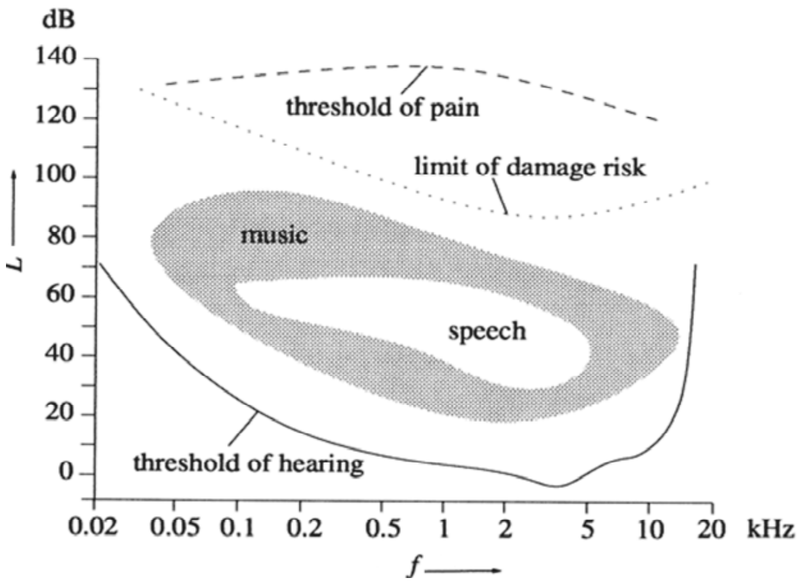


Fig. 2.1 One-third octave band pressure level L (in dB SPL) vs. frequency f for speech and music and the normal auditory range (Zwicker and Fastl, 1999).

Dynamic range is the difference between the hearing threshold level and the loudness discomfort level. Cochlear loss, caused mainly due to the loss of outer hair cells, results in the elevation of hearing thresholds with no corresponding increase in loudness discomfort level (pain level). Thus loudness recruitment reduces the available dynamic range and distorts loudness relationships among components of speech. It limits the benefit of linear amplification to compensate for the loss of audibility and it can severely affect the overall speech perception.

2.3.3 Temporal resolution and temporal masking

The minimum detectable gap between two successive signals is referred to as temporal resolution and it represents the ability of the auditory system to follow the temporal pattern of the sound. Temporal resolution (gap threshold) is about 2 – 3 ms for normal-hearing persons and about 8 ms for persons with sensorineural impairment (Fitzgibbons and Wightman, 1982; Florentine and Buus, 1984). Florentine and Buus (1984) and Shailer and Moore (1983) reported that persons with high frequency loss have larger gap detection threshold than normal-hearing persons. Several studies have reported elderly persons having larger gap thresholds (Moore et al., 1992; Schneider et al., 1994; Strouse et al., 1998; Larsby and Arlinger, 1998).

Poor temporal resolution is normally associated with forward and backward masking of weaker signal by the adjacent strong ones. The masker precedes the signal in forward masking. In backward masking, signal precedes the masker (Moore 1997, Gold and Morgan, 2002). Forward masking is attributed to the reduction in the sensitivity of recently stimulated cells or persistence in the pattern of neural activity evoked by the masker. The forward masking is found within 10 ms of the masker, and it reduces with increase in time (Danaher et al., 1978; Tyler et al., 1982). Backward masking occurs when the onset of the masker masks the offset of the previous signal segment (Oxenham and Plack, 1997). Backward masking is more severe than forward masking for very short masking intervals. The amount of backward masking decreases as the interval increases beyond 15 ms and occurs up to a temporal gap of 100 ms. Persons with sensorineural loss exhibit increased forward and backward masking. Elliot (1975) reported that the effect of backward masking extends up to 100 ms as compared to about 10 ms for normal-hearing persons. Danaher et al., (1978) showed that persons with sensorineural loss exhibit backward and forward masking over an interval 100 – 200 ms. Increased temporal masking results in a degradation of the detection of some of the acoustic events essential for speech perception.

2.3.4 Frequency selectivity

The ability to resolve spectral components in complex sounds is termed as frequency selectivity. The peripheral auditory system can be considered as a bank of overlapping band-pass filters known as the auditory filters (Moore, 1986; Rosen et al., 1998; Baker and Rosen, 2002). Fletcher (1953) studied the characteristics of auditory filters using a pure tone and band-pass filtered noise (centered at the tone frequency) as the masker, by finding the threshold of the tone as a function of the masker bandwidth. The bandwidth beyond which the increase in the threshold ceased was taken as the critical bandwidth. Different investigations have shown the critical bandwidths to be approximately constant for center frequencies below 500 Hz, and between 15-20 % of the center frequencies above 1 kHz (Zwicker, 1961; Pickles, 1982; Moore, 1986). Patterson (1976) determined the threshold for a sinusoid, centered in the spectral notch of broad-band noise, as a function of the width of the notch. Auditory filters were described in terms of an equivalent rectangular bandwidth (ERB) as a function of center frequency. The ERBs were 11 – 20 % of the center frequency. The ERBs and the critical bandwidth described by Zwicker (1961) are almost the same for frequencies above 1 kHz. Figure 2.2 shows the critical bandwidth, described by Zwicker (1961) along with the estimates of the ERB of the auditory filter. Florentine et al. (1980) measured critical bands in subjects with normal-hearing and with cochlear impairment and reported that the critical bands were relatively wider for the subjects with cochlear impairment. Persons with cochlear loss usually have auditory filters that are broader than normal (Glasberg and Moore, 1986; Tyler et al., 1983). This means that their ability to sense the spectral shapes of speech sounds, and to separate components of speech from background noise, is reduced. Investigations using simulated effect of reduced frequency selectivity (by smearing of the short-term spectrum)

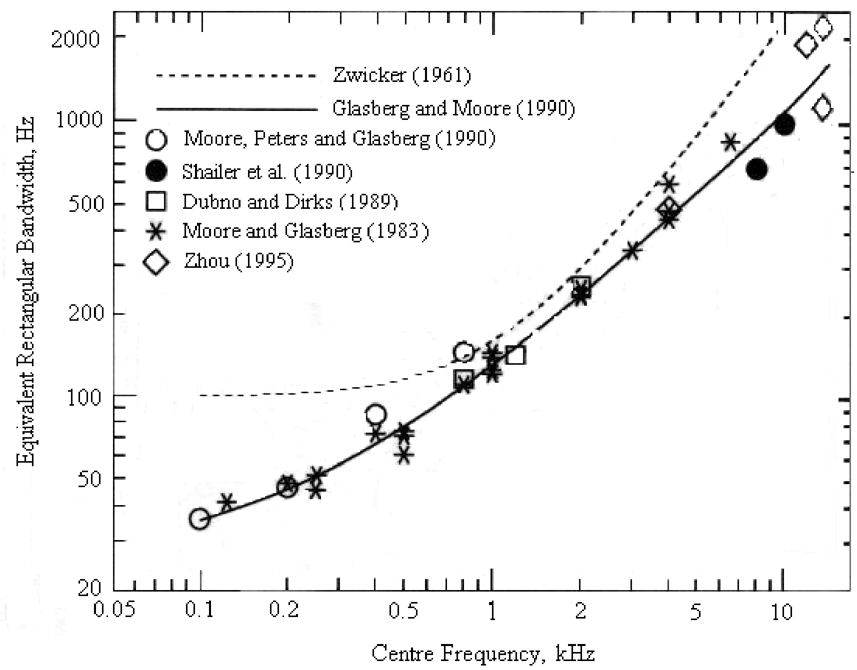


Fig. 2.2 The critical bandwidth (Zwicker, 1961) and the equivalent rectangular bandwidth (ERB) of the auditory filter (Moore, 1997).

have shown that it is a contributing factor in degraded speech perception (ter Keurs et al., 1992; Baer and Moore, 1993, 1994; Nejime and Moore, 1997).

2.4 Effects of sensorineural hearing loss on speech perception

The dynamic range of normal conversational speech is about 40 dB. Because of the elevated hearing thresholds and loudness recruitment, the available dynamic range for the listeners with sensorineural loss is much less (Stone and Moore, 1992a; Moore, 1997). A reduced dynamic range limits the perception of natural level variations of the normal conversational speech and it leads to a distorted loudness relationship among the components of speech sounds. It also leads to a distorted perception of amplitude modulation (Moore et al., 1996; Moore, 2003). Studies by Shannon et al. (1995) and Plomp (1988) have shown that the pattern of amplitude modulation of speech is important for intelligibility.

A major consequence of reduced frequency selectivity is greater susceptibility to spectral masking. In the normal ear, only a small band of masking noise around the signal frequency will pass through the auditory filters. In the impaired ear, auditory filters are relatively broader, allowing a larger band of masking noise (Moore, 1997). Poor frequency selectivity also affects the perception of spectral shape. Broader auditory filters produce a highly smoothed representation of the spectrum than normal auditory filters. The spectral features that are not prominent may be smoothed to such an extent that they become imperceptible (Moore, 2003). Addition of background noise to the speech fills in the valleys between the spectral peaks, further reducing their prominence. Another consequence of reduced frequency selectivity on speech perception in noise is connected with the temporal patterns at the outputs of individual auditory filters. The perceived frequency of a given formant may be partly cued by the time pattern at the outputs of the auditory filters tuned close to the formant frequency, and background noise disturbs this pattern (Young and Sachs, 1979; Miller et al., 1997). This effect is greater in persons with reduced frequency selectivity, since broader filters generally pass more background noise.

Vowels are characterized by well defined formant patterns and formant frequencies, with slowly changing spectrum. Vowels form the nucleus for the syllables and have relatively larger energy (French and Steinberg, 1947; Pickett, 1999; Stevens, 1980). Leek et al. (1987) reported that normal-hearing listeners, in quiet environment, needed a 2 dB peak-to-valley level difference at 75 % correct recognition score for vowel. For the same performance, they needed a level difference of 4 dB when tested by adding masking noise. For hearing-impaired listeners, this difference was 6 dB in a quiet environment. Smoothening of the second and higher formants due to spectral smearing and upward spread of masking may affect the perception of vowels, especially the back vowels having relatively lesser peak-to-valley level difference (Summers and Leek, 1994).

Degradation in temporal resolution and increase in temporal masking may severely affect the speech intelligibility. In speech signals, consonants carry major information crucial for speech intelligibility. As the consonantal segments have relatively less energy than vowels, they are susceptible to masking by their adjacent vowel segments (Flanagan, 1972; Dubno and Levitt, 1981; Sandlin, 1988, pp. 38-50; Ladefoged, 1982). The consonants are characterized by the articulatory features such as manner, voicing, duration, and place of articulation (Miller and Nicely, 1955; Ladefoged, 1982). The auditory system responds to these acoustic features in an integrated manner (Dubno and Dirks, 1989). More than one acoustic cues contribute to the discrimination of each of the features, and an acoustical cue may characterize more than one feature. The phonemes with a larger number of distinctive features in common are more likely to be confused. All the cues are influenced by the adjacent phonemes, the rate of speaking, and talker variability (Pickett, 1999; Dubno and Levitt, 1981).

The acoustic cues for manner discrimination are amplitude, duration, and formant structure. Based on this feature, consonants can be grouped as semivowels, nasals, stops and fricatives. (O’Shaughnessy, 1987). Semivowels have formant structure like the vowels, but the structure varies with time. Further, semivowels have relatively weaker energy, wider bandwidth, and higher concentration of energy at low frequencies as compared to vowels. Nasals are associated with formants and anti-resonances and they have greater low frequency energy (Stevens, 1980; O’Shaughnessy, 1987). Fricatives are characterized by noise-like spectrum and consist of turbulent noise in case of unvoiced fricatives (e.g. /s/) or turbulent noise plus glottal excitation in case of voiced fricatives (e.g. /z/). The fricative spectrum consists of peaks and valleys corresponding to the poles and zeros of the vocal tract filter. The relatively long duration of the noise in fricatives is an important cue for discriminating fricatives from stops. Stop consonants are characterized by rapid closing and opening of constriction somewhere in the vocal tract. Typically, stop bursts are approximately 30 dB weaker than the following vowels and are likely to be masked by the following vowel segments resulting in the poor recognition of the stop consonants. Voice-onset-time (VOT) is an important parameter for distinguishing voiced and unvoiced stops. Tyler et al. (1982) reported that the hearing-impaired listeners cannot effectively use VOT as a cue for discrimination of voicing feature due to poor temporal resolution.

In summary, poor speech recognition ability of persons with sensorineural loss can be attributed to loss of audibility in the parts of speech spectrum, abnormal relationship between intensity and loudness, and increased temporal and spectral masking. In the absence of any background noise, perception of vowels is normally not affected. However, in case of large amounts of spectral smearing, the perception of vowels may also get affected. Consonants are more susceptible to masking and hence they are easily confused. Listeners with high frequency loss can detect the voicing and manner of articulation cues. Since persons with flat loss are less affected by upward spread of masking compared to those with sloping high frequency loss, they have better discrimination of speech compared to those with sloping loss.

2.5 Speech processing techniques for sensorineural hearing loss

Hearing aids can be classified into five types, depending on the size and the way they are worn: body-worn hearing aid, eyeglass hearing aid, behind-the-ear (BTE) hearing aid, in-the-ear (ITE) hearing aid, and in-the-canal (ITC) hearing aid (CHABA, 1991). Amongst these, BTE hearing aids are most commonly used. For persons with sensorineural hearing impairment, frequency-selective amplification of the incoming acoustic signal by hearing aids may improve the audibility but may not improve the speech intelligibility, especially in noisy conditions. Dynamic range compression is commonly used to compensate for loudness recruitment and reduced dynamic range. Other speech processing techniques for improving the speech perception for persons with sensorineural loss include spectral transposition/compression, spectral or temporal enhancement, dichotic presentation of speech signals, etc.

2.5.1 Dynamic range compression

The objective of this technique is to compress the natural dynamic range of the speech signal to fit into the reduced dynamic range of the impaired ear. This is achieved using a compression amplifier. Compression amplification is characterized by compression and time constant parameters (Villchur, 1973; Sandlin, 1988). The compression parameters are: compression threshold, compression range, compression ratio, and compression linearity. Compression threshold is the input level above which gain begins to decrease. Compression range is the range of input level over which the compressor operates starting from the compression threshold. Compression ratio is the ratio of change in the input level to the corresponding change in the output level. Compression linearity refers to the uniformity of the compression ratio over the dynamic range. There are two time constants associated with compression: attack time (or rise time) and release time (or fall time). The attack time is the time interval between the moment the input signal is abruptly increased by a defined amount and the moment the output signal level comes within 2 dB of its new steady state level. Release time (or fall time) is the time interval between the moment when the input is abruptly decreased by the defined amount and the moment when the output signal level comes within 2 dB of its new steady state level.

Methods of compression amplification can be divided into three categories: (i) compression limiting, (2) long-term automatic volume control (AVC), and (3) syllabic compression (Sandlin, 1988). Each of these methods of compression can be implemented either within a single frequency band (wide band compression) or several frequency bands (multi-band compression).

Compression limiting is designed primarily for protection, and operates only at relatively intense sounds. The hearing aid behaves as a conventional amplifier for signals below the threshold of compression. When the compression threshold is exceeded, the gain of the amplifier is reduced accordingly, so that the output does not exceed the uncomfortable level. This scheme is implemented with a high compression ratio, high compression threshold, and short time constants. The amplification for the high level input sounds is restricted within the dynamic range of the hearing-impaired listeners. As a protection device, compression limiting is reported to be generally superior to simple peak clipping and the negative results obtained may be attributed to the poor choice of compression characteristics (Sandlin, 1988). Automatic volume control (AVC), or automatic gain control (AGC), is characterized by a low compression threshold, low-to-medium compression ratio, and long time constants (i.e., time constants much greater than the duration of individual syllables in the speech). It adjusts the gain for long-term variation in speech level so that more of the speech energy lies within the reduced dynamic range of hearing (Sandlin, 1988). In syllabic compression, the parameters of the amplification system are chosen such that relative intensities of the individual speech segments, which occur between or within syllable, are altered. This is achieved with low compression ratio, low-to-medium compression threshold, and short time constants. Syllabic compression enables the adjustment of gain for different speech syllables with the assumption that the relative amplification of weak segments can improve the intelligibility of speech.

An important development in the compression techniques is the introduction of multi-band amplitude compression. The speech is first filtered into a set of contiguous frequency bands. The filtered signals are then amplified using separate gain-compression amplifiers for each band. The outputs of all the amplifiers are summed up to obtain the compressed speech. Initial evaluations of multi-band compression with linear amplification and flat frequency-gain characteristic as a reference condition, were very favorable (Villchur, 1973). Subsequent evaluations of multi-band compression systems, with reference conditions corresponding to linear amplification and frequency-gain characteristic individually shaped, have given mixed results (Moore and Glasberg, 1984; Walker et al., 1984; Working Group on Communication Aids for Hearing Impaired, 1991). A larger number of channels should be effective in a variety of sound environments. However, an increase in the number of channels results in a reduction in temporal and spectral contrasts. Yund and Buckles (1994), evaluated the effect of the number of channels in multi-channel compression hearing aids. An improvement in the discrimination of speech in noise by the hearing-impaired persons occurred when the number of channels was increased from 4 to 8. The performance remained the same when the number of channels was further increased from 8 to 16.

Asano et al. (1991) investigated a digital processing method to compensate for the reduced dynamic range. The input signal was divided into 8 ms segments (corresponds to 128 samples, at 16 kHz sampling rate), and spectrum was computed as the average of 32-point FFT of the sub-segments of each segment. These were used to compute octave band spectra in the range 250 – 8000 Hz on a short-time basis. Relation between the loudness perceived by a normal listener and the hearing-impaired listener was used as a loudness compensation function to find the optimum gain for each octave band. With these frequency gain characteristics, the input signal was processed by an FIR filter. Experimental results indicated an improvement in the recognition scores, as compared to linear amplification, for 9 out of 13 sensorineural hearing-impaired subjects over a wide range of input signals. Tejero et al. (1991) employed a similar technique in which FIR filter was replaced by multi-band analysis-synthesis of speech. In their study, the input was divided into short segments of 25 ms with an overlap of 5 ms, and the magnitude and phase spectra were computed. The magnitude spectrum was modified according to the loudness compensation function, and the speech signal was resynthesized using the unmodified phase spectrum. Listening tests using 10 hearing-impaired subjects with 25 phonetically balanced words showed 10 – 30 % improvement in the recognition scores. Other studies on dynamic range compression for improving speech perception by persons with sensorineural loss have reported varying degree of improvements (Boike and Souza, 2000; Souza et al., 2005; Davies, 2009).

2.5.2 Temporal modification

A reduction in the ability to resolve the frequency components of complex sounds is one of the factors contributing to the difficulty in understanding the speech by persons with sensorineural loss, especially in the presence of background noise. Lorenzi et al. (2006) have shown that a reduced ability to process the temporal fine structure of sounds is also a factor contributing to this difficulty. Earlier studies on temporal modification employed slow playback, which also results in spectral compression, to improve speech perception by people with sensorineural loss (Garvey, 1953; Liberman et al., 1956; Fairbanks and Kodman, 1957; Klumpp and Webster, 1961; Tiffany and Bennett, 1961; Beasley et al., 1972; and Studebaker et al., 1987). Signal processing schemes enable temporal modifications without significant changes in its spectrum (Daniloff et al., 1968; Nagafuchi, 1976). Daniloff et al. (1968) carried out vowel identification tests in a / h/-vowel-/d / context with 20 normal-hearing subjects using temporally shortened signals. They reported a decrease in vowel identification scores only when the signal was 80 % time-compressed. Nagafuchi (1976) performed speech intelligibility tests with 160 normal-hearing children, aged between 4 – 11 years. Test material consisted of 20 nonsense monosyllables, temporally shortened between 75 % and 30 % or temporally prolonged between 150 % and 300 %. The recognition scores for temporally shortened speech signals decreased considerably for 75 % shortening. For 50 % temporally shortened signals, 90 % monosyllable recognition was observed. No significant decrease in recognition was observed for the prolongation of the monosyllables: 100% recognition was achieved, with 200 % temporally prolonged signal.

Several researchers have investigated the effect of enhancement of temporal cues in improving speech perception by persons with sensorineural loss (Bunnell and Martin, 1985; Picheny et al., 1985; Revoile et al., 1986a and 1986b). The perception of voicing for word-final consonants involves the detection and discrimination of both temporal and spectral cues. The temporal cues include the preceding vowel duration and the consonant closure duration, while the formant transitions in adjacent vowels and the presence of a voiced murmur during occlusion provide the spectral cues. Revoile et al. (1986b) investigated the effect of altering the vowel duration on the perception of the final stops, especially fricatives. Test material included spoken words /bæf, bæs, bæv, bæz/. The vowel duration was increased by 100 – 150 ms before the voiced fricatives, and reduced by the same amount before the unvoiced fricatives. Listening tests, conducted on severe-to-profound hearing-impaired subjects, showed an improvement of 20 – 40 % in recognition score.

The fact that consonants are of low intensity, and that they are generally preceded and/or succeeded by high intensity vowels, makes them more susceptible to masking. Many researchers have tried to improve the consonant perception by increasing the consonant-to-vowel intensity ratio (Ono et al. 1982; Gordan-Salant, 1986 and 1987; Montgomery and

Edge, 1988; Thomas, 1996; Thomas et al., 1996; Kennedy et al., 1998; Revoile et al., 2002). Gordon-Salant (1986, 1987), increased the consonant-to-vowel ratio (CVR) and consonantal duration, in order to improve the consonant perception. A set of 19 consonants in CV context were used as the test material. Consonantal duration was increased by 100 % and CVR by 10 dB. Listening tests were conducted on normal-hearing subjects in the presence of babble noise and hearing-impaired subjects in quiet. Both group of subjects showed improvement in recognition scores when CVR was increased.

Thomas et al. (1996) studied the effect of increase in CVR and duration of different acoustic subsegments using the synthesized speech signal, with nonsense syllables consisting of (i) consonants / p, t, k / and vowels / *a*, i, u / in both CV and VC context, and (ii) consonants / p, t, k, b, d, g / and vowel / *a* / in CV and VC context. Tests for CVR increase by 3, 6, 9, and 12 dB were conducted on five normal-hearing subjects by adding broad-band masking noise at different SNR values. Improvements in recognition scores occurred with increase in CVR, with relatively higher scores for the VC than the CV utterances. Analysis of the confusion matrices showed that overall information transmission and transmission of consonantal features increased with increase in CVR. A reduction in the response time also occurred with increase in CVR. Investigations assessing the effect of increase in the consonantal duration, by increasing the duration of acoustic subsegments like closure, closure release burst, formant transitions, and VOT showed that modification of the formant transition duration resulted in improved scores for all the subjects. Only one of the four subjects showed improvement for VOT modification. Other modifications did not result in improvements. In the study by Revoile et al. (2002), enhancements were applied to certain consonantal segments, known to be useful in the perception of consonants. The test material consisted consonants /k/, /t/, /g/, and /d/ located in the final position of the syllables in /bæ-C/ context. Voiced murmur segments during /d/ and /g/ and the release bursts of /t/ and /k/ were amplified above their natural levels. The results showed that stop voicing perception improved significantly (to at least 90 %) for 3 out of 4 hearing-impaired subjects.

2.5.3 Spectral modification

Several techniques involving spectral transposition, spectral reduction, and spectral enhancement have been reported for improving speech perception by persons with sensorineural loss.

In spectral transposition, the selected spectral components, which are important for speech perception, are transposed into the residual hearing frequency range of the sensorineural hearing-impaired listeners. A major concern in this scheme is the spectral distortion which affects the overall speech intelligibility. The simplest method of spectral transposition is the linear frequency compression, but the large frequency compression ratio required for compressing the speech spectrum up to 5 kHz within the spectral range of 1 or 2 kHz results in severe spectral distortion, leading to degraded speech perception. Earlier investigations on frequency compression have shown mixed results (Reed et al., 1983; Turner and Hurtig, 1999; McDermott et al., 1999; Sakamoto et al., 2000; Simpson et al., 2005, 2006; Robinson et al., 2007; Fraga et al., 2008). The study by Reed et al. (1983) was aimed at compressing the speech spectrum to within the hearing range of the hearing-impaired listeners. The processing involved segmentation, warping, dilation and time aliasing, and resynthesis. The best performance obtained for frequency compressed speech was equivalent to the performance obtained by low pass filtering with equivalent bandwidth. In the proportional frequency compression by Turner and Hurtig (1999), the frequency components in the complex spectrum were scaled by a constant factor preserving the ratio between the spectral components. Listening tests on 16 hearing-impaired subjects using nonsense syllables, showed an average improvement of 8 % for female voice and 4.7 % for male voice. McDermott et al. (1999) used nonlinear frequency compression to transpose the high frequency speech components into the lower frequency regions. Different frequency compression ratios were used for the voiced and unvoiced speech segments. However, as reported by Bashford (1987), switching between different frequency compression ratios and the use of the same spectral regions for principally different kinds of spectral information can be perceptually confusing. Sakamoto et al. (2000) reported a nonlinear frequency compression scheme that approximately preserved the naturalness and intelligibility of the speech signal up to certain compression ratio values. The main disadvantage of this compression scheme is the partial nonlinear frequency transposition of the voiced speech segments. The study reported 2 – 12 % improvement in monosyllabic word recognition for five subjects with severe-to-profound loss. In the study by Simpson et al. (2005), frequencies above 1.6 kHz were subjected to nonlinear frequency compression, with the compression increasing progressively with frequency. Monosyllabic word recognition tests using 17 subjects with moderate-to-profound sensorineural loss showed 13 – 17 % improvements in recognition scores.

Generally, the fine spectral details available in the natural speech may not be needed to understand the speech. Thus it may be possible to reduce the spectral content of the speech to suit individual hearing loss, without compromising its intelligibility. In spectral reduction technique, speech is band pass filtered into a number of bands. The signal is resynthesized as a sum of the noise bands or sine waves with frequencies equal to the center frequencies of the channels, amplitude modulated by the amplitude envelopes in the corresponding band (Kryter, 1961; Warren et al., 1995; Dorman et al., 1997; Shannon et al., 1998; Fu and Shannon, 1999; Loizou and Dorman, 1999; Friesen et al., 2001).

Kryter (1961) investigated the intelligibility of bandwidth compressed speech, by filtering the speech into a number of narrow bands with bandwidth of 25, 50, 300, and 500 Hz and with varying center frequencies. The intelligibility was found to be a joint function of number of bands, bandwidth, and center frequency of the narrowband filters. The most effective combination of filters consisted of five filters with 100 Hz bandwidth each and resulted in 80 % word intelligibility. Warren et al. (1995) investigated the speech perception abilities using sentences with bandwidth limited to one-third octave and one-twentieth octave. Stimuli, sampled at 20 kHz, were band pass filtered using nine one-third and one-twentieth octave bandwidth filters with different center frequencies: 370, 530, 750, 1100, 1500, 2100, 3000, 4200, and 6000 Hz. Listening tests on 420 normal hearing subjects showed high recognition scores of 95 % for one-third octave bandwidth filters with center frequencies of 1100, 1500, and 2100 Hz. Very low recognition scores of 23 % were reported for one-twentieth octave filters with center frequency of 370 Hz and 6000 Hz.

Several investigations have addressed the problem of the optimum number of channels needed to understand the spectrally reduced speech (Dorman et al., 1997; Fu and Shannon, 1999; Loizou and Dorman, 1999; Shannon et al., 1998). Dorman et al. (1997) and Loizou and Dorman (1999) reported that the number of channels required depends on the type of test material, with vowels requiring more channels than sentences. For 90 – 100 % recognition, 5 channels were required for sentences and 8 channels were required for vowels and consonants. Increasing the number of channels beyond 8 showed no improvements in the recognition scores. Investigations by Friesen et al. (2001) showed that for normal-hearing subjects, at SNR values of 15, 10, 5, and 0 dB, vowel, consonant, word and sentence recognition score improved with increase in the number of channels and observed 95 % recognition score when number of channels was increased to 20, indicating that the fine spectral information was more important in understanding the speech in noise.

One of the common methods, employed for spectral enhancement, involves enhancement of the short-term spectrum (Gordon-Salant, 1984; Kates, 1994; Summers and Leek, 1997; Moore, 2003). In this technique, the contrast between spectral peaks and valleys is increased, to counteract the level of the noise that fills in the spectral valleys and for making the excitation pattern in an impaired ear more like that evoked in a normal ear by the unprocessed stimuli. Gordon-Salant (1984) investigated a scheme where weak high frequency cues were enhanced while intense low frequency spectral cues were attenuated. The scheme was investigated for consonant identification on subjects with flat loss and high frequency loss. Better recognition scores were obtained for persons having high-frequency losses. Kates (1994) employed a signal processing technique in which most prominent spectral peaks were replaced by a small number of sinusoids, and reported no improvement in speech perception. In order to reduce the effect of upward spread of masking, Summers and Leek (1997) attenuated the energy of the first formant of the vowel in six synthetic consonant-vowel (CV) utterances: /b*a*/, /d*a*/, /g*a*/, /be/, /de/, and /ge/. Listening tests were conducted using normal-hearing and hearing-impaired subjects in quiet and with broad-band masking noise. The recognition scores improved when the first formant was attenuated up to 18 dB, particularly for vowel /*a*/. An improved performance was observed for hearing-impaired subjects, particularly for those with high hearing thresholds in the region between first and second formants.

Some other studies using moderate amounts of spectral enhancement have reported moderate improvements in speech intelligibility and quality (Simpson et al., 1990; Baer et al., 1993). Franck et al. (1999) investigated the separate and combined effect of spectral enhancement and syllabic compression on speech perception by subjects with sensorineural loss. Although the spectral enhancement improved vowel perception, no improvement in consonant recognition was observed. Use of single channel syllabic compression along with spectral enhancement resulted in no additional improvement in recognition scores. However, recognition scores decreased when spectral enhancement was used along with multi-channel syllabic compression indicating that the two processing schemes have opposite effects.

In all of these techniques, the processing involved enhancement of the spectral prominences which are perceptually significant. There may be errors in identification of the spectral prominences. Further, increase in the dynamic range of the speech signal may adversely affect the speech perception due to the reduced dynamic range of hearing associated with the sensorineural loss.

Multi-band frequency compression concentrates spectral energy towards the band centers, without introducing any spectral tilt or compression of the broad-band spectrum. In the multi-band frequency compression reported by Yasu et al. (2002) and Arai et al. (2004), the speech spectrum was divided into a number of bands corresponding to auditory critical bandwidths and the spectral samples in each band were compressed towards the center of the band along the frequency axis. The input speech was divided into frames (frame length: 512 samples, frame shift: 128 samples), and a DFT was computed for each frame after applying Hamming window. The magnitude spectrum was then compressed towards the center of each critical band along the frequency axis, and the resulting magnitude spectrum was combined with the original phase spectrum. Speech signal was resynthesized using the overlap-add method. Compression in the range 0.1 – 0.9 was used. Two experiments were conducted: (i) mean opinion score (MOS) test for a set of six sentences, and (ii) intelligibility scores for 50 vowel-consonant-vowel utterances produced by a male speaker, with two hearing-impaired subjects in each experiment. In the MOS test, the subjects made a pair-wise comparison of the unprocessed and processed speech and rated the processed speech on a 0 – 5 scale with number 3 assigned to the unprocessed. The best MOS scores were observed for the compression factor of 0.6 – 0.8, with the score improvements in the range 0.3 – 0.8. There was a modest improvement in the recognition score: 38.3 % for the processed speech as against 35.4 % for the unprocessed speech.

The main requirement for all the spectral modifications on the speech signal is the preservation of naturalness and intelligibility of the processed speech. Although sufficient acoustic information is presented in most of the techniques, the listener may not be able to interpret the spectrally altered speech into meaningful linguistic message, which may be because of a lack of knowledge and training. Studies have shown that training improved the speech perception ability for the frequency transposition hearing devices (De Filippo and Scott, 1978; Rosen et al., 1987; Dorman et al., 1997; Fu and Shannon, 1999).

2.5.4 Dichotic presentation

Persons with sensorineural loss have a greater difficulty in speech perception due to increased spectral and temporal masking, which primarily occurs at the cochlear level. The speech perception, at higher levels in the auditory system, takes place by combining the information received through both the ears. For reducing the effect of masking, the input speech signal can be processed to split it into two parts in a complementary manner and then given to the two ears for binaural dichotic presentation so that signal components likely to mask each other are presented to the different ears. Several such investigations have been reported (Lyregaard, 1982; Lunner et al., 1993; Chaudhari and Pandey, 1998a, 1998b, and 1999; Jangamashetti and Pandey, 2000a and 2000b; Jangamashetti et al. 2001; Cheeran and Pandey, 2004a and 2004b; and Murase et al. 2004).

In an effort to reduce the effect of temporal masking, earlier studies (Jangamashetti et al., 2000a and 2000b; Jangamashetti, 2003) investigated a scheme in which the speech signal was divided into time segments and alternate segments were presented to the left and right ears. The test material included 12 English consonants / p, t, k, b, d, g, m, n, s, z, f, v / and vowel / *a* / in VCV context. The scheme was evaluated on normal-hearing subjects (three and five respectively for step and trapezoidal fading functions) in the presence of broad-band masking noise at SNR values of ∞, 6, 3, 0, -3 dB. Intra-aural switching cycle of 20 ms and use of trapezoidal fading functions with 2 – 3 ms transitions resulted in highest improvements. At lower SNR, there was increase in recognition score in the range 11 – 23 % and decrease in response time in the range 0.13 – 0.26 s. Listening tests with five subjects with moderate sensorineural loss showed an improvement in recognition scores in the range 4 – 13 %.

In spectral splitting technique, the input speech signal is processed using a pair of comb filters, with complementary magnitude responses. The outputs of these filters are presented to the left and right ears for binaural dichotic presentation. The earlier studies, using different types of comb filters for improving the speech perception for persons with moderate bilateral sensorineural loss with residual hearing in both the ears have shown mixed results (Lyregaard 1982; Lunner et al., 1993; Chaudhari and Pandey, 1998a and 1998b; Cheeran and Pandey, 2005; and Murase et al., 2004).

The comb filters used by Lyregaard (1982) were based on constant bandwidth, realized using sum and difference of the input signal with its delayed version. These filters were designed for an efficient implementation, but the magnitude responses had a large overlap between the pass bands and stop bands. Three hearing-impaired subjects and two subjects with normal hearing participated in the listening tests. The test material consisted of two lists of 25 words, presented at signal-to-noise ratio of 12 and 4 dB for the hearing-impaired subjects and at 6 and -2 dB for the normal-hearing subjects. No significant improvement in the recognition score was reported for dichotic presentation.

Lunner et al. (1993) investigated spectral splitting using an 8-channel digital filter bank, with constant bandwidth of 700 Hz. The filters were designed with emphasis on realization efficiency and not on the separation of bands and perceptual distortion. The scheme was evaluated by conducting listening tests on three subjects with moderate bilateral sensorineural loss using a list of sentences with five words. An overall advantage of 2 dB in signal-to-noise-ratio was reported at 50 % correct recognition score.

Chaudhary and Pandey (1998a, 1998b) used linear-phase comb filters based on auditory critical bandwidths (Zwicker, 1961), with adjustable pass band gains and sharp transition bands. Listening tests on subjects with moderate bilateral sensorineural loss, with consonant-vowel and vowel-consonant-vowel utterances, showed improvements in recognition score in the range 5 – 9.5 %. Cheeran and Pandey (2004a, 2004b) used a pair of comb filters based on auditory critical bandwidth with complementary magnitude responses, designed for minimizing spectral distortion, for spectral splitting. These filters had transition bands in the range 78 – 117 Hz, pass band ripple of less than 1 dB, minimum stop band attenuation of 30 dB, and inter-band crossover gain in the range of -4 to -6 dB. Listening tests were conducted on five normal-hearing subjects with simulated hearing loss and on five subjects with moderate-to-severe bilateral sensorineural loss. The processing resulted in an SNR advantage of 9 dB for normal-hearing subjects. For the hearing impaired subjects, the improvements in recognition scores were in the range 7 – 20 %.

With an aim of overcoming the effect of both spectral and temporal masking simultaneously, earlier studies (Jangamashetti et al., 2001; Pandey et al., 2001; and Jangamashetti, 2003) have investigated combined splitting. In this scheme, a pair of time-varying comb filters with pre-calculated sets of coefficients was used. The scheme was implemented with *m* shiftings with each time-varying comb filter consisting of *m* sets of coefficients corresponding to *m* perceptually balanced filter pairs. The magnitude responses of these filters were such that the pass bands of each of these filter pairs were shifted along the frequency axis in a complementary manner. At any time, spectral splitting was achieved by processing the input speech using a pair of comb filters with complementary magnitude responses. Temporal splitting was then obtained by sweeping the bands. The bands were swept with a sweep cycle of 20 ms, and number of shiftings equal to 2, 4, 8, and 16. Twelve English consonants / p, t, k, b, d, g, m, n, s, z, f, v / and vowel / *a* / in VCV context were used as the test material for closed set identification. The scheme was evaluated on five subjects with normal hearing under simulated loss. The results showed an improvement in recognition score, response time, and information transmission for all shiftings. Maximum relative improvements were reported for SNR of -15 dB: 19 and 17 % for 4 and 8 shiftings respectively.

In a study comparing temporal, spectral, and combined splitting (Jangamashetti, 2003), on five subjects with moderate-to-severe bilateral sensorineural loss, the schemes provided different degrees of improvement, depending on the type of hearing loss. Temporal splitting was found to be useful for persons with high frequency loss. The persons with low frequency and gradual sloping symmetrical loss preferred combined splitting.

2.6 Research objectives

Persons with sensorineural hearing impairment have difficulty in understanding the speech, particularly in the presence of noise. The difficulty in speech perception can be attributed to several factors: (i) reduced audibility, wherein part of the speech spectrum becomes inaudible, (ii) loudness recruitment, which reduces the dynamic range and distorts loudness relationships among components of speech, (iii) increased temporal masking, which reduces the ability to use temporal cues important for speech perception, and (iv) increased spectral masking, which contributes to difficulty in discriminating spectral peaks due to masking by other spectral components of the speech signal and background noise.

Frequency-selective amplification can partly overcome the effects of reduced audibility. Several forms of dynamic range compression have been shown to be effective in compensating for the effects of loudness recruitment. Several spectral reduction and spectral transposition/compression techniques have been investigated to improve the speech perception for persons with high frequency hearing loss. The objective of these schemes is to either compress the speech spectrum within the residual hearing range for the hearing-impaired listener or to provide spectrally reduced but still critical auditory information to improve the speech perception ability. For overcoming the effects of increased temporal masking, several techniques, e.g., enhancement of consonant-to-vowel intensity ratio, duration modification of certain acoustic segments, have been reported with varying degrees of benefits. Many techniques, which involve splitting of speech into two parts in a complimentary manner, have been studied for binaural dichotic presentation. These techniques include temporal splitting, comb filter based spectral splitting, and combined splitting. The investigations with these schemes have shown varying degree of improvement in speech perception.

A review of the different speech processing techniques for improving speech perception by persons with sensorineural loss indicates scope for further investigations for developing and evaluating speech processing techniques, with optimal processing parameters, for reducing the effect of intraspeech spectral masking for both monaural and binaural hearing.

The objective of the research is to investigate two techniques for improving speech perception by persons with sensorineural loss: (i) spectral splitting scheme with binaural dichotic presentation for persons with moderate bilateral loss, and (ii) multi-band frequency compression of speech with monaural presentation for persons with residual hearing in only one ear.

In the earlier studies on spectral splitting, the issues of optimizing the filters with respect to bandwidth, and transition band responses for perceptual balance were not addressed. The investigations on monaural-binaural loudness balance, presented in Appendix A, showed that the gains of the two filters for left and right ears should be complementary on a linear scale. Based on this result, different pairs of comb filters are designed with their magnitude responses closely satisfying the requirement for perceptual balance. Investigations on binaural dichotic presentation using the comb filters for improving speech perception by normal-hearing subjects in the presence of masking noise, and by subjects with moderate bilateral sensorineural loss is presented in Chapter 3. One of the concerns in the use of dichotic presentation is its adverse effects on sound source localization. Investigations on the effect of spectral splitting on source localization, through listening tests conducted on normal-hearing and hearing-impaired subjects, are presented in Chapter 4.

In the investigation with monaural presentation, in Chapter 5, an analysis-synthesis technique for multi-band frequency compression, applied on the complex spectrum using overlap-add method, is implemented and optimized for (a) segmentation for analysis-synthesis, (b) bandwidth, (c) frequency mapping scheme, and (d) compression factor. Mean opinion score (MOS) tests, for quality assessment, are conducted on normal-hearing subjects in the presence of masking noise. Listening tests are then conducted for quantitative evaluation of speech intelligibility using normal-hearing subjects in the presence of broad-band masking noise and subjects with moderate-to-severe sensorineural loss.

Chapter 3

COMB FILTERS FOR BINAURAL DICHOTIC PRESENTATION

3.1 Introduction

Several investigations using binaural dichotic presentation, by spectrally splitting the speech signal using a pair of comb filters with complementary magnitude responses, have been reported (Lyregaard, 1982; Lunner et al., 1993; Chaudhari and Pandey, 1998b; Cheeran and Pandey, 2004b; Murase et al., 2004). In the spectral splitting scheme, alternate spectral bands are presented to the left and the right ears. These studies have reported mixed results: from no advantage to improvements in the recognition scores corresponding to an SNR advantage of 2 – 9 dB. All the filters used in these studies had linear phase responses but they used different bandwidths and filter realizations. The variations in the results reported may be due to the different bandwidths and magnitude responses of the filters in these studies. The magnitude responses of the comb filters used for spectral splitting should not result in variation in the loudness of different spectral components. The comb filters should have nearly flat response in the pass bands and large attenuation in the stop bands. The spectral components in the pass bands of each filter are presented to the corresponding ear. The spectral components in the transition bands are presented to both the ears. Therefore the two comb filters should have magnitude responses such that perceived loudness of the spectral components in the pass bands and transition bands are matched. It is further desirable that the processing and presentation do not result in lateralization of the sound and that the source localization ability using binaural hearing aids is not adversely affected.

The objectives of the investigations reported in this chapter are to design the comb filters with responses for perceptually balanced loudness and to assess the effect of bandwidth in improving speech perception by normal-hearing subjects in the presence of broad-band masking noise and subjects with moderate bilateral sensorineural loss. The next section provides a review of earlier investigations on spectral splitting, a comparison of the comb filters used in the earlier studies, and a description of the filters used in the present study. The

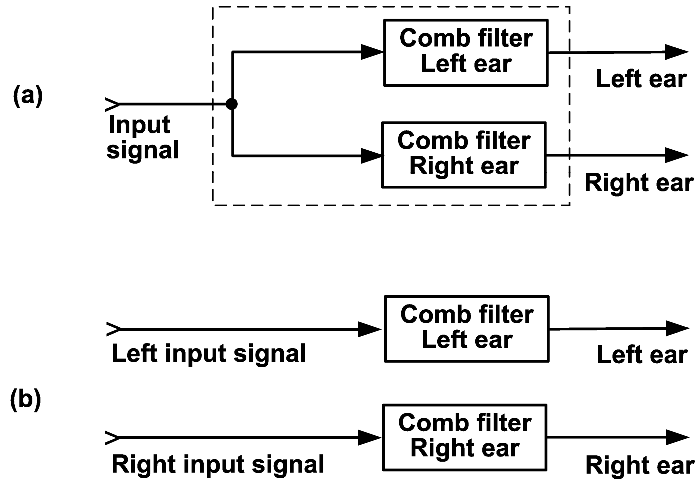


Fig. 3.1 Spectral splitting for binaural dichotic presentation: (a) single speech input, and (b) separate inputs for the left and the right ears.

subsequent sections present the listening tests for speech intelligibility, results, and discussion. Investigations on the effect of spectral splitting on localization of sound source are presented in the next chapter.

3.2 Spectral splitting for binaural dichotic presentation

Increased spectral masking in sensorineural loss reduces spectral contrast leading to degraded speech perception. Spectral splitting scheme with a single speech input (from a frontal microphone) is shown in Fig. 3.1(a). It uses a pair of comb filters with complementary pass bands and stop bands. It is based on the assumption that spectral masking takes place primarily at the peripheral level, and the speech perception process involves integration of information from the spectral components presented to the two ears. The neighboring spectral components which are likely to mask or get masked by each other are presented to different ears. The two comb filters may be implemented on a single digital signal processor with two outputs. An alternative scheme is shown in Fig. 3.1(b). Here the filters are applied on the input signals of the two hearing aids. This scheme should be implemented using separate processors for the left and the right ears so that no communication or coordination is needed between the two aids. In either of the two schemes, spectral splitting by comb filters may be combined with filters for adjustable frequency response and multi-band amplitude compression.

Lyregaard (1982) used a pair of constant bandwidth comb filters, with complementary magnitude responses, to separate the spectral components into two parts for binaural dichotic presentation. The two filters were realized using sum and difference of the input signal and its delayed version. Comb filters with bandwidth of 200, 500, and 800 Hz were realized by adjusting the delay. These filters were designed for an efficient realization, but the magnitude responses, as shown in Fig. 3.2(a), had a large overlap between the pass bands and stop bands. Listening tests were conducted using two lists of 25 words on two subjects with normal hearing and three hearing-impaired subjects. The stimuli were presented at signal-to-noise ratios of 6 and -2 dB for the normal-hearing subjects and 12 and 4 dB for the hearing-impaired subjects. No significant improvement in the recognition score was reported for dichotic over diotic presentation. Lack of significant improvement was attributed to three possible factors: improper filtering, insufficient listening experience by the subjects, and non-feasibility of binaural fusion of dichotic signals.

Lunner et al. (1993) employed an 8-channel digital filter bank, realized using complementary interpolated linear phase filters with constant bandwidth of 700 Hz, for spectral splitting. In case of dichotic presentation, odd numbered channels were presented to one ear and even numbered channels to the other ear. For diotic presentation, all the channels were given to both the ears. The filters were designed for an efficient realization and the gains were complementary on a linear scale. It was implemented for sampling rate of 11.6 kHz and the filters had a processing delay of about 4 ms. The magnitude responses (as shown in Fig. 3.2(b)) had much better separation of pass and stop bands as compared to the filters used by Lyregaard (1982). The gains of the channels were adjusted in accordance with the hearing thresholds of the subject. The test material, consisting of a list of sentences with five words, was presented in the presence of masking noise. The scheme was tested using three subjects with moderate bilateral sensorineural loss. An advantage of 2 dB in signal-to-noise-ratio was reported at 50 % recognition score.

The comb filters used by Chaudhari and Pandey (1998b) were based on auditory critical bandwidths (Zwicker, 1961) and were realized as a pair of 128-coefficient linear phase filters with sampling frequency of 10 kHz. The filters were designed for sharp transitions between pass bands and stop bands in the magnitude responses as shown in Fig. 3.2(c). Listening tests, conducted on subjects with moderate bilateral sensorineural loss, with consonant-vowel and vowel-consonant-vowel utterances as the test material showed an improvement in recognition scores in the range 5 – 9.5 %. Cheeran and Pandey (2004a, 2004b) used a pair of 256-cofficient linear phase FIR filters (sampling frequency = 10 kHz) with complementary magnitude responses designed for reducing the variation in the loudness of spectral components with frequency. The filter magnitude responses, shown in Fig. 3.2(d), had transition bandwidths in the range 78 – 117 Hz, pass-band ripple less than 1 dB, minimum stop-band attenuation of 30 dB, and inter-band crossover gain in the range -4 to -6 dB. Listening tests conducted with 12 nonsense vowel-consonant-vowel utterances, on 5 subjects with moderate to severe bilateral sensorineural hearing impairment, showed the improvements in recognition scores in the range 7 – 20 %.

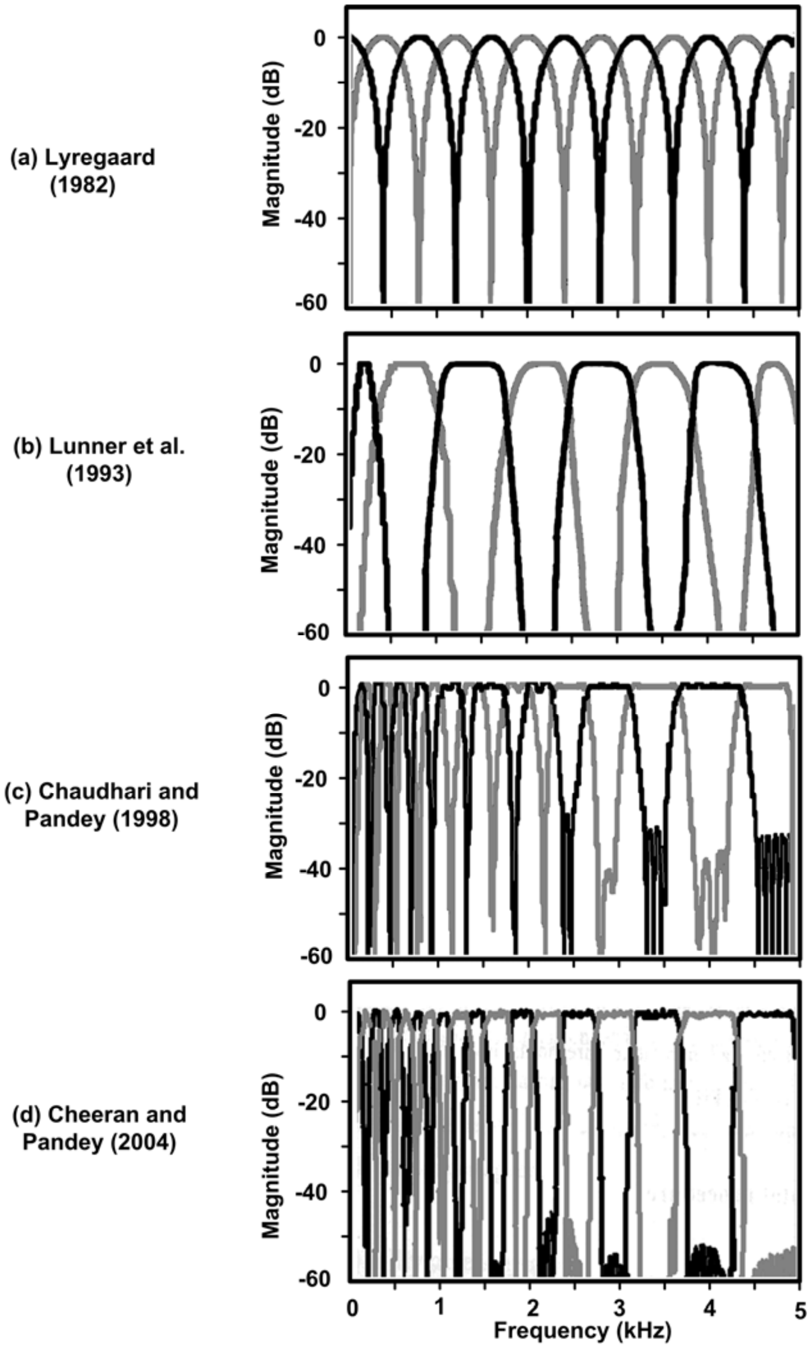


Fig. 3.2 Magnitude responses of comb filters used in earlier studies on binaural dichotic presentation: (a) Lyregaard (1982): obtained from the equations of filter realization, (b) Lunner et al., (1993): obtained after equalization of band peaks, (c) Chaudhari and Pandey (1998b), and (d) Cheeran and Pandey (2004). The black and gray traces show the magnitude responses for the left and right filters, respectively.

Earlier studies have used different types of filters without addressing the effects of filter bandwidths and filter responses. Listening tests were conducted to establish the relation between filter gains for perceptual balance of the loudness in the transition bands, and these investigations are presented in Appendix A. The tests conducted with pure tones of 250 Hz,

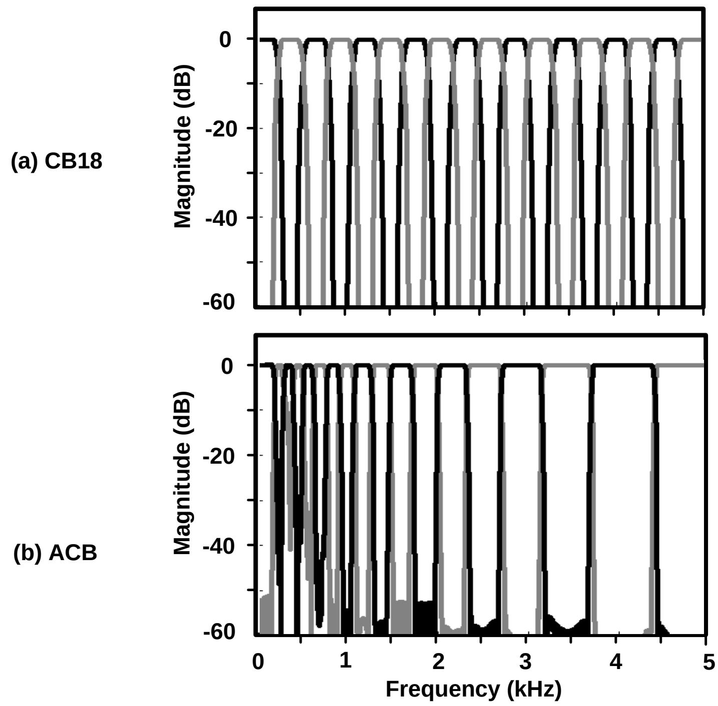


Fig. 3.3 Magnitude response of the pair of 513-coefficients perceptually balanced comb filters. Sampling frequency = 10 kHz, pass band ripple < 1 dB, band crossover gain: -5 to -6 dB. (a) CB18: stop-band attenuation > 64 dB, transition bandwidth = 75 – 80 Hz and, (b) ACB: stop-band attenuation > 29 dB, and transition bandwidth = 45 – 55 Hz. The black and gray traces show the magnitude responses for the left and right filters, respectively.

500 Hz, 1 kHz, and 2 kHz gave the value of binaural level difference of equal loudness (BLDEL) as 6 dB, which is in agreement with the range of values reported in the earlier studies (Scharf, 1968; Marks, 1978; Hall and Harvey, 1985; Hawkins et al., 1987; Zwicker and Henning, 1991; Epstein and Florentine, 2005; Whilby et al., 2006). The results presented in Appendix A further showed that the sum of the magnitude responses of the comb filters should be complementary on a linear scale. In another investigation, presented in Appendix B, the constant bandwidth filters with bandwidth smaller than about 350 Hz and filters based on auditory critical bandwidth showed no lateralization effect. On the basis of these results, two types of filters with perceptually balanced responses have been investigated: (i) filters based on constant bandwidth (CB), with 18 bands in the range 0 – 5 kHz (CB18) and (ii) filters based on auditory critical bandwidth (ACB). The magnitude responses of CB18 and ACB filters are shown in Fig. 3.3. These were designed as 513-coefficient linear phase FIR filters for sampling frequency of 10 kHz, by iterative application of frequency sampling technique (Rabiner and Gold, 1975; Proakis and Manolakis, 1992; Ifeachor and Jevis, 1997; Oppenheim et al., 1999) to get the desired magnitude responses. The transition bandwidths (-1 dB to -40 dB) of the filters were 75 – 80 Hz and 45 – 55 Hz for CB18 and ACB filters, respectively. Both types of filters had an inter-band crossover gain in the range of -5 to -6 dB

Table 3.1 List of bands in the two types of filters: (a) constan bandwidth (CB18), and (b) auditory critical bandwidth (ACB). f_c: center frequency, f_1: lower cutoff, f_2: upper cutoff.

(a) CB18

Band no.	f_c (kHz)	$f_1 - f_2$ (kHz)
1	0.139	0.000 - 0.278
2	0.417	0.278 - 0.556
3	0.694	0.556 - 0.833
4	0.972	0.833 - 1.111
5	1.250	1.111 - 1.389
6	1.528	1.389 - 1.667
7	1.806	1.667 - 1.944
8	2.083	1.944 - 2.222
9	2.361	2.222 - 2.500
10	2.639	2.500 - 2.778
11	2.917	2.778 - 3.055
12	3.195	3.055 - 3.333
13	3.472	3.333 - 3.611
14	3.750	3.611 - 3.889
15	4.028	3.889 - 4.167
16	4.306	4.167 - 4.444
17	4.583	4.444 - 4.722
18	4.861	4.722 - 5.000

(b) ACB

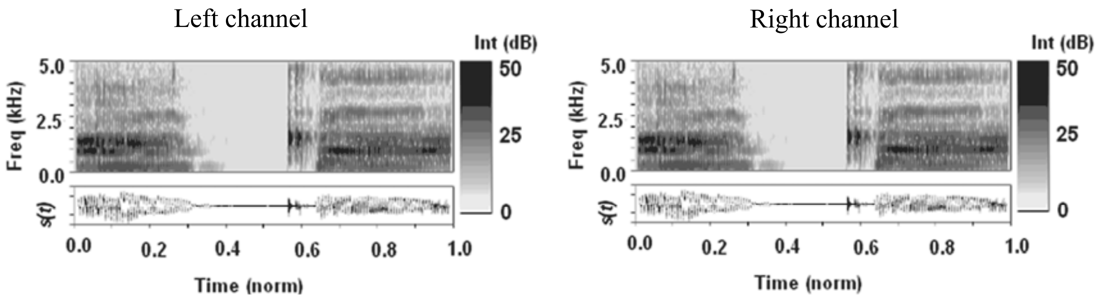
Band no.	f_c (kHz)	$f_1 - f_2$ (kHz)
1	0.13	0.01 – 0.20
2	0.25	0.20 – 0.30
3	0.35	0.30 – 0.40
4	0.45	0.40 – 0.51
5	0.57	0.51 – 0.63
6	0.70	0.63 – 0.77
7	0.84	0.77 – 0.92
8	1.00	0.92 – 1.08
9	1.17	1.08 – 1.27
10	1.37	1.27 – 1.48
11	1.60	1.48 – 1.72
12	1.86	1.72 – 2.00
13	2.16	2.00 – 2.32
14	2.51	2.32 – 2.70
15	2.92	2.70 – 3.15
16	3.42	3.15 – 3.70
17	4.05	3.70 – 4.40
18	4.70	4.40 – 5.00

and pass band ripple of less than 1 dB. The minimum stop band attenuation was 64 dB for CB18 and 29 dB for ACB filters.

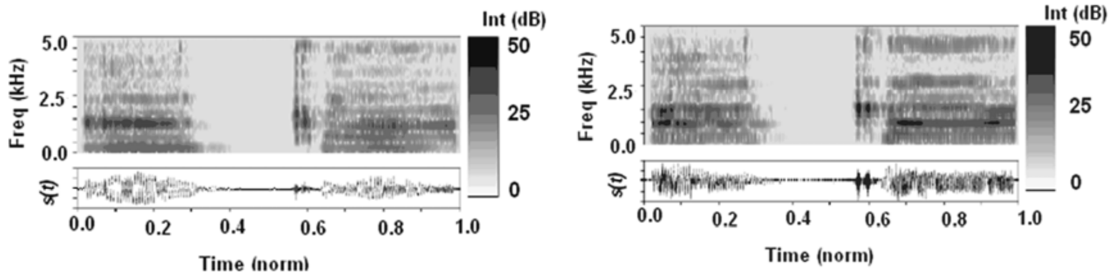
3.3 A comparative study of comb filter responses

The magnitude responses of the comb filters used in earlier studies and two types of filters used in the present study (CB18 and ACB) were examined for the condition of magnitude responses being complementary on a linear scale. The filters employed by Lyregaard (1982) had a periodically occurring maximum deviation of 0.4, indicating the possibility of a large variation in the loudness of spectral components, which could be one of the reasons for the reported lack of significant improvement in speech perception. The comb filters used by Lunner et al. (1993) used complimentary response linear-phasefilter, with the sum of the filter gains remaining constant with frequency. The filters used by Cheeran and Pandey (2004b) had a deviation of 0.26 at about 500 Hz and less than 0.18 for the rest of the band. The filters used in the current study have small deviations of less than 0.04, except for a deviation of approximately 0.2 at about 300 Hz for ACB filters.

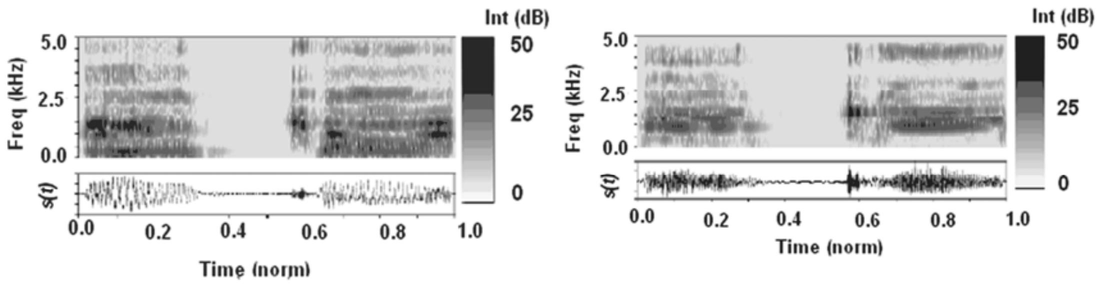
The center frequencies and the bandwidths (-6 dB points) are given in Table 3.1.The filter coefficients and impulse responses are given in Appendix F. The waveforms and wide-band spectrograms of VCV utterance /*aka*/ for unprocessed diotic presentation and for dichotic presentation using CB18 and ACB filter pairs are given in Fig. 3.4. It is observed that



(a) Unprocessed diotic



(b) Dichotic with CB18 filters



(c) Dichotic with ACB filters

Fig. 3.4 Waveforms and wide-band spectrograms (Δf = 300 Hz) of the VCV utterance /*aka*/ of 700 ms duration: (a) unprocessed diotic, (b) dichotic with CB18 filters, and (c) dichotic with ACB filters.

formant transitions are retained and harmonic structure in the voiced segments is preserved in the form of vertical striations.

3.4 Listening tests

Earlier studies involving evaluation of speech intelligibility have used different methods and test materials. In an open response test, the subject responds by repeating what is perceived. In a closed response test, subject responds by selecting the best match to what is perceived from a set of choices. Two of the commonly used intelligibility tests at word level are of the closed response type: (i) diagnostic rhyme test (DRT) and (ii) modified rhyme test (MRT). Both the tests assess the perception of consonants, which are considered to be more important for word intelligibility and are more likely to be confused than vowels. In DRT, only the initial consonants are tested and the subject has to select the correct word from a list of two words (Voiers, 1983). MRT provides a direct quantitative measure of speech intelligibility for words with both initial and final position variations of consonants, and requires minimal training for the listeners (House, et al., 1965; Kreul, et al., 1968; ANSI, 1989; Mackersie, et al., 1999; Yang and Hodgson, 2006; Yoo et al., 2007). The subject has to select the correct word from a list of six rhyming words. Another test often used, at word level, uses a set of phonetically balanced (PB) words, mostly presented for open-set response. Kryter (1965) compared the recognition scores obtained by MRT and PB word test by conducting listening tests on eight normal-hearing subjects and speech spectrum shaped noise as the masker. The scores for MRT and PB test with 200-word list were similar, but 25 % lower score was observed for 1000-word PB test.

The evaluation of the spectral splitting scheme in improving speech perception was carried out using modified rhyme test (MRT). In a multiple choice listening test, the response time provides a measure of the load on the perception process, and a decrease in the response time indicates an improved listening condition (Gatehouse and Gordon, 1990; Delogu et al., 1991; Baer et al., 1993; Meftah and Boudelaa, 1996; Apoux, et al., 2001). Hence the listening tests were conducted by recording the subject response as well as the response time.

The test material for MRT, as given in Appendix C, consisted of 50 sets of monosyllabic words of consonant-vowel-consonant (CVC) form. Each set consisted of six words with a vowel in the middle and either initial or final consonant remaining the same and the other consonant being different. Each of the words was preceded by a carrier phrase “Would you write ------”. All the sentences in the test material were recorded from a male speaker in an acoustically treated audiometry room, using a B&K microphone model 2210, at a sampling frequency of 10 kHz and with 16-bit quantization. All the 300 words (i.e. 50 sets × 6 words in each set) were arranged in 6 test lists (1x, 1y, 2x, 2y, 3x, 3y) of 50 words each. The words in each test list were selected by a two-level randomization process: (i) the set level, 1, 2, 3, and (ii) the word level, x, y within a set, in such a way that every word was presented once. The stimuli were presented binaurally through headphones. The presentation level was set at the most comfortable listening level as selected by the individual listener, and the amplifier gain setting was maintained across all the listening conditions. Hence there was no change in the signal level for the different processing and listening conditions.

The test was conducted using an automated test administration setup in an audiometry room. The subject, seated in front of the computer screen, clicked the “play” button on the test window, listened to the presentation, responded by selecting the best match out of the closed set of 6 response choices displayed on the screen. The order of the response choices on the screen was randomized to eliminate position bias. The test setup recorded the subject response and the response time. The presentation-response process was repeated for all of the 50 words in the test list.

Assessment of speech processing techniques and optimization of processing parameters involves listening tests on hearing-impaired listeners. As these tests are time consuming and may cause a fatigue, a simulation of hearing loss may be used for conducting the listening tests on normal-hearing subjects, for a preliminary evaluation and particularly for selecting the processing parameters. Different types of simulations have been reported to characterize the different aspects of sensorineural impairment (Vilchur, 1974; Vilchur, 1977; ter Keurs et al., 1992; Moore and Glasberg, 1993; Nejime and Moore, 1997). For an approximate simulation of intraspeech masking, the present study used addition of broad-band noise, band-limited to speech frequency range, at a specific SNR with respect to short-time (10 ms) energy of the signal. Different levels of loss were simulated by varying the SNR. The output signal $x(n)$ was obtained by adding stationary random noise $e(n)$ to the speech signal after scaling the noise for SNR = k dB for every 10-ms segment,

$$x(n) = s(n) + \beta e(n) \tag{3.1}$$

where $\beta = [\sigma_s^2 / \left(\sigma_e^2 \times 10^{0.1k}\right)]^{0.5}$ and $\sigma_s^2 = \sum s^2(n), \quad \sigma_e^2 = \sum e^2(n)$.

Since the noise was stationary, the scaling factor β depended on the energy of the input signal segment. The simulation did not add noise during silence segments. It resulted in the recognition scores for normal- hearing subjects similar to those for subjects with moderate-to-severe sensorineural loss (Jangamashetti et al., 2010). In consonant recognition test, it had maximum adverse effect on the reception of place and duration features. It may be noted that the final listening tests should be conducted on the hearing-impaired listeners.

Assessment of the effectiveness of the two types of comb filters (CB18 and ACB) in improving the speech perception involved two experiments. In the first experiment (Exp. IA), MRT was conducted on six normal-hearing subjects (4 male and 2 female, age: 35 – 45 years, pure-tone thresholds < 20 dB HL), with simulated sensorineural loss. Nine SNR values were used for the investigation: ∞ (no noise), 6, 3, 0, -3, -6, -9, -12, and -15 dB. Each subject responded for a total of 8100 presentations: 300 words × 3 processing conditions (diotic, dichotic with CB18 filters, and dichotic with ACB filters) × 9 SNR values. In order to minimize any bias due to practice, or fatigue, the presentation order for the different processing and listening conditions was randomized for each subject. On a given day, a subject participated in a maximum of two test sessions. A test session for each listening condition took approximately 45 min. For the six subjects and three processing conditions, total test period was spread over one month depending on the convenience and willingness of the subjects. In the second experiment (Exp. IB), MRT was conducted on 11 subjects (8 male and 3 female, age: 18 – 56 years) with moderate bilateral sensorineural loss (average pure tone threshold: 40 – 77 dB HL, asymmetry < 12 dB). All the subjects read and signed

Table 3.2 Hearing thresholds (dB HL) of the subjects with bilateral sensorineural hearing impairment. PTA: average pure tone threshold (dB HL), taken at 0.5, 1, and 2 kHz.

Sub. (sex, age)	Ear	Hearing threshold (dB HL)						PTA (dB)
		Frequency (kHz)						
		0.25	0.5	1	2	4	6	
IGH (M, 43)	L	40	45	55	50	55	60	50
	R	55	55	65	65	60	65	62
MNA (M, 21)	L	60	60	70	60	65	70	63
	R	65	65	60	65	70	75	63
NBC (F, 34)	L	50	50	60	60	65	65	57
	R	60	55	65	65	65	70	62
PAT (F, 56)	L	50	50	55	60	65	70	55
	R	55	55	55	60	60	70	57
PND (M, 18)	L	50	55	60	55	60	60	57
	R	55	55	50	60	65	65	55
PRP (M, 35)	L	70	75	75	80	85	90	77
	R	70	70	70	75	75	80	72
PSE (M, 34)	L	50	50	55	55	55	65	53
	R	60	60	60	65	70	75	62
RAJ (F, 50)	L	40	40	45	50	55	60	45
	R	50	50	60	65	65	65	58
SIY (M, 35)	L	50	60	65	65	60	70	63
	R	55	50	60	65	70	70	58
SKS (M, 28)	L	30	35	40	45	40	50	40
	R	40	40	50	45	55	55	45
SPB (M, 23)	L	45	45	50	55	65	65	50
	R	50	50	55	60	65	70	55

informed consent for participation in the listening tests and were compensated for travel and incidental expenses. The hearing thresholds of all the subjects are given in Table 3.2. As the results from Exp. IA showed that improvements in speech perception using ACB filters were much higher than those obtained using CB18, the listening tests in Exp. IB were conducted for the ACB filters only. The tests were conducted without the masking noise. The presentations were made binaurally through headphones and the subjects did not wear hearing aids. No frequency dependent gain or amplitude compression was used. Each subject responded for a total of 600 presentations: 300 words × 2 processing conditions (diotic, dichotic with ACB filters). As in the case of normal-hearing subjects, the presentation order for the different processing conditions was randomized for each subject in order to minimize the bias due to practice or fatigue. A test session for each processing condition was approximately of one hour duration. For the 11 subjects and with the two processing conditions, total test period was spread over one month depending on the convenience and willingness of the subjects.

3.5 Results

The results of the two experiments conducted for assessing the intelligibility of the processed speech, for normal-hearing and hearing-impaired subjects, are presented in this section.

Experiment IA: Tests on normal-hearing subjects

The results of MRT conducted on six normal-hearing subjects, with different levels of masking noise, are shown in Table 3.3. In addition to the recognition scores for the individual subjects, the mean score, the standard deviation (s.d.), the mean improvement, and one-tailed significance level (p) for paired (dichotic vs. diotic) t-test are also given in the table. Although no improvement in recognition scores was observed for higher SNR values, recognition scores increased for SNR values less than 0 dB. The increase in recognition scores ranged 4 – 18 % for CB18 and 7 – 28 % for ACB, and the improvements were statistically significant ($p < 0.001$) for both the filters. The improvements with ACB filters were higher than those with CB18 filters for all the subjects. Figure 3.5 gives a plot of percentage recognition score (averaged across the six subjects) as a function of SNR. It shows that the advantage of dichotic presentation increased monotonically with decrease in SNR. For all the SNR values, ACB filters gave higher improvement than the CB18 filters, increasing from 7 % at 0 dB to 28 % at -15 dB. As seen in Fig. 3.5, diotic presentation with SNR of -3 dB resulted in recognition score of approximately 75 %. For dichotic presentation with CB18 and ACB filters, the same recognition score occurred for SNR of about -9 dB and -15 dB, respectively. Thus the improvement in recognition scores for CB18 and ACB filters corresponded to SNR advantages of approximately 6 and 12 dB, respectively. A two-way repeated-measures analysis of variance (ANOVA) was conducted on recognition scores with processing and SNR as the main effects (Table E.1 in Appendix E). The effects of both the factors and their interaction were found to be statistically significant ($p < 0.001$). A one-way repeated-measures ANOVA was conducted on recognition scores with processing as the main effect, separately at each of the SNR values (Table E.2 in Appendix E) and the effect of processing was found to be significant ($p < 0.001$) at SNR values less than 0 dB. Tukey's honestly significant difference (HSD) test was conducted for pair-wise comparison of the scores (Table E.3 in Appendix E). It showed that the improvements in scores using both CB18 and ACB filters with respect to the unprocessed diotic presentation were significant ($p < 0.01$) for SNR < 0 dB. At these SNR values, the scores with ACB were significantly higher ($p < 0.01$) than those with CB18.

Table 3.3 Exp. IA: Recognition scores (%) for normal-hearing subjects, with 9 SNR conditions, and 2 types of comb filter pairs. s.d.: standard deviation, Impr.: improvement (averaged across the subjects), *p*: one tailed significance level for paired t-test (processed vs. unprocessed, $n = 6$, and df = 5).

Sub.	SNR= ∞			SNR= 6 dB			SNR= 3 dB		
	Unp. Diotic	Dichotic CB18	Dichotic ACB	Unp. Diotic	Dichotic CB18	Dichotic ACB	Unp. Diotic	Dichotic CB18	Dichotic ACB
DSJ	97.3	96.3	96.7	94.0	91.3	94.3	91.3	90.3	93.7
MKD	98.7	98.7	94.7	94.7	94.7	91.7	93.0	86.3	87.7
PNK	98.3	99.0	97.7	95.0	94.7	94.7	92.7	91.0	94.3
RSH	95.0	94.0	94.3	88.7	87.0	89.3	82.7	87.3	88.3
SGK	94.3	97.7	97.3	90.7	93.3	94.3	91.0	89.0	93.3
SPK	98.7	98.7	99.3	93.7	94.3	95.0	91.0	92.3	94.7
Mean	97.1	97.4	96.7	92.8	92.6	93.2	90.3	89.4	92.0
s.d.	1.9	1.9	1.9	2.5	3.0	2.2	3.8	2.3	3.1
Impr.		0.3	-0.4		-0.2	0.4		-0.9	1.7
p		n.s.	–		–	n.s.		–	n.s.

Sub.	SNR= 0 dB			SNR= -3 dB			SNR= -6 dB		
	Unp. Diotic	Dichotic CB18	Dichotic ACB	Unp. Diotic	Dichotic CB18	Dichotic ACB	Unp. Diotic	Dichotic CB18	Dichotic ACB
DSJ	85.7	87.7	88.0	77.3	87.7	91.0	73.3	84.3	89.7
MKD	86.0	88.3	92.0	80.3	85.7	87.7	70.0	86.3	88.3
PNK	84.7	88.3	92.7	77.0	85.0	89.7	67.7	78.0	87.7
RSH	75.3	83.7	87.3	69.3	83.7	89.3	63.7	79.3	82.0
SGK	83.3	89.3	90.7	72.3	87.3	89.0	69.0	84.3	89.3
SPK	86.3	89.7	92.3	78.0	86.0	89.7	73.3	78.3	88.3
Mean	83.6	87.8	90.5	75.7	85.9	89.4	69.5	81.8	87.6
s.d.	4.2	2.2	2.3	4.1	1.5	1.1	3.7	3.6	2.8
Impr.		4.2	6.9		10.2	13.7		12.3	18.1
p		<0.005	<0.001		<0.001	<0.001		<0.001	<0.001

Sub.	SNR= -9 dB			SNR= -12 dB			SNR= -15 dB		
	Unp. Diotic	Dichotic CB18	Dichotic ACB	Unp. Diotic	Dichotic CB18	Dichotic ACB	Unp. Diotic	Dichotic CB18	Dichotic ACB
DSJ	65.0	77.3	86.3	65.3	74.3	82.0	52.7	68.0	76.0
MKD	61.3	78.7	87.0	58.0	74.7	83.0	48.7	64.7	76.7
PNK	63.3	72.0	84.0	55.3	65.3	76.7	48.0	60.3	66.3
RSH	57.0	69.0	80.3	53.3	71.0	77.7	38.7	62.3	72.3
SGK	60.0	78.0	82.7	43.7	74.3	78.0	36.3	61.7	74.3
SPK	62.0	73.0	82.3	53.7	67.0	79.7	47.7	62.0	74.3
Mean	61.4	74.7	83.8	54.9	71.1	79.5	45.3	63.2	73.3
s.d.	2.8	3.9	2.5	7.1	4.1	2.5	6.4	2.8	3.7
Impr.		13.3	22.4		16.2	24.6		17.9	28
p		<0.001	<0.001		<0.001	<0.001		<0.001	<0.001

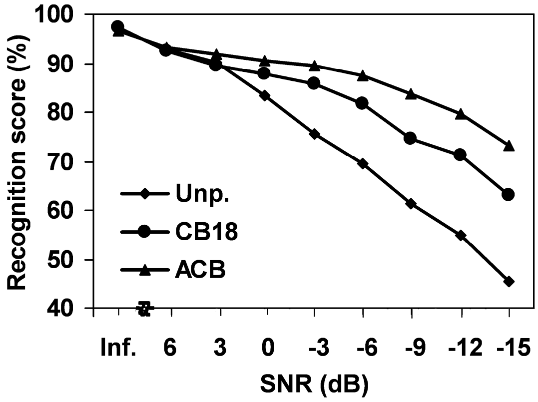


Fig. 3.5 Exp. IA: Recognition score (averaged across the six normal-hearing subjects) vs. SNR. Unp.: Unprocessed diotic presentation, CB18: dichotic presentation with CB18 filters, ACB: dichotic presentation with ACB filters.

Table 3.4 gives the response time for the unprocessed diotic presentation and for the dichotic presentation using CB18 and ACB filters. In addition to the response time for the individual subjects, the mean response time, the standard deviation (s.d.), the mean improvement, and one-tailed significance level (p) for paired (dichotic vs. diotic) t-test are also given in the table. The improvements in response time for CB18 were 0.14, 0.18, 0.18, and 0.04 s for SNR values of -3, -6, -9, and -12 dB, respectively. The corresponding improvements for ACB filters were 0.25, 0.33, 0.26, and 0.12 s, respectively. Figure 3.6 gives a plot of response time (averaged across the six subjects) as a function of SNR, for speech processed with the two filters. Although no significant decrease in response time was observed for higher SNR values, statistically significant ($p < 0.05$) decreases occurred for SNR values less than 0 dB for ACB filters. For CB18 filters, the decreases were significant ($p < 0.05$) for SNR values of -6 and -9 dB only.

A two-way repeated-measures ANOVA was conducted on response times with processing and SNR as the main effects (Table E.4 in Appendix E). Only the effect of SNR was found to be significant ($p < 0.001$). A one-way repeated-measures ANOVA was conducted on response times with processing as the main effect, separately at each of the SNR values (Table E.5 in Appendix E) and the effect of processing was found to be significant ($p < 0.05$) at SNR values of -6 and -9 dB. Tukey's HSD test conducted for pair-wise comparison (Table E.6 in Appendix E) showed that the improvement due to ACB filter was significant ($p < 0.05$) at SNR values of -6 and -9 dB.

Table 3.4 Exp. IA: Response time values (s) for normal-hearing subjects, with 9 SNR conditions, and two types of comb filter pairs. s.d.: standard deviation, Impr.: improvement (averaged across the subjects), *p*: one tailed significance level for paired t-test (processed vs. unprocessed, n = 6, and df = 5).

Sub.	SNR= ∞			SNR= 6 dB			SNR= 3 dB		
	Unp. Diotic	Dichotic		Unp. Diotic	Dichotic		Unp. Diotic	Dichotic	
		CB18	ACB		CB18	ACB		CB18	ACB
DSJ	1.52	1.43	1.70	2.29	1.89	2.07	2.33	1.76	2.12
MKD	2.65	3.31	2.97	2.87	3.12	3.17	3.26	3.47	3.26
PNK	3.36	3.33	3.47	3.60	3.43	3.63	3.68	3.71	3.63
RSH	2.66	3.01	2.78	2.93	2.88	2.92	3.17	3.17	3.02
SGK	2.93	1.90	2.09	2.05	2.42	1.90	2.54	1.96	1.88
SPK	2.69	3.10	3.20	3.24	3.32	3.29	3.54	3.59	3.49
Mean	2.64	2.68	2.70	2.83	2.84	2.83	3.09	2.94	2.90
s.d	0.61	0.81	0.68	0.58	0.59	0.69	0.54	0.86	0.73
Impr.		-0.04	-0.06		-0.01	0.00		0.15	0.19
p		–	–		–	n.s.		n.s.	n.s.

Sub.	SNR= 0 dB			SNR= -3 dB			SNR= -6 dB		
	Unp. Diotic	Dichotic		Unp. Diotic	Dichotic		Unp. Diotic	Dichotic	
		CB18	ACB		CB18	ACB		CB18	ACB
DSJ	2.56	2.08	2.50	2.73	1.98	2.49	2.47	2.01	2.38
MKD	3.38	3.66	3.43	3.43	3.51	3.49	3.68	3.50	3.38
PNK	4.16	4.08	3.85	3.90	3.90	3.71	3.97	4.09	3.62
RSH	3.38	3.37	3.17	3.49	3.46	3.31	3.45	3.38	3.13
SGK	2.45	2.17	2.12	2.68	2.70	2.19	2.76	2.39	2.38
SPK	3.57	3.78	3.51	3.88	3.74	3.39	3.95	3.86	3.39
Mean	3.25	3.19	3.10	3.35	3.21	3.10	3.38	3.20	3.05
s.d	0.64	0.86	0.66	0.54	0.73	0.61	0.63	0.83	0.54
Impr.		0.06	0.15		0.14	0.25		0.18	0.33
p		n.s.	<0.05		n.s.	<0.05		<0.05	<0.001

Sub.	SNR= -9 dB			SNR= -12 dB			SNR= -15 dB		
	Unp. Diotic	Dichotic		Unp. Diotic	Dichotic		Unp. Diotic	Dichotic	
		CB18	ACB		CB18	ACB		CB18	ACB
DSJ	2.49	2.14	2.40	2.19	2.37	2.52	2.29	2.31	2.50
MKD	3.65	3.38	3.31	3.42	3.46	3.37	3.37	3.48	3.36
PNK	4.26	4.04	3.72	4.19	4.26	4.08	4.29	4.45	4.41
RSH	3.23	3.19	3.28	3.10	3.46	3.37	3.07	3.56	3.49
SGK	2.73	2.92	2.61	3.72	2.75	2.76	3.67	3.24	2.79
SPK	4.03	3.62	3.52	3.99	4.11	3.82	4.00	3.95	3.92
Mean	3.40	3.22	3.14	3.44	3.40	3.32	3.45	3.50	3.41
s.d	0.71	0.65	0.52	0.72	0.74	0.60	0.72	0.72	0.71
Impr.		0.18	0.26		0.04	0.12		-0.050	0.04
p		<0.05	<0.05		<n.s.	<n.s.			<n.s.

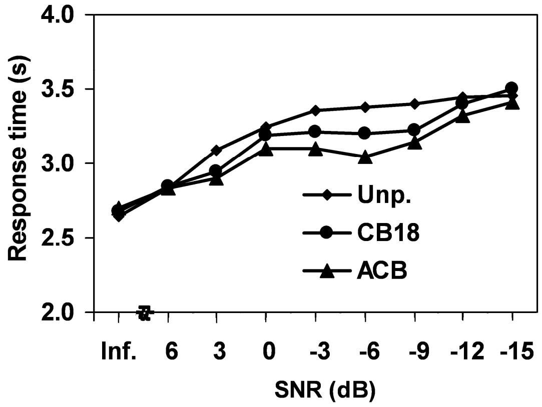


Fig. 3.6 Exp. IA: Response time (averaged across the six normal-hearing subjects) vs. SNR. Unp.: Unprocessed diotic presentation, CB18: dichotic presentation with CB18 filters, ACB: dichotic presentation with ACB filters.

Experiment IB: Tests on hearing-impaired subjects

The response data for the 11 subjects with bilateral sensorineural loss were analyzed to get the percentage recognition score and response time for the unprocessed diotic presentation and the dichotic presentation using the ACB based comb filter pair.

The recognition score and response time results are summarized in Table 3.5. In addition to the results for the individual subjects, it also gives the mean, s.d., the mean improvement, and one-tailed significance level (*p*) for paired t-test. The same results are shown as plots in Fig. 3.7 and Fig. 3.8. All the subjects had improvement in the recognition scores due to dichotic presentation and the improvements ranged 14 – 31 %, with a mean of 22 % ($p < 0.001$). Dichotic presentation resulted in a decrease in the response time for 10 of the 11 subjects, with a mean improvement of 0.26 s ($p < 0.001$).

3.6 Discussion

The effectiveness of spectral splitting scheme based on CB18 and ACB filters was assessed by listening tests using MRT for recognition of consonants. Listening tests were conducted on six normal-hearing subjects with sensorineural loss simulated by addition of broad-band masking noise (Exp. IA). Even though no improvement in recognition scores were observed for SNR values higher than 0 dB, improvement in percentage recognition scores were observed for lower values for both the types of filters. The improvement in recognition scores for CB18 were 10, 12, 13, 16, and 18 % for SNR values of -3, -6, -9, -12, and -15 dB, respectively. For ACB filters, the corresponding improvements in recognition scores were 14, 18, 22, 25, and 28 %, and these improvements were statistically significant ($p < 0.001$). At 75

Table 3.5 Exp. IB: Recognition scores (%) and response times (s) for the unprocessed speech and the speech processed with the ACB filters, for hearing impaired. Impr.: Improvement (%) in recognition score.

Subject.	Recog. score (%)		Response time (s)	
	Unp.	ACB	Unp.	ACB
IGH	57.3	84.0	4.10	3.68
MNA	63.3	77.3	3.57	3.66
NBC	62.3	79.3	3.99	3.95
PAT	60.7	87.3	3.68	3.63
PND	50.3	71.0	3.80	3.55
PRP	66.3	87.3	3.77	3.23
PSE	67.3	91.7	3.47	3.34
RAJ	59.7	89.0	3.75	3.18
SIY	56.0	87.0	3.86	3.52
SKS	62.3	83.3	3.74	3.56
SPB	63.0	77.7	4.09	3.62
Mean	60.8	83.2	3.80	3.54
s.d.	4.8	6.2	0.20	0.22
Impr.		22.4		0.26
p		<0.001		<0.001

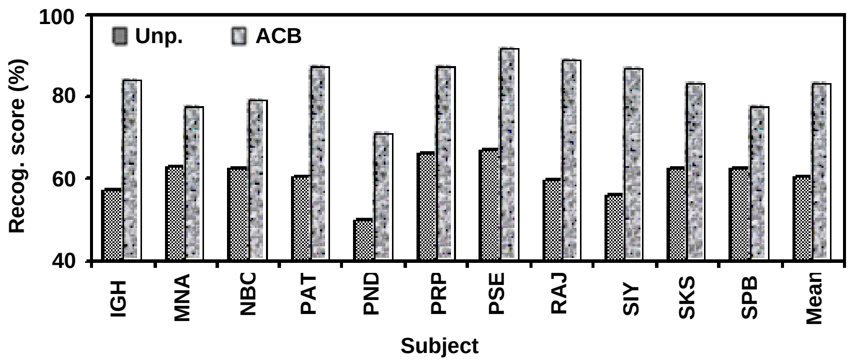


Fig. 3.7 Exp. IB: Recognition score (%) for 11 hearing-impaired subjects along with the mean recognition score, for unprocessed diotic presentation (Unp.) and the dichotic presentation with the ACB filters (ACB).

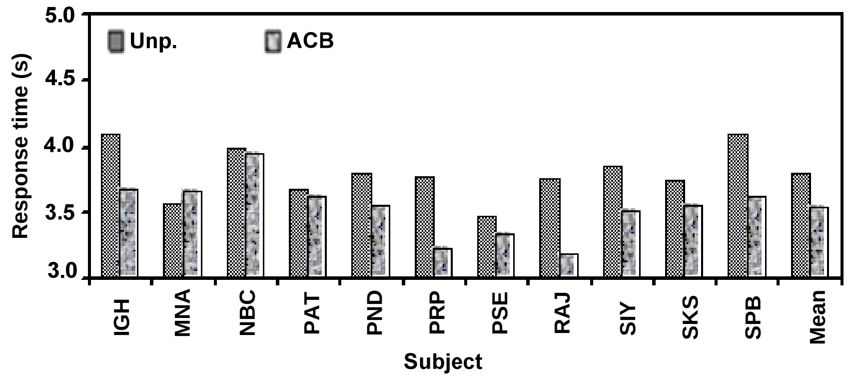


Fig. 3.8 Exp. IB: Response time (s) for 11 hearing-impaired subjects along with the mean response time, for unprocessed diotic presentation (Unp.) and the dichotic presentation with the ACB filters (ACB).

% recognition score, the improvements in recognition scores observed for CB18 and ACB filters were equivalent to an SNR advantage of approximately 6 and 12 dB, respectively. The improvements in response time for CB18 were 0.14, 0.18, 0.18, and 0.04 s for SNR values of -3, -6, -9, and -12 dB, respectively. The corresponding improvements for ACB filters were 0.25, 0.33, 0.26, and 0.12 s, respectively. Thus, the scheme of comb filter based spectral splitting, using both CB18 and ACB filters, helped in improving speech perception. The improvements with ACB filters were higher than those with CB18 filters and the differences were statistically significant at lower SNR values.

Based on the results of MRT on normal-hearing subjects, further evaluation of dichotic presentation using ACB was carried out by conducting MRT on 11 subjects with moderate bilateral sensorineural loss (Exp. IB). The subjects did not wear their hearing aids and no frequency dependent gain or amplitude compression was used. All the subjects showed improvement in recognition scores. The improvement ranged 14 – 31 % with a mean of 22 % ($p < 0.001$). The processing also resulted in a decrease in the response time with a mean of 0.26 s ($p < 0.001$), indicating a reduction on the perceptual load.

Thus the investigations showed that the scheme of spectral splitting, using perceptually balanced comb filters based on auditory critical bandwidths, helped in improving speech perception for subjects with moderate bilateral sensorineural loss. The pattern of improvement in recognition score and response time across the subjects with sensorineural loss did not show any specific relation to the audiograms or the recognition scores for diotic presentation. This may be because although all the subjects had moderate bilateral loss, the extent of masking for individual subjects may be different. Tests conducted on a larger number of subjects to evaluate the improvements due to processing and tests to measure the extent of masking may help in identifying the group most likely to benefit by the processing.

Chapter 4

EFFECT OF SPECTRAL SPLITTING ON SOURCE LOCALIZATION

4.1 Introduction

A concern in the use of binaural hearing aids with dichotic presentation is the possibility of its adverse effect on sound source localization. In spectral splitting, each band gets presented monaurally and hence source localization ability may get impaired for narrow band signals. For speech and broadband environmental sounds, listeners may be able to use the cues across the bands for source localization. This chapter presents investigations carried out to study the effect of spectral splitting on source localization.

The next section provides a review of some of the earlier research on sound source localization. The experimental method of the investigations to study the effect of spectral splitting on source localization is described in Section 4.3. The results are presented in the subsequent sections, followed by a discussion in the last section.

4.2 Sound source localization

Localization of a sound source involves perceptual integration of multiple acoustic cues: inter-aural time difference (ITD), inter-aural level difference (ILD), and spectral cues (Stevens and Newman, 1936; Moore, 1997; Hartmann, 1999; Chung et al., 2000; Langendijk and Bronkhorst, 2002; Best et al., 2005). ITD and ILD are the most important cues for source localization in the horizontal plane. According to the "duplex" theory of binaural localization, ITD is important for localization at low frequencies, while ILD is important at higher frequencies (Simon, 2005). Localization in vertical plane and discrimination of back and front primarily depends upon high frequency (> 5 kHz) spectral cues, caused by diffraction of sound by the pinna. ITD varies with the path length difference between the two ears, from a minimum value of 0 for a sound coming from straight ahead to a value of about 690 μs for a sound coming from a source located directly opposite to one ear. ILD varies over 0 – 20 dB (Middlebrooks, 1992; Moore, 1997; Best et al., 2005; and Van den Bogaert et al., 2006).

Several studies on the localization performance of normal-hearing and hearing-impaired subjects have been reported using different test material and processing conditions (Makous and Middlebrooks, 1990; Wightman and Kistler, 1992a, 1992b; Noble et al., 1994; Hofman and Van Opstal, 1998; Lorenzi et. al., 1999a and 1999b; Abel et al., 2000). Signals from two independently operating hearing aids, each with its own processing and introducing its own time delay, may distort the binaural cues needed for source localization (Dillon et al., 2003). Van den Bogaert et al. (2006) investigated horizontal source localization with bilateral hearing aids, using an array of 13 speakers placed at a distance of 1 m from the listener at angles varying from -90° to +90° with a spacing of 15°. The subject identified the speaker which was perceived as the sound source. For normal-hearing subjects, the mean rms error in recognition of the source direction was 6.8º, 13.5º, and 21.3º for telephone ring, 500 Hz tone, and 5000 Hz tone, respectively. The corresponding values for hearing-impaired subjects were 13º, 17º, and 22.4º when tested without hearing aids. Use of hearing aids resulted in an increase of 3º– 8º in the rms error.

In localization experiments involving presentation from an array of speakers, the subject needs to maintain a steady head position throughout the test. This problem can be overcome by simulating directionality using a pair of head related transfer functions (HRTFs), relating the acoustical signal from a source in the free field to the eardrum of the two ears of a listener (Begault, 1991; Weightman and Kistler., 1992b; Brungart and Rabinowitz, 1999; Algazi et al., 2002; Langedijk and Bronkhorst, 2002; Otani et al., 2009; Zhong and Xie, 2009). Murase et al. (2004) studied the effect of dichotic listening on source localization, by dividing the speech spectrum into two bands (based on the formant frequencies of Japanese vowels i.e. with frequency boundary of 0.8 kHz and 1.6 kHz), and presenting them dichotically to six normal-hearing and three hearing-impaired subjects. Directionality was simulated by processing the speech signals with HRTFs of a dummy head and torso obtained at five directions (-90º, - 45º, 0º, 45º, and 90º). Listening tests were conducted using three speech sentences as the stimuli presented through a pair of headphones. With diotic presentation, normal-hearing subjects were able to localize the sounds accurately, but a large spread of error in localization occurred for the hearing-impaired subjects. For normal-hearing subjects, the localization with dichotic presentation was poor as compared to that with the diotic presentation. For the hearing-impaired subjects, dichotic presentation resulted in lateralization of the sound to the side of the low frequency band, severely affecting the source localization.

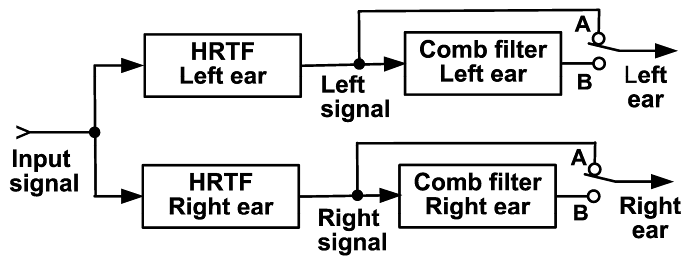


Fig. 4.1 Signal processing for studying source localization: switches in position A for diotic presentation and in position B for dichotic presentation

4.3 Experimental method

The objective of the investigation presented in this chapter was to study the effect of dichotic presentation using the comb filters based on auditory critical bandwidth, on source localization. A pair of HRTFs was used to generate spatial sounds for the investigation. CIPIC HRTF database is a public domain database, and it provides HRTFs for different combination of azimuth and elevation, along with the description of the detailed method of HRTF measurement and anthropometric parameters (Algazi et al., 2001; CIPIC HRTF database, 2001). The HRTFs for one of the subjects (subject 21, KEMAR manikin) from this database, for 0º elevation and frontal azimuth angle varying from -90º (extreme left) to +90º (extreme right) were used in the present study.

The signal processing scheme for studying source localization for diotic and dichotic presentations is shown in Fig. 4.1. For binaural diotic presentation (A), the two HRTF outputs were presented to the two ears. For dichotic presentation (B), the HRTF outputs were processed through the pair of ACB comb filters, with magnitude responses as shown earlier in Fig. 3.3(b). In the listening tests for intelligibility, in Chapter 3, binaural diotic presentation involved presentation of the same signal to both the ears. Here binaural diotic presentation involves two different signals to the two ears, but the signals are different only in terms of source direction cues.

Listening tests for studying the source localization were conducted on six normal-hearing subjects in the presence of broad-band masking noise, and on 11 subjects with moderate sensorineural loss. The two groups of subjects were the same as those who participated in the speech intelligibility tests presented in Chapter 3. For normal-hearing subjects, broad-band random noise was added as a masker to the processed stimuli with a constant SNR on short-time (10 ms) basis, at eight SNR values: ∞, 6, 3, 0, -3, -6, -9, and -12 dB. Hearing-impaired subjects were tested without adding broad-band masking noise.

Three tests were conducted on normal hearing-subjects: (i) left/center/right identification (Exp. IIA), (ii) left/center/right discrimination threshold (Exp. IIB), and (iii) source-direction identification (Exp. IIC). The test for source-direction identification was also conducted on hearing-impaired subjects (Exp. IID). In all the tests, the sound was presented binaurally at the most comfortable level as selected by the individual subjects.

4.3.1 Experiment IIA: Left/center/right identification

The objective of this experiment (Exp. IIA) was to compare left/center/right identification under the conditions of diotic and dichotic presentations. The stimuli included (i) band-pass filtered noise, center frequency of 1500 Hz and three bandwidths: one-sixth octave, one-third octave, and octave, (ii) broad-band noise, and (iii) sound of breaking glass as a broad-band environmental sound. The stimuli were processed through HRTFs for 0º elevation and azimuth angle varying from -90º (extreme left) to +90º (extreme right) in steps of 10º. For the sound of breaking glass, the processed stimuli were presented with different levels of broad-band noise added as a masker.

For each presentation, the subjects identified the perceived direction as (i) left, (ii) center, or (iii) right. Presentation was randomized with each angle repeated ten times. At the end of the test, the three percentage responses (i.e. left, center and right) were plotted against the azimuth angle of the source, for comparing the sharpness in the left-center and right-center transitions under the two types of presentation (A and B) at various SNR values. The total number of presentations for each subject was 1900 (19 angles × 10 repetitions × 2 processing conditions × 5 stimuli) in the absence of masking noise and 3040 (19 angles × 10 repetitions × 2 processing conditions × 8 SNR values) in the presence of masking noise.

4.3.2 Experiment IIB: Left/center/right discrimination threshold

The objective was to determine the smallest value of azimuth angle for the source to be localized on the left or right side of the center. Listening tests were conducted on six normal-hearing subjects with and without adding broad-band masking noise. The set of test materials was the same as in Exp. IIA. For the tests in the presence of masking noise, only the sound of breaking glass was used as the test stimuli. The azimuth angles were in ±90º range, with a step of 10º. Presentation was started from one of the extreme positions i.e. either -90° or 90°. The subject identified the presented sound as left, center, or right. For the next presentation, the angle was decreased by 10º in case of correct response, and increased by 20º for incorrect response. This procedure was continued until more than 50 % incorrect responses were observed for the same presentation angle. This angle was taken as the threshold of discrimination. Thresholds for left-center and right-center discrimination were tabulated for the two processing conditions and different SNR values.

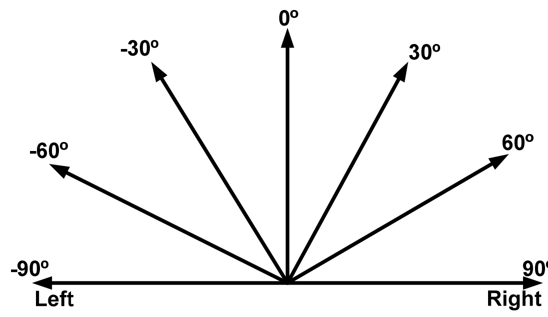


Fig. 4.2 Seven choice angles used in Exp. IIC and Exp. IID

4.3.3 Experiments IIC Exp. IID: Source direction identification

The objective of these experiments was to compare direction identification scores under the two processing conditions. The first set of listening tests (Exp. IIC) was conducted on six normal-hearing subjects in the presence of broad-band masking noise. The noise was added at eight SNR values of ∞ (no noise), 6, 3, 0, -3, -6, -9, and -12 dB. The second set of tests (Exp. IID) was conducted on 11 subjects with moderate bilateral sensorineural loss (the same set of subjects as in Chapter 3, with hearing thresholds as in Table 3.3) without adding masking noise. In both the tests, the sounds were processed with HRTFs corresponding to 0º elevation and azimuth angles of 0º, ±30º, ±60º, and ±90º. These directions were displayed in a chart, as shown in Fig. 4.2, kept in front of the subject. Stimuli processed for the different angles and the two types of processing were presented in a random order, with each angle repeated five times. The subject identified the perceived direction of the source as one of these seven angles. As in some of the earlier studies on localization (Giguere and Abel, 1993; Lorenzi et al., 1999a; Van Hoesel et al., 2002; Van den Bogaert et al., 2006), the responses were tabulated as entries in a stimulus-response matrix, and mean rms error (º) was calculated for each presentation angle, as a quantitative measure of localization error.

In Exp. IIC involving normal-hearing subjects, the test stimuli consisted of (i) broad-band noise, and (ii) sound of breaking glass as a broad-band environmental sound. For the tests in the presence of masking noise, only the sound of breaking glass was used as the test stimulus. Thus the total number of presentations per subject was 140 in the absence of noise (7 angles × 2 processing conditions × 5 repetitions × 2 stimuli), and 560 in the presence of masking noise (7 angles × 2 processing conditions × 5 repetitions × 8 SNR values). In Exp. IID involving the hearing-impaired subjects, the test stimuli included sound of breaking glass and broad-band noise. Each subject responded for a total of 140 presentations (7 angles × 2 processing conditions × 5 repetitions × 2 test stimuli).

4.4 Results of left/center/right identification (Exp. IIA)

The scores for left, center, and right identification (averaged across the six subjects) are plotted as a function of the azimuth angle in Fig. 4.3, for the noise band-pass filtered with different bandwidths. With diotic presentation, a transition in the response near 0º is observed indicating that the HRTF's successfully generated spatial sounds. The left-to-center and right-to-center transitions for the one-third octave, the octave, and the broad-band noise are sharper than those for the one-sixth octave noise. Dichotic presentation resulted in a smearing of responses, with a moderate decrease of about 20 % in the identification score at 10º.

Fig. 4.4 shows the plots of averaged identification scores for left, right, and center identification with the two types of presentation for the sound of breaking glass at various SNR values. The transition responses for these plots are almost similar to the responses obtained in the absence of noise, indicating that the presence of masking noise has a minimal effect on the response curves. At lower SNR values, as the left/center/right identification gets degraded, the difference in the diotic and dichotic scores become insignificant.

The crossover angles in the plots were taken as the left/center and right/center discrimination threshold and the mean of these two angles was taken as the discrimination threshold Φ. The threshold values for the diotic and dichotic presentations are given in Table 4.1, as Φ_A and Φ_B, respectively. Dichotic presentation resulted in an increase in the discrimination threshold (Φ_B - Φ_A) by 20, 9, 4, 4, and 3 degrees for 1/6-octave noise, 1/3-octave noise, 1-octave noise, broad-band noise, and sound of breaking glass, respectively. Only the increase for 1/6-octave noise was statistically significant ($p < 0.05$).

4.5 Results of left/center/right discrimination threshold (Exp. IIB)

The mean of the thresholds for left/center (L/C) and right/center (R/C) discrimination was taken as the discrimination threshold. The threshold values for the diotic (A) and dichotic (B) presentations are denoted as Φ_A and Φ_B, respectively. These threshold values, averaged across the subjects, are given in Table 4.2. The last column in the table gives the difference in the thresholds, i.e. increase in the threshold due to dichotic presentation along with standard deviation and significance level for one-tailed paired t-test.

The mean increase in the discrimination thresholds due to dichotic presentation ($\Delta\Phi = \Phi_B - \Phi_A$) was about 11, 7, 3, 3, and 5 degrees for 1/6-octave noise, 1/3-octave noise, 1-octave noise, broad-band noise, and sound of breaking glass, respectively. The values of discrimination thresholds and the changes in the thresholds due to dichotic presentation are comparable to those obtained in Exp. IIA. The maximum increase in thresholds due to dichotic presentation is observed for one-sixth-octave band noise. Only a relatively moderate

Table 4.1 Exp. IIA: Averaged (across the six subjects) discrimination thresholds (mean of L/C and R/C crossover angles) in degrees under the two processing conditions, for different test material. Φ_A: threshold (deg.) for diotic presentation, Φ_B: threshold (deg.) for dichotic presentation.

Test material	SNR (dB)	Φ_A		Φ_B		$\Phi_B - \Phi_A$		
		Mean	s.d.	Mean	s.d.	Mean	s.d.	*p*
1/6-oct. noise	∞	13.0	5.8	33.0	4.9	20.0	16.7	<0.050
1/3-oct noise	∞	8.0	3.8	17.0	4.1	9.0	3.8	n.s.
1-oct. noise	∞	6.0	4.2	10.0	8.2	4.0	7.7	n.s.
BB noise	∞	7.0	6.3	11.0	2.7	4.0	2.0	n.s.
Breaking glass	∞	5.8	2.0	9.2	2.0	3.3	2.6	n.s.
	6	5.8	2.0	15.0	6.3	9.2	5.8	<0.010
	3	10.0	4.5	15.0	4.5	5.0	3.2	<0.010
	0	5.8	2.0	15.8	3.8	10.0	3.2	<0.001
	-3	10.0	4.5	13.3	6.8	3.3	4.1	n.s.
	-6	6.7	2.6	12.5	6.1	3.3	4.1	<0.050
	-9	11.7	6.1	15.0	3.2	3.3	4.1	n.s.
	-12	13.3	4.1	15.0	6.3	1.7	2.6	n.s.

Table 4.2 Exp. IIB: Averaged (across the six subjects) discrimination thresholds (mean of L/C and R/C thresholds) in degrees under the two processing conditions, for different test material. Φ_A: threshold (deg.) for diotic presentation, Φ_B: threshold (deg.) for dichotic presentation.

Test material	SNR (dB)	Φ_A		Φ_B		$\Phi_B - \Phi_A$		
		Mean	s.d.	Mean	s.d.	Mean	s.d.	*p*
1/6-oct. noise	∞	17.0	6.1	26.0	9.7	10.8	8.6	<0.050
1/3-oct noise	∞	10.8	2.0	18.0	6.1	7.0	5.1	<0.010
1-oct. noise	∞	10.8	2.0	14.0	3.8	3.0	4.0	n.s.
BB noise	∞	11.0	2.0	14.0	3.8	3.0	4.0	n.s.
Breaking glass	∞	11.0	2.0	16.0	5.8	5.0	6.3	n.s.
	6	11.0	2.0	22.0	6.8	11.0	6.6	<0.001
	3	14.0	3.8	21.0	5.8	7.0	5.2	<0.010
	0	11.0	2.0	18.0	2.6	7.0	2.7	<0.000
	-3	13.0	2.7	21.0	8.0	8.0	8.2	<0.050
	-6	13.0	4.1	19.0	7.4	6.0	5.8	<0.050
	-9	14.0	4.9	19.0	8.0	5.0	6.8	n.s.
	-12	16.0	3.7	21.0	7.3	5.0	5.4	<0.050

degradation in the left/right discrimination ability is observed for the signals with bandwidth more than one-third octave and the differences are not statistically significant.

To study the localization performance in the presence of broad-band masking noise, further evaluation was carried out to study the change in the thresholds of discrimination at different SNR values using sound of breaking glass as the test material. The mean of these threshold values, averaged across the subjects, are also given in Table 4.2. Presence of masking noise only moderately affected the left/right discrimination performance. The difference in averaged thresholds for the diotic and dichotic presentations was less than 10º.

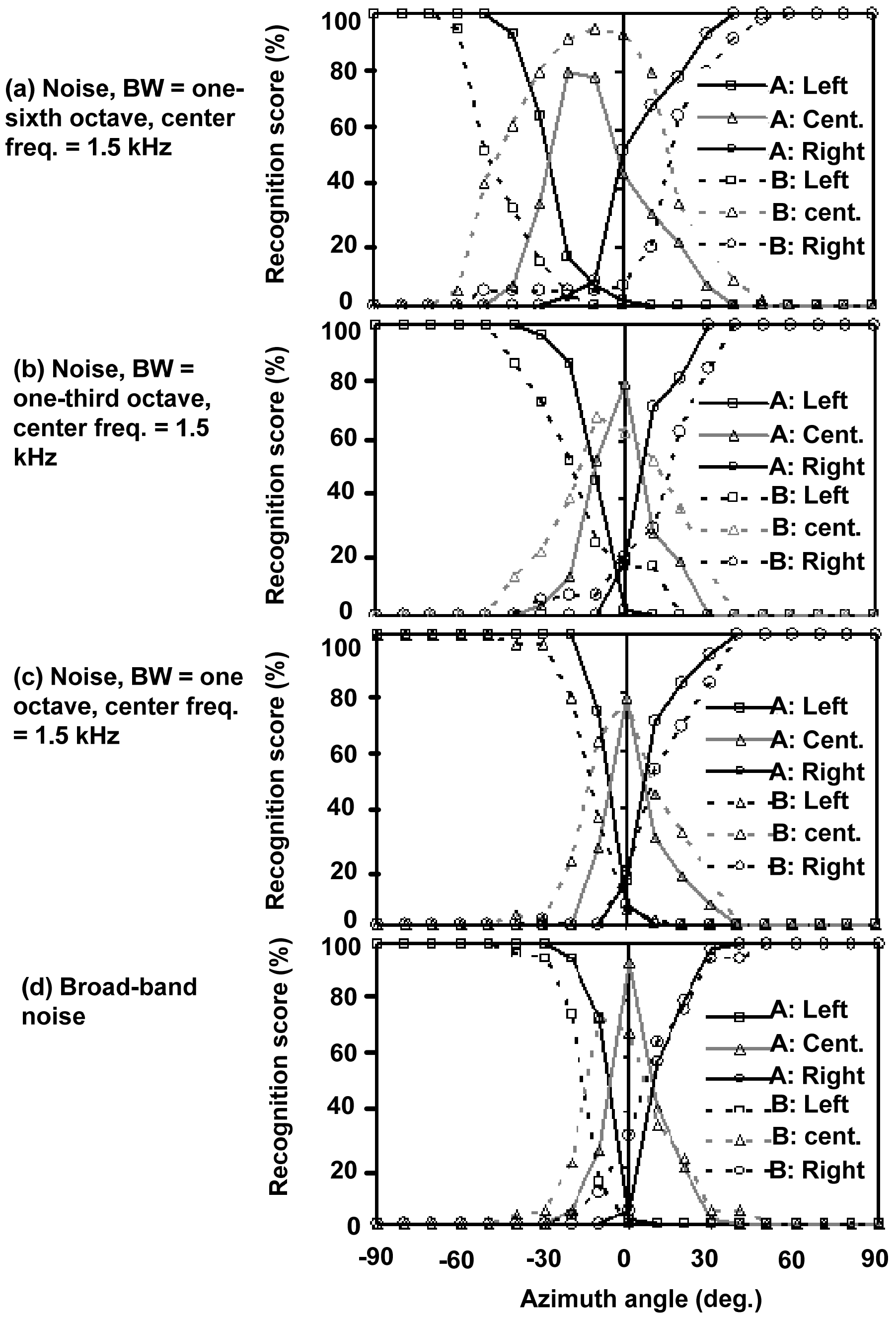


Fig. 4.3 Exp. IIA: Percentage recognition score (averaged across the six subjects) under the two processing conditions (A: diotic and B: dichotic) for band-pass filtered noise (center frequency = 1500 Hz): (a) BW = one-sixth octave, (b) BW = one-third octave, and (c) BW = one octave, and (d) Broad-band noise.

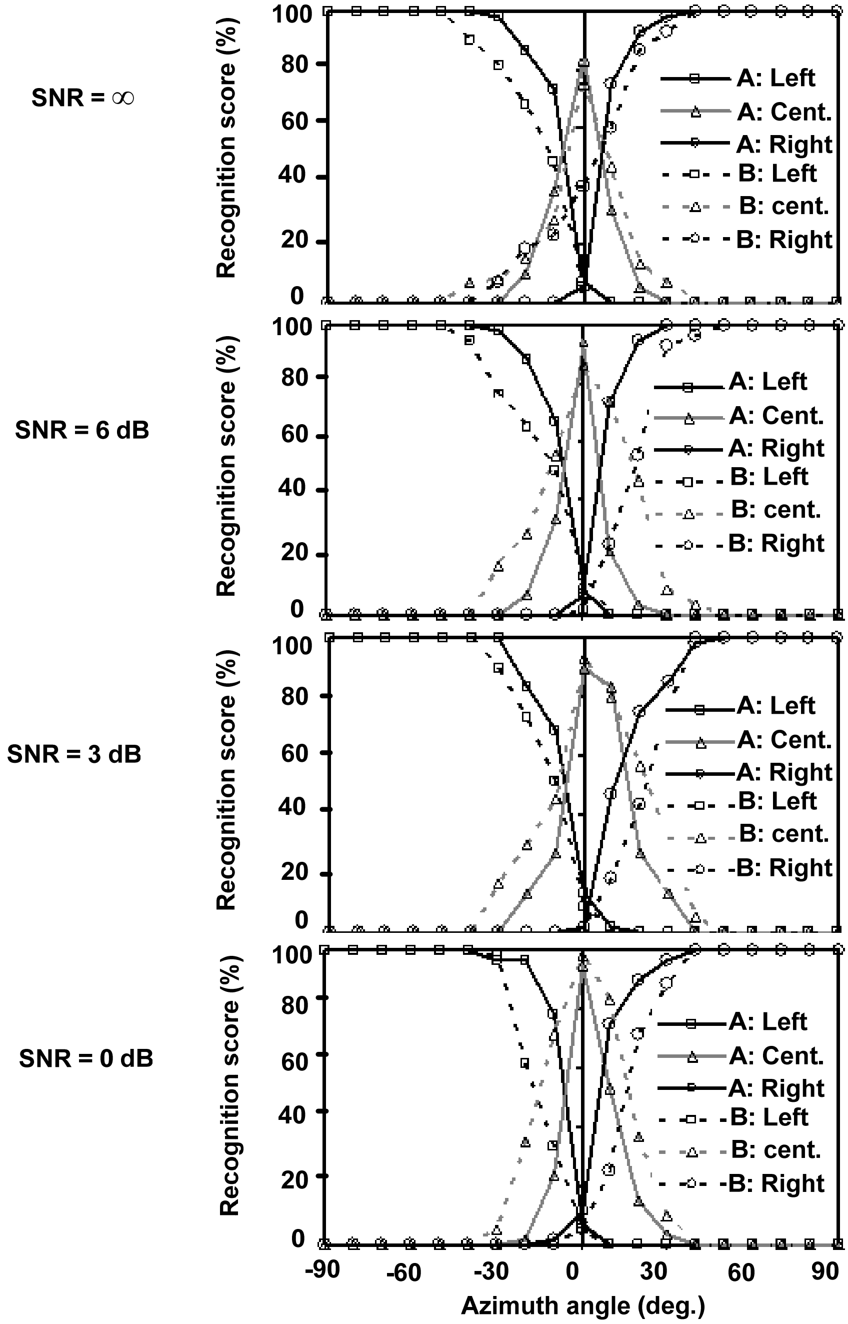


Fig. 4.4 A Exp. IIA: Percentage recognition score (averaged across the six subjects) under the two processing conditions (A: diotic and B: dichotic) for sound of breaking glass with masking noise of (a) ∞, (b) 6 dB, (c) 3 dB SNR, and (d) 0 dB values.

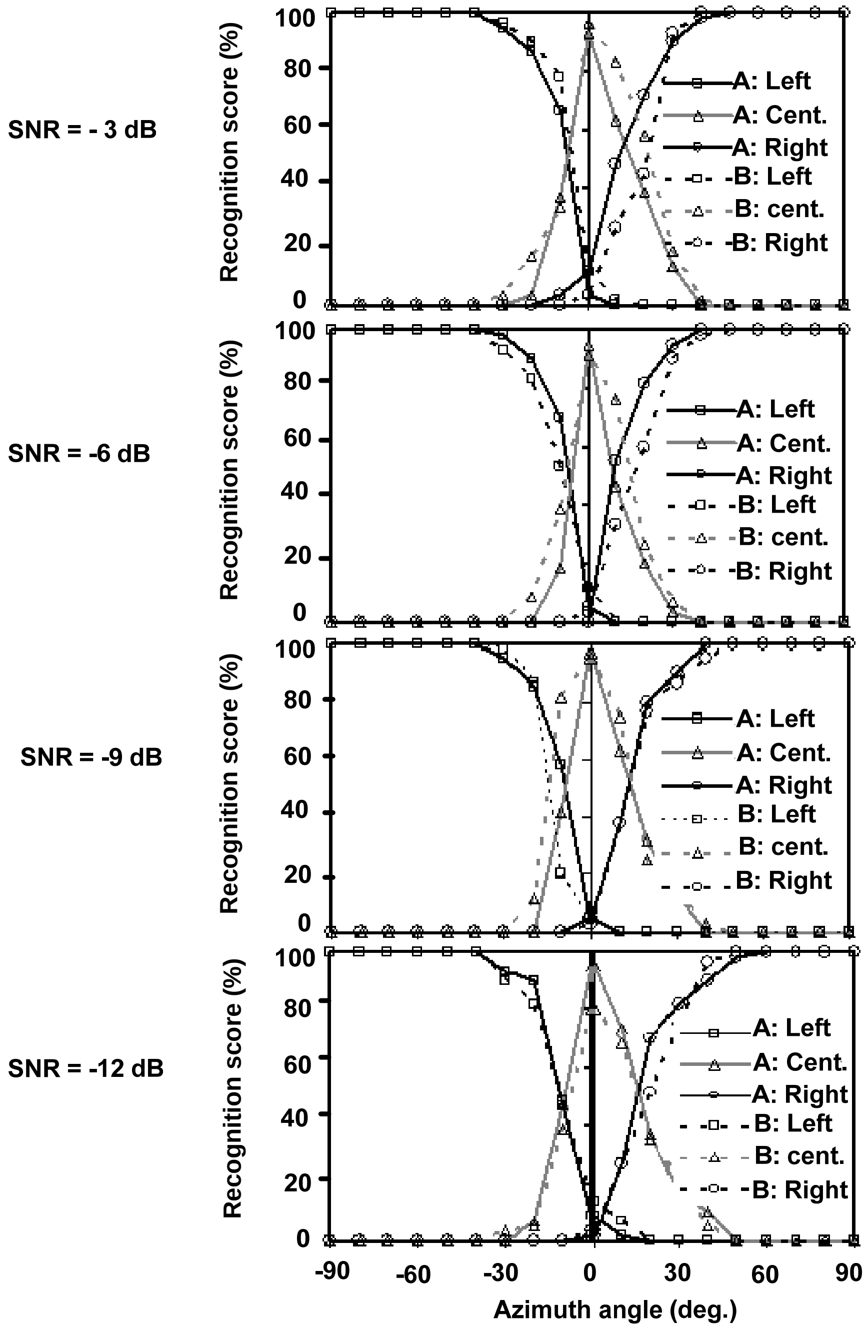


Fig. 4.4 B Exp. IIA: Percentage recognition score (averaged across the six subjects) under the two processing conditions (A: diotic and B: dichotic) for sound of breaking glass with masking noise of (a) -3 dB, (b) -6 dB, (c) -9 dB, and (d) -12 dB SNR values.

4.6 Results of source direction identification by normal-hearing subjects (Exp. IIC)

The stimulus-response matrix, with responses by all the subjects merged together is shown in Table 4.3 for sound of breaking glass. For diotic presentation, the responses were mainly along the diagonal, showing that HRTFs were able to successfully generate spatial perception. With dichotic presentation, there was a larger spread in responses. Similar patterns were observed in the stimulus-response matrices for the broad-band noise. The mean of the direction identification scores for the diotic presentation were 73 % for the sound of breaking glass and 62 % for broad-band noise. Under dichotic presentation, the corresponding scores were 63 % and 54 %.

Averaged (across the six subjects) rms errors for each presentation angle and SNR value, under diotic and dichotic presentation are given in Table 4.4. The errors for different angles were similar. Increase in the rms error due to dichotic presentation ranged 1º – 8º.

4.7 Results of source direction identification by the hearing-impaired subjects (Exp. IID)

Stimulus-response matrix, with responses by all the eleven subjects merged together is shown in Table 4.5 for the sound of breaking glass. For diotic presentation, the responses were mainly along the diagonal, while the dichotic presentation resulted in a larger spread in responses. The stimulus-response matrix for broad-band random noise as the stimuli showed a similar pattern. The direction identification scores for all the subjects under the two presentation conditions are given in Table 4.6. With diotic presentation, the mean identification scores were 62.4 % for the sound of breaking glass and broad-band noise. The scores were similar to those obtained by normal-hearing subjects under -9 dB SNR. The corresponding scores for dichotic presentation were 57.8 and 60.3 %, respectively. Thus, the dichotic presentation resulted in a moderate decrease in the scores: 4.6 % for the sound of breaking glass and 2.1 % for broad-band noise. The rms errors in the angle identification are given in Table 4.7 for the two types of presentation. The errors for different angles were approximately the same. The mean (across the angles) rms error for diotic presentation of breaking glass and broad-band noise were 18.4º and 19.3º, respectively. For dichotic presentation, the mean rms error increased to 19.4º for both the stimuli. Thus there was no significant difference in the rms error under the two types of presentation for both the test stimuli.

Table 4.3 Exp. IIC: Identification scores for presentation angle vs. perceived angle. No. of total presentations for each angle = 30 (5 presentations×6 subjects).Test material: sound of breaking glass with no masking noise.

Pres. angle (deg.)	A: Diotic presentation							B: Dichotic presentation						
	Perceived angle (deg.)							Perceived angle (deg.)						
	-90	-60	-30	0	30	60	90	-90	-60	-30	0	30	60	90
-90	**20**	9	1					**18**	8	4				
-60	4	**20**	6					15	**10**	3	2			
-30		12	**18**					4	13	**12**	1			
0			6	**23**	1				2	4	**23**	1		
30				5	**16**	9					6	**13**	10	1
60					7	**19**	4				2	7	**10**	11
90					4	14	**12**				4	2	9	**15**

Table 4.4 Exp. IIC: Averaged (across the six subjects) rms error (º) for different SNR values under two processing conditions (A and B) for the sound of breaking glass. A: diotic presentation, B: dichotic presentation.

Angle (deg)	SNR (dB)															
	∞		6		3		0		-3		-6		-9		-12	
	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B
-90	15	17	14	21	16	17	17	21	14	21	15	23	23	29	27	23
-60	16	23	20	20	18	20	19	21	19	20	19	21	19	21	23	24
-30	17	26	21	20	20	21	19	20	18	22	23	23	20	23	23	20
0	11	18	13	12	12	12	14	16	15	16	15	15	13	18	20	15
30	19	20	20	26	20	23	18	20	23	23	20	22	20	20	22	25
60	17	25	17	23	18	18	22	22	21	18	18	19	20	21	19	22
90	12	15	16	20	20	20	20	16	17	19	20	20	20	19	19	22
-90	15	21	17	20	18	19	18	19	18	20	19	20	19	22	22	22
Mean	15	17	14	21	16	17	17	21	14	21	15	23	23	29	27	23
s.d.	2.9	4.2	3.1	4.3	2.9	3.5	2.5	2.4	3.2	2.4	2.9	2.8	3.0	3.6	2.9	3.3

Table 4.5 Exp. IID: Identification scores for presentation angle vs. perceived angle. No. of total presentation for each angle = 55 (5 presentations ×11 subjects). Test material: sound of breaking glass.

Pres. angle (deg.)	A: Diotic presentation							B: Dichotic presentation						
	Perceived angle (deg.)							Perceived angle (deg.)						
	-90	-60	-30	0	30	60	90	-90	-60	-30	0	30	60	90
-90	**37**	18						**32**	23					
-60	5	**31**	19					8	**28**	19				
-30	1	11	**30**	13					15	**31**	9			
0			9	**41**	5					3	**40**	12		
30				12	**30**	13					3	**28**	23	1
60					16	**29**	10					9	**30**	16
90						14	**41**					1	20	**34**

Table 4.6 Exp. IID: Angle identification score (%) for 11 hearing-impaired subjects under diotic (A) and dichotic (B) presentations.

Subject	Test material			
	Breaking glass		BB noise	
	A	B	A	B
IGH	57	54	63	60
MNA	60	57	57	51
NBC	57	57	60	57
PRP	63	54	60	63
PSE	69	60	69	57
PAT	63	60	60	63
PND	71	60	77	69
RAJ	60	63	54	60
SPB	54	51	63	60
SIY	63	60	57	66
SKS	69	60	66	57
Mean	62.4	57.8	62.4	60.3
s.d.	5.5	3.6	6.4	4.9
p		<0.005		n.s.

Table 4.7 Exp. IID: Average (across the 11 subjects) rms error (°) under diotic (A) and dichotic (B) conditions.

Pres. Angle (deg.)	Test material			
	Breaking glass		BB noise	
	A	B	A	B
-90	17	19	20	16
-60	20	21	22	20
-30	21	20	20	22
0	15	16	16	16
30	20	22	16	22
60	21	20	20	19
90	15	18	18	22
Mean	18.4	19.4	19.3	19.4
s.d.	2.7	2.0	1.7	2.1
p		n.s.		n.s.

4.8 Discussion

Listening tests were conducted to study the effect of dichotic presentation on source localization. The tests were conducted on six normal-hearing subjects and 11 subjects with moderate sensorineural loss. Normal-hearing subjects were tested in the presence of broad-band masking noise, at different values of SNR while the hearing-impaired subjects were tested without masking noise. Head related transfer functions (HRTFs) were used to generate spatial sounds in the frontal azimuth plane. Overall evaluation involved three experiments: (i) left/center/right identification, (ii) left/center/right discrimination threshold, and (iii) source direction identification. All the three experiments were conducted on six normal-hearing subjects, while only the third experiment was conducted on 11 hearing-impaired subjects.

In the tests for left/center/right identification (Exp. IIA), diotic presentation resulted in a sharp transition in the response with azimuth angle, indicating that the HRTF's successfully generated spatial sounds. Dichotic presentation reduced the sharpness in transition responses and it resulted in a moderate decrease in the identification scores. The decrease was significant only for one-sixth octave noise and not for one-third octave noise and other broad-band stimuli. The tests conducted in the presence of masking noise showed a moderate adverse effect of dichotic presentation in the presence of masking noise. In the test for left/center/right discrimination thresholds (Exp. IIB), the mean increase in discrimination thresholds due to dichotic presentation was 11, 7, 3, 3, and 5 degrees for 1/6-octave noise, 1/3-octave noise, 1-octave noise, broad-band noise, and sound of breaking glass, respectively. The results of listening tests for the sound of breaking glass in the presence of masking noise showed that the threshold increase due to dichotic presentation was less than 10º for all SNR values, indicating that perception of source direction was only marginally affected.

The tests for direction identification conducted on normal-hearing subjects in the presence of broad-band masking noise (Exp. IIC) showed that dichotic presentation resulted in an increase of 1° – 8° in the mean rms error. In the tests involving hearing-impaired subjects with bilateral senorineural loss (Exp. IID), dichotic presentation resulted in a small increase in rms error due to dichotic presentation: 1.0° for sound of breaking glass and 0.1° for broad-band noise.

Thus the three tests using normal hearing subjects and the direction identification test using hearing-impaired subjects showed that the broad-band sound sources could be localized during dichotic presentation. The ACB based comb filters had only a small effect on source localization for broadband stimuli, and it may be inferred that the subjects were able to use the ITD and ILD cues across the bands for perceiving the source direction.

Chapter 5

MULTI-BAND FREQUENCY COMPRESSION

5.1 Introduction

Several studies have investigated the usefulness of spectral contrast enhancement schemes in improving the intelligibility of speech in noise for normal-hearing subjects and for subjects with sensorineural hearing loss (Bunnel, 1990; Stone and Moore, 1992b; Baer et al., 1993; Miller et al., 1999; Yang et al., 2003; Cohen, 2006). The processing involves enhancement of the spectral prominences which are perceptually significant. There may be errors in identification of the spectral prominences. Further, enhancement of spectral contrast may increase the dynamic range of the speech signal and hence may adversely affect speech perception due to the reduced dynamic range associated with the sensorineural loss. Another technique that can be used for reducing the effect of spectral masking in monaural hearing is multi-band frequency compression (Yasu et al., 2002; Arai et al., 2004; Kulkarni et al., 2007). In this technique, the speech spectrum is divided into a number of analysis bands. Spectral samples in each of these bands are compressed towards the band center by a constant compression factor, resulting in presentation of the speech energy in relatively narrow bands for reducing the masking by adjacent spectral components. The processing does not introduce any spectral tilt or compression of the broad-band spectrum, and it approximately preserves the harmonic structure in case of voiced speech and randomness in case of unvoiced speech, and the listeners do not need practice to adapt to the processed sound.

The objective of the study reported in this chapter is to find the best combination of processing parameters for multi-band frequency compression and to evaluate the effectiveness of the technique for improving speech perception for listeners with sensorineural loss. After a review of frequency compression techniques in Section 5.2, our signal processing technique is described in Section 5.3. The subsequent sections present the listening tests carried out for optimization of multi-band frequency compression parameters along with the results of speech intelligibility tests, and discussion.

5.2 Frequency compression

Earlier investigations on frequency compression for improving speech perception by the hearing-impaired listeners have shown mixed results (Reed et al., 1983; Turner and Hurtig, 1999; McDermott and Dean, 2000; Sakamoto et al., 2000; Simpson et al., 2005, 2006; Robinson et al., 2007; Fraga et al., 2008). The main objective of these studies was to compress the speech spectrum along the frequency axis, to improve speech perception by listeners with high-frequency loss. In the study by Reed et al. (1983), the speech spectrum was compressed to fit it into the reduced frequency range of the hearing impaired listener. The technique involved segmentation, warping, dilation and time aliasing, and resynthesis. With the frequency range of 1250 Hz and 2500 Hz for the compressed speech, the best performance obtained for frequency compressed speech was equivalent to the performance obtained by lowpass filtering with equivalent bandwidth. Turner and Hurtig (1999) investigated proportional frequency compression preserving the ratio between the spectral samples. The frequency components in the complex spectrum were scaled by a constant factor (0.5, 0.6, 0.7, 0.8, and 0.9). It resulted in scaling of the fundamental frequency of voicing by the same factor in the synthesized speech, while temporal envelope and duration were not significantly altered. In listening tests using nonsense syllables, recognition scores for normal-hearing subjects dropped significantly for compression factors below 0.7. Performance varied across the 16 hearing-impaired subjects, with an average improvement of 8 % for female voice and 4.7 % for male voice. Using a similar signal processing technique and a compression factor of 0.6, McDermott and Dean (2000) conducted listening tests using six hearing-impaired subjects with steeply sloping high frequency loss using monosyllabic words as the test material. Despite intensive training of the subjects, no improvements in recognition scores were observed.

Sekimoto and Saito (1980) used nonlinear frequency compression based on partial correlation (PARCOR) analysis-synthesis (Itakura et al., 1972). The original speech was low pass filtered with cut-off frequency of 5 kHz, and then sampled at 10 kHz with 10-bit quantization. A preliminary PARCOR analysis was carried out on speech frames (20 ms duration with 15 ms of overlap) to obtain 12th order linear prediction coefficients (LPC), pitch period, and voicing decision. The spectral envelope was obtained from the LPC coefficients by computing a 256-point DFT. An autocorrelation function, using a 256-point IDFT, was computed after modifying the spectral envelope by a predefined nonlinear compression function $f' = G(f)$. The final PARCOR analysis was then carried out on the first 13 terms of the autocorrelation function to obtain a new set of LPC coefficients. With these LPC coefficients, linear compression was achieved by PARCOR synthesis, at a lower sampling frequency. Compression factors of 1 (no compression), 0.8, 0.6, 0.5, and 0.4 were achieved by using sampling frequencies of 10, 8, 6, 5, and 4 kHz, respectively. Using this analysis-synthesis method, it was possible to (i) separately adjust the compression for voiced and unvoiced segments, (ii) adjust the fundamental frequency, and (iii) adjust the spectral envelope. However, the study did not report results from listening tests. Sakomoto et al. (2000) assessed the benefit of the scheme proposed by Sekimoto and Saito (1980) with the compression factor for both voiced and unvoiced speech adjusted in the range of 0.1 to 0.9 in steps of 0.1, by conducting listening tests on subjects with severe-to-profound loss (> 100 dB for frequencies of 2 kHz and higher). Out of 11 subjects, five preferred the processed speech and these five subjects participated in the intelligibility tests with test material consisting of monosyllabic and disyllabic Japanese words. An improvement of 2 - 12 % in recognition scores was observed for nonsense monosyllabic words. No improvement in recognition scores were observed for disyllabic words.

In the study by Simpson et al. (2005), frequencies above 1.6 kHz were subjected to nonlinear frequency compression, with the compression progressively increasing with frequency. There was no overlap of the lower uncompressed and the upper compressed parts of the spectrum, but the frequency ratios in the compressed part of the spectrum were not preserved. Seventeen subjects with moderate-to-profound sensorineural loss participated in monosyllabic word recognition tests. Improvements in recognition scores, in the range 13 – 17 %, were observed ($p < 0.001$). Another study by Simpson et al. (2006), using the same signal processing scheme and subjects with steeply sloping loss, showed mean recognition scores of 56 % for no compression and 52 % for compression. Robinson et al. (2007) evaluated a frequency transposition technique which was adapted to the subject's high-frequency dead region. Spectral samples from well within the dead regions were transposed to just within the dead region, without applying frequency compression. Low-frequency spectral samples were amplified, but were not affected by the transposition. A consonant discrimination test was carried out using VCV stimuli and the detection of the word-final consonants /*s*/ and /*z*/ was assessed using word pairs. Seven subjects with high-frequency dead regions participated in the test. For VCV tests, two subjects showed improvements in recognition scores. Averaged across the seven subjects, there was an improvement in the detection of word-final consonants /*s*/ and /*z*/.

Fraga et al. (2008) used piecewise linear frequency compression in an attempt to improve the perception of fricative consonants for persons with high-frequency loss. The input speech signal, sampled at 16 kHz, was divided into 50-ms frames with 75 % overlap, and a 2048-point DFT was computed on each frame. Since the compression was to be applied only to fricatives and affricates, the frames were classified into noise-like and tone-like frames based on a spectral flatness measure (SFM). The speech spectrum was divided into three bands: 0 – 0.5 kHz, 0.5 – 3 kHz, and 3 – 8 kHz. No compression was applied for the first band (0 - 0.5 kHz) in order to preserve the pitch perception of the voiced fricatives. The frequencies in the second band were heavily compressed (compression factor of 0.2), as the band is not relevant for fricative discrimination. Since most of the cues for fricative discrimination are present in the third band, these frequencies were compressed by a factor of 0.67, resulting in the mapping of 3 – 8 kHz to 1 – 4.33 kHz. The test material consisted of 24 monosyllables formed by the combination of six frequently used Brazilian Portuguese fricatives, with 19 monosyllables being known words in Portuguese, and the remaining 5 being nonsense syllables. Listening tests were conducted using ten normal-hearing subjects with the simulated loss above 1500 and 2000 Hz, by low-pass filtering the speech signal. For the simulated loss above 1500 Hz, the processing resulted in improvement in the recognition scores by 13 % and 9 % for male and female voice, respectively. For the simulated loss above 2000 Hz, the corresponding improvements were 4 % and 13 %.

Baskent and Shannon (2006) investigated the effect of frequency transposition around the dead regions in the cochlea, using a noise band vocoder simulating the stimulation pattern of a cochlear implant (Shannon et al., 2002; Shannon et al., 1995). Speech was processed by two types of processing: (i) dropped mapping by elimination of speech energy in the dead region, and (ii) frequency transposition by evenly redistributing the speech energy in the dead region into the remaining non-dead regions. Listening tests, with speech material consisting of 12 medial vowels in /*h*/-V-/*d*/ context and 14 consonants in /a/-C-/a/ context, were conducted using seven normal-hearing subjects with dead region simulated about center frequencies of 1, 2.4, and 5 kHz respectively. No significant improvement in speech perception was observed for frequency transposition under the different simulated dead region conditions, and the authors have attributed the lack of improvement to the possible spectral distortion due to frequency transposition.

In the multi-band frequency compression reported by Yasu et al. (2002) and Arai et al. (2004), the speech spectrum was divided into a number of bands corresponding to the auditory critical bandwidths and the spectral samples in each band were compressed towards the center of the band along the frequency axis. The input speech was divided into frames (with 75 % overlap), and a DFT was computed for each frame after applying Hamming window. The magnitude spectrum was then compressed towards the center of each critical band along the frequency axis, and the resulting magnitude spectrum was combined with the original phase spectrum. Speech signal was resynthesized using the overlap-add method. Compression in the range 0.1 – 0.9 was used. Two experiments were conducted: (i) mean opinion score (MOS) test for a set of six sentences, and (ii) intelligibility scores for 50 vowel-consonant-vowel utterances produced by a male speaker, with two hearing-impaired subjects in each experiment. In the MOS test, the subjects made a pair-wise comparison of the unprocessed and processed speech and rated the processed speech on a 0 – 5 scale with number 3 assigned to the unprocessed. The best MOS scores were observed for the compression factor of 0.6 – 0.8, with score improvements in the range of 0.3 – 0.8. There was a modest improvement in the recognition score: 38.3 % for the processed speech as against 35.4 % for the unprocessed speech.

Multi-band frequency compression concentrates spectral energy towards the band centers, without introducing any spectral tilt or compression of the broad-band spectrum. The quality and intelligibility of the resynthesized speech signal depends on the frequency mapping scheme used for compression, the type of bands along the frequency axis, the segmentation used for analysis-synthesis, and the compression factor. The objective of the present study is to find the best combination of these processing parameters for improving speech perception for persons with sensorineural loss. Three different frequency mapping schemes were investigated: (i) sample-to-sample mapping, (ii) spectral sample superimposition, and (iii) spectral segment mapping. Three different bandwidths were considered: (i) constant bandwidth, (ii) 1/3-octave bandwidth, and (iii) auditory critical bandwidth (ACB). These schemes for multi-band frequency compression were investigated using two types of segmentation for analysis-synthesis: (i) fixed frame and (ii) pitch synchronous. The selection of frequency mapping scheme, bandwidth, and segmentation was based on subjective evaluation of the perceived quality as assessed through MOS tests conducted on the normal-hearing subjects with increased masking due to sensorineural loss simulated by adding broad-band masking noise. The effectiveness of the scheme, with processing parameters as selected after the MOS tests, in improving speech intelligibility was evaluated by conducting modified rhyme test (MRT), on normal-hearing subjects with simulated loss and on subjects with moderate-to-severe sensorineural loss.

5.3 Signal processing for multi-band frequency compression

Signal processing for multi-band frequency compression involves three steps: (i) segmentation and spectral analysis, (ii) spectral modification, and (iii) resynthesis. In our scheme, the multi-band compression is carried out on the complex spectrum. It is computationally more efficient as compared to the earlier reported scheme involving computation of the magnitude and phase spectra (Yasu et al., 2002; Arai et al., 2004). The input speech signal, sampled at 10 kHz, is divided into segments with 50 % overlap. Each segment is zero padded to the length of N, and the N-point DFT is computed on it. The frequency axis is divided into a number of bands, in accordance with the type of bandwidth selected for multi-band compression. The complex spectral samples falling in each of the bands are compressed by a constant compression factor towards the center of the corresponding band, as shown in Fig. 5.1. The resulting complex spectrum is converted back to the time domain by N-point IDFT, and modified speech is resynthesized by the overlap-add method without any change in the rms value of the signal (Rabiner and Schafer, 1978; Proakis

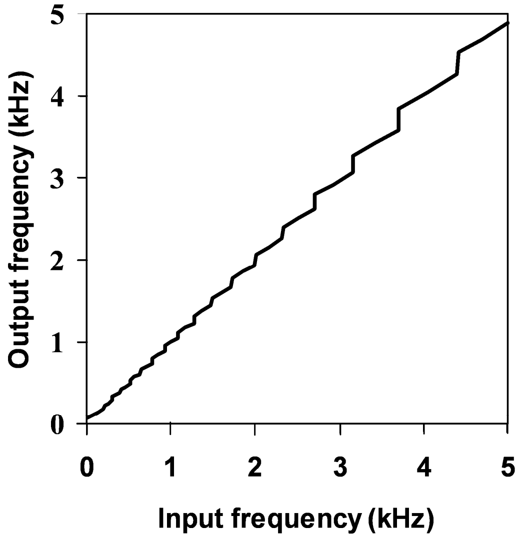


Fig. 5.1 Frequency mapping for multi-band frequency compression, with auditory critical bandwidths and compression factor of 0.6

and Manolakis, 1992; Mitra, 1998). In the present study, $N = 1024$ was used, as increasing it further did not result in a decrease in the perceived distortion in the resynthesized speech for various compression factors.

5.3.1 Segmentation

The frequency compression scheme was implemented using two types of segmentations for analysis-synthesis: (1) fixed-frame and (2) pitch-synchronous frames. In fixed-frame segmentation, the frame size was 20 ms with an overlap of 50 %. The pitch-synchronous segmentation used the frame length of two local pitch periods with an overlap of one pitch period. The processing involved a voicing decision followed by detection of glottal closure instants (GCIs), using the algorithm by Childers and Hu (1994). For a voiced segment, analysis was carried out using an analysis frame spanning from the previous GCI to the next GCI. For an unvoiced segment, the analysis frame width was the same as that of the last voiced frame.

5.3.2 Bandwidth

With the objective of finding the optimum bandwidth, three types of bandwidth were investigated: (i) constant bandwidth (CB) with number of bands equal to 18, corresponding to bandwidth 278 Hz, in the 0 – 5 kHz frequency range, (ii) 19 bands of 1/3-octave bandwidth in the frequency range of 0.07 to 5 kHz, and (iii) 18 bands based on auditory critical bandwidth (Zwicker, 1961). Table 5.1 lists the bands in the three types of bandwidths.

Table 5.1 List of bands in the three types of bandwidths: (a) constant bandwidth, (b) 1/3-octave bandwidth, and (c) auditory critical bandwidth. f_c: center frequency, f_l: lower cutoff, f_2: upper cutoff.

(a) Constant bandwidth

Band no.	f_c (kHz)	$f_1 - f_2$ (kHz)
1	0.139	0.000 - 0.278
2	0.417	0.278 - 0.556
3	0.694	0.556 - 0.833
4	0.972	0.833 - 1.111
5	1.250	1.111 - 1.389
6	1.528	1.389 - 1.667
7	1.806	1.667 - 1.944
8	2.083	1.944 - 2.222
9	2.361	2.222 - 2.500
10	2.639	2.500 - 2.778
11	2.917	2.778 - 3.055
12	3.195	3.055 - 3.333
13	3.472	3.333 - 3.611
14	3.750	3.611 - 3.889
15	4.028	3.889 - 4.167
16	4.306	4.167 - 4.444
17	4.583	4.444 - 4.722
18	4.861	4.722 - 5.000

(b) 1/3-octave bandwidth

Band no.	f_c (kHz)	$f_1 - f_2$ (kHz)
1	0.08	0.07 – 0.09
2	0.10	0.09 – 0.11
3	0.13	0.11 – 0.14
4	0.16	0.14 – 0.19
5	0.20	0.18 – 0.22
6	0.25	0.22 – 0.28
7	0.32	0.28 – 0.35
8	0.40	0.35 – 0.45
9	0.50	0.45 – 0.56
10	0.64	0.56 – 0.71
11	0.80	0.71 – 0.89
12	1.01	0.89 – 1.12
13	1.27	1.12 – 1.41
14	1.60	1.41 – 1.78
15	2.01	1.78 – 2.24
16	2.53	2.24 – 2.82
17	3.19	2.82 – 3.55
18	4.01	3.55 – 4.47
19	4.74	4.47 – 5.00

(c) Auditory critical bandwidth

Band no.	f_c (kHz)	$f_1 - f_2$ (kHz)
1	0.13	0.01 – 0.20
2	0.25	0.20 – 0.30
3	0.35	0.30 – 0.40
4	0.45	0.40 – 0.51
5	0.57	0.51 – 0.63
6	0.70	0.63 – 0.77
7	0.84	0.77 – 0.92
8	1.00	0.92 – 1.08
9	1.17	1.08 – 1.27
10	1.37	1.27 – 1.48
11	1.60	1.48 – 1.72
12	1.86	1.72 – 2.00
13	2.16	2.00 – 2.32
14	2.51	2.32 – 2.70
15	2.92	2.70 – 3.15
16	3.42	3.15 – 3.70
17	4.05	3.70 – 4.40
18	4.70	4.40 – 5.00

5.3.3 Frequency mapping

The quality and intelligibility of frequency compressed speech was found to depend on the type of frequency mapping employed. Three frequency mapping techniques were investigated: (i) sample-to-sample mapping, (ii) spectral sample superimposition and, (iii) spectral segment mapping.

(1) Sample-to-sample mapping: The relationship between compressed spectrum Y and the original spectrum X is given as

$$Y(k') = X(k) \tag{5.1}$$

The spectral sample k' of the compressed spectrum is related to the frequency sample k falling in the i th analysis band of the original spectrum by the following relation

$$k' = k_{ic} + \text{round}\ (c(k - k_{ic})) \tag{5.2}$$

where c = compression factor (0 – 1), and k_{ic} = center frequency of the i th analysis band, given by

$$k_{ic} = \text{round}\ (0.5(k_{is} + k_{ie})) \tag{5.3}$$

where k_{is} and k_{ie} are the starting and ending indices for the i th band. The frequency mapping is illustrated in Fig. 5.2, for the ACB bands of 0.2 – 0.3 kHz and 1.27 – 1.48 kHz with the center frequency of 0.25 kHz and 1.37 kHz respectively, and compression factor of 0.6. For f_s of 10 kHz and 1024-point DFT, spectral samples from 21 to 31 with 26 as the center point fall

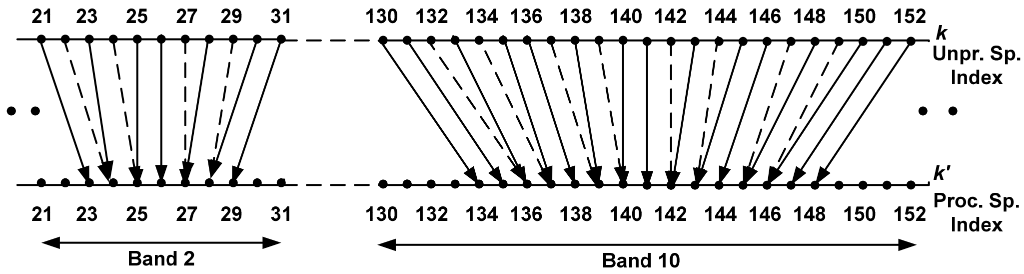


Fig. 5.2 Sample-to-sample mapping

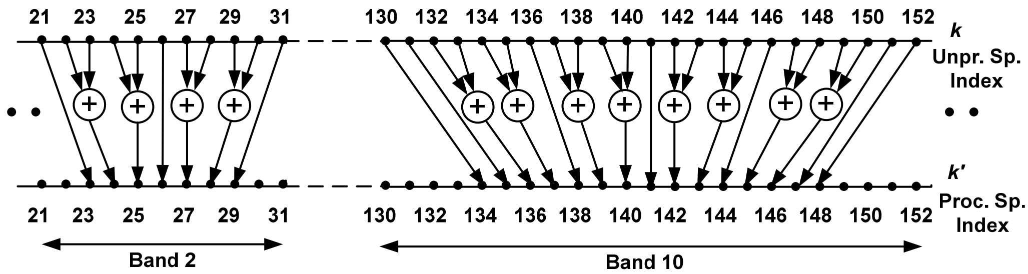


Fig. 5.3 Spectral sample superimposition

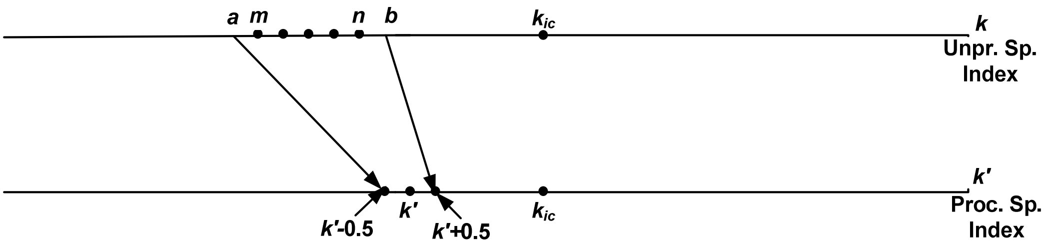


Fig. 5.4 Spectral segment mapping

in the first band and samples from 130 to 152 with 141 as the center point fall in the second band. In this mapping, if two or more spectral samples are mapped on to the same point, then only the one with the largest index amongst them is retained, resulting in an irregular variation in the spectrum and signal energy.

(2) Spectral sample superimposition: This mapping addresses the problem of missing components in the earlier mapping, by adding the spectral samples which map to the same index, as shown in Fig. 5.3. Reduction in the energy of the processed signal, as observed in the sample-to-sample mapping gets partly compensated. However, variation in the number of spectral samples contributing to the mapping causes some irregular variation in the spectrum of the resynthesized speech.

(3) Spectral segment mapping: This mapping achieves frequency compression without irregular variation in the spectrum. Let k' be the spectral sample index of the compressed spectrum $Y(k')$. As shown in Fig. 5.4, the spectral segment from a to b in the

unprocessed spectrum contributes to the sample k' on the compressed scale. The values of a and b are given by

$$a = k_{ic} - [(k_{ic} - (k' - 0.5))/ c] \tag{5.4}$$

$$b = a + 1/c \tag{5.5}$$

where c is the compression factor and k_{ic} is the center frequency of the i th band. Let m and n be the indices of the first and the last spectral samples, respectively, falling in the segment from a to b. Index m is the lowest integer higher than a and n is the highest integer lower than b. The processed spectrum is then given by

$$Y(k') = (m-a)X(m) + \sum_{j=m+1}^{n-1} X(j) + (b-n)X(n) \tag{5.6}$$

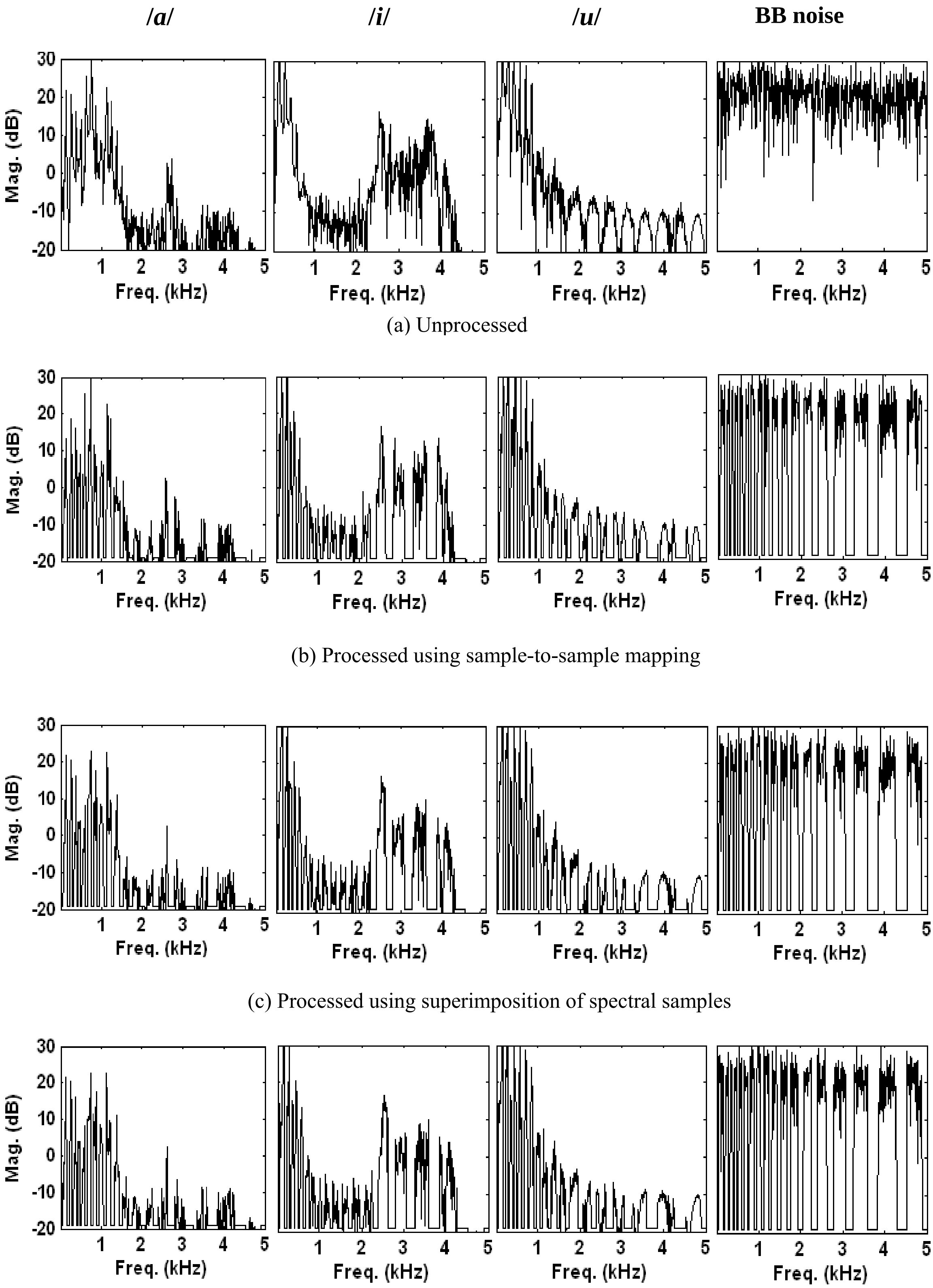
In this scheme, all the samples of the original spectral segment uniformly contribute to the compressed spectrum. As a result, the desired frequency compression is achieved without introducing any irregular variation.

A comparison of the mapping schemes was carried out using short-time spectra and spectrograms of unprocessed and frequency-compressed speech. For compression factor c = 1, there were no changes in magnitude and phase spectra. For other compression factors, irregular spectral variations were observed for the first two schemes, but no such variations were observed for the spectral segment mapping. Figure 5.5 shows the spectra of the 100 ms segments of vowel /*a*/, /*i*/, /*u*/, and broad-band noise processed using the three frequency mapping schemes with compression factor of 0.6. It is observed that the output from segment mapping has least spectral distortion. Figure 5.6 shows the wide-band spectrogram for the VCV utterance /*aka*/, processed using auditory critical bandwidth based spectral segment mapping, fixed-frame segmentation, and c = 0.6. It is observed that formant transitions are retained, with moderate shifts (less than 1/3-octave) in the formant locations. Harmonic structure is preserved in the form of vertical striations.

5.4 Listening tests

The listening tests were conducted for evaluation for the quality of the frequency compressed speech and evaluation for the intelligibility of the processed speech. Mean opinion score (MOS) test was conducted for quality assessment and the modified rhyme test (MRT) was used for speech intelligibility assessment. The overall evaluation involved three experiments: MOS tests on normal-hearing subjects with simulated sensorineural loss, MRT on normal-hearing subjects with simulated sensorineural loss, and MRT on hearing-impaired subjects.

The objective of the first experiment (Exp. IIIA) was to select the optimal combination of segmentation for analysis-synthesis, bandwidth, and frequency mapping scheme. In this experiment, evaluation of the perceived quality of the compressed speech for



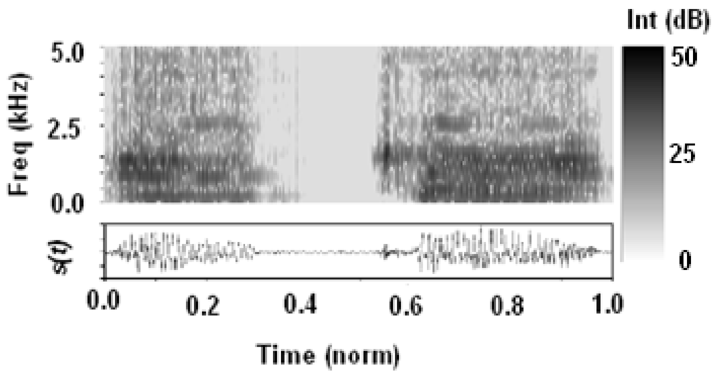
(a) Unprocessed

(b) Processed using sample-to-sample mapping

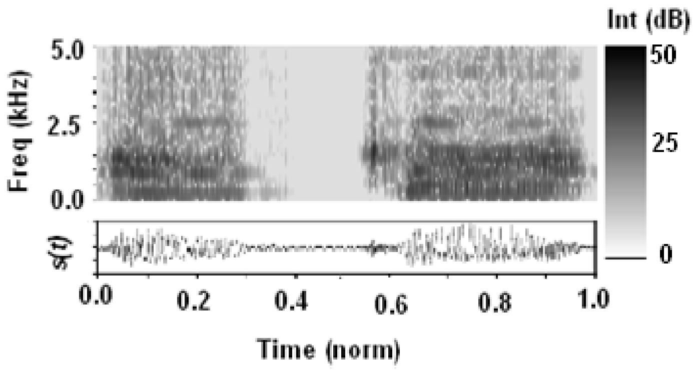
(c) Processed using superimposition of spectral samples

(d) Processed using spectral segment mapping

Fig. 5.5 Spectra of vowels /a/, /i/, /u/, and broad-band noise (segment length = 100 ms): unprocessed and processed using the three mappings for multi-band compression. Bandwidth: ACB, segmentation: fixed-frame, compression factor = 0.6.



(a) Unprocessed



(b) Processed

Fig. 5.6 Waveforms and wide-band spectrograms (Δf =300 Hz) of the VCV utterance /*aka*/ of 700 ms duration: (a) unprocessed and (b) frequency compressed with segment mapping scheme. Bandwidth: ACB, Segmentation: Fixed-frame, compression factor: 0.6.

different segmentation, bandwidths, and frequency mapping schemes was carried out through MOS tests, conducted on normal-hearing subjects with increased masking due to sensorineural loss simulated by broad-band masking noise as described in Section 3.4. Informal listening tests showed maximum benefit for compression factor of 0.6 (Kulkarni and Pandey, 2008), and hence MOS tests were carried out with compression factor of 0.6 for different segmentation, bandwidths and frequency mapping schemes. Broad-band random noise was added as a masker to the processed speech, keeping the SNR constant on a short-time (10 ms) basis, as described earlier in Section 3.4.

Based on the result of MOS test, further evaluation of the scheme with optimal combination of segmentation, bandwidth, and frequency mapping was carried out for consonant recognition using MRT. In the second experiment (Exp. IIIB), listening tests were conducted on normal-hearing subjects in the presence of broad-band masking noise for compression factors of 1.0, 0.8, 0.6, and 0.4. In the third experiment (Exp. IIIC), MRT was conducted to assess the effectiveness of the scheme for subjects with moderate-to-severe sensorineural loss.

5.4.1 Exp. IIIA: Mean opinion score (MOS) test

To assess the quality of multi-band frequency compression for various mapping schemes, bandwidths, and segmentations, mean opinion score (MOS) tests (Rothauser et al., 1969; Nakatsui and Mermelstein, 1982; Polkosky and Lewis, 2003) were conducted using six subjects with normal hearing (3 male and 3 female, age: 35 to 45 years, pure-tone thresholds < 20 dB HL). Test material included sustained vowels /*a i u*/ and the sentence "*we were away a year ago*". The stimuli were recorded from a male speaker in an acoustically treated audiometry room, using B&K microphone model 2210, at 10 kHz of sampling frequency and with 16-bit quantization.

Earlier studies have reported that the use of reference sound in MOS tests helps in normalizing the scores and the scores from such tests conducted at different times and places can be compared in a more reliable manner (Goodman et al., 1976; Nakatsui and Mermelstein, 1982; Kitawaki et al., 1984). The test procedure, employed in the current investigation, is similar to the one used by Yasu et al. (2004). Each presentation had two observation intervals: reference sound (unprocessed) and test sound (processed) separated by 0.5 s of silence in between. Subjects made pair-wise comparison of quality of the processed sound relative to the unprocessed on a 0 – 5 point scale with higher number indicating the better perceived quality. The number 3 was assigned for the reference sound (unprocessed). To assess the usefulness of the scheme in the increased spectral masking, MOS tests were performed by adding broad-band masking noise to the processed speech, at SNR values of 6, 0, and -3 dB.

5.4.2 Exp. IIIB and IIIC: Modified rhyme test (MRT)

The effectiveness of processing scheme in improving the recognition of consonants was assessed by conducting modified rhyme test (House, et al., 1965; Kreul, et al., 1968; ANSI, 1989; Yang and Hodgson, 2006). The method and test material are the same as described earlier in Chapter 3. The test was conducted using an automated test administration setup in an audiometry room. The stimuli were presented monaurally through headphones. The presentation level was set at the most comfortable listening level as selected by the individual listener, and the amplifier gain setting was maintained across all the listening conditions. Hence there was no change in the signal level for the different processing and listening conditions. In a multiple choice listening test, the response time provides a measure of the load on the perception process, and a decrease in the response time indicates an improved listening condition (Gatehouse and Gordon, 1990; Delogu et al., 1991; Baer et al., 1993; Meftah and Boudelaa, 1996; Apoux, et al., 2001). Hence the listening tests were conducted by recording the subject response as well as the response time.

The processing concentrated the spectral energy towards the band centers, without introducing a spectral tilt, or compression of the broad-band spectrum. It approximately preserved the harmonic structure in case of voiced speech and randomness in case of unvoiced speech, and it was verified that the listeners did not need practice to adapt to the processed sound. The processing did not result in a change in the signal level, because it used a scaling factor to compensate for the 50 % overlap-add in the analysis-synthesis. The RMS values of the processed stimuli were found to be within 0.3 dB with respect to the corresponding unprocessed stimuli.

In Exp. IIIB, the listening tests were conducted on six normal-hearing subjects (3 male and 3 female, age: 35 to 45 years, pure-tone thresholds < 20 dB HL). Tests were conducted for speech processed with compression factors of 1.0 (unprocessed), 0.8, 0.6, and 0.4. The stimuli were presented in the presence of broad-band masking noise at 9 SNR values: ∞ (no noise), 6, 3, 0, -3, -6, -9, -12, and -15 dB. Each subject responded to a total of 10,800 presentations (300 words × 4 compression factors × 9 SNR values), with a test run for each of the listening conditions taking about 40 min. In order to minimize any bias due to practice or fatigue, the presentation order for the different processing and listening conditions was randomized for each subject. On a given day, a subject participated in a maximum of two tests. Test sessions were spread over a period of one month depending on the availability and convenience of the subjects.

In Exp. IIIC, the tests were conducted on eight subjects with moderate-to-severe sensorineural loss (6 male and 2 female, age: 32–66 years, average of the pure-tone thresholds at 0.5, 1, and 2 kHz : 45–88 dB HL, and hearing thresholds as given in Table 5.2). All the subjects read and signed informed consent for participation in the listening tests and were compensated for travel and incidental expenses. During the tests, subjects did not wear their hearing aids. The speech signal was presented through headphones and no frequency-dependent amplification was used. The speech was processed with compression factors of 1.0, 0.8, 0.6, and 0.4. Each subject responded to a total of 1,200 presentations (300 words × 4 compression factors), and a test session for each listening condition took approximately one hour. The presentation order for the different processing conditions was randomized for each subject in order to minimize the bias due to practice or fatigue. The test sessions were spread over one month as per the convenience and willingness of the subjects.

5.5 Results

In this section, the results of all the three experiments conducted for assessing the quality and intelligibility of the speech processed by multi-band frequency compression are presented.

Table 5.2 Hearing thresholds of the test ear for the subjects in Exp. IIIC. PTA: Average of the pure tone thresholds at 0.5, 1, and 2 kHz.

Subject (Sex, age)	Test ear	Hearing threshold (dB HL)						PTA (dB HL)
		Frequency (kHz)						
		0.25	0.5	1	2	4	6	
KNR (M, 32)	R	50	55	55	55	60	60	55
MNR (M, 45)	R	80	85	80	100	100	110	88
PKR (M, 66)	R	80	80	90	90	95	110	87
PAL (F, 56)	L	50	50	55	60	65	70	55
PPR (M, 35)	R	70	70	70	75	75	80	72
PEL (M, 34)	L	50	50	55	55	55	65	54
RJL (F, 56)	L	40	40	45	50	55	60	45
SSR (M, 54)	R	85	85	85	85	90	100	85

5.5.1 Results of Experiment IIIA (MOS test conducted on normal-hearing subjects)

The results of MOS tests, conducted on six normal-hearing subjects, for studying the effects of different types of segmentation, bandwidth, and frequency mapping schemes on the perceived quality, are summarized in Table 5.3. In all the three cases, the compression factor *c* was 0.6. The values shown are the difference between the MOS (averaged across the six subjects) for the processed speech and that for the unprocessed speech, with a positive cell entry indicating an improvement in the perceived quality of the processed sound. The standard deviations are given in parentheses. The statistically significant increases are indicated along with the significance level (p) for one-tailed paired t-test. In the absence of masking noise, processing resulted in a decrease in the scores but it resulted in improved scores in the presence of masking noise. The patterns of improvement for the vowel sequence and the sentence as the test material were the same, values being higher in case of sentence. For a detailed comparison we will consider the sentence scores.

Table 5.3(a) shows the improvement in the score, for the two segmentation schemes, with spectral segment mapping and ACB bandwidth. For no-noise condition (SNR = ∞), processing using either of the segmentations degraded the quality, and decrease in the score due to fixed-frame segmentation was higher. The processing resulted in improvement in scores at lower SNR values. With pitch-synchronous segmentation, the improvement was visible at SNR of 6 dB and increased further at lower SNR values. At SNR = -3 dB, both segmentation techniques showed statistically significant (p < 0.005) improvements, and improvements with pitch-synchronous segmentation was higher than that with fixed-frame segmentation. Across all the SNR conditions, the scores with pitch-synchronous segmentation

Table 5.3 Exp. IIIA: Effect of different segmentations, bandwidths, and mappings on the increase in MOS, with respect to the unprocessed speech, averaged across the six subjects. Standard deviations are given in the parentheses.

(a) Effect of segmentation. $c = 0.6$, BW: ACB, Mapping: M3

Test mat.	SNR (dB)	Segmentation	
		Fixed-frame	Pitch-synch.
/aiu/	∞	-0.47 (0.16)	-0.22 (0.18)
	6	-0.17 (0.19)	0.05 (0.16)
	0	0.04 (0.12)	0.13 (0.10) *
	-3	0.01 (0.10)	0.36 (0.19) †
Sent.	∞	-0.32 (0.10)	-0.07 (0.08)
	6	0.01 (0.15)	0.33 (0.20) ††
	0	0.30 (0.15) ††	0.56 (0.16) ††
	-3	0.35 (0.10) ††	0.68 (0.17) ††

(b) Effect of bandwidth. $c = 0.6$. Seg.: fixed frame, Mapping: M3.

Test mat.	SNR (dB)	Bandwidth		
		Const.	1/3 oct.	ACB
/aiu/	∞	-0.50 (0.15)	-0.22 (0.08)	-0.12 (0.10)
	6	-0.47 (0.14)	-0.12 (0.18)	0.00 (0.09)
	0	-0.44 (0.15)	-0.05 (0.20)	0.06 (0.12)
	-3	-0.28 (0.18)	0.05 (0.16)	0.20 (0.08) ††
Sent.	∞	-0.37 (0.12)	-0.05 (0.14)	0.00 (0.11)
	6	-0.10 (0.19)	0.18 (0.12) †	0.38 (0.12) ††
	0	0.05 (0.14)	0.32 (0.15) ††	0.53 (0.08) ††
	-3	0.18 (0.08) ††	0.31 (0.13) ††	0.71 (0.19) ††

(c) Effect of mappings. $c = 0.6$, Seg.: fixed frame, BW: ACB.

Test mat.	SNR (dB)	Mapping		
		Sample-to-sample	Super impo.	Spect. segment
/aiu/	∞	-0.62 (0.15)	-0.2 (0.13)	-0.07 (0.08)
	6	-0.37 (0.19)	-0.12 (0.13)	-0.02 (0.08)
	0	-0.07 (0.16)	0.05 (0.14)	0.13 (0.10) *
	-3	0.06 (0.12)	0.18 (0.17) *	0.33 (0.15) ††
Sent.	∞	-0.49 (0.12)	-0.22 (0.08)	-0.12 (0.12)
	6	-0.18 (0.18)	0.13 (0.21)	0.40 (0.15) ††
	0	-0.18 (0.15)	0.33 (0.15)	0.53 (0.14) ††
	-3	0.00 (0.14)	0.43 (0.14) ††	0.71 (0.17) ††

$^{*}p < 0.05$ $\quad$ $^{\dagger}p < 0.01$ $\quad$ $^{\dagger\dagger}p < 0.005$

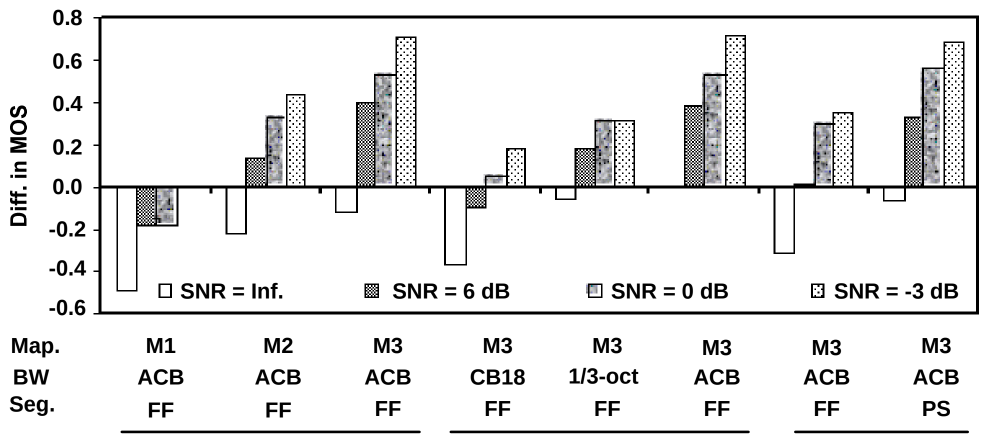


Fig. 5.7 Exp. IIIA: Difference in MOS (averaged across the six subjects) for different processing conditions. Segmentation: fixed-frame (FF) and pitch-synchronous (PS), Bandwidth: ACB, 1/3-octave (1/3-oct.), and CB18, Mappings: sample-to-sample (M1), superimposition of spectral samples (M2), and spectral segment mapping (M3). Compression factor (c) = 0.6, Test material: sentence.

were higher by 0.25 – 0.33 as compared to the fixed-frame segmentation. Table 5.3(b) shows the effect of different bandwidths. The tests were conducted for spectral segment mapping with fixed-frame segmentation. For SNR = -3 dB, processing with all the three bandwidths resulted in statistically significant improvements ($p < 0.005$). Across all the SNR conditions, the scores with ACB were consistently higher than with the other two bandwidths. Table 5.3(c) shows the effect of different mapping schemes on the perceived quality of the processed speech for fixed-frame segmentation and ACB bandwidth. The quality of the compression scheme using spectral segment mapping was higher than the other two mappings, for all the test materials and SNR values. At SNR = -3 dB, it gave an improvement of 0.71 ($p < 0.005$).

The differences in MOS for different combinations of the processing conditions, and sentence as the test material, are plotted in Fig. 5.7. The improvements were higher at lower SNR values, and the maximum improvement occurred for processing using ACB based compression with spectral segment mapping and pitch-synchronous segmentation.

5.5.2 Results of Experiment IIIB (MRT on normal-hearing subjects)

The MRT was conducted to evaluate the improvement in speech intelligibility and to investigate the effects of compression factor c. Processing was carried out with ACB based bandwidths and pitch-synchronous segmentation, as this combination resulted in maximum improvement in MOS in Exp. IIIA.

Table 5.4 Exp. IIIB: Recognition score (%) for subjects with normal hearing, for 9 SNR conditions, and 3 compression factors (*c*). s.d.: standard deviation, Impr.: improvement (averaged across the subjects), *p*: one tailed significance level for paired t-test (processed vs. unprocessed, *n* = 6, and df = 5).

Sub-ject	SNR= ∞ dB				SNR= 6 dB				SNR= 3 dB			
	Unp.	*c*			Unp.	*c*			Unp.	*c*		
		0.8	0.6	0.4		0.8	0.6	0.4		0.8	0.6	0.4
DSJ	97.0	93.3	90.6	85.6	91.0	89.7	86.6	81.6	90.3	86.7	83.0	84.3
MKD	96.3	96.0	90.3	84.0	92.3	90.0	84.3	73.0	85.3	88.0	81.0	78.3
PNK	98.3	99.0	88.6	83.6	95.0	93.0	82.3	80.0	92.6	81.3	78.0	77.6
RSH	93.0	89.7	87.0	79.6	87.0	85.7	78.0	70.6	80.6	81.0	80.3	73.0
SGK	96.3	93.7	90.6	84.6	93.0	90.3	84.3	81.0	90.6	87.7	80.0	78.3
SPK	93.3	94.0	83.0	84.6	86.3	85.3	77.6	78.0	81.3	75.7	74.6	75.6
Mean	95.7	94.3	88.4	83.7	90.7	89.0	82.2	77.4	86.7	83.4	79.5	77.9
s.d.	2.1	3.1	3.0	2.1	3.4	3.0	3.7	4.5	5.1	4.9	2.9	3.8
Impr.		-1.4	-7.3	-12.0		-1.7	-8.5	-13.3		-3.3	-7.2	-8.8
p		–	–	–		–	–	–		–	–	–

Sub-ject	SNR= 0 dB				SNR= -3 dB				SNR= -6 dB			
	Unp.	*c*			Unp.	*c*			Unp.	*c*		
		0.8	0.6	0.4		0.8	0.6	0.4		0.8	0.6	0.4
DSJ	75.3	81.0	87.6	80.6	72.6	81.3	87.0	79.6	70.0	75.3	81.0	78.3
MKD	74.0	81.0	83.0	77.0	71.6	80.0	84.0	75.0	70.0	82.0	83.0	71.3
PNK	84.6	78.3	85.3	80.6	77.0	74.3	82.6	78.3	67.6	64.7	79.3	76.0
RSH	69.3	80.3	80.6	70.3	67.3	72.7	75.0	67.3	62.3	68.7	70.6	63.6
SGK	77.3	81.7	88.6	81.0	73.3	76.0	83.3	79.0	62.3	66.3	80.3	76.3
SPK	69.0	70.3	79.6	76.0	69.0	63.7	77.3	71.0	64.3	62.7	78.6	72.6
Mean	74.9	78.8	84.1	77.6	71.8	74.7	81.5	75.0	66.0	70.0	78.8	73.0
s.d.	5.8	4.3	3.7	4.1	3.4	6.3	4.5	5.0	3.6	7.3	4.3	5.3
Impr.		3.9	9.2	2.7		2.9	9.7	3.2		4.0	12.8	7.0
p		n.s.	<0.001	n.s.		n.s.	<0.001	0.01		n.s.	<0.001	<0.01

Sub-ject	SNR= -9 dB				SNR= -12 dB				SNR= -15 dB			
	Unp.	*c*			Unp.	*c*			Unp.	*c*		
		0.8	0.6	0.4		0.8	0.6	0.4		0.8	0.6	0.4
DSJ	65.6	71.7	78.3	74.0	57.0	61.7	71.3	68.3	48.0	54.3	62.3	57.0
MKD	61.3	73.7	80.6	70.0	58.0	62.7	70.6	62.6	48.6	55.3	63.3	48.0
PNK	63.3	66.0	74.6	68.3	55.3	55.3	69.6	60.6	48.0	47.3	58.6	50.6
RSH	55.0	61.3	68.3	60.6	46.6	69.0	62.6	53.3	38.3	57.7	62.0	40.0
SGK	57.3	60.7	78.0	69.0	50.0	54.0	73.3	61.0	44.3	51.0	61.0	51.0
SPK	54.6	57.3	76.6	66.3	46.0	57.7	68.6	56.6	41.3	50.7	64.6	45.6
Mean	59.5	65.1	76.1	68.0	52.2	60.1	69.3	60.4	44.8	52.7	62.0	48.7
s.d.	4.6	6.5	4.3	4.4	5.3	5.6	3.7	5.1	4.2	3.8	2.1	5.7
Impr.		5.6	16.6	8.5		7.9	17.1	8.2		7.9	17.2	3.9
p		<0.01	<0.001	<0.001		<0.05	<0.001	<0.001		<0.01	<0.001	<0.05

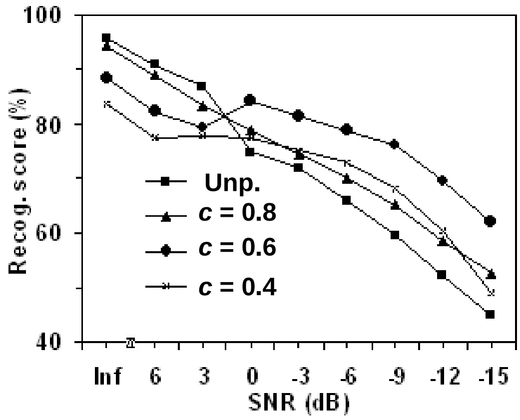


Fig. 5.8 Exp. IIIB: Recognition score (averaged across the six subjects) vs. SNR for unprocessed and frequency compressed with compression factor (*c*) of 0.8, 0.6, and 0.4.

The recognition scores from the MRT conducted on the six normal-hearing subjects, for the 9 SNR values of the masking noise, are shown in Table 5.4. In addition to the scores for the individual subjects, the mean score, the standard deviation (s.d.), the mean improvement, and one-tailed significance level (*p*) for a paired t-test (processed vs. unprocessed) are also given. Figure 5.8 gives a plot of percentage recognition score (averaged across the six subjects) as a function of SNR, for speech processed with the three values of the compression factor *c*.

The processing resulted in a reduction in the recognition score at the SNR values of ∞, 6, and 3 dB. However, an increase in the recognition score occurred for the SNR values lower than 0 dB. The improvements with *c* value of 0.8 and 0.4 were smaller than those with 0.6. Table 5.4 shows that the increase in the recognition scores for all the SNR values less than 3 dB for $c = 0.6$ were statistically significant ($p < 0.001$), and the increase was approximately 17 % for SNR < -6 dB. For unprocessed speech, the mean recognition score was about 60 % for SNR of -9 dB. For the speech signal processed with $c = 0.6$, the same recognition score was obtained for SNR of -15 dB, indicating an SNR advantage of approximately 6 dB.

A two-way repeated-measures analysis of variance (ANOVA) was conducted on recognition scores with processing and SNR as the main effects (Table E.7 in Appendix E). The effects of both the factors and their interaction were found to be statistically significant ($p < 0.001$). A one-way repeated-measures ANOVA was conducted on recognition scores with processing as the main effect, separately at each of the SNR values (Table E.8 in Appendix E) and the effect of processing was found to be significant ($p < 0.001$) at all SNR values. Tukey's honestly significant difference (HSD) test was conducted for pair-wise comparison of the scores (Table E.9 in Appendix E). At all SNR values, $c = 0.6$ resulted in the highest

Table 5.5 Exp. IIIB: Response time values (s) for subjects with normal hearing, for 9 SNR conditions, and 3 compression factors (c). s.d.: standard deviation, Impr.: improvement (averaged across the subjects), p: one tailed significance level for paired t-test (processed vs. unprocessed, n = 6, and df = 5).

Sub-ject	SNR= ∞ dB				SNR= 6 dB				SNR= 3 dB			
	Unp.	c			Unp.	c			Unp.	c		
		0.8	0.6	0.4		0.8	0.6	0.4		0.8	0.6	0.4
DSJ	2.33	2.02	2.76	2.57	2.32	2.18	2.59	2.55	2.46	2.58	2.44	2.53
MKD	2.73	2.96	3.03	4.57	3.38	2.80	2.68	4.94	2.98	2.79	2.74	4.50
PNK	2.56	3.33	3.32	3.28	3.10	3.38	2.89	3.44	3.06	3.41	2.95	3.27
RSH	2.41	3.25	3.10	3.36	2.58	3.04	2.93	2.82	2.68	3.13	2.88	3.09
SGK	2.45	3.94	3.22	3.30	2.93	3.34	2.80	3.44	3.06	3.24	2.82	3.20
SPK	3.47	3.01	4.51	3.10	3.73	2.83	4.09	3.34	4.20	2.87	4.48	4.17
Mean	2.66	3.08	3.32	3.37	3.01	2.93	3.00	3.42	3.07	3.00	3.05	3.46
s.d.	0.42	0.63	0.61	0.66	0.51	0.44	0.55	0.83	0.60	0.31	0.72	0.73
Impr.		-0.42	-0.66	-0.71		0.08	0.01	-0.41		0.07	0.02	-0.39
p		–	–	–		n.s.	n.s.	–		n.s.	n.s.	–

Sub-ject	SNR= 0 dB				SNR= -3 dB				SNR= -6 dB			
	Unp.	c			Unp.	c			Unp.	c		
		0.8	0.6	0.4		0.8	0.6	0.4		0.8	0.6	0.4
DSJ	2.50	2.45	2.30	2.54	2.86	2.59	2.17	2.51	2.30	2.32	2.15	2.39
MKD	3.25	2.81	2.74	4.68	3.48	2.98	2.50	4.31	3.08	2.56	2.24	3.83
PNK	3.40	3.58	2.90	3.31	3.37	3.36	2.86	3.19	3.39	3.57	2.80	3.54
RSH	2.61	3.11	3.05	3.00	2.70	2.98	2.80	2.91	2.94	2.94	2.61	2.89
SGK	3.32	3.22	2.80	3.31	3.21	3.24	2.68	3.19	3.28	3.24	2.72	3.54
SPK	4.19	2.87	4.17	3.80	4.37	2.86	4.38	4.02	4.39	2.86	3.58	3.46
Mean	3.21	3.01	2.99	3.44	3.33	3.00	2.90	3.36	3.23	2.91	2.68	3.28
s.d.	0.61	0.39	0.63	0.74	0.59	0.27	0.77	0.68	0.69	0.45	0.51	0.53
Impr.		0.2	0.22	-0.23		0.33	0.43	-0.03		0.32	0.55	-0.05
p		n.s.	n.s.	–		n.s.	<0.05	–		n.s.	<0.005	–

Sub-ject	SNR= -9 dB				SNR= -12 dB				SNR= -15 dB			
	Unp.	c			Unp.	c			Unp.	c		
		0.8	0.6	0.4		0.8	0.6	0.4		0.8	0.6	0.4
DSJ	2.47	1.52	1.77	2.22	2.57	1.60	1.54	2.16	2.70	1.85	2.07	2.14
MKD	3.17	2.85	1.78	3.67	3.13	2.85	2.95	4.03	2.93	2.46	2.57	4.28
PNK	3.78	3.67	3.17	3.48	3.90	3.56	2.78	3.39	3.85	3.70	3.38	3.38
RSH	3.07	2.93	2.19	3.03	2.85	2.85	2.33	2.90	3.11	2.89	2.84	3.02
SGK	3.57	3.25	3.17	3.48	3.73	3.45	2.78	3.39	3.69	3.51	3.38	3.38
SPK	4.24	2.82	2.90	3.71	4.40	2.76	3.36	3.93	4.49	2.81	3.59	4.29
Mean	3.38	2.84	2.50	3.26	3.43	2.84	2.62	3.30	3.46	2.87	2.97	3.41
s.d.	0.62	0.72	0.66	0.56	0.70	0.70	0.63	0.69	0.67	0.68	0.58	0.81
Impr.		0.54	0.88	0.12		0.59	0.81	0.13		0.59	0.49	0.05
p		<0.05	<0.001	n.s.		<0.05	<0.05	n.s.		0.05	<0.005	n.s.

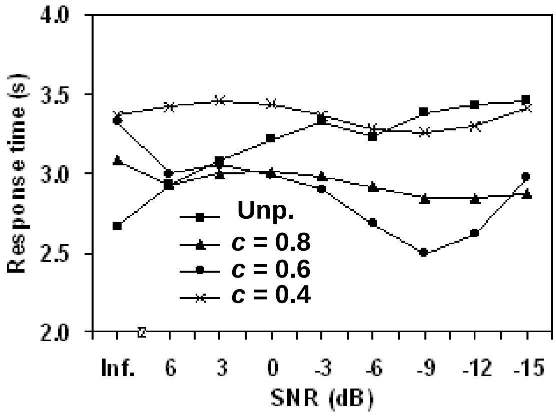


Fig. 5.9 Exp. IIIB: Response time (averaged across the six subjects) vs. SNR for unprocessed and frequency compressed with compression factor (*c*) of 0.8, 0.6, and 0.4.

improvement and the improvements were significant (p < 0.01) for SNR < 0 dB. The improvements for compression factor 0.8 and 0.4 were significant for SNR values less than -9 and -6 dB, respectively. For SNR < 0 dB, the scores for c = 0.6 were significantly higher (p < 0.01) than those with the other two compression factors.

Table 5.5 gives the response times for unprocessed and speech processed with the three c values. The mean response time, standard deviation, the mean improvement, and one-tailed p value for a paired t-test are also given in the table. For the unprocessed speech, the response times increased as SNR decreased for all the subjects indicating an increased perceptual load. A reduction in the response time was observed for processed speech at lower SNR values, indicating a reduced perceptual load. Figure 5.9 shows the mean response time as a function of SNR. At SNR values < 0 dB, the processing reduced the response time for all the values of c, with maximum reduction observed for c = 0.6. The decrease in response time at the SNR values of 3, 0, -3, -6, -9, -12, and -15 dB were 0.02, 0.22, 0.43, 0.55, 0.88, 0.81, and 0.49 s, respectively.

A two-way repeated-measures ANOVA was conducted on response times with processing and SNR as the main effects (Table E.10 in Appendix E). Only the effect of interaction was found to be significant (p < 0.001). A one-way repeated-measures ANOVA was conducted on response times with processing as the main effect, separately at each of the SNR values (Table E.11 in Appendix E) and the effect of processing was found to be significant (p < 0.05) at SNR values of -6, -9, and -12 dB. Tukey's HSD test conducted for pair-wise comparison (Table E.12 in Appendix E) showed that the improvement for compression factor of 0.6 was significant (p < 0.01) at SNR values of -9 and -12 dB.

Table 5.6 Exp. IIIC: Recognition scores for the hearing-impaired subjects. s.d.: standard deviation, Impr.: improvement, *p*: significance level (one tailed) for paired t-test (unprocessed vs. processed, n = 8, and df = 7).

Sub.	Unp.	*c*		
		0.8	0.6	0.4
KNR	63.3	64.7	78.3	58.3
MNR	51.0	56.0	69.7	67.0
PKR	46.7	54.7	65.7	60.3
PAL	64.3	69.0	85.0	67.0
PPR	70.7	72.3	80.0	67.3
PEL	70.3	62.5	91.7	63.7
RJL	67.7	70.0	81.3	72.7
SSR	63.3	60.3	77.7	63.0
Mean	62.2	63.7	78.7	64.9
s.d.	8.8	6.5	8.2	4.6
Impr.		1.5	16.5	2.7
p			<0.001	

Thus the recognition scores and the response time results from MRT show that multi-band frequency compression improved speech perception in the presence of masking noise, and maximum improvement was observed for c = 0.6.

5.5.3 Results of Experiment IIIC (MRT on hearing-impaired subjects)

In this experiment, speech intelligibility was assessed by conducting MRT on subjects with moderate-to-severe sensorineural loss. Eight subjects participated in the listening tests. The response data were analyzed to get the percentage recognition score and response time for unprocessed and processed speech with compression factor c = 0.8, 0.6, and 0.4.

The recognition score results are summarized in Table 5.6. In addition to the scores for the individual subjects, it also gives the mean, s.d., the mean improvement, and one-tailed p value for a paired t-test. The recognition score for all the subjects are also shown in Fig. 5.10. Only a small improvement in recognition score was observed for c of 0.8 and 0.4. For c = 0.6, the improvement ranged 9 – 21 % (mean = 16.5 %, $p < 0.001$). A one-way repeated-measures ANOVA, conducted on recognition scores (Table E. 13 in Appendix E), showed a significant effect ($p < 0.001$) of processing. Tukey's HSD test for pair-wise comparison (Table E. 15 in Appendix E) showed that the improvements in recognition scores were highest for compression factor of 0.6 and the differences with reference to the other compression factors were statistically significant ($p < 0.01$).

The response times for the unprocessed speech and the speech processed with the three compression factors for all the subjects are given in Table 5.7. The mean, s.d, the mean improvement, and one-tailed p value for a paired t-test are also given in the table. The

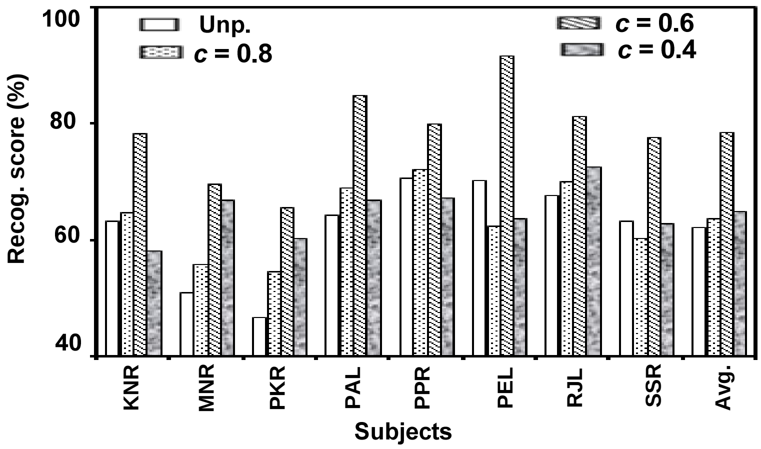


Fig. 5.10 Exp. IIIC: Recognition score (%) for the hearing impaired subjects for unprocessed and frequency compressed speech with compression factor (c) of 0.8, 0.6, and 0.4.

Table 5.7 Exp. IIIC: Response time values (s) for the hearing-impaired subjects. s.d.: standard deviation, Impr.: improvement, p: significance level (one tailed) for paired t-test (unprocessed vs. processed, n = 8, and df = 7).

Sub.	Unp.	c		
		0.8	0.6	0.4
KNR	4.82	4.63	3.77	3.95
MNR	4.19	4.11	3.67	3.91
PKR	4.54	4.08	3.49	3.71
PAL	4.23	5.08	3.63	3.86
PPR	4.18	4.54	3.80	3.78
PEL	4.97	4.92	3.57	4.03
RJL	4.45	3.76	3.26	3.68
SSR	3.89	3.32	3.00	3.82
Mean	4.41	4.31	3.52	3.84
s.d.	0.36	0.60	0.27	0.12
Impr.		0.1	0.89	0.57
p			<0.001	<0.001

response times are also shown in Fig. 5.11. The processing reduced the response time for all the subjects, with the maximum improvement observed for c = 0.6. Improvements across the subjects ranged 0.38 – 1.41 s, with a mean improvement of 0.89 s (p < 0.001). These improvements are very similar to those observed for the listeners with normal hearing at lower SNR values. A one-way repeated measures ANOVA conducted on response times (Table E.14 in Appendix E) showed a significant (p < 0.001) effect for the processing. Tukey's HSD test for pair-wise comparison (Table E.15 in Appendix E) showed that the improvements in response time were highest for compression factor of 0.6 and the differences with respect to the other compression factors were statistically significant (p < 0.01).

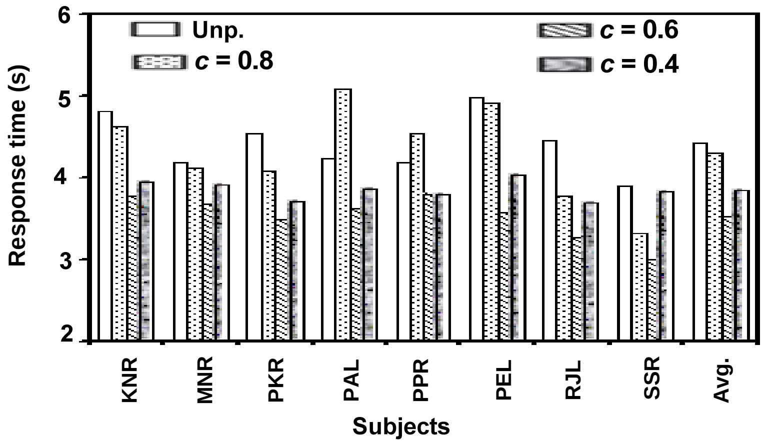


Fig. 5.11 Exp. IIIC: Response time (s) for the hearing impaired subjects for unprocessed and frequency compressed speech with compression factor (c) of 0.8, 0.6, and 0.4.

5.6 Discussion

The multi-band frequency compression was carried out on the complex spectrum and hence it did not involve computation of magnitude and phase spectra. We investigated the effects of frequency mapping, bandwidth, segmentation for analysis-synthesis, and compression factor, to find their best combination, and assessed the effectiveness of the scheme in improving speech perception for monaural presentation.

The MOS test was conducted for subjective evaluation of the quality of processed speech, for finding the effects of different types of frequency mapping scheme, bandwidth, and segmentation for analysis-synthesis. Listening tests were conducted on six subjects with normal hearing, with broad-band masking noise added to simulate increased spectral masking. The pitch-synchronous segmentation resulted in better scores than the fixed-frame segmentation. Out of the three bandwidths investigated, maximum improvement in MOS was observed for auditory critical bandwidth based compression, for all the test materials and SNR values. The compression scheme using spectral segment mapping was rated higher than the other two mappings. An overall investigation showed that the highest scores were observed for auditory critical band based frequency compression using spectral segment mapping and pitch-synchronous segmentation for analysis-synthesis.

The effectiveness of multi-band frequency compression in improving recognition of consonants by normal-hearing subjects in the presence of broad-band masking noise and by subjects with moderate-to-severe sensorineural loss was assessed using MRT for recognition of consonants in the word-initial and word-final positions. Processing involved pitch-synchronous segmentation with 50 % overlap, spectral segment mapping for compression of complex spectral samples in auditory critical bandwidth based analysis bands, and overlap-add method for resynthesis. Average response time was also measured to provide an indication of the load on the perception process.

In the listening tests using six normal-hearing subjects, the maximum improvement in the recognition score was observed for the compression factor of 0.6. With this compression factor, the improvement in recognition scores (with respect to the unprocessed) averaged across the subjects was 9.7, 12.8, 16.6, 17.1, and 17.2 % for SNR values of -3, -6, -9, -12, and -15 dB, respectively, and the improvements were statistically significant ($p < 0.001$). At lower SNR values, the improvement in the scores was equivalent to an SNR advantage of 6 dB, indicating that the processing helped in improving speech intelligibility in the presence of increased masking. Analysis of response time indicated that the processing helped in reducing the response time at lower SNR values, with the maximum reductions in the response time for the compression factor of 0.6. With this compression factor, the mean improvement in response time was 0.43, 0.55, 0.88, and 0.81 s for SNR values of -3, -6, -9, and -12 dB, respectively.

Further evaluation of the processing scheme was carried out using eight subjects with moderate-to-severe sensorineural loss. For the compression factor of 0.8, a moderate increase in the recognition score in the range 1 – 8 % was observed for 6 out of the 8 subjects. For the compression factor of 0.6, improvement in recognition score in the range 9 – 21 % (mean = 16.5 %, $p < 0.001$) occurred for all the subjects. For the compression factor of 0.4, there was only a moderate improvement in the range 3 – 16 % for four subjects. The processing resulted in a decrease in the response time for all the subjects: 0.10, 0.89, and 0.57 s for compression factors of 0.8, 0.6, and 0.4, respectively. The compression factor giving the maximum improvement in the tests on hearing-impaired subjects was 0.6, the same value as in the tests on normal-hearing subjects. The pattern of improvement in recognition score and response time across the individual subjects did not show any specific relation to the audiograms or to the scores for the unprocessed speech. This may be because although all the subjects had moderate sensorineural loss in the test ear, the extent of masking across them may be very different. Tests conducted on a larger number of subjects to evaluate the improvements due to processing and tests to measure the extent of masking may help in identifying the group most likely to benefit by the processing.

Thus the investigation showed that the scheme of multi-band frequency compression using pitch-synchronous segmentation, auditory critical bandwidths, and spectral segment mapping, helped in improving speech perception for subjects with moderate-to-severe sensorineural loss. The maximum improvement was observed for the compression factor of 0.6. This compression factor also resulted in maximum reduction in response time. The scheme needs to be evaluated in conjunction with frequency selective amplification and multi-band amplitude compression.

Chapter 6

SUMMARY AND CONCLUSIONS

6.1 Introduction

Sensorineural hearing loss is characterized by frequency dependent shift in hearing thresholds, loudness recruitment and reduced dynamic range, poor temporal resolution and increased temporal masking, and poor frequency resolution and increased spectral masking (Pickles, 1982; Carney and Nelson, 1983; Glasberg and Moore, 1986; Moore, 1997; Pickett, 1999). Increased spectral masking results into smearing of spectral peaks, adversely affecting the speech perception. Frequency-selective amplification by hearing aids can make the sound audible but may not be very useful in improving the speech perception. Several signal processing techniques for improving speech perception by persons with sensorineural loss have been investigated with varying degree of success.

Since masking takes place primarily at the peripheral level, while integration of binaural information takes place at higher levels in the auditory system, the speech signal can be split into two parts with each part containing complementary spectra for binaural dichotic presentation to the left and right ears. In this scheme, the spectral components likely to mask each other are presented to the different ears, for reducing the effect of spectral masking. Earlier studies on binaural dichotic presentation by spectral splitting of speech signal using a pair of complementary comb filters, for improving speech perception by persons with moderate bilateral sensorineural loss, have shown mixed results: from no advantage to improvements in recognition scores corresponding to an SNR advantage of 2 – 9 dB (Lyregaard, 1982; Lunner et al., 1993; Chaudhari and Pandey, 1998a, b; Cheeran and Pandey, 2004b; Murase et al., 2004). The filters used in the earlier studies had different bandwidths and realizations. The effectiveness of the comb filter based spectral splitting scheme, in reducing the effects of intraspeech spectral masking, depends on the comb filter responses. The comb filters should have (i) nearly flat response in pass bands, (ii) sharp transition bands, and (iii) large attenuation in stop bands. The spectral components in the pass band are presented to the corresponding ear, but those in the transition bands are presented to both the ears. Therefore the two filters should have magnitude responses such that perceived loudness of the spectral components in speech remains balanced. The mixed results reported in earlier studies may be attributed to the magnitude responses of the filters used. Further, it is desirable that the processing and presentation does not result in lateralization of the sound and the source localization ability with binaural hearing aids is not affected. This necessitates an investigation to optimize the comb filter responses for improving speech perception, and to study the effect of dichotic presentation on source localization.

The spectral splitting scheme can be used only for persons with moderate bilateral sensorineural loss and using binaural hearing aids. For monaural hearing, several studies have investigated the usefulness of spectral contrast enhancement schemes for improving the intelligibility of speech in noise for normal-hearing subjects and for subjects with sensorineural loss (Bunnel, 1990; Stone and Moore, 1992; Baer et al., 1993; Miller et al., 1999; Yang et al., 2003; Cohen, 2006). The processing involved enhancement of the spectral prominences which are perceptually significant. There may be errors in identification of the spectral prominences, and increase in the dynamic range of the speech signal may adversely affect the speech perception due to the reduced dynamic range of hearing associated with the sensorineural loss. In the multi-band frequency compression, investigated by Yasu et al. (2002) and Arai et al. (2004), the speech spectrum was divided into a number of bands based on auditory critical bandwidth, and spectral components in each of the bands were compressed towards the center of the corresponding band. The study reported a moderate improvement in speech perception by persons with sensorineural loss. The main advantage of this scheme is that the processing compresses the spectrum within a band and overall spectral shape of the speech signal is largely unaffected. Thus, harmonic structure for the voiced sounds and randomness for the unvoiced sounds is retained with relatively less spectral distortion. It may be possible to further improve the effectiveness of this scheme by investigating the scheme with respect to segmentation for analysis-synthesis, bandwidth, frequency mapping, and compression factor.

The overall objective of the research reported in the thesis was to investigate two speech processing schemes for reducing the effect of intraspeech spectral masking in sensorineural loss: (i) spectral splitting for binaural dichotic presentation, and (ii) multi-band frequency compression for monaural presentation.

In spectral splitting scheme, the perceived loudness of different spectral components in the speech signal should be balanced, especially for the components in transition bands which get presented to both the ears. Listening tests for investigating monaural-binaural loudness balance (Appendix A) showed that the sum of the amplitudes of the left and right tones in binaural presentation being equal to that of the monaural tone resulted in monaural-binaural loudness match. This result indicated that the magnitude response of the comb filters used for dichotic presentation should be complementary on a linear scale. Comb filter pairs based on different constant bandwidths and auditory critical bandwidth were designed with magnitude responses closely satisfying the requirement for perceptual balance. Listening tests were conducted, using modified rhyme test (MRT) for consonant recognition, on normal-hearing subjects with different levels of increased masking simulated by broad-band masking noise and on subjects with moderate bilateral sensorineural loss (Chapter 3). To quantify the effect of spectral splitting on source localization, investigations were carried out by conducting listening tests on normal-hearing and hearing-impaired subjects (Chapter 4). Head related transfer functions (HRTFs) for one of the subject in the CIPIC HRTF database (Algazi et al., 2001; CIPIC HRTF database, 2001), in the frontal azimuth plane, were used to generate spatial sounds for the investigations. Listening tests for studying the source localization were conducted on six normal-hearing subjects in the presence of broad-band masking noise and on 11 subjects with moderate sensorineural loss.

An analysis-synthesis technique for multi-band frequency compression, applied on the complex spectrum using overlap-add method, was implemented and optimized for (a) segmentation for analysis-synthesis, (b) bandwidth, and (c) frequency mapping scheme (Chapter 5). The evaluation was first carried out through listening tests for the quality of the processed speech using mean opinion score (MOS) tests, conducted on normal-hearing subjects in the presence of broad-band masking noise. Further, multi-band frequency compression with optimal processing parameters has been investigated for recognition of consonants, using MRT, on normal hearing subjects with increased masking simulated by broad-band masking noise and on subjects with moderate-to-severe sensorineural loss.

The summary of investigations, conclusions drawn on the basis of the results and some suggestions for further research are given in the following sections.

6.2 Summary of the investigations

The research reported in the thesis involved three investigations which can be summarized as the following.

1) *Comb filters for binaural dichotic presentation*. Perceptually balanced comb filter pairs were designed with different bandwidths: constant bandwidth filters with *n* bands (CB*n*) and filters based on auditory critical bandwidth (ACB). The filters were designed as 513-coefficient linear phase FIR filters with sampling frequency of 10 kHz, using iterative application of frequency sampling technique. The filters had pass band ripple < 1 dB, high stop band attenuation > 30 dB, and small transition bands (< 80 Hz). For constant bandwidth filters with bandwidth less than 350 Hz and filters based on auditory critical bandwidth, listening tests showed no lateralization of the sound for broad-band stimuli.

The effectiveness of spectral splitting using CB18 and ACB filters was assessed by conducting listening tests, for recognition of consonants, using modified rhyme test (MRT). In Exp. IA, the tests were conducted on six normal-hearing subjects with sensorineural loss simulated by adding broad-band masking noise with SNR constant on a short-time basis. Even though no improvement in recognition scores was observed for SNR values higher than 0 dB, improvement in recognition scores was observed for lower SNR values for both the types of filters. The improvement in recognition scores for CB18 were 10, 12, 13, 16, and 18 % for SNR values of -3, -6, -9, -12, and -15 dB, respectively. The corresponding improvements for ACB filters were 14, 18, 22, 25, and 28 %, respectively, and these improvements were statistically significant. At 75 % recognition score, the improvements in recognition scores observed for CB18 and ACB filters were equivalent to an SNR advantage of approximately 6 and 12 dB, respectively. Response time was also recorded and analyzed as an indicator of the load on perception process. These results showed that processing with both the types of filters decreased the response time at lower SNR values. The improvements in response time for CB18 were 0.14, 0.18, 0.18, and 0.04 s for SNR values of -3, -6, -9, and -12 dB, respectively. The corresponding improvements for ACB filters were 0.25, 0.33, 0.26, and 0.12 s, respectively. Thus the scheme of comb filter based spectral splitting, using both CB18 and ACB filters, helped in improving the speech perception. The improvements observed for the spectral splitting using comb filters based on auditory critical bandwidths was higher than those obtained with the constant bandwidth based comb filters.

In Exp. IB, further evaluation of spectral splitting using ACB filters was carried out by conducting MRT on 11 subjects with moderate sensorineural loss, without using any frequency-dependent gain or amplitude compression. All the subjects showed statistically significant improvement in recognition scores in the range 14 – 31 % ($p = 0.001$), indicating that the processing helped in improving speech perception. The processing also resulted in a mean decrease of 0.26 s in the response time ($p < 0.001$).

2) *Effect of spectral splitting on source localization*. Effect of spectral splitting on source localization was investigated by conducting listening tests on six normal-hearing subjects in the presence of broad-band masking noise and 11 subjects with moderate bilateral sensorineural loss. Head related transfer functions (HRTFs) were used to generate spatial sounds in the frontal azimuth plane. In the left/center/right identification tests (Exp. IIA), dichotic presentation resulted in a moderate reduction in the identification scores (less than 20 % at 10°) for broad-band sounds. In the test for left/center/right discrimination thresholds (Exp. IIB), dichotic presentation resulted in an increase in the threshold by less than 10º for the sounds with bandwidth greater than 1/3-octave. In the source direction (in the frontal azimuth plane) identification tests conducted on normal-hearing subjects in the presence of

masking noise (Exp. IIC), a small increase (1º – 8º) in the mean rms error was observed for dichotic condition. In a similar test conducted on subjects with bilateral sensorineural loss (Exp. IID), dichotic presentation resulted in a very small increase in the mean (averaged across the subjects) rms error due to dichotic processing: 1.0° for breaking glass and 0.1° for broad-band noise. Thus the study showed that the broad-band sound sources could be localized during dichotic presentation. The ACB based comb filters had only a small effect on source localization for broadband stimuli, and it may be inferred that the subjects were able to use the binaural cues across the bands for perceiving the source direction.

3) Multi-band frequency compression: The objective of this part of the study was to investigate the scheme of multi-band frequency compression for improving speech perception. The compression was applied on the complex spectrum and hence it did not involve computation of magnitude and phase spectra. MOS tests conducted on normal-hearing subjects with simulated sensorineural loss (Exp. IIIA) showed the highest scores for the compression using pitch-synchronous segmentation with 50 % overlap, spectral segment mapping, and auditory critical bandwidth based analysis bands. The effectiveness of the scheme for improving recognition of consonants was assessed by conducting MRT with consonants in the word-initial and word-final positions. Average response time was also measured to provide an indication of the load on the perception process.

MRT conducted on six normal-hearing subjects with sensorineural loss simulated by adding broad-band masking noise (Exp. IIIB) showed maximum improvement in the recognition score for a compression factor of 0.6. With this compression factor, the mean improvement in recognition scores was 9.7, 12.8, 16.6, 17.1, and 17.2 % for SNR values of -3, -6, -9, and -12 dB, respectively, and the improvements were statistically significant ($p <$ 0.001). At about 60 % recognition score, the improvement in recognition score for compression factor of 0.6 was equivalent to an SNR advantage of 6 dB. The processing helped in reducing the response time with maximum reduction observed for compression factor of 0.6. With this compression factor, the mean improvement in response time was 0.43, 0.55, 0.88, and 0.81 s for SNR values of -3, -6, -9, and -12 dB, respectively. The listening tests conducted on eight subjects with moderate-to-severe sensorineural loss (Exp. IIIC) showed a moderate increase in recognition scores in the range 1 – 8 % for compression factor of 0.8 for six out of eight subjects. For compression factor of 0.6, all the subjects showed an increase in recognition scores in the range 9 to 21 %. For compression factor of 0.4, there was an improvement of 3 – 16 % for four subjects. The processing resulted in a decrease in response time for all the subjects with mean of 0.10, 0.89, and 0.57 s for compression factors of 0.8, 0.6, and 0.4, respectively. As in the case of normal-hearing subjects, maximum improvement in recognition scores and response time was observed for compression factor of 0.6.

6.3 Conclusions

The research involved investigations on (i) spectral splitting for binaural dichotic presentation for persons using binaural hearing aids, and (ii) multi-band frequency compression for persons using monaural hearing aids.

For binaural dichotic presentation, the comb filter pairs used in the present study were designed to closely satisfy the requirement for perceptual balance. Listening tests conducted on normal-hearing subjects with sensorineural loss simulated by masking noise showed that both types of filter pairs (CB18 and ACB) improved speech perception at SNR values lower than 0 dB and the improvements were higher for ACB filters than those with CB18 filters. At 75 % recognition scores, the improvements in consonant recognition scores was equivalent to an SNR advantage of 6 and 12 dB for CB18 and ACB filters, respectively. Listening tests on hearing-impaired subjects with moderate-to-severe bilateral loss were conducted using only the ACB filters. These tests showed the improvements in the range 14 – 31 % with a mean of 22 % in recognition score and a mean decrease in response time of 0.26 s. Investigations were also carried out to study the effect of dichotic presentation on source localization. Even though source localization of narrowband sounds was moderately affected by dichotic presentation, the subjects were able to perceive direction for broadband sounds, by using the binaural cues across the bands.

For monaural presentation, the scheme of multi-band frequency compression was optimized with respect to segmentation for analysis-synthesis, bandwidth, and frequency mapping through MOS tests. Multi-band frequency compression applied on the complex spectrum using pitch-synchronous segmentation, auditory critical bandwidths, and spectral segment mapping, was found to be optimal. Listening tests for speech intelligibility showed maximum improvement for a compression factor of 0.6. The tests with normal-hearing subjects with sensorineural loss simulated by masking noise showed an SNR advantage of 6 dB. For the hearing-impaired subjects, improvement in recognition scores was in the range 9 – 21 % with a mean of 16.5 % and the mean reduction in the response time was 0.89 s.

The conclusions based on the results of the investigations may be summarized as the following

(i) Dichotic presentation using a pair of comb filters with magnitude responses complementary on a linear scale improved speech perception for normal-hearing persons with simulated loss and for persons with moderate bilateral sensorineural loss. Filters based on auditory critical bandwidth resulted in a higher improvement than those based on constant bandwidth. It was further found that subjects were able to localize the broad-band sounds using binaural cues across the bands. The technique may be useful for improving speech perception for persons with moderate bilateral sensorineural loss, who can use binaural hearing aids.

(ii) For monaural presentation, multi-band frequency compression using compression on complex spectrum with pitch-synchronous segmentation, auditory critical bandwidths, and spectral segment mapping improved speech perception for normal-hearing persons with simulated loss and for persons with moderate sensorineural loss. A compression factor of 0.6 resulted in maximum improvement. The technique may be useful for improving speech perception for persons with sensorineural loss using monaural hearing aids.

The pattern of improvement in recognition scores and response times across the individual subjects, in both the schemes, did not show any specific relation to the audiograms or to the scores for the unprocessed speech. This may be because although all the subjects had moderate sensorineural loss, the extent of masking across them may be very different. In our listening tests, the order of presentation with different processing conditions was randomized across the subjects in order to minimize the bias due to practice or fatigue. Tests conducted on a larger number of subjects to evaluate the improvements due to processing and tests to measure the extent of masking may help in identifying the group most likely to benefit by the processing.

6.4 Suggestions for future work

In our study on spectral splitting and multi-band frequency compression, listening tests were conducted for consonant identification using modified rhyme test. Both the schemes need to be evaluated using different types of test material and a larger number of subjects with different types of loss characteristics. These studies will help in establishing an optimal choice of bandwidths used in the comb filters for binaural dichotic presentation and an optimal value of compression factor for multi-band frequency compression, which may vary across the subjects depending upon the extent of increase in spectral masking.

The study was carried out for evaluating the schemes in reducing the effect of intraspeech masking. Their effectiveness in improving speech perception in the presence of noise needs to be evaluated. Further, both the schemes need to be evaluated by incorporating frequency-selective gain and multi-band amplitude compression in accordance with the loss characteristics of the individual subjects.

Appendix A

PERCEPTUAL BALANCE
IN BINAURAL PRESENTATION

A.1 Introduction

Several studies have investigated binaural dichotic presentation using spectral splitting for improving speech perception by persons with moderate bilateral sensorineural hearing loss (Lyregaard, 1982; Lunner et al., 1993; Lunner, 1997; Chaudhari and Pandey, 1998a, 1998b; Cheeran and Pandey, 2004b; Murase et al., 2004). A pair of comb filters, with complementary magnitude responses, are used to present alternate bands to left and right ears. As the filters used in these studies had linear phase responses, the variations in the results reported may be attributed to the different magnitude responses. The comb filters used for spectral splitting should have a small ripple in the pass band and a large attenuation in the stop band. As filters have finite transition bands between pass and stop bands, the spectral components of the speech signal in the transition bands are presented to both the ears. The perceived loudness of different spectral components in the speech signal should be balanced, especially for components in the transition bands which get presented to both the ears. Therefore the two filters should have magnitude responses such that perceived loudness for spectral components in the transition bands is the same as that in the pass bands.

The objective of the investigation presented in this appendix is to study the perceptual balance in binaural hearing, i.e., finding a relationship between the signal amplitudes in the left and the right ear in binaural presentation which will evoke the same loudness as a monaural presentation. Earlier studies on binaural level difference for equal loudness (BLDEL), and binaural summation of loudness are reviewed in the next section. Subsequent sections describe the investigation on perceptual balance.

A.2 Loudness of binaural presentation

Several studies comparing the loudness of binaurally and monaurally presented sounds have been reported (Scharf, 1968; Marks, 1978; Hall and Harvey, 1985; Hawkins et al., 1987; Zwicker and Henning, 1991; Epstein and Florentine, 2005; Whilby et al., 2006).

A study by Scharf (1968) reported the binaural level difference for equal loudness (BLDEL) to be about 5 dB at low presentation levels, 7 dB at moderate presentation levels, and about 6 dB at high presentation levels. In another study by Scharf (1969), involving dichotic presentation of two tones, the subjects perceived two distinct auditory images. There was no change in the perceived loudness for the tones, when the frequency separation between the tones presented to the two ears was varied over a wide frequency range. In the investigation by Hall and Harvey (1985) involving presentation of 2 kHz pure tone at 70 and 80 dB SPL, the BLDEL was found to be 3 – 4 dB for hearing-impaired subjects and 8 – 9 dB for normal-hearing subjects. For 2 kHz tone, at 90 dB, both the groups had BLDEL of about 9 dB. For tone of 500 Hz, presented at the three levels, the BLDEL was about 9 dB for both the groups. In a study by Hawkins et al. (1987) using 4 kHz pure tone with a presentation level ranging within listener's most comfortable level to the discomfort level, the BLDEL for impaired listeners was in the range of 5 – 12 dB and it was not significantly different from that for listeners with normal hearing.

Zwicker and Henning (1991) conducted listening tests to match the loudness of monaurally and binaurally presented tone bursts. The presentation consisted of four binaural bursts of 60 ms each alternating with four similar bursts presented monaurally. The subjects were asked to adjust the level of one type of bursts to match the perceived loudness of the other type. The test tones were of frequency 250 Hz, 710 Hz, and 2 kHz. The binaural presentation had same intensity in both the ears. An increment of 10 dB was needed for the monaural sound to match the perceived loudness of the binaural sound. This difference was almost the same across the three test frequencies and presentation levels. Similar experiment was conducted by adding the stimuli to low pass filtered noise with cutoff frequency of 840 Hz, 1.5 kHz, and 4 kHz for the test tone of 250 Hz, 710 Hz, and 2 kHz respectively. The stimuli were added with (i) in-phase noise and (ii) out-of- phase noise and then presented binaurally with varying inter-aural phase difference. For both the noise conditions, the inter-aural phase difference had a significant effect on perceived loudness for 250 Hz, a moderate effect for 710 Hz, and minimal effect for 2 kHz tone bursts.

Cheeran (2005) conducted listening tests to find the difference in levels of monaural and binaural presentations, such that they evoke the same perceived loudness. The stimuli used were four pure tones (0.25, 1, 2, and 4 kHz) of 1 s duration each, sustained vowel /a/, and broad-band noise. Five normal-hearing subjects participated in the listening tests. The stimulus was presented monaurally and binaurally, one after the other, with an inter-stimulus interval of 1 s. The monaural intensity was fixed at 85 dB and binaural intensity was varied in the range 70 – 84 dB. The task of the subject was to mark each binaural sound as "high", "same", or "low" depending on the perceived loudness with respect to the monaural sound. The results showed that the perceived loudness matched when binaural level was 4 – 12 dB lower than the monaural level. Whilby et al. (2006) investigated BLDEL for normal and hearing-impaired listeners using 1 kHz pure tone of 5 ms and 200 ms duration. They used loudness matching procedure: (i) monaural level was fixed and binaural level was varied and (ii) binaural level was fixed and monaural level was varied for equal loudness. The fixed level ranged from 10 to 90 dB SL. The BLDEL for normal-hearing subjects ranged from 2 to 15 dB, and for hearing-impaired listeners it was 1.5 to 12 dB.

Marks (1978) investigated binaural summation of loudness using pure tone stimuli of frequency 0.1, 0.4, and 1 kHz. A set of nine SPLs were used for left and right ears, with a total of 81 combinations of binaural stimuli. Fourteen normal-hearing subjects participated in the test for estimation of perceived loudness. The study showed a linear additivity of the numerical responses for loudness, for all the test tone frequencies. In the same study (Marks, 1978), an experiment was conducted to obtain a set of equal loudness curves at four presentation levels, using 1 kHz test tone for finding various combinations of sound pressure levels to the left and right ears that produced a given level of loudness. The standard tone was presented binaurally, with equal intensity in both the ears, at the presentation levels of 20, 30, 40, or 50 dB SPL. The variable tone was also binaural, set to give a fixed intensity ratio at the two ears. However, in every match, subjects controlled the absolute levels of the left or the right ear components. During matching process, the stimulus sequence was continuous: the standard tone of 1 s duration, 1 s of silence, variable tone of 1s duration, 1 s of silence and so on. The subject matched the loudness of the variable tone to that of the standard tone, by controlling either the left or right ear components. Three normal-hearing subjects participated in the listening test. Figure A.1 shows the equal loudness curves obtained for four presentation levels of the standard tone. The shape of the curves for different presentation level was same with a small inter-subject variation. A monaural sound needed to be 5 – 7 dB above the binaural sound for it to evoke same loudness as that of binaural.

The main objective of most of these studies was to find BLDEL, for different stimuli and presentation levels, for normal-hearing and hearing-impaired listeners. However, for the design of comb filters with perceptually balanced response in the transition band, we need to know the relation between the gains of the two filters (for left and right ears) such that there are no irregular variations in the perceived loudness of spectral components in the transition bands. Therefore the aim of our investigation is to study the relation between the two amplitude scaling factors for binaural presentation which will result in a match of the loudness of the binaural presentation to that of the monaural presentation.

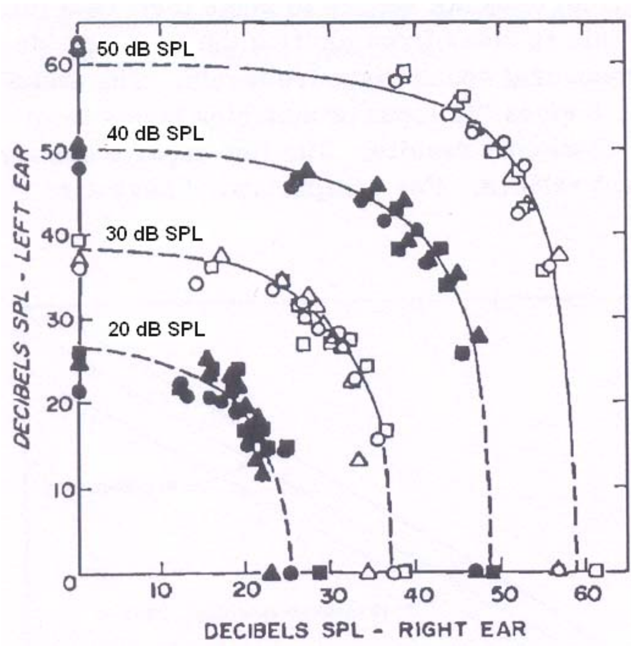


Fig. A.1 Equal loudness curves obtained for four presentation levels: 20. 30, 40, and 50 dB SPL, as reported by Marks (1978).

A.3 Experimental method

As shown in Fig. A.2, the input signal was scaled by scaling factor α for the left ear and by β for the right ear for presentation through a pair of headphones. Signal amplitudes were scaled to compensate for any unbalance in the response of the two headphones at the test tone frequencies.

The listening tests were carried out for obtaining the relation between the amplitude scaling factors, for the left and the right ears, so that the binaural presentation evokes the same loudness as that of the monaural presentation. The overall investigation involved two experiments. In the first experiment (Exp. I), perceptual balance was investigated for pure tones of frequencies 250 Hz, 500 Hz, 1 kHz and 2 kHz, presented at the most comfortable level (MCL) for the individual listener. The second experiment (Exp. II) was conducted to examine the effect of presentation level on perceptual balance. The test involved tone of 500 Hz presented at three presentation levels: MCL – 6 dB, MCL, and MCL + 6 dB. The values of α and β were in the range 0 – 1, in steps of 0.1.

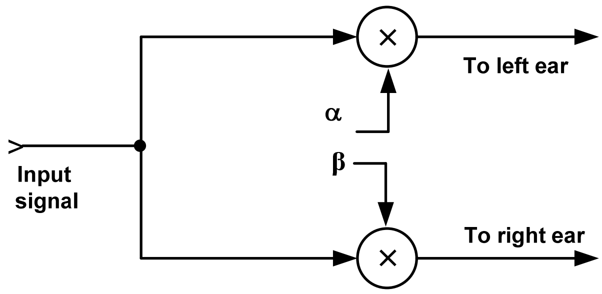


Fig. A. 2 Scheme for perceptual balance test

The listening tests were conducted using three-interval, three-alternative forced choice (3I-3AFC) paradigm (Kortekaas and Kohlrausch, 1999). Each presentation had three observation intervals: reference (monaural), test (binaural), and reference (monaural), separated by 0.5 s silences. Depending on whether the perceived loudness of the binaural sound was lower than, equal to, or higher than that of the monaural sound, the subject marked the response as L, E, or H on the response sheet. The subject could listen to the sounds more than once before finalizing the response. Combinations of α and β in binaural presentation were selected randomly.

In Exp. I, there were a total of 484 presentations for each subject: 4 test frequencies × 11 values of α × 11 values of β. A total of eight normal-hearing subjects participated in the listening tests. In Exp. II, there were a total of 363 presentations for each subject: 3 presentation levels × 11 values of α × 11 values of β. These tests were conducted on six normal-hearing subjects. In both the experiments, a mean of values of β which correspond to a monaural-binaural loudness balance was calculated, for each value of α.

A.4 Results and discussion

The results of Exp. I are summarized in Table A.1. It gives the β values (averaged across the eight subjects) for perceptual balance for each of the values, for the four frequencies. The standard deviations (given in parentheses) are small, indicating only a small inter-subject variation in the β values. Figure A.3 shows a β vs. α plot. The plots indicate an approximately linear relationship for all the four frequencies.

The results of Exp. II are given in Table A.2. It gives the β values obtained for perceptual balance, for three presentation levels: MCL – 6 dB, MCL, and MCL + 6 dB. The standard deviations are given in parentheses. Figure A.4 gives a plot of values of α vs. β, obtained for perceptual balance at different presentation levels. For all the three presentation levels, the plots indicate an approximately linear relationship. A plot of scaling factors on dB scale, with MCL as the reference, is shown in Fig. A.5. The shape of the curves for loudness balance is similar to those reported in the study by Marks (1978) as shown in Fig A.1.

Table A.1 Exp. I: Mean values (s. d. in parentheses, $n = 8$) of β, obtained for perceptual balance, for four test tone frequencies.

α	Frequency			
	250 Hz	500 Hz	1 kHz	2 kHz
0.0	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
0.1	0.91 (0.04)	0.89 (0.06)	0.89 (0.08)	0.91 (0.05)
0.2	0.86 (0.05)	0.83 (0.08)	0.82 (0.08)	0.78 (0.07)
0.3	0.76 (0.07)	0.71 (0.07)	0.76 (0.07)	0.70 (0.11)
0.4	0.63 (0.08)	0.57 (0.09)	0.64 (0.09)	0.61 (0.12)
0.5	0.46 (0.09)	0.45 (0.06)	0.55 (0.10)	0.50 (0.12)
0.6	0.37 (0.07)	0.40 (0.09)	0.45 (0.11)	0.41 (0.15)
0.7	0.29 (0.10)	0.27 (0.06)	0.36 (0.14)	0.28 (0.11)
0.8	0.20 (0.06)	0.20 (0.06)	0.21 (0.09)	0.19 (0.09)
0.9	0.13 (0.04)	0.09 (0.03)	0.11 (0.06)	0.12 (0.08)
1.0	0.10 (0.03)	0.08 (0.03)	0.06 (0.05)	0.07 (0.06)

Table A.2 Exp. II: Mean values (s.d. in parentheses, $n = 6$) of β, obtained for perceptual balance, for three presentation levels. Test tone frequency: 500 Hz.

α	Presentation level		
	MCL – 6 dB	MCL	MCL + 6 dB
0.0	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
0.1	0.88 (0.08)	0.91 (0.06)	0.91 (0.07)
0.2	0.78 (0.06)	0.83 (0.10)	0.85 (0.08)
0.3	0.65 (0.06)	0.72 (0.08)	0.70 (0.10)
0.4	0.59 (0.11)	0.58 (0.10)	0.59 (0.07)
0.5	0.48 (0.13)	0.46 (0.06)	0.48 (0.05)
0.6	0.30 (0.13)	0.41 (0.10)	0.36 (0.11)
0.7	0.19 (0.10)	0.26 (0.07)	0.23 (0.06)
0.8	0.06 (0.06)	0.20 (0.07)	0.04 (0.04)
0.9	0.02 (0.03)	0.10 (0.03)	0.02 (0.03)
1.0	0.00 (0.00)	0.09 (0.02)	0.02 (0.03)

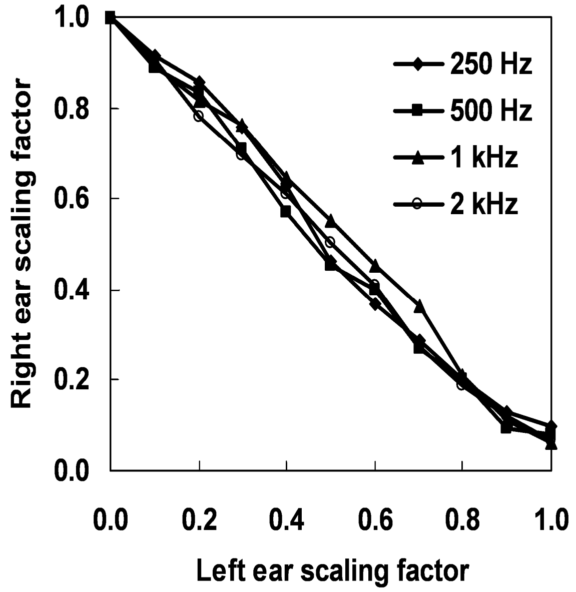


Fig. A.3 Exp. I: Relation between two scaling factors (α and β) for perceptual balance, shown for four test tone frequencies, presented at MCL.

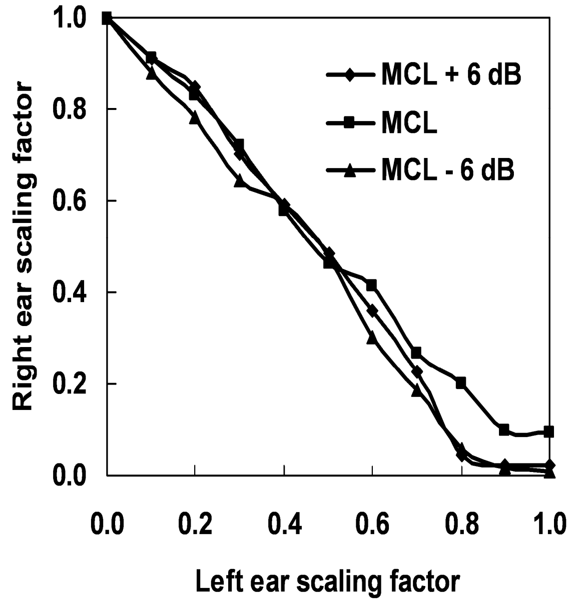


Fig. A.4 Exp. II Relation between two scaling factors (α and β) for perceptual balance, for 500 Hz tone at three presentation levels.

Earlier studies have shown that loudness generally grows as a power function of sound pressure (Stevens, 1956; Reynolds and Stevens, 1960; Stevens and Guirao, 1964; Scharf and Fishken, 1970). Assuming that the loudness of binaural sound is a power law summation of the two individual sounds (Fletcher and Munson, 1933; Hellman and Zwislocki, 1963; Marks, 1978), the scaling factors α and β should have following relationship for perceptual balance,

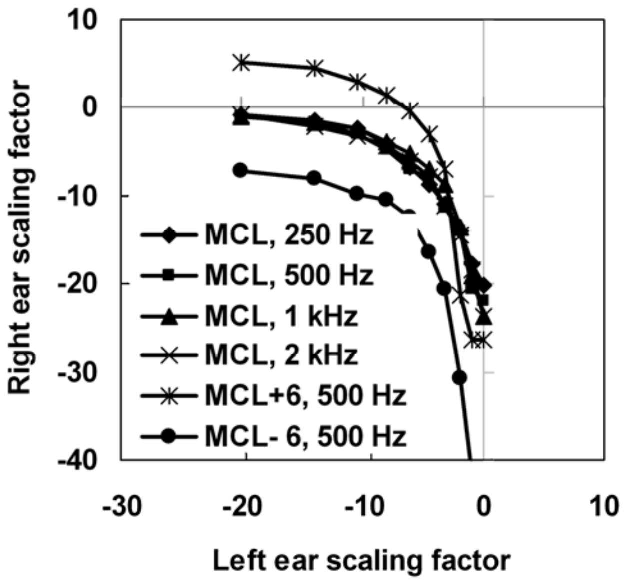


Fig. A.5 Relation between the two scaling factors (α and β) on dB scale, for perceptual balance, for three presentation levels and four frequencies.

$$(\alpha)^{\rho} + (\beta)^{\rho} = 1 \tag{A.1}$$

where, ρ is the power relating amplitude to the loudness. For finding an approximate fit to the observed values of β, its values were computed for different values of ρ (0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6, 1.8, and 2), from Eqn. A.1. Table A.3 gives the observed values of β for 500 Hz test tone, the computed values of β, and the approximation error. Similar analysis was carried out for all the four test tone frequencies. Figure A.6 shows the RMS error, in the approximation of β using the power law addition model, as a function of ρ. For all the four frequencies, minimum error was observed for $\rho \approx 1$, indicating that perceptual balance is achieved by $\alpha + \beta \approx 1$.

Various ranges for the BLDEL have been reported in the earlier investigations: 5 – 7 dB (Scharf, 1969), 8 – 9 dB (Hall and Harvey, 1985), 5 – 12 dB (Hawkins et al., 1987), 2 – 15 dB (Whilby et al., 2006), and 4 – 12 dB (Cheeran, 2006). In the current investigation, perceptual balance is obtained for the binaural sound when the two amplitude scaling factors are nearly linearly related which means a BLDEL of about 6 dB. Thus, the BLDEL obtained in the current investigation nearly falls in the middle of the various ranges of BLDEL reported in the previous investigations.

Table A. 3 Error in approximation of β (observed, and computed from Eqn. A.1 for different value of ρ), for each α, in perceptual balance experiment. Test tone frequency: 500 Hz.

α	β (obs.)	β, computed for different values of ρ								
		0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0
0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.10	0.89	0.28	0.62	0.81	0.90	0.95	0.97	0.98	0.99	1.00
0.20	0.83	0.16	0.45	0.67	0.80	0.88	0.92	0.95	0.97	0.98
0.30	0.71	0.09	0.33	0.55	0.70	0.80	0.86	0.91	0.93	0.95
0.40	0.57	0.05	0.24	0.44	0.60	0.71	0.79	0.85	0.89	0.92
0.50	0.45	0.03	0.17	0.34	0.50	0.62	0.71	0.78	0.83	0.87
0.60	0.40	0.01	0.11	0.26	0.40	0.52	0.62	0.69	0.75	0.80
0.70	0.27	0.01	0.06	0.18	0.30	0.42	0.51	0.59	0.66	0.71
0.80	0.20	0.00	0.03	0.10	0.20	0.30	0.39	0.47	0.54	0.60
0.90	0.09	0.00	0.01	0.04	0.10	0.17	0.24	0.31	0.38	0.44
1.00	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
max. error		0.67	0.38	0.16	0.08	0.17	0.26	0.33	0.39	0.44
rms error		0.42	0.26	0.11	0.03	0.11	0.17	0.23	0.27	0.31

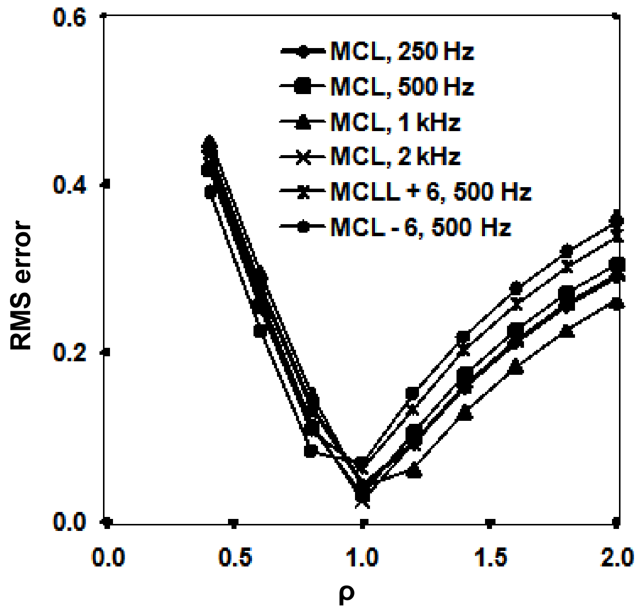


Fig. A.6 RMS error in approximation of β as computed from Eqn. A.1 shown as a function of ρ.

It may thus be concluded that perceptual balance in comb filter based spectral splitting scheme is maintained when the gains of the pair of comb filters are complementary on linear scale. This result is particularly significant with respect to the design of comb filters with perceptually balanced responses, especially in the transition bands.

Appendix B

LATERALIZATION IN BINAURAL DICHOTIC PRESENTATION

B.1 Introduction

Lateralization is a process in which perceived loudness in one ear tends to dominate the other and it describes the apparent location of sound source within the head (Noble and Byrne, 1990; Moore, 1997; Brungart and Rabinowitz, 1999). In comb filter based spectral splitting, spectral components in a band get presented to only one ear, causing the possibility of lateralization of the sound, which may offset the benefit obtained by dichotic presentation and may have adverse effect on source localization. This appendix presents our investigation on sound lateralization for the two types of comb filters for binaural dichotic presentation: filters with (i) constant bandwidths with number of bands varying from 2 – 18 (CB*n*) and (ii) auditory critical bandwidths (ACB).

B.2 Methodology

The input signal is processed using a pair of comb filters, to get the two outputs to be presented to the left and the right ears through a pair of headphones for binaural dichotic presentation. The filters for processing were designed and implemented for a sampling frequency of 10 kHz. Two types of comb filters were used: (i) constant bandwidth based comb filters with different number of bands (CB*n*), and (ii) comb filters based on auditory critical bandwidths (ACB). The comb filters were designed as 513-tap linear phase FIR filters, using an iterative application of frequency sampling technique (Rabiner et al., 1970; Rabiner and Gold, 1975; Rabiner and Schafer, 1978; Proakis and Monolakis, 1992). The investigation presented in Appendix A showed that the magnitude responses of the two comb filters should be complementary on a linear scale, and the filters used in this study were designed accordingly.

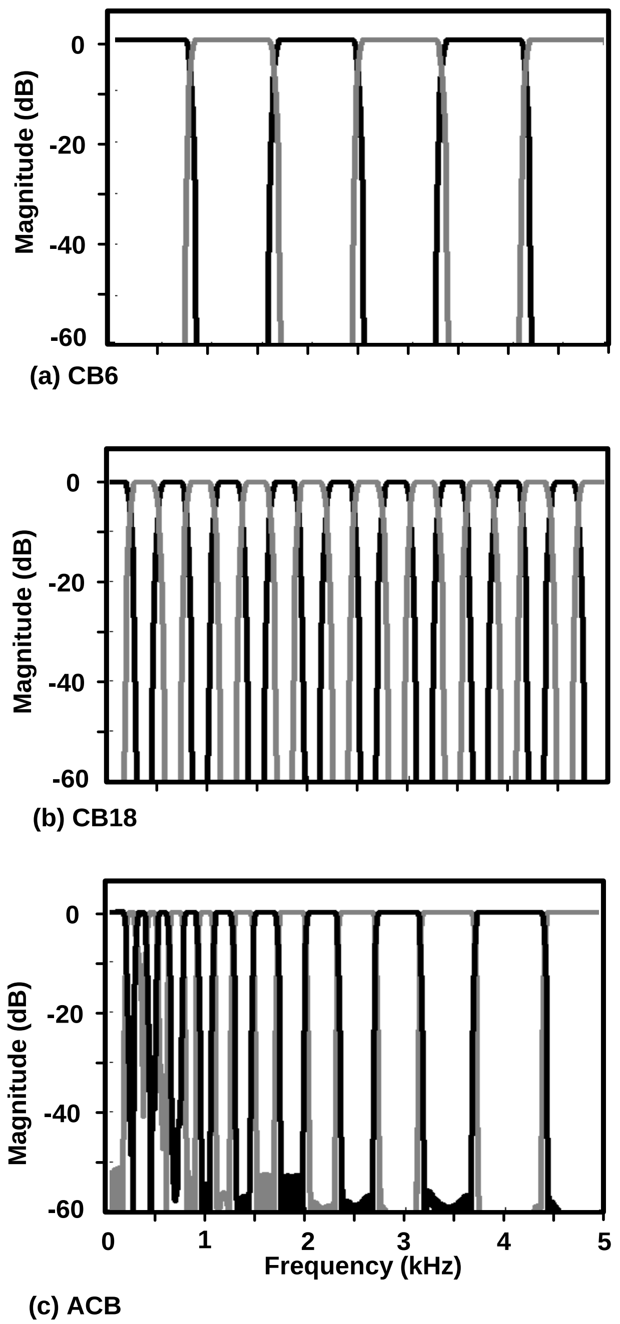


Fig. B.1 Overlapped magnitude response of comb filter pairs: (a) CB6, (b) CB18, and (c) ACB. The black and gray traces correspond to the responses of the left and right filters. The black and gray traces show the magnitude responses for the left and right filters, respectively.

The constant bandwidth filters were designed having the total number of bands n = 2, 4, .., 18, with bandwidth given as 5/n kHz. These filters had pass band ripple of less than 1 dB, the minimum stop band attenuation of 64 dB and inter-band crossover gains in the range of -5 to - 6

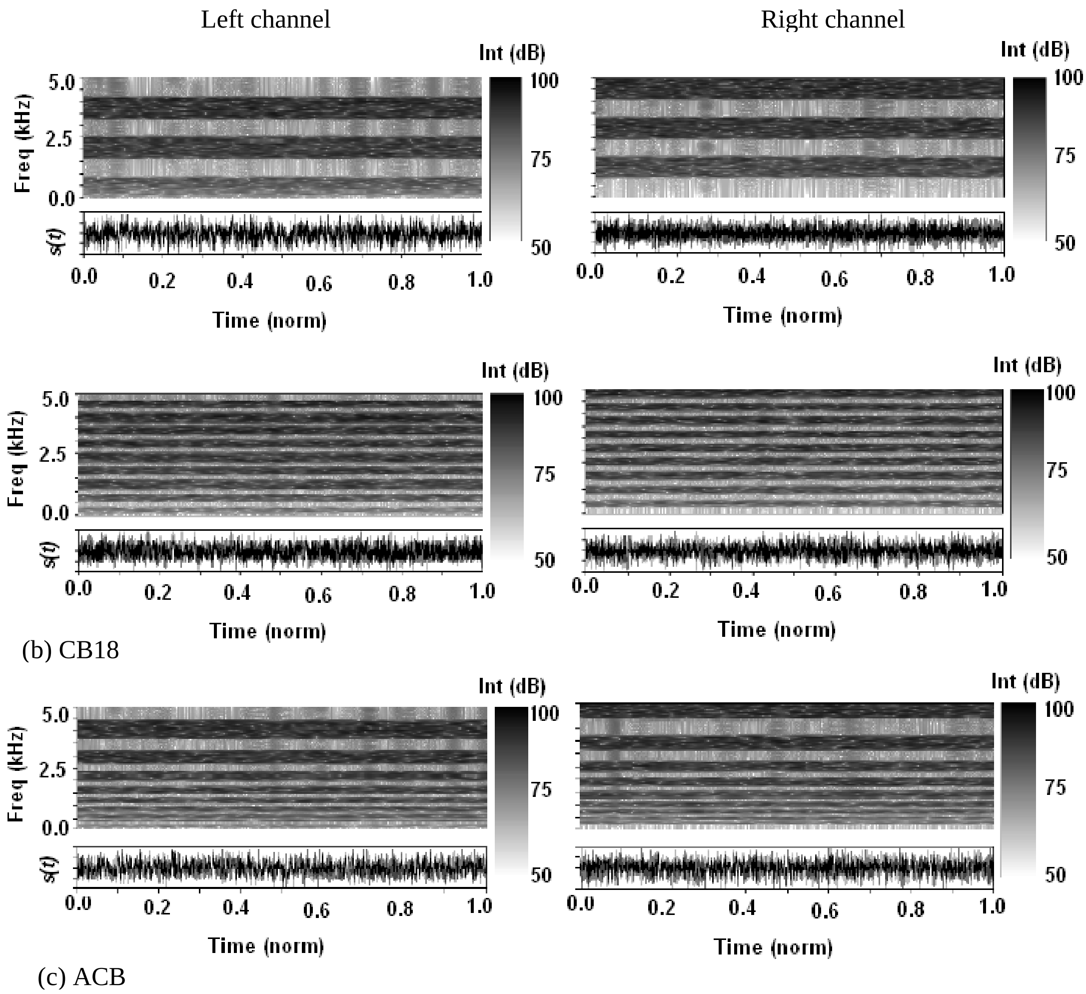


Fig. B.2 Narrow band spectrograms of broadband noise (500 ms, 10 k samples/s) processed using different pairs of comb filters: (a) CB6, (b) CB18, (c) ACB.

dB. The magnitude responses of constant bandwidth filters with 6 and 18 bands (CB6 and CB18) are shown in Fig. B.1(a) and (b). Auditory critical bandwidth (ACB) based comb filters, with the magnitude response shown in Fig. B.1(c), had pass band ripple less than 1 dB, minimum stop band attenuation of 29 dB, and inter-band crossover gain in the range of -4 to -6 dB. Figure B.2 shows the narrow band spectrograms of broad-band noise, processed using filters with different bandwidths. These spectrograms show that the signal presented to the two ears have complementary bands with minimal inter-band overlap.

All the filters were assessed through informal listening tests, using six normal-hearing subjects, for perceptual balance of the perceived loudness, when the signal switches between the two ears due to change in the frequency in the inter-band crossovers. Sinusoidal tones with the frequency linearly swept (i) from 50 Hz to 5 kHz over 40 s, (ii) from 100 Hz to 300 Hz over 30 s, and (iii) from 3 kHz to 3.5 kHz over 30 s, were used as the test stimuli. For each presentation, the subjects responded verbally as either “yes” or “no”, depending on whether they noted any change in the perceived loudness when the signal switched between the two ears during the sweeping of the tone frequency. For all the comb filter pairs and test tones, no change in the perceived loudness was observed.

To study the lateralization effect in spectral splitting scheme, listening tests were conducted on six subjects with normal hearing. The test material included vowels /*a*/, /*i*/, /*u*/, and the sentence “*we were away a year ago*”. The test stimuli were recorded from a male speaker in an audiometry room, using B&K microphone model 2210, at 10 kHz of sampling frequency and with 16-bit quantization. Subjects were asked to use a 0 – 10 loudness scale, and split the total loudness of 10 for the loudness perceived in the two ears. Thus the number assigned to an ear gave a measure of its perceived loudness with respect to that in the other ear. For each subject, there were a total of 40 presentations (4 stimuli × 10 filters). Correspondence between the left and the right filter outputs and the headphones were maintained across the subjects.

B.3 Results

The numerical values of the perceived loudness in the two ears (on the 0 – 10 scale), averaged across the six subjects, are given in Table B.1. As the results for the three vowels were similar, the numerical values of loudness were averaged. The table also shows the difference between loudness values for the left and right ears and the standard deviation of the differences across the subjects. In case of constant bandwidth filters with small number of bands, the perceived loudness in the left ear dominated that in the other ear. This dominance of the perceived loudness in the left ear can be attributed to the fact that the low frequency band is always presented to the left ear. However, the subjects perceived the same loudness in both the ears when the number of bands in the constant bandwidth filters was 16 or higher. One-tailed paired t-test for the loudness values between the two ears showed statistically significant (p < 0.001) difference for constant bandwidth filters with number of bands lesser than 16. The differences were statistically not significant for CB16, CB18 and ACB filters.

Table B.1 Perceived loudness in the two ears (averaged across the six subjects)

Filter pair	Test material					
	Vowels			Sentence		
	Left	Right	L-R (s.d)	Left	Right	L-R (s.d)
CB2	9.9	0.1	9.8 (0.7) ††	9.7	0.3	9.4 (0.8) ††
CB4	9.7	0.3	9.4 (0.8) ††	9.3	0.7	8.6 (0.8) ††
CB6	8.5	1.5	7.0 (2.2) ††	8.4	1.6	6.8 (0.8) ††
CB8	7.7	2.3	5.4 (1.2) ††	6.7	3.3	3.4 (0.5) ††
CB10	6.5	3.5	3.0 (0.8) ††	6.1	3.9	2.9 (0.6) ††
CB12	6.4	3.6	2.8 (0.9) ††	5.9	4.1	1.8 (0.8) ††
CB14	6.0	4.0	2.0 (0.7) ††	5.7	4.3	1.4 (0.5) ††
CB16	5.1	4.9	0.2 (0.5) †	5.0	5.0	0.0 (0.0)
CB18	5.0	5.0	0.0 (0.0)	5.0	5.0	0.0 (0.0)
ACB	5.0	5.0	0.0 (0.0)	5.0	5.0	0.0 (0.0)

† $p < 0.05$ †† $p < 0.001$

B.4 Discussion

The objective of the investigation, presented in this appendix, was to study the lateralization in spectral splitting using comb filter pairs based on different constant bandwidths and auditory critical bandwidth. In order to assess the lateralization with these filters, listening tests were conducted on six normal-hearing subjects using vowels and a sentence as the test material. The subjects rated the perceived loudness in the two ears on a 0 – 10 scale, such that the sum of the numerical responses for the two ears equaled 10. In case of constant bandwidth filters, filter bandwidths larger than 350 Hz (number of bands less than 16) resulted in lateralization of sounds. There was no lateralization for filters based on auditory critical bandwidth. Thus in order to avoid lateralization of the sound source, the comb filters should either have constant bandwidth of less than 350 Hz or the filters should be based on auditory critical bandwidth.

Appendix C

TEST MATERIAL FOR
MODIFIED RHYME TEST (MRT)

The 300 CVC words used as the test material for the MRT (House, et al., 1965; ANSI, 1989) conducted for consonant recognition are listed in Table C.1 and C.2. The words are arranged as 50 groups, each group consisting of six rhyming words. The words in each group in the first table have different consonants in the initial position, while those in the second table differ in the final consonants. Some of the words in both the lists have double consonants in the final positions.

Table C.1: MRT words groups with different initial consonants

Sl. No	Six rhyming words					
01	went	sent	bent	dent	tent	rent
02	hold	cold	told	fold	sold	gold
03	kit	bit	fit	hit	wit	sit
04	must	bust	gust	rust	dust	just
05	bed	led	fed	red	wed	shed
06	pin	sin	tin	fin	din	win
07	not	tot	got	pot	hot	lot
08	vest	test	rest	best	west	nest
09	way	may	say	pay	day	gay
10	pig	big	dig	wig	rig	fig
11	shop	mop	cop	top	hop	pop
12	coil	oil	soil	toil	boil	foil
13	same	name	game	tame	came	fame
14	peel	reel	feel	eel	keel	heel
15	hark	dark	mark	bark	park	lark
16	thaw	law	raw	paw	jaw	saw
17	pen	hen	men	then	den	ten
18	heat	neat	feat	seat	meat	beat
19	dip	sip	hip	tip	lip	rip
20	hang	san	bang	rang	fang	gang
21	took	cook	look	hook	shook	book
22	fill	kill	will	hill	till	bill
23	bale	gale	sale	tale	pale	male
24	wick	sick	kick	lick	pick	tick
25	fun	sun	bun	gun	run	nun

Table C.2: MRT words groups with different final consonants

Sl. No	Six rhyming words					
01	pat	pad	pan	path	pack	pass
02	lane	lat	late	lake	lace	lame
03	teak	team	teal	teach	tear	tease
04	din	dill	dim	dig	dip	did
05	dug	dung	duck	dud	dub	dun
06	sum	sun	sung	sup	sub	sud
07	seep	seen	seethe	seek	seem	seed
08	pig	pill	pin	pip	pit	pick
09	back	bath	bad	bass	bat	ban
10	pale	pace	page	pane	pay	pave
11	cane	case	cape	cake	came	cave
12	tan	tang	tap	tack	tam	tab
13	fit	fib	fizz	fill	fig	fin
14	heave	hear	heat	heal	heap	heath
15	cup	cut	cud	cuff	cuss	cud
16	puff	puck	pub	pus	pup	pun
17	bean	beach	beat	beak	bead	beam
18	kill	kin	kit	kick	king	kid
19	mass	math	map	mat	man	mad
20	ray	raze	rate	rave	rake	race
21	save	same	sale	sane	sake	safe
22	sill	sick	sip	sing	sit	sin
23	peace	peas	peak	peach	peat	peal
24	bun	bus	but	bug	buck	buff
25	sag	sat	sass	sack	sad	sap

Appendix D

TEST INSTRUCTIONS AND FORMS

D.1 Introduction

Before the commencement of the listening test session, the subject was verbally informed about the research aimed at developing speech processing schemes for improving speech perception by persons with sensorineural hearing loss. He/she was then given a sheet with test instructions to read, and subsequently these instructions were also verbally explained. The tests were conducted after the subject agreed to participate and signed the consent form.

D.2 Instructions for MRT

1. You will be seated in front of a computer terminal with a mouse to click the appropriate button on the screen. The level of the sound presented will be adjusted to the most comfortable level for you. You will have a trial test to become familiar with the procedure and sounds. Be relaxed and attentive throughout the test.
2. The display on the screen will show the following buttons
    - *"Play"* button – to listen to the sound
    - *Response panel with six words* – to mark your choice
    - *"Next"* button – to move to the next presentation
3. For every presentation, you will hear a sentence "W*ould you write* --" followed by a word to be recognized by you.
4. Your task is to click "play" button to listen to the sound presented over the headphone, and then click the best matching word amongst the six similarly rhyming words displayed on the response panel. The sound is presented once only and if you can not recognize the word, you can guess.

5. Click the “next” button for the next presentation. This procedure will be repeated until all the 50 words in the selected test list have been presented.
6. There are a total of 6 test lists and you will have a break of about 5 minutes in between the tests.

D.3 Instructions for localization tests

Exp. I: Left/center/right identification

1. You will be seated in a quiet room. The level of the sound presented will be adjusted to the most comfortable level for you. You will have a trial test to become familiar with the procedure and sounds. Be relaxed and attentive throughout the test.
2. For every presentation, you will hear a sound through the pair of headphones, processed to change the apparent direction of the sound source in a random order.
3. Your task is to identify the perceived direction of the sound as left, center, or right. The sound is presented once only and if you cannot identify the direction, you can guess.
4. The above procedure (2, 3) will be repeated until the sound for all the 19 angles are presented ten times. The duration of the test session will be approximately 30 min.
5. A total of 26 test sessions will be conducted, depending on your convenience and willingness with not more than two test sessions in a day.

Exp. II: Left/center/right discrimination threshold

1. You will be seated in a quiet room. The level of the sound presented will be adjusted to the most comfortable level for you. You will have a trial test to become familiar with the procedure and sounds. Be relaxed and attentive throughout the test.
2. For every presentation, you will hear a sound through the pair of headphones, processed to change the apparent direction of the sound source.
3. Your task is to identify the perceived direction of the sound as left, center or right. The sound is presented once only and if you cannot identify the direction, you can guess.
4. The above procedure (2, 3) will be repeated until the two thresholds of discrimination (left/center and right/center) are obtained. The duration of the test session will be approximately 15 – 20 min.
5. A total of 26 test sessions will be conducted depending on your convenience and willingness with not more than three test sessions in a day.

Exp. III: Source direction identification

1. You will be seated in a quiet room. The level of the sound presented will be adjusted to the most comfortable level for you. You will have a trial test to become familiar with the procedure and sounds. Be relaxed and attentive throughout the test.
2. For every presentation, you will hear a sound through the pair of headphones, processed to change the apparent direction of the sound source.
3. Your task is to identify the perceived direction of the sound as one of the seven choices, as indicated in the chart displayed in front of you. The sound is presented once only and if you cannot identify the direction, you can guess.
4. The above procedure (2, 3) will be repeated until all the seven angles are presented five times. The duration of the test session will be approximately 15 min.
5. The test sessions will be conducted depending on your convenience and willingness with not more than three test sessions in a day.

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

SUBJECT BACKGROUND INFORMATION

Date __/__/____

Name ________________________________ Code ____________

Address ____

____

Phone (   ) ___ _______

____

Sex _______ Age _____________

Occupation: ____

Place of birth: ____

First language: ____

Other languages: ____

Handedness: Left / Right

History of noise exposure: ______________________________

History of hearing problems: ______________________________

____

Other remarks: ____

D.5 Form for subject's willingness to participate

CONSENT FORM

I have carefully read the test instructions provided by P. N. Kulkarni (Ph.D. Scholar, IIT Bombay) for participation in listening experiments for evaluation of speech processing schemes. I am willing to participate in the tests conducted by him. I understand that I can discontinue the participation at any time and that the data obtained from the tests will be used for research without identifying me.

Signature: ______________________________

Name: ______________________________

Address: ______________________________

____

Date: ______________________________

Appendix E

STATISTICAL ANALYSIS OF TEST RESULTS

E.1 Comb filters for binaural dichotic presentation

Table E.1 Comb Filter Exp. IA: Two-way repeated-measures ANOVA on recognition scores (%), with processing (unprocessed diotic, dichotic CB18, dichotic ACB) and SNR (9 levels) as the sources of variation.

Source	df	SS	MS	*F*	*p*
Subject	5	554.8	110.9		
Proc.	2	4544.4	2272.2	101.48	< 0.001
SNR	8	21005.0	2625.6	271.5	< 0.001
Proc. × SNR	16	3054.3	190.9	40.98	< 0.001
Error					
Proc.	10	223.9	22.4		
SNR	40	386.8	9.7		
Proc. × SNR	80	372.6	4.6		
Total	161	30142			

Table E.2 Comb Filter Exp. IA: One-way repeated-measures ANOVA on recognition scores (%), with processing (unprocessed diotic, dichotic CB18, dichotic ACB) as the source of variation, conducted separately at each SNR value.

SNR	Source	df	MS	*F*	*p*
∞	Subject	5	7.48		
	Proc.	2	0.807	0.44	n.s.
	Error	10	1.80		
6	Subject	5	16.35		
	Proc.	2	0.68	0.33	n.s.
	Error	10	2.07		
3	Subject	5	19.77		
	Proc.	2	10.72	2.17	n.s.
	Error	10	4.92		
0	Subject	5	21.35		
	Proc.	2	73.76	23.9	<0.001
	Error	10	3.08		
-3	Subject	5	8.1		
	Proc.	2	303.9	51.2	<0.001
	Error	10	5.94		

SNR	Source	df	MS	*F*	*p*
-6	Subject	5	22.74		
	Proc.	2	509.5	88.1	< 0.001
	Error	10	5.78		
-9	Subject	5	21.25		
	Proc.	2	756.7	186.7	< 0.001
	Error	10	4.05		
-12	Subject	5	37.26		
	Proc.	2	940.62	53.2	< 0.001
	Error	10	17.68		
-15	Subject	5	34.0		
	Proc.	2	1202.6	84.1	< 0.001
	Error	10	14.3		

Table E.3 Comb Filter Exp. IA: Paired comparison of recognition scores (%), using Tukey's HSD test, based on the ANOVA results in Table E.2, for different SNR values. The significance level (p) is in parentheses below the difference in % score. $q|_{p=0.05}$ = 3.67, $q|_{p=0.01}$ = 4.84.

SNR = ∞ HSD = 2.01, p = 0.05 2.65, p = 0.01		
Proc.	CB18	ACB
Unp.	0.3 (n.s.)	-0.4
CB18		0.7 (n.s.)

SNR = 6 dB HSD = 2.15, p = 0.05 2.84, p = 0.01		
Proc	CB18	ACB
Unp.	-0.2	0.4 (n.s.)
CB18		0.6 (n.s.)

SNR = 3 dB HSD = 3.32, p = 0.05 4.38, p = 0.01		
Proc	CB18	ACB
Unp.	-0.9	1.7 (n.s.)
CB18		2.6 (n.s.)

SNR = 0 dB HSD = 2.63, p = 0.05 3.47, p = 0.01		
Proc.	CB18	ACB
Unp.	4.2 (0.01)	6.9 (0.01)
CB18		2.7 (0.05)

SNR = -3 dB HSD = 3.65, p = 0.05 4.81, p = 0.01		
Proc	CB18	ACB
Unp.	10.2 (0.01)	13.7 (0.01)
CB18		3.5 (n.s.)

SNR = -6 dB HSD = 3.60, p = 0.05 4.75, p = 0.01		
Proc	CB18	ACB
Unp.	12.3 (0.01)	18.1 (0.01)
CB18		5.8 (0.01)

SNR = -9 dB HSD = 3.01, p = 0.05 3.97, p = 0.01		
Proc.	CB18	ACB
Unp.	13.3 (0.01)	22.4 (0.01)
CB18		9.1 (0.01)

SNR = -12 dB HSD = 6.30, p = 0.05 8.31, p = 0.01		
Proc	CB18	ACB
Unp.	16.2 (0.01)	24.6 (0.01)
CB18		8.4 (0.01)

SNR = -15 dB HSD = 5.66, p = 0.05 7.47, p = 0.01		
Proc	CB18	ACB
Unp.	17.9 (0.01)	28.0 (0.01)
CB18		10.1 (0.01)

Table E.4 Comb Filter Exp. IA: Two-way repeated-measures ANOVA on response times (s), with processing (unprocessed diotic, dichotic CB18, dichotic ACB) and SNR (9 levels) as the sources of variation.

Source	df	SS	MS	F	p
Subject	5	55.29	11.06		
Proc.	2	0.54	0.27	1.48	ns
SNR	8	9.39	1.17	15.87	< 0.001
Proc. × SNR	16	0.47	0.028	0.93	ns
Error					
Proc.	10	1.81	0.181		
SNR	40	2.96	0.074		
Proc. × SNR	80	2.52	0.03		
Total	161	72.99			

Table E.5 Comb Filter Exp. IA: One-way repeated-measures ANOVA on response times (s), with processing (unprocessed diotic, dichotic CB18, dichotic ACB) as the source of variation, conducted separately at each SNR value.

SNR	Source	df	MS	*F*	*p*
∞	Subject	5	1.27		
	Proc.	2	0.007	0.06	n.s.
	Error	10	0.11		
6	Subject	5	1.10		
	Proc.	2	0.003	0.01	n.s.
	Error	10	0.03		
3	Subject	5	1.5		
	Proc.	2	0.06	1.57	n.s.
	Error	10	0.04		
0	Subject	5	1.52		
	Proc.	2	0.04	1.21	n.s.
	Error	10	0.03		
-3	Subject	5	1.11		
	Proc.	2	0.097	2.23	n.s.
	Error	10	0.043		

SNR	Source	df	MS	*F*	*p*
-6	Subject	5	1.31		
	Proc.	2	0.17	5.93	<0.02
	Error	10	0.02		
-9	Subject	5	1.14		
	Proc.	2	0.11	3.99	<0.05
	Error	10	0.03		
-12	Subject	5	1.27		
	Proc.	2	0.02	0.27	n.s.
	Error	10	0.07		
-15	Subject	5	1.41		
	Proc.	2	0.01	0.203	ns
	Error	10	0.056		

Table E.6 Comb Filter Exp. IA: Paired comparison of response times (s), using Tukey's HSD test, based on the ANOVA results in Table E.5, for different SNR values. The significance level (p) is in parentheses below the difference in response time values (s). $q|_{p=0.05} = 3.67$, $q|_{p=0.01} = 4.84$.

SNR = ∞ HSD = 0.50, *p* = 0.05 0.65, *p* = 0.01		
Proc.	CB18	ACB
Unp.	-0.04	-0.06
CB18		-0.13

SNR = 6 dB HSD = 0.26, *p* = 0.05 0.34, *p* = 0.01		
Proc	CB18	ACB
Unp.	-0.01	0.00 (n.s.)
CB18		0.01 (n.s.)

SNR = 3 dB HSD = 0.30, *p* = 0.05 0.39, *p* = 0.01		
Proc	CB18	ACB
Unp.	0.15 (n.s.)	0.19 (n.s.)
CB18		0.04 (n.s.)

SNR = 0 dB HSD = 0.26, *p* = 0.05 0.34, *p* = 0.01		
Proc.	CB18	ACB
Unp.	0.06 (n.s.)	0.15 (n.s.)
CB18		0.09 (n.s.)

SNR = -3 dB HSD = 0.31, *p* = 0.05 0.41, *p* = 0.01		
Proc	CB18	ACB
Unp.	0.14 (n.s.)	0.25 (n.s.)
CB18		0.12 (n.s.)

SNR = -6 dB HSD = 0.21, *p* = 0.05 0.28, *p* = 0.01		
Proc	CB18	ACB
Unp.	0.18 (n.s.)	0.33 (0.01)
CB18		0.29 (0.01)

SNR = -9 dB HSD = 0.26, *p* = 0.05 0.34, *p* = 0.01		
Proc.	CB18	ACB
Unp.	0.18 (n.s.)	0.26 (0.05)
CB18		0.13 (n.s.)

SNR = -12 dB HSD = 0.40, *p* = 0.05 0.52, *p* = 0.01		
Proc	CB18	ACB
Unp.	0.04 (n.s.)	0.12 (n.s.)
CB18		0.12 (n.s.)

SNR = -15 dB HSD = 0.35, *p* = 0.05 0.47, *p* = 0.01		
Proc	CB18	ACB
Unp.	-0.05	0.04 (n.s.)
CB18		0.01 (n.s.)

E.2 Multi-band frequency compression

Table E.7 Multi-band compression Exp. IIIB: Two-way repeated-measures ANOVA on recognition scores (%), with processing (c = 1.0, 0.8, 0.6, 0.4) and SNR (9 levels) as the sources of variation.

Source	df	SS	MS	*F*	*p*
Subject	5	1673.5	334.7		
Proc.	3	1606.1	535.4	12.26	< 0.001
SNR	8	28637	3579.6	338.6	< 0.001
Proc. × SNR	24	3747.7	156.15	22.61	< 0.001
Error					
Proc.	15	655.2	43.68		
SNR	40	422.8	10.57		
Proc. × SNR	120	828.7	6.9		
Total	215	37571			

Table E.8 Multi-band compression Exp. IIIB: One-way repeated-measures ANOVA on recognition scores (%), with processing (c = 1.0, 0.8, 0.6, 0.4) as the source of variation, conducted separately at each SNR value.

SNR	Source	df	MS	*F*	*p*
∞	Subject	5	15.98		
	Proc.	3	185.34	49.22	<0.001
	Error	15	3.76		
6	Subject	5	38.51		
	Proc.	3	230.67	42.73	<0.001
	Error	15	5.39		
3	Subject	5	48.29		
	Proc.	3	96.67	11.87	<0.001
	Error	15	8.14		
0	Subject	5	53.95		
	Proc.	3	89.64	9.37	<0.001
	Error	15	9.56		
-3	Subject	5	71.15		
	Proc.	3	101.47	12.48	<0.001
	Error	15	8.13		

SNR	Source	df	MS	*F*	*p*
-6	Subject	5	61.71		
	Proc.	3	172.95	10.10	<0.001
	Error	15	17.11		
-9	Subject	5	74.7		
	Proc.	3	285.37	31.88	<0.001
	Error	15	8.95		
-12	Subject	5	35.21		
	Proc.	3	295.89	13.91	<0.001
	Error	15	21.27		
-15	Subject	5	19.15		
	Proc.	3	326.59	19.68	<0.001
	Error	15	14.3		

Table E.9 Multi-band compression Exp. IIIB: Paired comparison of recognition scores (%), using Tukey's HSD test, based on the ANOVA results in Table E.8, for different SNR values. The significance level (p) is in parentheses below the difference in % score. $q|_{p=0.05}$ = 3.96, $q|_{p=0.01}$ = 5.02.

SNR = ∞ HSD = 3.13, p = 0.05 3.97, p = 0.01			
Proc.	0.8	0.6	0.4
1.0	-1.4 (n.s.)	-7.3 (0.01)	-12.0 (0.01)
0.8		5.9 (0.01)	10.6 (0.01)
0.6			4.7 (0.01)

SNR = 6 dB HSD = 3.75, p = 0.05 4.76, p = 0.01			
Proc.	0.8	0.6	0.4
1.0	-1.7 (n.s.)	-8.5 (0.01)	-13.3 (0.01)
0.8		6.8 (0.01)	12.4 (0.01)
0.6			4.8 (0.01)

SNR = 3 dB HSD = 4.61, p = 0.05 5.85, p = 0.01			
Proc.	0.8	0.6	0.4
1.0	-3.3 (n.s.)	-7.2 (0.01)	-8.8 (0.01)
0.8		3.9 (n.s.)	5.5 (0.05)
0.6			1.6 (n.s.)

SNR = 0 dB HSD = 5.00, p = 0.05 6.34, p = 0.01			
Proc.	0.8	0.6	0.4
1.0	3.9 (n.s.)	9.2 (0.01)	2.7 (n.s.)
0.8		5.3 (0.05)	1.2 (n.s.)
0.6			6.5 (0.01)

SNR = -3 dB HSD = 4.61, p = 0.05 5.38, p = 0.01			
Proc.	0.8	0.6	0.4
1.0	2.9 (n.s.)	9.7 (0.01)	3.2 (n.s.)
0.8		6.8 (0.01)	0.3 (n.s.)
0.6			6.5 (0.01)

SNR = -6 dB HSD = 6.69, p = 0.05 8.48, p = 0.01			
Proc.	0.8	0.6	0.4
1.0	4.0 (n.s.)	12.8 (0.01)	7.0 (0.05)
0.8		18.8 (0.01)	3.0 (n.s.)
0.6			5.8 (n.s.)

SNR = -9 dB HSD = 4.84, p = 0.05 6.13, p = 0.01			
Proc.	0.8	0.6	0.4
1.0	5.6 (0.05)	16.6 (0.01)	8.5 (0.01)
0.8		11.0 (0.01)	3.1 (n.s.)
0.6			8.1 (0.01)

SNR = -12 dB HSD = 7.45, p = 0.05 9.45, p = 0.01			
Proc.	0.8	0.6	0.4
1.0	7.9 (0.05)	17.1 (0.01)	8.2 (0.05)
0.8		9.2 (0.05)	0.3 (n.s.)
0.6			8.9 (0.05)

SNR = -15 dB HSD = 6.58, p = 0.05 8.35, p = 0.01			
Proc.	0.8	0.6	0.4
1.0	7.9 (0.05)	17.2 (0.01)	3.9 (n.s.)
0.8		9.3 (0.01)	4.0 (n.s.)
0.6			13.3 (0.01)

Table E.10 Multi-band compression Exp. IIIB: Two-way repeated-measures ANOVA on response time values, with processing (c = 1.0, 0.8, 0.6, 0.4) and SNR (9 levels) as the sources of variation.

Source	df	SS	MS	F	p
Subject	5	38.6	7.7		
Proc.	3	7.97	2.66	1.92	n.s.
SNR	8	0.8	0.1	0.84	n.s.
Proc. × SNR	24	5.98	0.25	5.44	< 0.001
Error					
Proc.	15	20.72	1.38		
SNR	40	4.78	0.12		
Proc. × SNR	120	5.49	0.05		
Total	215	84.37			

Table E.11 Multi-band compression Exp. IIIB: One-way repeated-measures ANOVA on response times (s), with processing (unprocessed c = 1.0, 0.8, 0.6, 0.4) as the source of variation, conducted separately at each SNR value.

SNR	Source	df	MS	F	p
∞	Subject	5	0.57		
	Proc.	3	0.63	2.33	n.s.
	Error	15	0.27		
6	Subject	5	0.67		
	Proc.	3	0.30	1.16	n.s.
	Error	15	0.26		
3	Subject	5	0.87		
	Proc.	3	0.26	1.23	n.s.
	Error	15	0.21		
0	Subject	5	0.78		
	Proc.	3	0.26	1.16	n.s.
	Error	15	0.23		
-3	Subject	5	0.86		
	Proc.	3	0.32	1.57	n.s.
	Error	15	0.20		
-6	Subject	5	0.80	5.77	
	Proc.	3	0.46	3.36	<0.04
	Error	15	0.14		
-9	Subject	5	1.32	11.53	
	Proc.	3	0.99	8.67	<0.001
	Error	15	0.11		
-12	Subject	5	1.46	11.75	
	Proc.	3	0.86	6.9	<0.003
	Error	15	0.12		
-15	Subject	5	1.34	6.7	
	Proc.	3	0.55	2.86	n.s.
	Error	15	0.19		

Table E.12 Multi-band compression Exp. IIIB: Paired comparison of response time (s), using Tukey's HSD test, based on the ANOVA results in Table E.11, for different SNR values. The significance level (p) is in parentheses below the difference in response time values (s). $q|_{p=0.05}$ = 3.96, $q|_{p=0.01}$ = 5.02.

SNR = ∞
HSD = 0.84, p = 0.05
1.06, p = 0.01

Proc.	0.8	0.6	0.4
1.0	-0.42 (n.s.)	-0.66 (n.s.)	-0.71 (n.s.)
0.8		0.24 (n.s.)	0.29 (n.s.)
0.6			0.05 (n.s.)

SNR = 6 dB
HSD = 0.82, p = 0.05
1.06, p = 0.01

Proc.	0.8	0.6	0.4
1.0	0.08 (n.s.)	0.01 (n.s.)	-0.41 (n.s.)
0.8		0.07 (n.s.)	0.49 (n.s.)
0.6			0.42 (n.s.)

SNR = 3 dB
HSD = 0.74, p = 0.05
0.94, p = 0.01

Proc.	0.8	0.6	0.4
1.0	0.07 (n.s.)	0.02 (n.s.)	-0.39 (n.s.)
0.8		0.05 (n.s.)	0.46 (n.s.)
0.6			0.41 (n.s.)

SNR = 0 dB
HSD = 0.77, p = 0.05
0.98, p = 0.01

Proc.	0.8	0.6	0.4
1.0	0.20 (n.s.)	0.22 (n.s.)	-0.23 (n.s.)
0.8		0.02 (n.s.)	0.43 (n.s.)
0.6			0.45 (n.s.)

SNR = -3 dB
HSD = 0.72, p = 0.05
0.92, p = 0.01

Proc.	0.8	0.6	0.4
1.0	0.33 (n.s.)	0.43 (n.s.)	-0.03 (n.s.)
0.8		0.1 (n.s.)	0.36 (n.s.)
0.6			0.46 (n.s.)

SNR = -6 dB
HSD = 0.60, p = 0.05
0.76, p = 0.01

Proc.	0.8	0.6	0.4
1.0	0.32 (n.s.)	0.55 (n.s.)	-0.05 (n.s.)
0.8		0.23 (n.s.)	0.37 (n.s.)
0.6			0.6 (0.05)

SNR = -9 dB
HSD = 0.54, p = 0.05
0.68, p = 0.01

Proc.	0.8	0.6	0.4
1.0	0.54 (0.05)	0.88 (0.01)	0.12 (n.s.)
0.8		0.34 (n.s.)	0.42 (n.s.)
0.6			0.76 (0.01)

SNR = -12 dB
HSD = 0.56, p = 0.05
0.71, p = 0.01

Proc.	0.8	0.6	0.4
1.0	0.59 (0.05)	0.81 (0.01)	0.13 (n.s.)
0.8		0.22 (n.s.)	0.46 (n.s.)
0.6			0.68 (0.05)

SNR = -15 dB
HSD = 0.70, p = 0.05
0.89 p = 0.01

Proc.	0.8	0.6	0.4
1.0	0.59 (n.s.)	0.49 (n.s.)	0.05 (n.s.)
0.8		0.10 (n.s.)	0.13 (n.s.)
0.6			0.23 (n.s.)

Table E.13 Multi-band compression Exp. IIIC: One-way repeated-measures ANOVA on recognition scores (%), with processing (c = 1.0, 0.8, 0.6, 0.4) as the source of variation

Source	df	SS	MS	F	p
Subject	7	1003.5	143.35		
Proc.	3	1396.2	465.4	21.88	<0.001
Error	21	446.6	21.27		

Table E.14 Multi-band compression Exp. IIIC: One-way repeated-measures ANOVA on response times (s), with processing (c = 1.0, 0.8, 0.6, 0.4) as the source of variation

Source	df	SS	MS	F	p
Subject	7	2.23	0.32		
Proc.	3	4.08	1.36	15.98	<0.001
Error	21	1.79	0.08		

Table E.15 Multi-band compression Exp. IIIC: Paired comparison of (a) recognition scores (%) and (b) response times (s) using Tukey's HSD test, based on the ANOVA results in Table E.13 and E.14, respectively. The significance level (p) is in parentheses. $q|_{p=0.05}$ = 3.87, $q|_{p=0.01}$ = 4.85.

(a) Recognition score
HSD = 6.31, p = 0.05
7.91, p = 0.01

Proc.	0.8	0.6	0.4
Unp.	1.5 (n.s.)	16.5 (0.01)	2.7 (n.s.)
0.8		15.0 (0.01)	1.2 (n.s.)
0.6			13.8 (0.01)

(b) Response time
HSD = 0.39, p = 0.05
0.48, p = 0.01

Proc.	0.8	0.6	0.4
Unp.	0.1 (n.s.)	0.89 (0.01)	0.57 (0.01)
0.8		0.79 (0.01)	0.48 (0.05)
0.6			0.32 (n.s.)

Appendix F

COMB FILTER COEFFICIENTS

The coefficients of the 513-coefficient filters in the CB18 and ACB comb filter pairs are given in Table F.1 and Table F.2, respectively. The corresponding impulse responses are plotted in Fig. F.1 and Fig. F.2, respectively.

Table F.1: CB18 comb filter pair: filter coefficients ($b_0 – b_{512}$) for the left and right ear filters.

Left ear filter				Right ear filter			
b_0 - b_{63}	b_{64} - b_{127}	b_{128} - b_{191}	b_{192} - b_{255}	b_0 - b_{63}	b_{64} - b_{127}	b_{128} - b_{191}	b_{192} - b_{255}
1.2813e-4	3.7927e-4	-4.9381e-3	-7.7912e-3	-9.7154e-5	3.8413e-6	5.7214e-3	1.0521e-2
-1.6650e-4	-5.3532e-5	1.6194e-3	1.9885e-2	1.9435e-4	4.2655e-4	-6.9715e-4	-1.6908e-2
1.9651e-4	1.7189e-3	2.1405e-3	-4.4447e-3	-1.7270e-4	-1.3751e-3	-1.1104e-3	7.5512e-3
4.6792e-4	1.3951e-6	2.5862e-3	-2.6224e-3	-4.4926e-4	2.9426e-4	-1.4874e-3	5.7203e-3
5.0253e-4	4.6138e-4	3.3066e-3	-5.8612e-4	-4.9021e-4	-2.3086e-4	-2.1836e-3	3.5236e-3
2.3479e-5	1.3906e-3	5.8928e-3	-1.7517e-2	-1.8773e-5	-1.2392e-3	-4.7926e-3	2.0140e-2
5.4724e-4	2.1103e-3	3.3035e-3	-2.1811e-2	-5.5136e-4	-2.0477e-3	-2.2729e-3	2.3973e-2
4.9602e-4	7.0250e-4	-3.7098e-3	-4.5634e-3	-5.1004e-4	-7.3336e-4	4.6267e-3	6.1407e-3
2.4474e-6	3.5746e-6	5.8749e-3	2.9003e-3	-2.7198e-5	-1.2709e-4	-5.1113e-3	-2.0002e-3
3.7918e-4	-8.9218e-4	-1.7453e-3	-1.8912e-2	-4.1513e-4	6.8232e-4	2.3216e-3	1.9082e-2
3.7805e-5	-9.6648e-4	-4.2520e-3	-2.2099e-2	-8.4948e-5	6.8164e-4	4.6132e-3	2.1529e-2
2.5981e-4	-3.7926e-4	3.7228e-3	-2.2592e-2	-3.1761e-4	3.5055e-5	-3.5981e-3	2.1316e-2
1.0131e-4	5.4353e-5	4.3259e-3	-3.0173e-4	-1.6860e-4	-4.3915e-4	-4.4534e-3	-1.6083e-3
-1.7446e-4	5.6704e-4	-6.3470e-4	-5.4024e-3	9.9469e-5	-9.7181e-4	2.4537e-4	2.9646e-3
-5.2386e-5	-2.7064e-3	-9.2881e-3	1.0931e-2	-2.7866e-5	2.3027e-3	8.6334e-3	-1.3767e-2
2.3977e-4	-1.4355e-3	-1.0669e-2	-1.1179e-3	-3.2228e-4	1.0531e-3	9.7515e-3	-1.9728e-3
-7.0798e-5	-2.1023e-3	-7.0013e-3	-2.3897e-2	-1.0511e-5	1.7592e-3	5.8328e-3	2.0699e-2
-1.9724e-4	-1.7246e-3	2.1808e-3	2.4021e-2	1.2087e-4	1.4354e-3	-3.5813e-3	-2.7180e-2
3.3197e-4	5.5167e-4	-1.7715e-3	2.5264e-2	-3.9956e-4	-7.7618e-4	1.6877e-4	-2.8250e-2
1.3684e-4	2.1410e-4	1.5140e-3	2.6089e-2	-1.9193e-4	-3.6778e-4	-3.2777e-3	-2.8784e-2
1.1125e-4	3.0775e-3	2.2123e-3	2.2808e-2	-1.5045e-4	-3.1588e-3	-4.0840e-3	-2.5109e-2
-5.9853e-4	3.3508e-3	3.4904e-3	2.8628e-2	5.7808e-4	-3.3623e-3	-5.4053e-3	-3.0454e-2
-5.1679e-4	-7.3292e-4	4.1433e-3	4.2942e-2	5.1730e-4	7.8470e-4	-6.0264e-3	-4.4233e-2
-8.2967e-4	-4.2032e-3	-3.1731e-3	5.3135e-2	8.5256e-4	4.3093e-3	1.4049e-3	-5.3853e-2
-9.8796e-4	-9.1502e-4	4.6139e-3	1.8288e-2	1.0338e-3	1.0648e-3	-6.1791e-3	-1.8417e-2
-3.6385e-4	-1.5964e-3	-4.0468e-3	9.9852e-4	4.3251e-4	1.7788e-3	2.7729e-3	-5.4598e-4
-1.2484e-3	-2.8476e-3	-5.1140e-3	2.7821e-2	1.3388e-3	3.0525e-3	4.2153e-3	-2.6816e-2
-1.0235e-3	-2.4032e-3	-3.5795e-3	-1.4122e-2	1.1338e-3	2.6223e-3	3.1298e-3	1.5628e-2
3.1707e-4	-1.2426e-3	-1.2063e-2	-2.7550e-2	-1.8944e-4	1.4696e-3	1.2120e-2	2.9486e-2
1.2421e-3	1.7717e-3	-1.0156e-2	-4.7105e-2	-1.1005e-3	-1.5404e-3	1.0758e-2	4.9384e-2
8.0569e-4	-1.9935e-4	7.5575e-3	-1.9005e-2	-6.5413e-4	4.3335e-4	-6.3988e-3	2.1525e-2
5.9960e-4	2.0412e-3	1.2315e-2	-2.3037e-2	-4.4289e-4	-1.8046e-3	-1.0617e-2	2.5690e-2
-1.2590e-6	-3.2309e-4	2.1852e-3	-3.4860e-2	1.5766e-4	5.6250e-4	3.3035e-6	3.7534e-2
-4.4211e-5	-1.6855e-3	9.3624e-3	3.7320e-2	1.9424e-4	1.9272e-3	-6.7632e-3	-3.4735e-2
3.8990e-4	8.5897e-4	5.5058e-3	3.5155e-2	-2.5275e-4	-6.1712e-4	-2.6054e-3	-3.2761e-2
-4.3830e-5	1.7416e-4	9.7468e-3	3.1393e-2	1.6142e-4	6.2992e-5	-6.6796e-3	-2.9282e-2
9.2951e-4	-9.0501e-4	4.5985e-3	3.5365e-2	-8.3805e-4	1.1299e-3	-1.5174e-3	-3.3614e-2
6.3303e-4	-1.6461e-3	2.9200e-3	1.4265e-2	-5.7372e-4	1.8485e-3	1.2601e-5	-1.2939e-2
3.3086e-4	-2.3624e-3	4.6557e-3	-5.3728e-2	-3.0875e-4	2.5304e-3	-2.0330e-3	5.4583e-2
-3.1644e-4	-3.0236e-3	-8.8788e-3	-3.9453e-2	2.9771e-4	3.1446e-3	1.1043e-2	3.9804e-2
-4.2149e-4	2.6332e-3	-3.6471e-3	-4.7706e-2	3.6006e-4	-2.5710e-3	5.2258e-3	4.7538e-2
1.4054e-4	1.3870e-3	-4.4129e-4	-3.0422e-2	-2.4450e-4	-1.3929e-3	1.3422e-3	2.9736e-2
-7.8045e-4	3.3172e-3	1.1690e-2	5.4029e-2	6.3625e-4	-3.3973e-3	-1.1519e-2	-5.5215e-2
-7.3063e-4	3.8791e-3	1.2587e-2	1.7865e-2	5.5052e-4	-4.0352e-3	-1.3154e-2	-1.9514e-2
-1.7339e-4	-1.0843e-3	4.2222e-4	2.3337e-2	-3.6506e-5	8.5472e-4	-1.6889e-3	-2.5391e-2
-3.1735e-4	1.2102e-3	-7.2093e-3	-6.2437e-2	8.5191e-5	-1.5068e-3	5.3214e-3	6.0056e-2
-1.2036e-3	3.4388e-3	4.6923e-3	-7.5571e-2	9.5766e-4	-3.7929e-3	-7.0885e-3	7.2960e-2
-5.7671e-6	6.7253e-3	-3.3874e-4	-1.2539e-2	-2.4519e-4	-7.1257e-3	-2.4287e-3	9.8112e-3
-4.1522e-4	1.8020e-3	-1.3379e-2	4.4348e-2	1.6814e-4	-2.2370e-3	1.0389e-2	-4.7068e-2
-3.7723e-4	1.5056e-3	2.7729e-3	4.6180e-2	1.4250e-4	-1.9638e-3	-5.8348e-3	-4.8764e-2
9.3773e-4	5.8466e-3	-6.0013e-3	-8.1926e-2	-1.1522e-3	-6.3177e-3	3.0067e-3	7.9603e-2
-3.9504e-5	4.6179e-3	5.4793e-3	-6.0048e-2	-1.4755e-4	-5.0926e-3	-8.2853e-3	5.8094e-2
4.7134e-5	-9.6318e-5	1.2145e-2	4.4162e-2	-2.0017e-4	-3.7278e-4	-1.4664e-2	-4.5659e-2
-3.4179e-4	8.2741e-4	2.1601e-3	4.5553e-2	2.2875e-4	-1.2814e-3	-4.3191e-3	-4.6536e-2
5.8305e-4	2.0482e-3	-1.0737e-2	-1.2786e-1	-6.5065e-4	-2.4757e-3	8.9894e-3	1.2742e-1
4.1971e-4	-3.7432e-3	-1.7206e-2	3.7283e-2	-4.3702e-4	3.3563e-3	1.5902e-2	-3.7189e-2
-5.3038e-4	-2.0617e-3	-2.5930e-5	1.9913e-2	5.6743e-4	1.7327e-3	-8.1608e-4	-1.9320e-2
-3.6645e-4	-8.2756e-3	-1.8161e-3	-1.0117e-1	4.6082e-4	8.0251e-3	1.4469e-3	1.0221e-1
6.9265e-4	-5.5947e-3	-1.5702e-3	7.4986e-2	-5.3959e-4	5.4453e-3	1.6796e-3	-7.3584e-2
1.6258e-3	5.8926e-4	4.7458e-3	-8.1166e-2	-1.4148e-3	-6.1453e-4	-4.1551e-3	8.2858e-2
6.1510e-4	-8.1412e-4	1.5778e-2	3.9865e-2	-3.4979e-4	9.3407e-4	-1.4708e-2	-3.7958e-2
-4.5539e-4	-3.5842e-4	-2.7665e-3	-4.6069e-2	7.6840e-4	6.4000e-4	4.3063e-3	4.8125e-2
-1.1842e-3	-5.0595e-3	-7.3354e-3	1.3779e-2	1.5349e-3	5.5119e-3	9.3217e-3	-1.1629e-2
1.4272e-3	-5.9397e-3	-9.8165e-3	-3.4055e-2	-1.0522e-3	6.5629e-3	1.2207e-2	3.6256e-2
b_{256} = 5.0732e-1 $b_i = b_{512-i}$, $257 \leq i \leq 512$				b_{256} = 4.9489e-1 $b_i = b_{512-i}$, $257 \leq i \leq 512$			

Table F.2: ACB comb filter pair: filter coefficients ($b_0 – b_{512}$) for the left and right ear filters.

Left ear filter				Right ear filter			
b_0 - b_{63}	b_{64} - b_{127}	b_{128} - b_{191}	b_{192} - b_{255}	b_0 - b_{63}	b_{64} - b_{127}	b_{128} - b_{191}	b_{192} - b_{255}
1.2813e-4	3.7927e-4	-4.9381e-3	-7.7912e-3	-9.7154e-5	3.8413e-6	5.7214e-3	1.0521e-2
-1.6650e-4	-5.3532e-5	1.6194e-3	1.9885e-2	1.9435e-4	4.2655e-4	-6.9715e-4	-1.6908e-2
1.9651e-4	1.7189e-3	2.1405e-3	-4.4447e-3	-1.7270e-4	-1.3751e-3	-1.1104e-3	7.5512e-3
4.6792e-4	1.3951e-6	2.5862e-3	-2.6224e-3	-4.4926e-4	2.9426e-4	-1.4874e-3	5.7203e-3
5.0253e-4	4.6138e-4	3.3066e-3	-5.8612e-4	-4.9021e-4	-2.3086e-4	-2.1836e-3	3.5236e-3
2.3479e-5	1.3906e-3	5.8928e-3	-1.7517e-2	-1.8773e-5	-1.2392e-3	-4.7926e-3	2.0140e-2
5.4724e-4	2.1103e-3	3.3035e-3	-2.1811e-2	-5.5136e-4	-2.0477e-3	-2.2729e-3	2.3973e-2
4.9602e-4	7.0250e-4	-3.7098e-3	-4.5634e-3	-5.1004e-4	-7.3336e-4	4.6267e-3	6.1407e-3
2.4474e-6	3.5746e-6	5.8749e-3	2.9003e-3	-2.7198e-5	-1.2709e-4	-5.1113e-3	-2.0002e-3
3.7918e-4	-8.9218e-4	-1.7453e-3	-1.8912e-2	-4.1513e-4	6.8232e-4	2.3216e-3	1.9082e-2
3.7805e-5	-9.6648e-4	-4.2520e-3	-2.2099e-2	-8.4948e-5	6.8164e-4	4.6132e-3	2.1529e-2
2.5981e-4	-3.7926e-4	3.7228e-3	-2.2592e-2	-3.1761e-4	3.5055e-5	-3.5981e-3	2.1316e-2
1.0131e-4	5.4353e-5	4.3259e-3	-3.0173e-4	-1.6860e-4	-4.3915e-4	-4.4534e-3	-1.6083e-3
-1.7446e-4	5.6704e-4	-6.3470e-4	-5.4024e-3	9.9469e-5	-9.7181e-4	2.4537e-4	2.9646e-3
-5.2386e-5	-2.7064e-3	-9.2881e-3	1.0931e-2	-2.7866e-5	2.3027e-3	8.6334e-3	-1.3767e-2
2.3977e-4	-1.4355e-3	-1.0669e-2	-1.1179e-3	-3.2228e-4	1.0531e-3	9.7515e-3	-1.9728e-3
-7.0798e-5	-2.1023e-3	-7.0013e-3	-2.3897e-2	-1.0511e-5	1.7592e-3	5.8328e-3	2.0699e-2
-1.9724e-4	-1.7246e-3	2.1808e-3	2.4021e-2	1.2087e-4	1.4354e-3	-3.5813e-3	-2.7180e-2
3.3197e-4	5.5167e-4	-1.7715e-3	2.5264e-2	-3.9956e-4	-7.7618e-4	1.6877e-4	-2.8250e-2
1.3684e-4	2.1410e-4	1.5140e-3	2.6089e-2	-1.9193e-4	-3.6778e-4	-3.2777e-3	-2.8784e-2
1.1125e-4	3.0775e-3	2.2123e-3	2.2808e-2	-1.5045e-4	-3.1588e-3	-4.0840e-3	-2.5109e-2
-5.9853e-4	3.3508e-3	3.4904e-3	2.8628e-2	5.7808e-4	-3.3623e-3	-5.4053e-3	-3.0454e-2
-5.1679e-4	-7.3292e-4	4.1433e-3	4.2942e-2	5.1730e-4	7.8470e-4	-6.0264e-3	-4.4233e-2
-8.2967e-4	-4.2032e-3	-3.1731e-3	5.3135e-2	8.5256e-4	4.3093e-3	1.4049e-3	-5.3853e-2
-9.8796e-4	-9.1502e-4	4.6139e-3	1.8288e-2	1.0338e-3	1.0648e-3	-6.1791e-3	-1.8417e-2
-3.6385e-4	-1.5964e-3	-4.0468e-3	9.9852e-4	4.3251e-4	1.7788e-3	2.7729e-3	-5.4598e-4
-1.2484e-3	-2.8476e-3	-5.1140e-3	2.7821e-2	1.3388e-3	3.0525e-3	4.2153e-3	-2.6816e-2
-1.0235e-3	-2.4032e-3	-3.5795e-3	-1.4122e-2	1.1338e-3	2.6223e-3	3.1298e-3	1.5628e-2
3.1707e-4	-1.2426e-3	-1.2063e-2	-2.7550e-2	-1.8944e-4	1.4696e-3	1.2120e-2	2.9486e-2
1.2421e-3	1.7717e-3	-1.0156e-2	-4.7105e-2	-1.1005e-3	-1.5404e-3	1.0758e-2	4.9384e-2
8.0569e-4	-1.9935e-4	7.5575e-3	-1.9005e-2	-6.5413e-4	4.3335e-4	-6.3988e-3	2.1525e-2
5.9960e-4	2.0412e-3	1.2315e-2	-2.3037e-2	-4.4289e-4	-1.8046e-3	-1.0617e-2	2.5690e-2
-1.2590e-6	-3.2309e-4	2.1852e-3	-3.4860e-2	1.5766e-4	5.6250e-4	3.3035e-6	3.7534e-2
-4.4211e-5	-1.6855e-3	9.3624e-3	3.7320e-2	1.9424e-4	1.9272e-3	-6.7632e-3	-3.4735e-2
3.8990e-4	8.5897e-4	5.5058e-3	3.5155e-2	-2.5275e-4	-6.1712e-4	-2.6054e-3	-3.2761e-2
-4.3830e-5	1.7416e-4	9.7468e-3	3.1393e-2	1.6142e-4	6.2992e-5	-6.6796e-3	-2.9282e-2
9.2951e-4	-9.0501e-4	4.5985e-3	3.5365e-2	-8.3805e-4	1.1299e-3	-1.5174e-3	-3.3614e-2
6.3303e-4	-1.6461e-3	2.9200e-3	1.4265e-2	-5.7372e-4	1.8485e-3	1.2601e-5	-1.2939e-2
3.3086e-4	-2.3624e-3	4.6557e-3	-5.3728e-2	-3.0875e-4	2.5304e-3	-2.0330e-3	5.4583e-2
-3.1644e-4	-3.0236e-3	-8.8788e-3	-3.9453e-2	2.9771e-4	3.1446e-3	1.1043e-2	3.9804e-2
-4.2149e-4	2.6332e-3	-3.6471e-3	-4.7706e-2	3.6006e-4	-2.5710e-3	5.2258e-3	4.7538e-2
1.4054e-4	1.3870e-3	-4.4129e-4	-3.0422e-2	-2.4450e-4	-1.3929e-3	1.3422e-3	2.9736e-2
-7.8045e-4	3.3172e-3	1.1690e-2	5.4029e-2	6.3625e-4	-3.3973e-3	-1.1519e-2	-5.5215e-2
-7.3063e-4	3.8791e-3	1.2587e-2	1.7865e-2	5.5052e-4	-4.0352e-3	-1.3154e-2	-1.9514e-2
-1.7339e-4	-1.0843e-3	4.2222e-4	2.3337e-2	-3.6506e-5	8.5472e-4	-1.6889e-3	-2.5391e-2
-3.1735e-4	1.2102e-3	-7.2093e-3	-6.2437e-2	8.5191e-5	-1.5068e-3	5.3214e-3	6.0056e-2
-1.2036e-3	3.4388e-3	4.6923e-3	-7.5571e-2	9.5766e-4	-3.7929e-3	-7.0885e-3	7.2960e-2
-5.7671e-6	6.7253e-3	-3.3874e-4	-1.2539e-2	-2.4519e-4	-7.1257e-3	-2.4287e-3	9.8112e-3
-4.1522e-4	1.8020e-3	-1.3379e-2	4.4348e-2	1.6814e-4	-2.2370e-3	1.0389e-2	-4.7068e-2
-3.7723e-4	1.5056e-3	2.7729e-3	4.6180e-2	1.4250e-4	-1.9638e-3	-5.8348e-3	-4.8764e-2
9.3773e-4	5.8466e-3	-6.0013e-3	-8.1926e-2	-1.1522e-3	-6.3177e-3	3.0067e-3	7.9603e-2
-3.9504e-5	4.6179e-3	5.4793e-3	-6.0048e-2	-1.4755e-4	-5.0926e-3	-8.2853e-3	5.8094e-2
4.7134e-5	-9.6318e-5	1.2145e-2	4.4162e-2	-2.0017e-4	-3.7278e-4	-1.4664e-2	-4.5659e-2
-3.4179e-4	8.2741e-4	2.1601e-3	4.5553e-2	2.2875e-4	-1.2814e-3	-4.3191e-3	-4.6536e-2
5.8305e-4	2.0482e-3	-1.0737e-2	-1.2786e-1	-6.5065e-4	-2.4757e-3	8.9894e-3	1.2742e-1
4.1971e-4	-3.7432e-3	-1.7206e-2	3.7283e-2	-4.3702e-4	3.3563e-3	1.5902e-2	-3.7189e-2
-5.3038e-4	-2.0617e-3	-2.5930e-5	1.9913e-2	5.6743e-4	1.7327e-3	-8.1608e-4	-1.9320e-2
-3.6645e-4	-8.2756e-3	-1.8161e-3	-1.0117e-1	4.6082e-4	8.0251e-3	1.4469e-3	1.0221e-1
6.9265e-4	-5.5947e-3	-1.5702e-3	7.4986e-2	-5.3959e-4	5.4453e-3	1.6796e-3	-7.3584e-2
1.6258e-3	5.8926e-4	4.7458e-3	-8.1166e-2	-1.4148e-3	-6.1453e-4	-4.1551e-3	8.2858e-2
6.1510e-4	-8.1412e-4	1.5778e-2	3.9865e-2	-3.4979e-4	9.3407e-4	-1.4708e-2	-3.7958e-2
-4.5539e-4	-3.5842e-4	-2.7665e-3	-4.6069e-2	7.6840e-4	6.4000e-4	4.3063e-3	4.8125e-2
-1.1842e-3	-5.0595e-3	-7.3354e-3	1.3779e-2	1.5349e-3	5.5119e-3	9.3217e-3	-1.1629e-2
1.4272e-3	-5.9397e-3	-9.8165e-3	-3.4055e-2	-1.0522e-3	6.5629e-3	1.2207e-2	3.6256e-2
b_{256} = 5.0732e-1 $b_i = b_{512-i}$, $257 \leq i \leq 512$				b_{256} = 4.9489e-1 $b_i = b_{512-i}$, $257 \leq i \leq 512$			

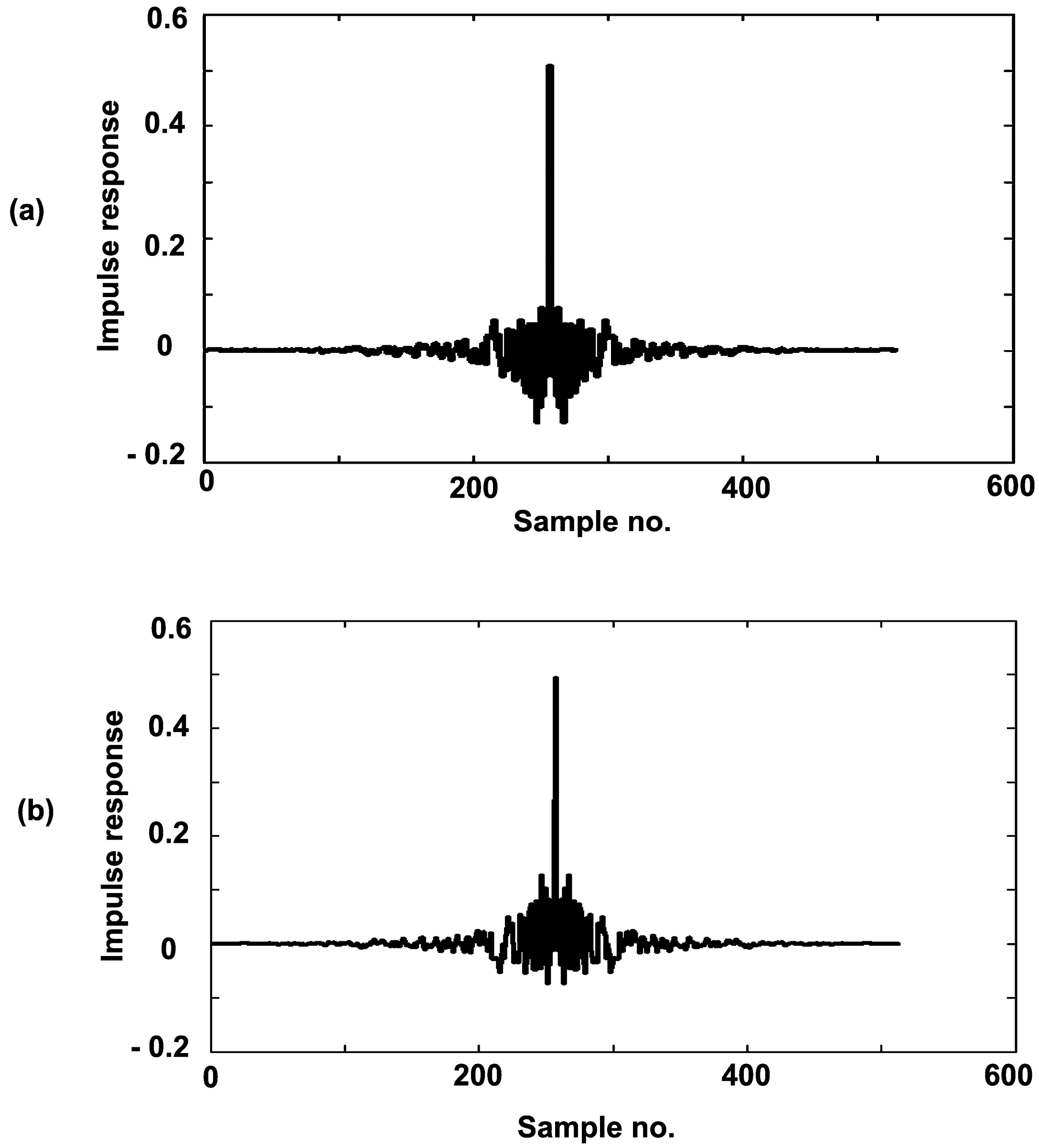


Fig. F.1 Impulse response of CB18 filters used for spectral splitting: (a) left filter impulse response (b) right filter impulse response.

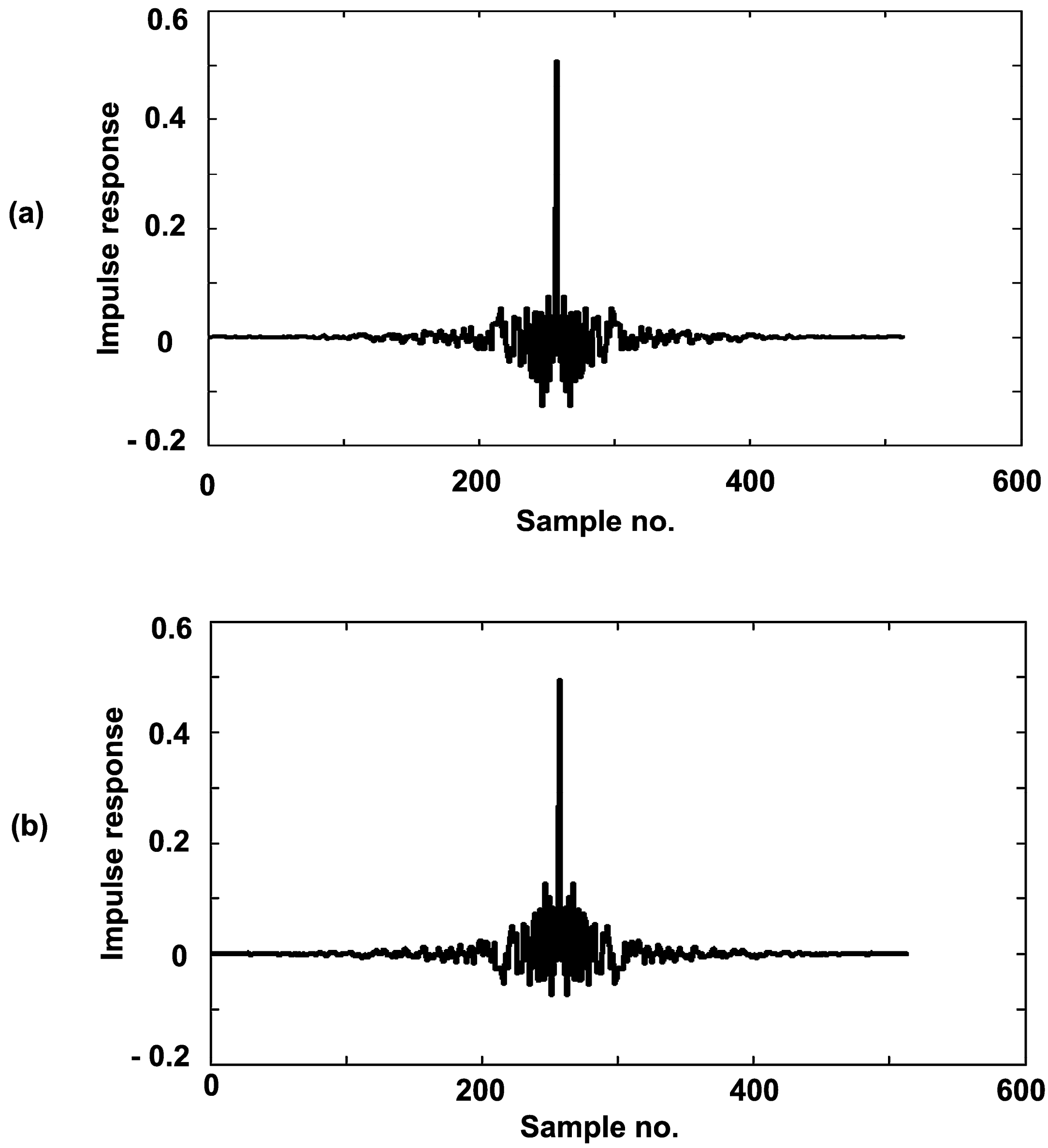


Fig. F.2 Impulse response of ACB filters used for spectral splitting: (a) left filter impulse response (b) right filter impulse response.

REFERENCES

Abel, S. M., Giguere, C., Consoli, A., and Papsin, B. C. (**2000**). “The effect of aging on horizontal plane sound localization,” J. Acoust. Soc. Am. **108**, 743 – 752.

Algazi, V. R., Duda, R. O., and Thompson, D. W. (**2001**). “The CIPIC HRTF database,” *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics,* Mohonk, New York, 99 – 102.

Algazi, V. R., Duda, R. O., Duraiswami, R., Gumerov, N. A., and Tang, Z. (**2002**). “Approximating the head related transfer functions using simple geometric models of head and torso,” J. Acoust. Soc. Am. **112**, 2053 – 2063.

ANSI (**1989**). Methods for measuring the intelligibility of speech over communication systems. *Revised Standards Report ANSI S3.2-1989*, American National Standards Institute, New York.

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

Arai, T., Yasu, K., and Hodoshima, N. (**2004**). “Effective speech processing for various impaired listeners,” *Proc. 18th Int. Congress Acoust.* (ICA 2004, Kyoto, Japan), 1389 – 1392.

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

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

Baer, T., Moore, B. C. J., and Gatehouse, S. (**1993**). “Spectral contrast enhancement of speech in noise for listeners with sensorineural hearing impairment: effects on intelligibility, quality, and response times,” J. Rehabil. Res. Dev. **30**, 49 – 72.

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

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

Bashford, J. A. and Warren, R. M. (**1987**). "Effects on spectral alternation on the intelligibility of words and sentences," Perception & Psychophysics. **42**, 431 – 438.

Baskent, D. and Shannon, R. V. (**2006**). "Frequency transposition around dead regions simulated with a noise band vocoder," J. Acoust. Soc. Am. **119**, 1156 – 1163.

Beasley, D. S., Schwimmer, S., and Rintelman, W. F. (**1972**). "Intelligibility of time compressed CNC monosyllables," J. Speech and Hear. Res. **15**, 340 – 350.

Begault, D. R. (**1991**). "Challenges to the successful implementation of 3-D sound," J. Audio Eng. Soc. **39**, 864 – 870.

Best, V., Carlile, S., Jin, C., and Van Schaik, A. (**2005**). "The role of high frequencies in source localization," J. Acoust. Soc. Am. **118**, 353 – 363.

Boike, K. T. and Souza, P. E. (**2000**). "Effect of compression ratio on speech recognition and speech-quality ratings with wide dynamic range compression amplification," J. Speech, Language, Hearing Res. **43**, 456 – 468.

Brungart, D. S. and Rabinowitz, W. R. (**1999**). "Auditory localization of nearby sources. Head related transfer functions," J. Acoust. Soc. Am. **106**, 1465 – 1479.

Bunnell, T. and Martin, J. (**1985**). "Toward improving speech intelligibility," J. Acoust. Soc. Am. Suppl. **77**, S106.

Bunnel, H. T. (**1990**). "On enhancement of spectral contrast in speach for hearing-impaired listeners," J. Acoust. Soc. Am. **88**, 2546-2556.

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

CIPIC (**2001**). CIPIC HRTF Database, University of California, Davis, Cal. Available online http://interface.cipic.ucdavis.edu/ CIL_html/CIL_HRTF_database.htm

CHABA (**1991**). "Speech-perception aids for hearing-impaired people: Current status and needed research," J. Acoust. Soc. Am. **90**, 637 – 683.

Chaudhari, D. S. and Pandey, P. C., (**1998a**). "Dichotic presentation of speech signal with critical band filtering for improving speech perception," *Proc. IEEE ICASSP*, 3601 – 3604.

Chaudhari, D. S. and Pandey, P. C. (**1998b**). "Dichotic presentation of speech signal using critical filter band for bilateral sensorineural hearing impairment," *Proc. 16th Int. Congress Acoust.* (ICA 1998, Seattle, Wash.), AE 3.1.

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

Cheeran, A. N. and Pandey, P. C. (**2004a**). "Evaluation of speech processing schemes using binaural dichotic presentation to reduce the effect of masking in hearing-impaired listeners," *Proc. 18th Int. Congress Acoust.* (ICA 2004, Kyoto, Japan), II, 1523 – 1526.

Cheeran, A. N. and Pandey, P. C. (**2004b**). "Speech processing for hearing aids for moderate bilateral sensorineural hearing loss," *Proc. IEEE ICASSP*, IV-17-20.

Cheeran, A. N. (**2005)**. Speech processing with dichotic presentation for binaural hearing aids for moderate bilateral sensorineural loss*,* Ph.D. Thesis, School of Biosciences and Bioengineering, Indian Institute of Technology Bombay, India.

Childers, D. G. and Hu, T. H., (**1994**). "Speech synthesis by glottal excited linear prediction," J. Acoust. Soc. Am. **96**, 2026 – 2036.

Chung, W., Carlile, S., and Leong, P. (**2000**). "A performance adequate computational model for auditory localization," J. Acoust. Soc. Am. **107**, 432 – 445.

Cohen, I. (**2006**). "Speech spectral modeling and spectral enhancement based on autoregressive conditional heteroscedasticity models," Signal Processing. **86**, 698-709.

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

Daniloff, R. G., Shriner, T. H., and Zemlin, W. R. (**1968**). "Intelligibility of vowels altered in duration and frequency," J. Acoust. Soc. Am. **44,** 700 – 707.

Davies, V. E., Souza, P. E., Brennan, M., and Stecker, G. C. (**2009**). "Effects of audibility and multichannel wide dynamic range compression on consonant recognition for listeners with severe hearing loss," Ear Hear. **30**, 494 – 504.

De Filippo, C. L. and Scott, B. L. (**1978**). "A method for training and evaluating the reception of ongoing speech," J. Acoust. Soc. Am. **63,** 1186 – 1192.

Delogu, C., Paoloni, A., Pocci, P., and Sementina, C. (**1991**). "Quality evaluation of text-to speech synthesizers using magnitude estimation, categorical estimation, pair comparison and reaction time methods", *Proc. EUROSPEECH,* Genova, 353 – 355**.**

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

Deutsch, 1. J. and Richards, A. M. (**1979**). *Elementary Hearing Science.* (Allyn Bacon, Boston, Mass.).

Dillon, H., Keidser, G., and Silberstein, H. (**2003**). "Sound quality comparisons of advanced hearing aids," Hearing J. **56**, 1 – 6.

Dorman, M. F., Loizou, P. C., and Rainey, D. (**1997**). "Speech intelligibility as a function of number of channels of simulation for signal processors using sine wave and noise-band output," J. Acoust. Soc. Am. **102**, 2403 – 2411.

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

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

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.

Elliott, L. L. (**1975**). "Temporal and masking phenomena in persons with sensorineural hearing loss," J. Audiology. **14**, 336-353.

Epstein, M. and Florentine, M. (**2005**). "A test of equal loudness ratio hypothesis using cross modality matching functions," J. Acoust. Soc. Am. **99**, 907 – 913.

Evans, E. F. (**1975**). "The sharpening of cochlear frequency selectivity in the normal and abnormal cochlea," Audiology. **14**, 419 – 442.

Fairbanks, G. and Kodman, F. (**1957**). "Word intelligibility as a function of time compression," J. Acoust. Soc. Am. **29**, 636 – 644.

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

Flanagan, J. L. (**1972**). *Speech Analysis Synthesis and Perception.* (Springer-Verlag, New York), pp. 9 - 22.

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

Fletcher, H. and Munson, W. A. (**1933**). "Loudness, its definition, measurement and calculation," J. Acoust. Soc. Am. **5**, 82 – 108.

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

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

Fowler, E. P. (**1936**). "A method for the early detection of otosclerosis," Arch. Otolaryngol. **24**, 731 – 741.

Franck, B. A. M., Sidonne, C., Van Kreveld-Bos, G. M., Dreschler, W. A., and Verschuure, H. (**1999**). "Evaluation of spectral enhancement in hearing aids, combined with phonemic compression," J. Acoust. Soc. Am. **106**, 1452 – 1464.

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

Friesen, L. M., Shannon, R. V., Baskent, D., and Wang, X. (**2001**). "Speech recognition in noise as a function of number of spectral channels: Comparison of acoustic hearing and cochlear implants," J. Acoust. Soc. Am. **110**, 1150 – 1163.

Fraga, F. J., Prates, L. P. C. S, and Iorio, M. C. M. (**2008**). "Frequency compression / trans-position of fricative consonants for the hearing impaired with high frequency dead regions," *Proc. Interspeech*, Brisbane, Australia, 2238 – 2241.

Fu, Q. J. and Shannon, R. V. (**1999**). "Recognition of spectrally degraded and frequency-shifted vowels in acoustic and electric hearing," J. Acoust. Soc. Am. **105**, 1889 – 1900.

Garvey, W. D. (**1953**). "The intelligibility of speeded speech," J. Exp. Psychol. **45,** 102 – 106.

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

Giguere, C. and Abel, S. M. (**1993**). "Sound localization: effects of reverberation time, speaker array, stimulus frequency, and stimulus rise/decay," J. Acoust. Soc. Am. **94**, 769 – 776.

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

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

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

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

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

Gold, B. and Morgan, N. (**2002**). *Speech and Audio Signal Processing*. (Wiley, Singapore), pp. 210-212.

Goodman, D. J., McDermott, B. J., and Nakatani, L. H. (**1976**). "Subjective evaluation of PCM coded speech," Bell System Technical Journal. **55**, 1087 – 1109.

Guyton, A. C. (**1986**). *Textbook of Medical Physiology* (Sounders, Philadelphia, PA), pp. 633 – 642.

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

Hartmann, W. M. (**1999**). "How we localize sound," Phys. Today. **11**, 24 – 29.

Hawkins, D. B., Prosek, R. A., Walden, B. E., and Montgomery, A. A. (**1987**). "Binaural loudness summation in hearing impaired," J. Speech Hear. Res. **30**, 37 – 43.

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

Hofman, P. M. and Van Opstal, J. (**1998**). "Spectro-temporal factors in two-dimensional human sound localization," J. Acoust. Soc. Am. **103**, 2634 – 2648.

House, A. S., Williams, C. E., Hecker, M. H. L., and Kryter, K. D. (**1965**). "Articulation testing methods: consonantal differentiation with closed-response set," J. Acoust. Soc. Am. **37**, 158 – 166.

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

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

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

Itakura, F., Saito, S., Koike, Y., Sawabe, H., and Nishikawa, M. (**1972**). "An audio response unit based on partial correlation," IEEE Trans. Communication. **20**, 792 – 797.

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

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

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

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

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

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

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

Kitawaki, N., Honda, M., and Itoh, K. (**1984**). "Speech quality assessment methods for speech coding systems," IEEE Communications Magazine. **22,** 26 – 33.

Klumpp, R. B. and Webster, J. C. (**1961**). "Intelligibility of time-compressed speech," J. Acoust. Soc. Am. **33**, 265 – 267.

Kortekaas, R. W. L. and Kohlrausch, A. (**1999**). "Psychoacoustical evaluation of PSOLA II. Ddouble-formant stimuli and the role of vocal perturbation," J. Acoust. Soc. Am. **105**, 522-535.

Kreul, K. J., Nixon, J. C., Kryter, K. D., Bell, D. W., Lang, J. S., and Schubert, E. D. (**1968**). "A proposed clinical test of speech discrimination," J. Speech Hearing Res. **11**, 536-552.

Kryter, K. D. (**1961**). “Bandwidth compression of speech by spectrum sampling,” J. Acoust. Soc. Am. **33**, 1663.

Kryter, K. D. and Whitman, E. C. (**1965**). “Some comparisons between rhyme and PB-word intelligibility tests,” J. Acoust. Soc. Am. **37**, 1146.

Kulkarni, P. N., Pandey, P. C., and Jangamashetti, D. S. (**2006**). “Perceptually balanced filter response for binaural dichotic presentation to reduce the effect of spectral masking,” (abstract) J. Acoust. Soc. Am. **120**, 3253.

Kulkarni, P. N. and Pandey, P. C. (**2008**). “Frequency mapping for multi-band frequency compression for improving speech intelligibility,” *Proc. 14th National Conf. Communications*, IIT Bombay, Mumbai, India. 437 – 441.

Ladefoged, P. (**1982**). *A Course in Phonetics.* 2nd Ed. (Harcourt Brace Jovanovich, New York), pp 47-162.

Langedijk, E. H. A. and Bronkhorst, A. W. (**2002**). “Contribution of spectral cues for human sound localization,” J. Acoust. Soc. Am. **112**, 1583 – 1596.

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

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

Levitt, H., Pickett, J. M., and Houde, R. A., Eds. (**1980**). *Sensory Aids for the Hearing Impaired.* (IEEE Press, New York), pp. 3-10.

Liberman, A. B., Delattre, P., Gerstman, L. J., and Cooper, F. S. (**1956**). “Tempo of frequency change as a cue for distinguishing classes of speech sounds,” J. Experimental Psychology. **52**, 127 – 137.

Lorenzi, C. S., Gatehouse, S., and Lever, C. (**1999a**). “Sound localization in noise in hearing impaired listeners,” J. Acoust. Soc. Am. **105**, 3454 – 3463.

Lorenzi, C. S., Gatehouse, S., and Lever, C. (**1999b**). “Sound localization in noise in normal hearing listeners,” J. Acoust. Soc. Am. **105**, 1810 – 1820.

Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., and Moore, B. C. J. (**2006**). “Speech perception problems of the hearing impaired reflect inability to use temporal fine structures,” Proc. National Aacademy of Science of the United States of America (PNAS), **103**, 1866 – 1869.

Loizou, P. C. and Dorman, M. F. (**1999**). “On the number of channels needed to understand speech,” J. Acoust. Soc. Am., **106**, 2097 – 2103.

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. (**1997**). A digital filter bank hearing aid, Doctoral thesis, Dept. of Neuroscience and Locomotion, Linkoping Univ., Linkoping, Sweden.

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

Mackersie, C., Neuman, A., and Levitt, H. (**1999**). "A comparison of response time and word recognition measures using a word monitoring and closed set word identification task" Ear and Hear. **20,** 140 – 148.

Makous, J. C., and Middlebrooks, J. C. (**1990**). "Two-dimensional sound localization by human listeners," J. Acoust. Soc. Am. **87**, 2188 – 2200.

Marks, L. E. (**1978**). "Binaural summation of the loudness of pure tones," J. Acoust. Soc. Am. **64**, 107 – 113.

McDermott, H. J., Dorkos, V. P., Dean, M. R., and Ching, T.Y.C. (**1999**). "Improvements in speech perception with use of the AVR transonic frequency transposing hearing aid," J. Speech, Language, Hear. Res. **42**, 1323 – 1335.

McDermott, H. J. and Dean, M. R. (**2000**). "Speech perception with steeply sloping hearing loss," British J. Audiology. **34**, 353 – 361.

Meftah, M. and Boudelaa, S. (**1996**). "How facilitatory can lexical information be during word recognition? Evidence from Moroccon Arabic," *Proc. Int. Conf. Spoken Language Processing (ICSLP)*, Philadelphia, Penn. **1**, 74 – 77.

Middlebrooks, J. C. (**1992**). "Narrow-band sound localization related to external ear acoustics," J. Acoust. Soc. Am. **92**, 2607 – 2624.

Miller, G. A. and Nicely, P. E. (**1955**). "An analysis of perceptual confusions among some English consonants," J. Acoust. Soc. Am. **72**, 338 – 352.

Miller, R. L., Schilling, J. R., Franck, K. R., and Young, E. D. (**1997**). "Effects of acoustic trauma on the representation of the vowel /ε/ in cat auditory nerve fibers," J. Acoust. Soc. Am. **101**, 3602 – 3616.

Miller, R. L., Calhoun, B. M., and Young, E. D. (**1999**). "Contrast enhancement improves the perception of /ε/ -like vowels in the hearing-impaired auditory nerve," J. Acoust. Soc. Am. **106**, 2693-2708.

Mitra, S. K. (**1998**). *Digital Signal Processing, a Computer-Based Approach* (McGraw-Hill, Singapore), pp. 423 – 472.

Moore, B. C. J. and Glasberg, B. R. (**1983**). "Suggested formulae for calculating auditory filter bandwidths and excitation patterns," J. Acoust. Soc. Am. **74**, 750 – 753.

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

Moore, B. C. J. (**1986**). *Frequency Selectivity in Hearing.* (Academic, London).

Moore, B. C. J, Peters, R. W., and Glasberg, B. R. (**1990**). "Auditory filter shapes at low center frequencies," J. Acoust. Soc. Am. **88**, 132-140.

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

Moore, B. C. J. and Glasberg, B. R. (**1993**). "Simulation of the effects of loudness recruitment and threshold elevation on the intelligibility of speech in quiet and in a background of speech," J. Acoust. Soc. Am. **94**, 2050 – 2062.

Moore, B. C. J., Wojtczak, M., and Vickers, D. A. (**1996**). "Effect of loudness recruitment on the perception of amplitude modulation," J. Acoust. Soc. Am. **100**, 481 – 489.

Moore, B. C. J. (**1997**). *An Introduction to the Psychology of Hearing*. 4th ed. (Academic, London), pp. 66 – 107.

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

Moore, B. C. J. (**2002**). "Hearing loss in elderly and its compensation with hearing aids," Gerontechnology. **1**, 140 – 152.

Moore, B. C. J. (**2003**). "Speech processing for the hearing impaired: successes, failures, and implications for speech mechanisms," Speech Communication. **41**, 81 – 91.

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.

Murase, A., Nakajima, F., Sakamoto, S., Suzuki, Y., and Kawase, T. (**2004**). "Effect and sound localization with dichotic-listening hearing aids," *Proc. 18th Int. Congress Acoust.* (ICA 1988, Kyoto), II-1519 – 1522.

Nagafuchi M. (**1976**). "Intelligibility of distorted speech sounds shifted in frequency and time in normal children," Audiology. **15,** 326 – 337.

Nakatsui, M. and Mermelstein, P. (**1982**). "Subjective speech-to-noise ratio as a measure of speech quality for digital waveform coders," J. Acoust. Soc. Am. **72**, 1136 – 1144.

Nejime, Y. and Moore, B. C. J. (**1997**). "Simulation of the effect of threshold elevation and loudness recruitment combined with reduced frequency selectivity on the intelligibility of speech in noise," J. Acoust. Soc. Am. **102**, 603 – 615.

Nelson, D. A., Chargo, S. J., Kopun, J. G., and Freyman, R. L. (**1990**). "Effect of forward-masked psychophysical tuning curves in quiet and noise," J. Acoust. Soc. Am. **88**, 2143 – 2151.

Noble, W., Byrne, D., and Lepage, B. (**1994**). "Effects on sound localization of configuration and type of hearing impairment," J. Acoust. Soc. Am. **95**, 992 – 1005.

Noble, W., and Byrne, D. (**1990)**. “A comparison of different binaural hearing aid systems for sound localization in the horizontal and vertical planes,” Br. J. Audiol. **24**, 335 – 346.

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

Oppenheim, A. V., Schafer, R. W., and Buck, J. R. (**1999**). *Discrete-time Signal Processing* (Prentice-Hall, Englewood Cliffs, New Jersey), pp. 465 – 507.

O’Shaughnessy, D. (**1987**). *Speech Communication: Human and Machine.* (Addison-Wesley. Reading, Mass.), pp. 39 – 78.

Otani, M., Hirahara, T., and Ise, S. (**2009**). “Numerical study on source-distance dependency of head related transfer functions,” J. Acoust. Soc. Am. **125**, 3253 – 3261.

Oxenham, A. J. and Plack, C. J. (**1997**). “A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing,” J. Acoust. Soc. Am. **101**, 3666 – 3675.

Pandey, P. C., Jangamashetti, D. S., and Cheeran, A. N. (**2001**). “Binaural dichotic presentation to reduce the effect of increased temporal and spectral masking in sensorineural hearing impairment,” (abstract), J. Acoust. Soc. Am., **110**, 2705.

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

Picheny, M. A., Durlach, N. I., and Draida, L. D. (**1985**). “Speaking clearly for the hard of hearing I**.** Intelligibility differences between clear and conversational speech,” J. Speech. Hear. Res. **28**, 96 – 103.

Pick, G. F., Evans, E. F., and Wilson, J. P. (**1977**). *Psychophysics and Physiology of Hearing* (Academic, London).

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

Pickles, J. O. (**1982**). *An Introduction to Physiology of Hearing*, (Academic, London), pp. 286 - 300.

Plomp, R. (**1988**). “The negative effect of amplitude compression in multichannel hearing aids in the light of the modulation transfer function,” J. Acoust. Soc. Am. **83**, 2322 – 2327.

Polkosky, M. D. and Lewis, J. R. (**2003**). “Expanding the MOS: Development and psychometric evaluation of the MOS-R and MOS-X,” Int. J. Speech Tech. **6**, 161 – 182.

Proakis, J. G. and. Manolakis, D. G. (**1992**). *Digital Signal Processing: Principles, Algorithms, and Applications* (Macmillan, New York), pp. 620 – 662.

Rabiner, L. R. and Gold, B. (**1975**). *Theory and Application of Digital Signal Processing* (Prentice-Hall, Englewood Cliffs, New Jersey), pp. 75 – 121.

Rabiner, L. R. and Schafer, R. W. (**1978**). *Digital Processing of Speech Signals* (Prentice-Hall, Englewood Cliffs, New Jersey), pp. 274 – 277.

Rabiner, L. R., Gold, B., and McGonegal, C.A. (**1970**). "An approach to the approximation problem for nonrecursive digital filters," IEEE Trans. Audio Electroacoust. **AU-16**, 83 – 105.

Reed, C. M., Hicks, B. L., Braida, L. D., and Duriach, N. I. (**1983**). "Discrimination of speech processed by low pass filtering and pitch invariant frequency lowering," J. Acoust. Soc. Am. **74**, 409 – 419.

Rintelmann, W. E. (**1991**). *Hearing Assessment* 2nd Ed. (Allyn Bacon, Massachusetts).

Revoile, S. G., Holden-Pitt, L., Edward, D., and Pickett, J. M. (**1986a**). "Some rehabilitative considerations for future speech-processing hearing aids,". J. Rehabil. Res. Dev. **23**, 89 – 94.

Revoile, S. G., Holden-Pitt, L., Pickett, J. M, and Brandt, F. (**1986b**). "Speech cue enhancement for the hearing impaired: Altered vowel duration for perception of final fricative voicing," J. Speech. Hear. Res. **29**, 240 – 255.

Revoile, S. G., Holden-Pitt, L., Edward, D., Pickett, J. M., and Brandt, F. (**2002**). "Speech-cue enhancement for the hearing impaired: Amplification of burst/murmur cues for improved perception of final stop voicing," J. Rehabil. Res. Dev. **24**, 207 – 216.

Reynolds, G. S. and Stevens, S. S. (**1960**). "Binaural summation of loudness," J. Acoust. Soc. Am. **32**, 1337 – 1344.

Robinson, J. D., Baer, T., and Moore, B. C. J. (**2007**). "Using transposition to improve consonant discrimination and detection for listeners with severe high-frequency hearing loss," Int. J. Audiology. **46,** 293 – 308.

Rosen, S., Walliker, J. R., Fourcin, A., and Ball, V. (**1987**). "A microprocessor-based acoustic hearing aid for the profoundly impaired listener," J. Rehabilitation Res. and Dev. **24**, 239 – 260.

Rosen, S., Baker, R. J., and Darling, A. (**1998**). "Auditory filter nonlinearity at 2 kHz in normal hearing listeners," J. Acoust. Soc. Am. **103**, 2539 – 2550.

Rothauser, E. H., Chapman, W. D., Guttman, N., Hecker, M. H. L., Nordby, K. S., Silbiger, H. R., Urbanek, G. E., and Weinstock, M.(**1969**). "IEEE recommended practice for speech quality measurements," IEEE Trans. Audio Electroacoust. **17**, 227 – 246.

Sakamoto, S., Goto, K., Tateno, M., and Kaga, K. (**2000**). "Frequency compression hearing aid for severe-to-profound hearing impairment," Auris Nasus Larynx. **27**, 327 – 334.

Sandlin, R. E. Eds. (**1988**). *Handbook of Hearing Aid Amplification, Vol. 1, Theoretical and Technical Considerations.* (College-Hill, Boston, Mass.), pp. 38-71.

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

Scharf, B. (**1968**). "Binaural loudness summation as a function of bandwidth," *Proc. 6th Int. Congress Acoustics (ICA 1968*, Tokyo), 25 – 28.

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

Scharf, B. and Fishken, D. (**1970**). "Binaural summation of loudness: Reconsidered," J. Exp. Psychol. **86**, 374 – 379.

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

Sekimoto, S. and Saito, S. (**1980**). "Nonlinear frequency compression speech processing based on PARCOR analysis-synthesis technique," Ann Bull Rilp. **14**, 65 – 72.

Shailer, M. J. and Moore, B. C. J. (**1983**). "Gap detection as a function of frequency, bandwidth, and level," J. Acoust. Soc. Am. **74**, 467 – 473.

Shailer, M. J., Moore, B. C. J., Glasberg, B. R., Watson, N., and Harris, S. (**1990**). "Auditory filter shapes at 8 and 10 kHz," J. Acoust. Soc. Am. **88**, 141-148.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science. **270**, 303 – 304.

Shannon, R. V., Fan-Gang, Z., and Wygonski, J. (**1998**). "Speech recognition with altered spectral distribution of envelope cues," J. Acoust . Soc. Am. **104**, 2467 – 2476.

Shannon, R. V., Galvin, J. J., and Baskent D. (**2002**). "Holes in hearing," J. Assoc. Res. Otolaryngol. **3**, 185 – 199.

Simon, H. J. (**2005**). "Bilateral amplification and sound localization: then and now," J. Rehabilitation Res. and Dev. **42,** 117 – 132.

Simpson, A. M., Moore, B. C. J., and Glasberg, B. R., (**1990**). "Spectral enhancement to improve the intelligibility of speech in noise for hearing-impaired listeners," Acta Otolaryngol. Suppl. **469**, 101 – 107.

Simpson, A., Hersbach, A. A., and McDermott, H. J. (**2005**). "Improvements in speech perception with an experimental nonlinear frequency compression hearing device," Int. J. Audiology. **44**, 281 – 292.

Simpson, A., Hersbach, A. A., and McDermott, H. J. (**2006**). "Frequency-compression outcomes in listeners with steeply sloping audiograms," Int. J. Audiology. **45**, 619 – 629.

Steinberg, J. C. and Gardner, M. B. (**1937**). "The dependency of hearing impairment on sound intensity," J. Acoust. Soc. Am. **9**, 11 – 23.

Souza, P. E., Jenstad, L. M., and Filino, R. (**2005**). "Using multichannel wide-dynamic range compression in severely hearing-impaired listeners: effects on speech recognition and quality," Ear Hearing. **26**, 120-131.

Stevens, S. S. and Newman, E. B. (**1936**). "The localization of actual sources of sound," Am. J. Psychol. **48**, 297 – 306.

Stevens, S. S. (**1956**). "The direct estimation of sensory magnitudes - loudness," Am. J. Psychol. **69**, 1 – 15.

Stevens, J. C. and Guirao, M. (**1964**). "Individual loudness functions," J. Acoust. Soc. Am. **36**, 2210 – 2213.

Stevens, K. N. (**1980**). "Acoustic correlates of some phonetic categories," J. Acoust. Soc. Am. **68**, 836-842.

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

Stone, M. A. and Moore, B. C. J. (**1992b**). "Spectral feature enhancement for people with sensorineural hearing impairment: effects on speech intelligibility and quality," J. Rehabil. Res. Dev. **29**, 39 – 56.

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

Studebaker, G. A., Pavlovic, C. V., and Sherbecoe, R. L. (**1987**). "A frequency importance function for continuous discourse," J. Acoust. Soc. Am. **81**, 1130 – 1138.

Summers, V and Leek, M. R. (**1994**). "The internal representation of spectral contrast in hearing-impaired listeners," J. Acoust. Soc. Am. **95**, 3518 – 3528.

Summers, V. and Leek, M. R. (**1997**). "Intraspeech spread of masking in normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **101**, 2866 – 2876.

Tejero, J. C., Bernal, S., Hidaldo, J. A., Fernandez, J., Urquiza, R., and Gago, A. (**1991**). "A digital hearing aid that compensates for sensorineural hearing impairments," *Proc. IEEE ICASSP*, 2991 – 2994.

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

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

Thomas, T. G. Pandey, P. C., and Agashe, S. D. (**1996**). "Effect of consonantral intensity and duration modification on speech perception by listeners with simulated hearing impairment," J. Acoust. Soc. India. **24**, VI – 4.1 – 4.5.

Tiffany, W. R. and Bennett, D. A. (**1961**). "Intelligibility of slow-played speech," J. Speech and Hear. Res. **4,** 248 – 258.

Troost, B. T. and Waller, M. A. (**1998**). *Diagnostic Principals in Neuro-Otology: The Auditory System,* 2nd Ed. (Wiley, New York).

Turner, C. W. and Hurtig, R. R. (**1999**). "Proportional frequency compression of speech for listeners with sensorineural hearing loss," J. Acoust. Soc. Am. **106**, 877 – 886.

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

Tyler, R. S., Summerfield, Q., Wood, E. J., and Fernandes, M. A. (**1982**). “Psychoacoustic and phonetic temporal processing in normal and hearing-impaired listeners,” J. Acoust. Soc. Am. **72**, 740 – 752.

Van den Bogaert, T., Klasen, T. J., Moonen, M., Van Deun, L., and Wouters, J. (**2006**). “Horizontal localization with bilateral hearing aids: without is better than with,” J. Acoust. Soc. Am. **119**, 515 – 526.

Van Hoesel, R., Ramsden, R., and O’Driscoll, M. (**2002)**. “Sound direction identification, interaural time delay discrimination, and speech intelligibility advantages in noise for a bilateral cochlear implant user,” Ear Hear. **23**, 137 – 149.

Villchur, E. (**1973**). “Signal processing to improve speech intelligibility in perceptive deafness,” J. Acoust. Soc. Am. **53**, 1646 – 165.

Villchur, E. (**1974**). “Simulation of the effect of recruitment on loudness relationships in speech,” J. Acoust. Soc. Am. **56**, 1601-1611..

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.

Voiers, W. D. (**1983).** “Evaluating processed speech using the diagnostic rhyme test,” Speech Tech. **55**, 30 – 39.

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

Wall, L. G. (**1995**). *Hearing for the Speech-Language Pathologist and Health Care Professional,* (Butterworth-Heinemann, Boston, Mass.).

Warren, R. M., Reiner, K. R., Bashford, J. A., and Brubaker, B. S. (**1995**). “Spectral redundancy: Intelligibility of sentences heard through narrow spectral slits,” Perception & Psychophysics. **57**, 175 – 182.

Wightman, F. L. and Kistler, D. J. (**1992a**). “The dominant role of low frequency interaural time differences in sound localization,” J. Acoust. Soc. Am. **91**, 1648 – 1661.

Wightman, F. L. and Kistler, D. J. (**1992b**). “A model of HRTFs based on principal component analysis and minimum phase reconstruction,” J. Acoust. Soc. Am. **91**, 1637 – 1647.

Whilby, S., Florentine, M., Wagner, E., and Marozeau, J. (**2006**). “Monaural and binaural loudness of 5 and 200 ms tones in normal and impaired hearing,” J. Acoust. Soc. Am. **119**, 3931 – 3939.

Working Group on Communication Aids for the Hearing Impaired (**1991**). "Speech perception aids for hearing impaired people: current status and needed research," J. Acoust. Soc. Am. **90,** 637 – 685.

Yang, J., Luo, F. L., and Nehorai, A. (**2003**). "Spectral contrast enhancement: algorithms and comparisons," Speech Communication. **39**, 33-46.

Yang, W. and Hodgson, M. (**2006**). "Auralization study of optimum reverberation times for speech intelligibility for normal and hearing-impaired listeners in classrooms with diffuse sound fields," J. Acoust. Soc. Am. **120**, 801 – 807.

Yasu, K., Kobayashi, K., Shinohara, K., Hishitani, M., Arai, T., and Murahara, Y. (**2002**). "Frequency compression of critical band for digital hearing aids," *Proc. China-Japan Joint Conf. on Acoustics*. 159 – 162.

Yasu, K., Hishitani, M., Arai, T., and Murahara, Y. (**2004**). "Critical-band based frequency compression for digital hearing aids," Acoust. Sci. & Tech. **25,** 61 – 63.

Yost, W. A. (**1994**). *Fundamentals of Hearing: An Introduction*, (Academic, New York).

Yoo, S. D., Boston, J. R., Jaroudi, A., and Li, C. C. (**2007**). "Speech signal modification to increase intelligibility in noisy environment," J. Acoust. Soc. Am. **122**, 1138 – 1149.

Young, E. D. and Sachs, M. B. (**1979**). "Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers," J. Acoust. Soc. Am. **66**, 1381 – 1403.

Yund, E. W. and Buckles, K. M. (**1994**). "Multichannel compression hearing aids: effects of number of channels in speech discrimination in noise," J. Acoust. Soc. Am. **97**, 1206 – 1240.

Zhou, B. (**1995**). "Auditory filter shapes at high frequencies," J. Acoust. Soc. Am. **98**, 1935-1942.

Zhong, X. L. and Xie, B. S. (**2009**). "Maximum azimuthal resolution needed in measurement of head related transfer functions," J. Acoust. Soc. Am. **125**, 2209 – 2220.

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

Zwicker, E. and Henning, G. B. (**1991**). "On the effect of interaural phase differences on loudness," Hearing Research. **53**, 141 – 152.

Zwicker, E. and Fastl, H. (**1999**). *Psychoacoustics,* 2nd Ed., (Springer-Verlag, Berlin), pp. 8-19.

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

Author's Resume

Pandurangarao N. Kulkarni received the B.E degree in Electronics and Communication Engineering from Karnatak University, Dharawad, India, in 1988 and the M.Tech. degree in Digital Electronics and Advanced Communication from Mangalore University, Mangalore, India, in 1995. Presently, he is working as an Assistant Professor in Basaveshwar Engineering College, Bagalkot, India, and he is pursuing his Ph.D. in Electrical Engineering at the Indian Institute of Technology Bombay, India. His research interests include speech processing for the hearing impaired, digital signal processing and applications, and digital communication.

Thesis Related Publications

Submitted to journals

Kulkarni, P. N., Pandey, P. C., and Jangamashetti, D. S. “Binaural dichotic presentation to reduce the effects of spectral masking in moderate bilateral sensorineural hearing loss,” submitted to Int. J. Audiology.

Kulkarni, P. N., Pandey, P. C., and Jangamashetti, D. S. “Multi-band frequency compression for improving speech perception by listeners with moderate sensorineural hearing loss,” submitted to Speech Communication.

Journal papers

Kulkarni, P. N., Pandey, P. C., and Jangamashetti, D. S. (**2007**). “Multi-band frequency compression for reducing the effects of spectral masking,” Int. J. Speech Tech. **10**, 219-227.

Kulkarni, P. N., Pandey, P. C., and Jangamashetti, D. S. (**2006**). “Perceptually balanced filter response for binaural dichotic presentation to reduce the effect of spectral masking,” (abstract) J. Acoust. Soc. Am. **120**, 3253.

Conference Proceedings: International

Kulkarni, P. N., Pandey, P. C., and Jangamashetti, D. S. (**2010**). “Study of perceptual balance for designing comb filters for binaural dichotic presentation,” Proc. 20th Int. Cong. Acoust. (ICA 2010, Sydney, Australia, 23-27 Aug. 2010), Paper no. 556.

Jangamashetti, D. S., Cheeran, A. N., Kulkarni, P. N., and Pandey, P. C. (**2010**). “Simulation of increased masking in sensorineural hearing loss for a preliminary evaluation of speech processing schemes,” Proc. 20th Int. Cong. Acoust. (ICA 2010, Sydney, Australia, 23-27 Aug. 2010), Paper no. 980.

Kulkarni, P. N., Pandey, P. C., and Jangamashetti, D. S. (**2009**). “Multi-band frequency compression for sensorineural hearing impairment,” Proc. 16th Int. Conf. Digital Signal Processing, (DSP 2009, Santorini, Greece), Paper S4P.1.

Kulkarni, P. N., and Pandey, P. C. (**2007**). “Effect of binaural dichotic presentation with critical bandwidth based comb filters on source localization,” Proc. 19th Int. Cong. Acoust. (ICA 2007, Madrid, Spain), paper PPA-09-005.

Conferences Proceedings: National

Kulkarni, P. N. and Pandey, P. C. (**2008**). “Optimizing the comb filters for spectral splitting of speech to reduce the effect of spectral masking,” Proc. IEEE Int. Conf. Signal Processing Networking (ICSCN-08, Chennai, India), 69-73.

Kulkarni, P. N., and Pandey, P. C. (**2008**). “Frequency mapping for multi-band frequency compression for improving speech intelligibility,” Proc. 14th National Conference on Communications *(*NCC 2008, Indian Institute of Technology Bombay, Mumbai, India), pp. 437-441.

Kulkarni, P. N., Pandey, P. C., and Jangamashetti, D. S. (**2008**). “Study of perceptual balance in comb filter based spectral splitting of speech signal to reduce the effect of frequency masking,” Proc. Frontiers of Research in Speech and Music (FRSM 2008, Jadavpur University, Kolkata, India), pp. 101-105.

Acknowledgements

With deep regards and profound respect, I take this opportunity to express my gratitude and indebtedness to my supervisor Prof. P. C. Pandey for support, inspiring guidance, motivation, and constructive criticism which have made this work possible. My sincere thanks and gratitude to my co-supervisor Prof. D. S. Jangamashetti for her guidance and support.

I am thankful to Prof. V. M. Gadre and Prof. Preeti Rao, members of the research progress committee for my thesis, for their valuable suggestions and encouragements at various stages of the work.

I am thankful to all the friends in the Signal Processing & Instrumentation (SPI) Lab at IIT Bombay for making my time enjoyable. My special thanks go to Alice, Milind, Parveen, Vinod, Jayan, Mohan, Vidyadhar, Rajath, and Nataraj for their encouragement and support. It indeed has been an honor and privilege to work and live in the wonderful environment of the institute. So many people have enriched my life here and I am surely blessed to have met them all, although I am unable to thank them all.

I am grateful to Dr. Kiran Kalburgi, Dr. Doddamani, and authorities of S. N. Medical College, Bagalkot, for extending all the support in carrying out listening tests. I am thankful to all the subjects who participated in the listening tests.

For sponsoring my doctoral studies, I wholeheartedly thank Sri. V. C. Charantimath, Chairman, B. V. V. Sangha, Bagalkot and Dr. R. N. Herkal, Principal, Basaveshwar Engineering College, Bagalkot. I am indebted to Prof. S. M. Iddalagi, Head of E&C Department for all his encouragement and support. I am also thankful to Dr. S. G. Hiremath and Dr. S. S. Manvi, former heads of E&C Department for their help. Special thanks to my friend Shankar Kambalimath for the encouragement, help, and moral support throughout the course of my research. I express my sincere thanks to all the faculty members and office staff of my department for their help and good wishes. My special thanks to those who participated in the listening tests.

I am grateful to my mother, mother-in-law, brother Arvind and his wife Vijayalaxmi for their blessings and moral support. I am also thankful to Pavan and Prasanna for their love and affection. My sincere thanks to brother-in-laws Dr. Ravi and Krishna for their best wishes and encouragement. Words can’t convey my feelings towards my wife Shailaja and daughters Meghana and Anjana for their patient toleration of my disposition during the period of my research work. I have deprived them many happy moments but they strove to give me joy, support, and encouragement at the home front.

The research work was partly supported by a project grant under the National Programme on Perception Engineering, sponsored by the Department of Information Technology, MCIT, Government of India.

Pandurangarao N. Kulkarni