34

SPEECH PROCESSING FOR BINAURAL DICHOTIC PRESENTATION

A dissertation
submitted in partial fulfillment of the requirements
for the degree of
Master of Technology

by

Munish S. Ratanpal
(Roll no. 98307014)

Guide
Prof. P. C. Pandey

Department of Electrical Engineering
Indian Institute of Technology, Bombay
January 2000

Indian Institute of Technology, Bombay

Dissertation Approval

The thesis entitled "Speech processing for binaural dichotic presentation", by Shri Munish S. Ratanpal (Roll. No. 98307014) is approved for award of the degree of **Master of Technology** in Electrical Engineering with specialization in Control & Instrumentation.

Supervisor	--	(Prof. P. C. Pandey)
Internal examiner	VGadre --	(Prof. V. M. Gadre)
External examiner	vkmadan --	(Dr. V. K. Madan)
Chairperson	BBandyopadhyay --	(Prof. B. Bandyopadhyay)

Date : 7th February 2000

Munish S. Ratanpal: *Speech processing for binaural dichotic presentation*, M.Tech. dissertation, Department of Electrical Engineering, I.I.T Bombay, Jan. 2000,
Guide: Prof. P.C.Pandey

Abstract

Persons with sensorineural hearing impairment experience degraded speech perception due to spectral masking along the cochlear partition. Splitting the speech signal by filtering it with a filter bank and adding signals from alternate bands is likely to reduce the effect of spectral masking and thus help in increasing speech intelligibility.

A scheme for dichotic presentation by splitting speech into two signals with complementary spectra on the basis of critical band filters, has been developed and evaluated earlier by D. S. Chaudhari in 1999. This project involves the selection of optimal filter coefficients for the speech processing scheme, implementation and testing of the scheme on a PC interfaced to sound/multimedia card, and development of tools for spectrographic analysis.

ACKNOWLEDGMENT

I would like to take this opportunity to thank my esteemed guide Dr. P. C. Pandey for his valuable and consistent help throughout the course of this work. I am grateful for the help and encouragement of Shri D. S. Chaudhari, on whose research work the present project is based. I also express my sincere thanks to my friends and colleagues in the SPAI Lab for their company and support during project work.

14th Jan, 2000 Munish S. Ratanpal

IIT Bombay

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1
INTRODUCTION

1.1 Problem overview

Sensorineural hearing impairment is characterized by an increase in threshold of hearing, reduction in dynamic range of hearing, and increased temporal and spectral masking.

The ability to perceptually combine the binaurally received signals from the two ears improves speech perception under adverse listening conditions. Splitting the speech signal into two complementary parts on the basis of frequency and presenting it binaurally might reduce the effect of spectral masking and thus improve speech intelligibility. Thus, the binaural hearing aid which can split the speech signal on the basis of critical bands, can be helpful to hearing impaired persons. Research on binaural dichotic presentation has been carried out earlier by Lyregaard [1], Lunner *et al.* [2], and Chaudhari & Pandey [3].

1.2 Project objectives

Chaudhari has investigated a scheme for splitting speech into two signals with complementary spectra on the basis of critical bands to lessen the effect of reduced auditory frequency selectivity and thus improve speech intelligibility [3].

This project aims to improve the implementation of speech-processing scheme by selection of optimal filter coefficients for an appropriate trade-off between roll-off rate at band transitions and sidelobes in stopbands. The filters used for speech processing will be designed using a graphical user interface – based software for frequency sampling method of FIR filter design. Also, the dichotic presentation technique is to be implemented using a PC and sound card, which can facilitate the recording and playback of sound. The processing scheme should provide adjustable filter gains as a way of partial matching of the filter response to the frequency characteristics of the individual subjects hearing loss.

For analysis of speech signals, a spectrographic analysis package is required, which can display the time-varying spectrum of speech. The software should be able to run on a PC that has a sound card and a microphone, without the need for any special signal acquisition hardware or DSP peripherals.

1.3 Outline of the dissertation

Chapter 2 gives an overview of binaural dichotic presentation. Chapter 3 describes the spectrographic analysis package. Chapter 4 describes filter optimization for reduced amplitude of sidelobes. Chapter 5 describes the implementation of binaural dichotic presentation using sound card. Chapter 6 gives a summary of work done and suggestions for future work.

Chapter 2
BINAURAL DICHOTIC PRESENTATION

2.1 Introduction

This chapter provides an overview of various types of hearing impairment, the perceptual effects of sensorineural hearing loss and a review of binaural dichotic presentation schemes for improving speech perception in case of sensorineural hearing loss.

2.2 Hearing impairment

Hearing impairment can be classified into four major categories: conductive loss, sensorineural loss, central loss, and functional loss. Conductive loss is caused due to an abnormality before the cochlea, typically in the middle ear, ear drum or ear canal structures. It leads to attenuation of the incoming sounds, resulting in increase in hearing thresholds. However frequency selectivity is not affected [4].

Sensorineural loss is most commonly caused by a defect in the cochlea and is then known as cochlear loss. But it may also result due to defects in the auditory nerve or higher centers in the auditory system and is then known as retrocochlear loss. Unlike conductive loss, which is often medically or surgically treatable, sensorineural loss is typically permanent and in certain cases becomes progressively worse with time [5].

Central loss may be due to damage to the auditory cortex by cerebral hemorrhage, meningitis, etc. It causes reduction in auditory comprehension ability. The causes of functional deafness are more psychological rather than physiological [4].

2.3 Effect of sensorineural loss on speech perception

Threshold of hearing is the minimum level at which a tone can be perceived in the absence of other sounds. The loudness discomfort level is the highest level to which a tone can be raised before a person finds it uncomfortable. The dynamic range is the difference between the threshold and the loudness discomfort level.

Persons with sensorineural hearing loss experience an increase in hearing thresholds. But the loudness discomfort level is unaffected, thus reducing the dynamic range [6]. Also, there is an abnormal increase in perceived loudness with increase in sound level, which is called loudness recruitment. The reduction in dynamic range is illustrated in Fig. 2.1. The curves A, B and C show the typical hearing thresholds for persons suffering from mild-to-moderate, severe and profound impairment respectively. As the loudness discomfort level is similar for both normal as well as sensorineurally hearing impaired persons, it is shown by a single dashed line at the top. The area between the threshold curve and the loudness discomfort level is known as residual hearing area. It can be seen that the residual hearing area becomes progressively smaller with increasing hearing loss. Normal conversational speech covers a range of at least 30 dB [7], which may be more than the dynamic range of hearing for sensorineurally hearing impaired persons.

Sensorineural loss also results in degradation of temporal resolution and increase in temporal masking [8]. Temporal resolution is the minimum detectable gap between two successive signals, while temporal masking refers to the effect of intense sounds on preceding and following sounds.

Widening of auditory filters resulting in an increase in spectral masking is also one of the important effects of sensorineural loss [9]. This leads to loss of frequency selectivity. Spectral masking refers to the phenomenon of adjacent frequency components of the signal suppressing the perception of each other. Frequency selectivity refers to the ability to detect a tone in the presence of a complex stimulus. The effect of poor frequency selectivity is to reduce the difference in amplitudes of the spectral peaks and troughs in the perceived sound. This makes hearing more susceptible to interfering sounds, and may degrade speech perception due to confusion in identification of consonantal segments [10] [11].

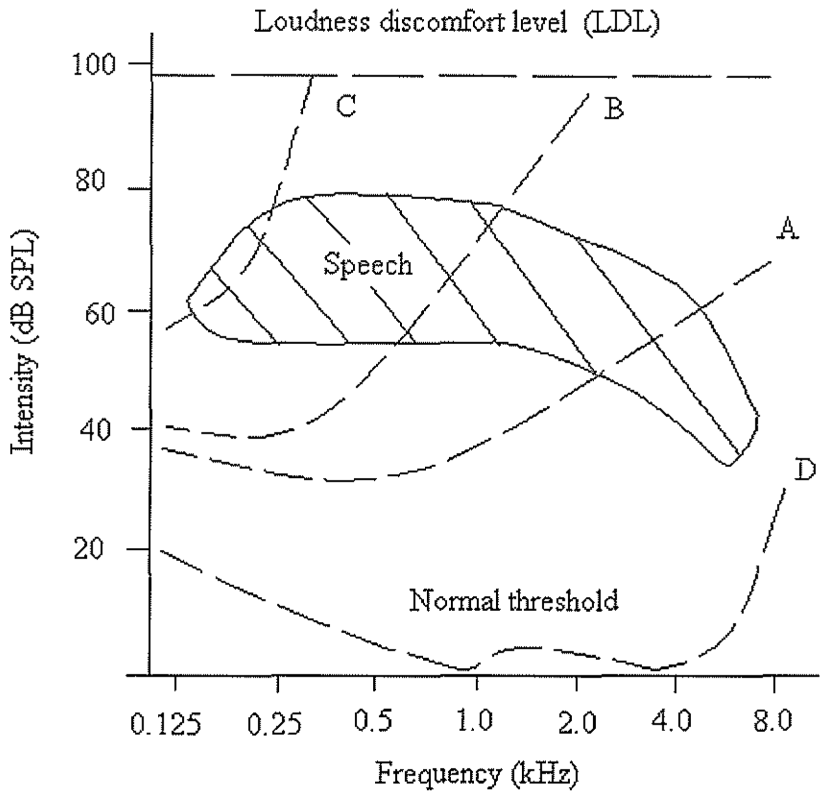


Fig. 2.1) Illustrates the dynamic range of impaired auditory system. Curves A, B and C are hearing thresholds for sensorineural impairment of various degrees. Hatched portion indicates area within which typical speech spectra lie. Source: [6]

2.4 Review of earlier dichotic presentation schemes

Most of the speech processing schemes for hearing aids involve monaural listening which refers to sound presentation to one ear only, whereas binaural listening involves both the ears. Binaural listening could be "diotic" with the same signal presented to the two ears or it could be "dichotic" with different signals presented to the two ears.

It was suggested by Fletcher [12] that the peripheral auditory system behaves like a bank of bandpass filters, with overlapping passbands. These filters are called auditory filters or critical bands. Fletcher's experiment used a sinusoidal signal masked by a bandpass noise which had the same center frequency as the signal. The power density of the masker was held constant and the threshold of the signal as a function of the bandwidth of the noise was measured. The bandwidth at which the threshold did not increase any further was the critical bandwidth. Subsequently, different researchers have used various methods for estimating the critical bands. In 1976, Patterson [13] introduced a method which used a fixed frequency signal masked by a noise with a bandstop centered at the signal frequency. The shape of the auditory filter was estimated by taking the first derivative of the function relating the tone threshold to the width of the bandstop. The Fig. 2.2 shows the variation of critcal bands with frequency based on traditional estimates and more recent estimates using the notched-noise method. It is seen that the bandwidth of these filters increases with increasing frequency. for frequencies above 1 kHz, the traditional and more recent estimates match and the bandwidth is approximately 1/3rd octave. For frequencies below 1kHz, the traditional estimates of the bandwidth show it to be constant at about 100 Hz, while the more recent estimates show that the bandwidth is linearly related to frequency [14] [15].

As stated earlier, sensorineural loss is associated with widening of the auditory filter bandwidths and reduced frequency selectivity. We can take advantage of the fact that the reduced frequency selectivity takes place at the peripheral auditory system. Thus, it may be possible to improve speech perception by filtering the speech using a bank of critical band filters and adding outputs of alternate bands to obtain two signals for binaural dichotic presentation.

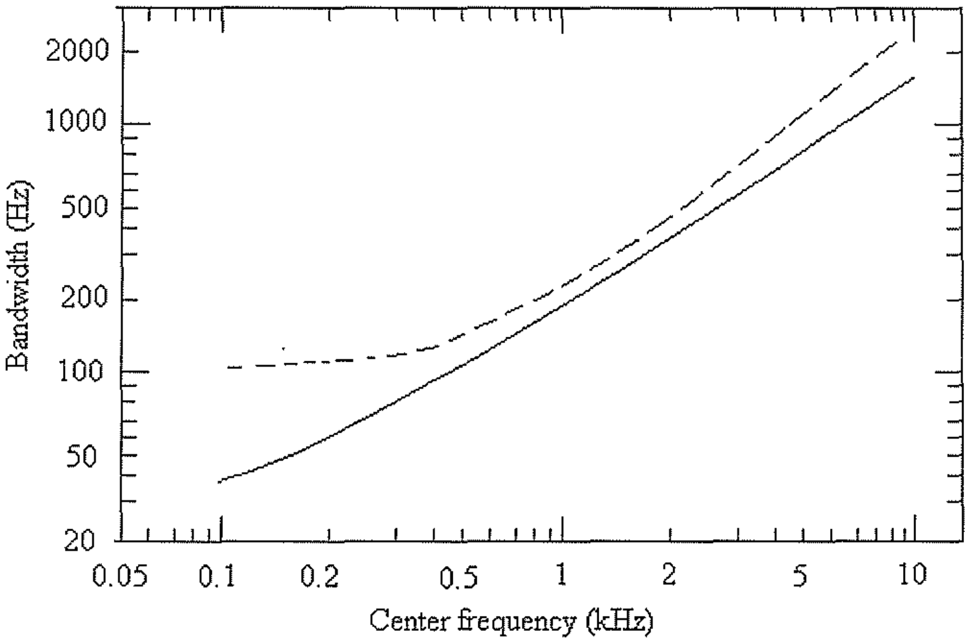


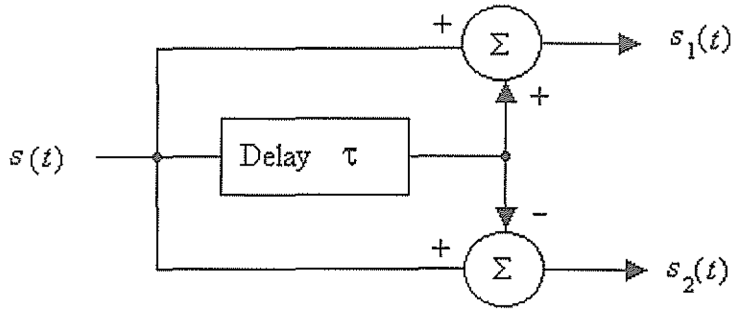
Fig. 2.2) Critical bandwidth as a function of the center frequency. The dashed curve shows the traditional estimate [17], while the solid curve shows more recent estimates of equivalent rectangular bandwidth using the notched noise method. Adapted from [5] Fig. 3.10

Lyregaard evaluated a scheme for dichotic presentation to improve speech intelligibility in noise due to reduced frequency selectivity [1]. The scheme, shown in Fig. 2.3 used comb filters realized using an analog delay, so as to present signals containing different frequency components to the two ears. The listening tests were conducted on three hearing impaired subjects with both diotic as well as dichotic presentation

The test material consisted of two lists of 25 words each, and a speech-spectrum shaped background noise was presented at S/N ratios of 12 and 4 dB speech intelligibility. Experiments were conducted with three values of comb filter bandwidths- 200, 500 and 800 Hz. However, improvements in the scores for dichotic over diotic presentation were not statistically significant.

Lunner *et al.* tested the use of an 8-channel digital filter bank in diotic and dichotic modes [2]. The set-up is shown in Fig. 2.4. The bank consisted of complementary interpolated linear-phase FIR filters, so that most of the coefficients were zero-valued. The filter bandwidths were approximately 700 Hz., and had individually adjustable gains to match the frequency response with the individual user's needs. The scheme achieved spectral splitting, but the filters did not have sharp transition bands. The system was implemented using TI TMS320C25 processor and a PC. A portable version was also built using the TMS320E25 processor. Dichotic presentation was done by combining the signals from alternate bands. The evaluation of the scheme was done by finding the speech-to-noise ratio which satisfied the 50 % correct word recognition criterion. The results indicated an improvement in speech-to-noise ratio of 2 dB for the dichotic conditions over diotic.

In 1996, Mithal as part of his B.Tech project at IIT Bombay [16] conducted experiments involving spectral splitting of speech into two signals using filters with 1/3 octave bandwidth as an approximation to critical bands. The listening tests were conducted under two conditions- (1) diotic presentation in which all bands are presented to both ears, (2) dichotic presentation of even numbered bands to the left and odd numbered bands to the right. Hearing loss was simulated using white noise, band-limited to 4.6 kHz, as a masker for obtaining different signal-to-noise ratios. The improvement in



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

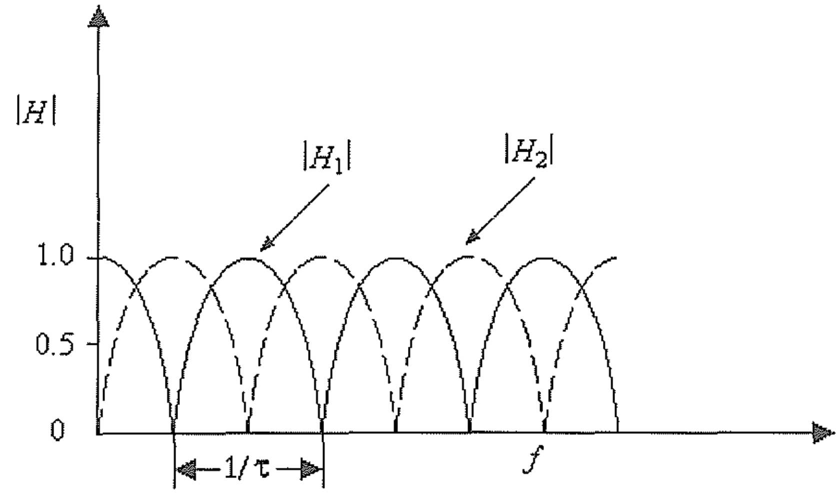


Fig. 2.3) A schematic of the comb-filtering scheme used by Lyregaard [1]. $s(t)$ is the input and $s_1(t)$ and $s_2(t)$ are the output signals presented to the two ears. $|H_1(f)|$ and $|H_2(f)|$ are the magnitude responses of the two channels. Source:[23]

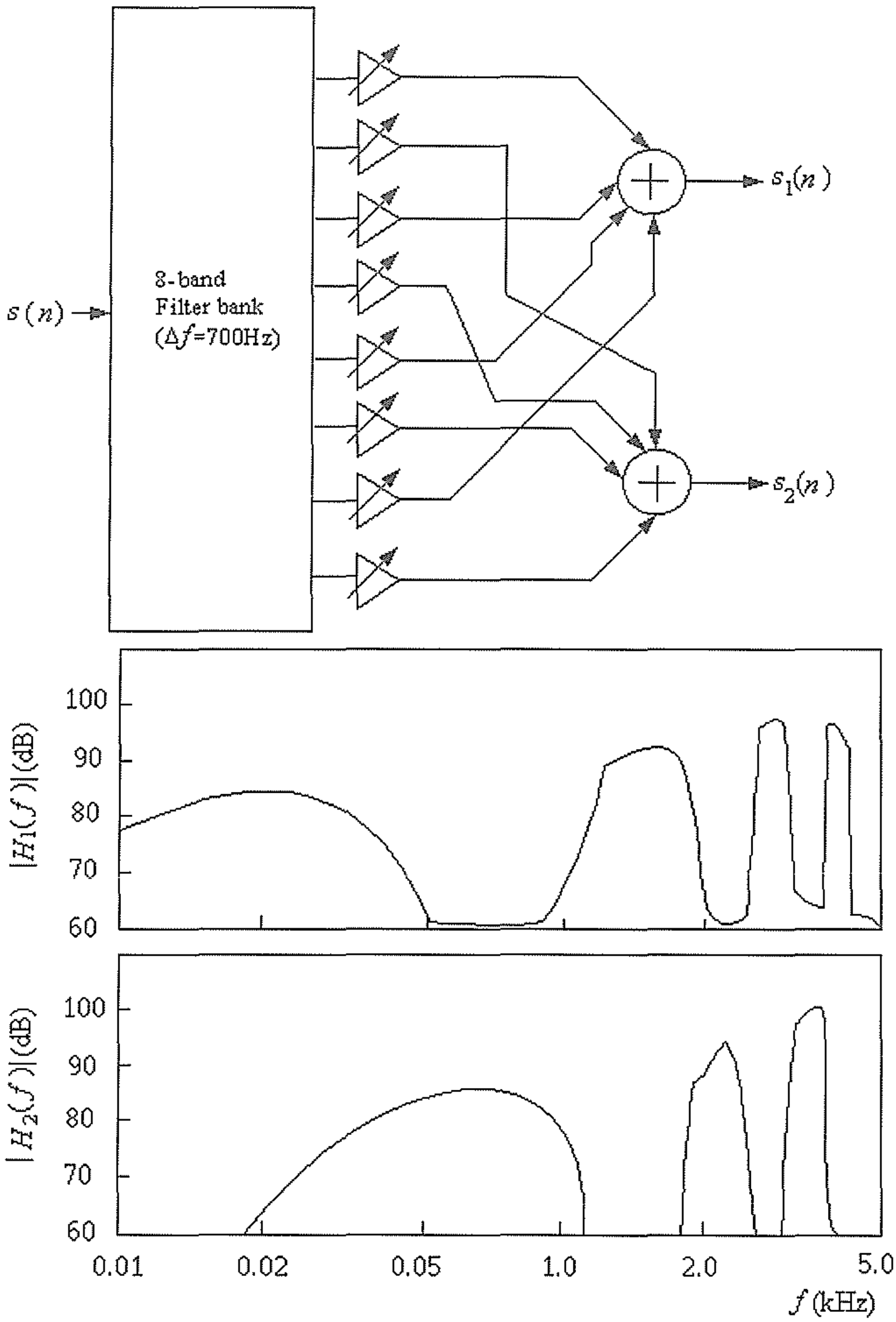


Fig. 2.4) Schematic diagram and magnitude response of the TMS320C25 processor-based filter bank for dichotic presentation developed by Lunner *et al.* [2]. $s(n)$ is the input and $s_1(n)$ and $s_2(n)$ are the output waveforms for presenting to the two ears. Based on description given in Lunner et al [2].

speech recognition scores due to processing for dichotic presentation was in the range of 1.3–3.7 % and 3-4.7 % for SNR of 3 and 6 dB respectively.

These studies formed the basis of a speech-processing scheme for binaural dichotic presentation based on critical band filtering. This scheme is described in the next section, and forms the basis for the current project.

2.5 Dichotic presentation based on critical band filtering

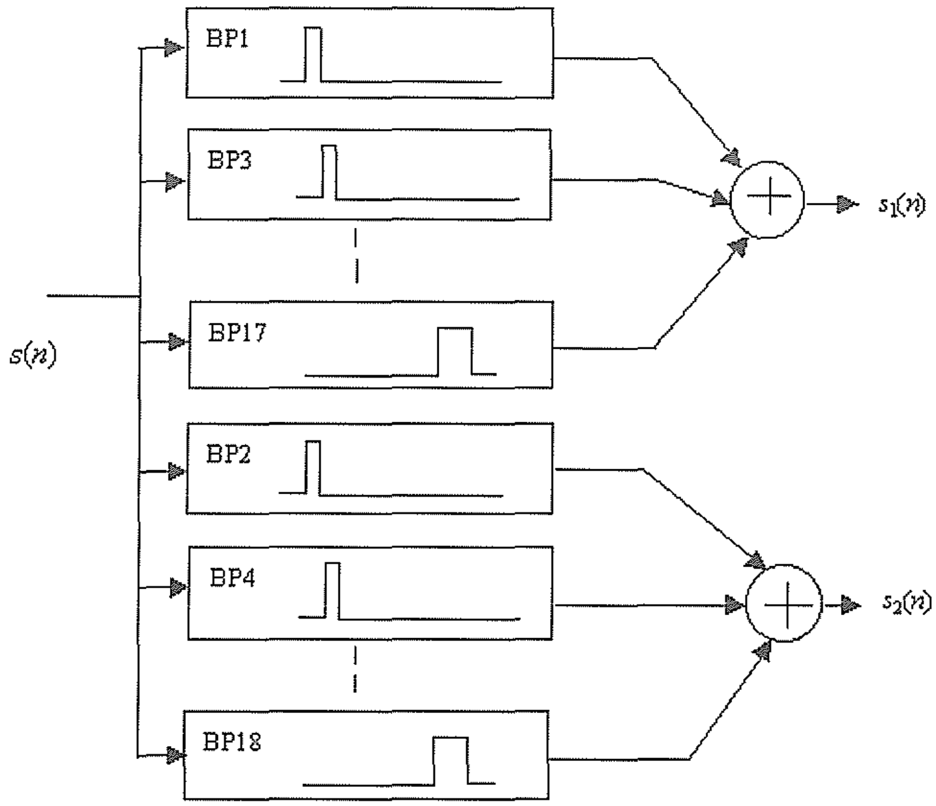
In view of the earlier studies, Chaudhari, as part of his Ph. D. research at IIT Bombay has implemented an audio-signal processing scheme for splitting speech for binaural dichotic presentation, using eighteen critical bands corresponding to auditory filters as described by Zwicker [17] [5] and combining the odd and even numbered critical bands. Speech processing for dichotic presentation can be implemented in three different ways:

1] Fig. 2.5 shows a scheme with two parallel filter banks, each with nine band pass filters. The two output signals are obtained by adding the outputs of the filter in the bank. But in this scheme, there is a possibility of notches in the magnitude at crossover frequencies due to phase shifts in the adjacent filters.

2] Fig. 2.6 shows a scheme in which the splitting of the two signals is done using cascade combination of band reject filters. The phase shifts of individual filters do not affect the magnitude response in this case.

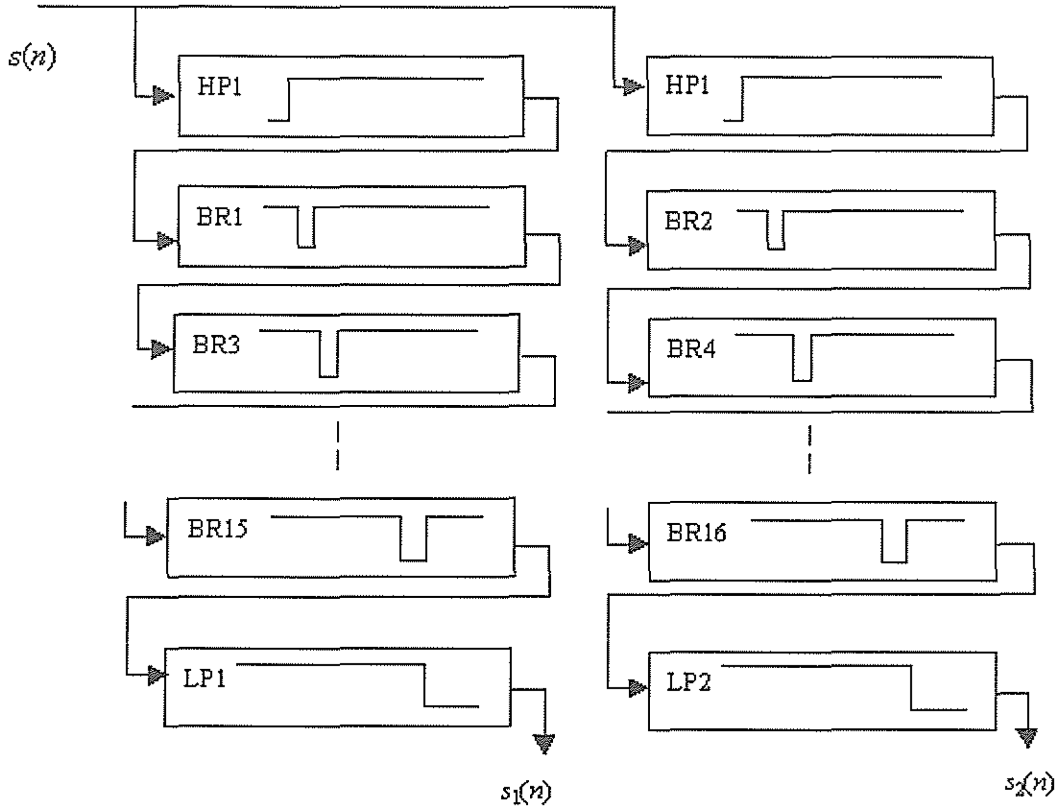
3] Fig. 2.7 shows two filters with the desired comb filter response. This results in overall efficiency of realization.

The scheme for dichotic presentation was implemented for both off-line and real-time processing. For off-line processing, the cascade combination of band reject filters shown in Fig. 2.5 was selected. The signal acquisition was done using analog interface circuit of a TI / TMS320C50 based DSP board. The processing scheme was implemented off-line by means of a C program. Linear-phase FIR filters with 255 coefficients were used to obtain sharp cutoff. As shown in Fig. 2.8, the processed speech was output using the two D/A channels of PCL 208 data acquisition card. The D/A outputs were passed through a pair of lowpass filters and a pair of audio amplifiers. Since processing was



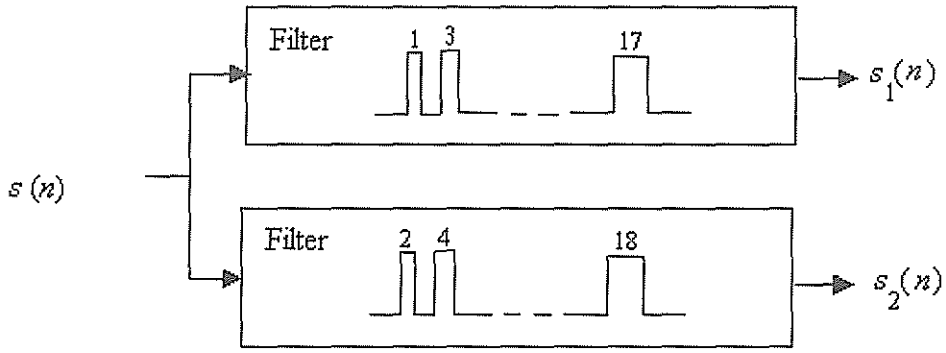
Filter	Freq(kHz)	Filter	Freq(kHz)
BP1	0.10– 0.20	BP2	0.20-0.30
BP3	0.30-0.40	BP4	0.40-0.51
BP5	0.51-0.63	BP6	0.63-0.77
BP7	0.77-0.92	BP8	0.92-1.08
BP9	1.08-1.27	BP10	1.27-1.48
BP11	1.48-1.72	BP12	1.72-.2.00
BP13	2.00-2.32	BP14	2.32-2.70
BP15	2.70-3.15	BP16	3.15-3.70
BP17	3.70-4.40	BP18	4.40-5.00

Fig. 2.5) Splitting of speech signal using two banks of bandpass filters. Each filter block has its own magnitude response. Source: [23]



Filter	Freq(kHz)	Filter	Freq(kHz)
HP1	0.07	HP2	0.20
BR1	0.20-0.30	BR2	0.30-0.40
BR3	0.40-0.51	BR4	0.51-0.63
BR5	0.63-0.77	BR6	0.77-0.92
BR7	0.92-1.08	BR8	1.08-1.27
BP9	1.27-1.48-	BR10	1.48-.1.72
BP11	1.72-2.00	BR12	2.00-2.32
BP13	2.32-2.70-	BR14	2.70-3.15
BP15	3.15-3.70	BR16	3.70-4.40
LP1	4.40	LP2	5.00

Fig. 2.6) Splitting of speech signal using cascade combination of band-reject filters. Each filter block has its own magnitude response. Source: [23]



Filter	Freq(kHz)	Filter	Freq(kHz)
BP1	0.10 -0.20	BP2	0.20-0.30
BP3	0.30-0.40	BP4	0.40-0.51
BP5	0.51-0.63	BP6	0.63-0.77
BP7	0.77-0.92	BP8	0.92-1.08
BP9	1.08-1.27	BP10	1.27-1.48
BP11	1.48-1.72	BP12	1.72-.2.00
BP13	2.00-2.32	BP14	2.32-2.70
BP15	2.70-3.15	BP16	3.15-3.70
BP17	3.70-4.40	BP18	4.40-5.00

Fig. 2.7) Schematic representation for splitting of speech signal using two comb filters. Each block has its own magnitude response. Source: [23]

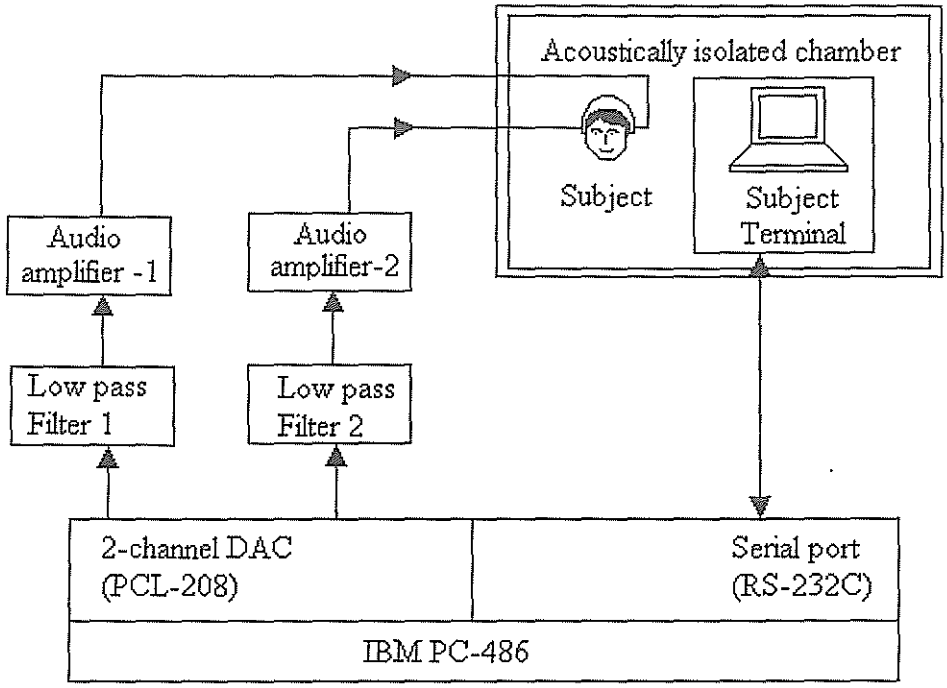


Fig. 2.8) Experimental set-up used in the computerized test administration system for listening tests used in off-line processing. Source: [23].

being done off-line, the higher computational complexity of this scheme was not considered important.

For real-time processing, two DSP boards using TI/TMS320C50 16-bit fixed-point were used. The set-up is shown in Fig. 2.9. The stimuli in the form of pre-recorded speech files were output at a rate of 10 kSa/s through one D/A port of the PCL-208 data acquisition card. Diotic presentation was done by passing the D/A output through lowpass filter and a pair of audio amplifiers. For dichotic presentation, the lowpass filter output was given to the two DSP boards which had filter programs running on them. The magnitude response was approximated with 128 coefficients using frequency-sampling technique of linear phase FIR filter design. One of the serial ports of the computer was used for loading the filter program to the DSP board through a switch, while the other was used for communicating with the subject terminal. The real time processing was implemented in two ways. In the first method, the gains of all filter bands were kept constant. The second implementation provided adjustable gains in the range of –3 to +3 dB to compensate for the frequency characteristics of the subjects' hearing loss.

The test material for the processing schemes consisted of twelve consonants /p, b, t, d, k, g, m, n, s, f, v/ in vowel-consonant–vowel (VCV) and consonant-vowel (CV) context, the vowel being /a/. For off-line implementation, experiments were carried out both with normal subjects with hearing loss simulated by broadband noise, and with hearing impaired subjects. For normal hearing subjects, it was found that the recognition scores for processed speech were higher for a particular level of masking noise. Hearing impaired subjects showed significant improvement in recognition scores, with average improvements in place feature of 29 % and 25% in VCV and CV contexts respectively.

Real-time implementation was carried out in two ways - PS-CG, in which the gains of all filter bands were constant and PS-AG, in which the gains of the filters were varied in the range of $\pm$ 3 dB. The percentage relative improvements in scores were highly significant and ranged from 9.2 to 23.6 and 14.4 to 19.2 in VCV and CV contexts respectively, with nearly maximum improvement for place feature. With adjustable gain, the percentage relative improvement in scores ranged from 2.2 to 6.4 and 1.6 to 7.8 in VCV and CV contexts respectively. It was seen that PS-CG improved quality of speech

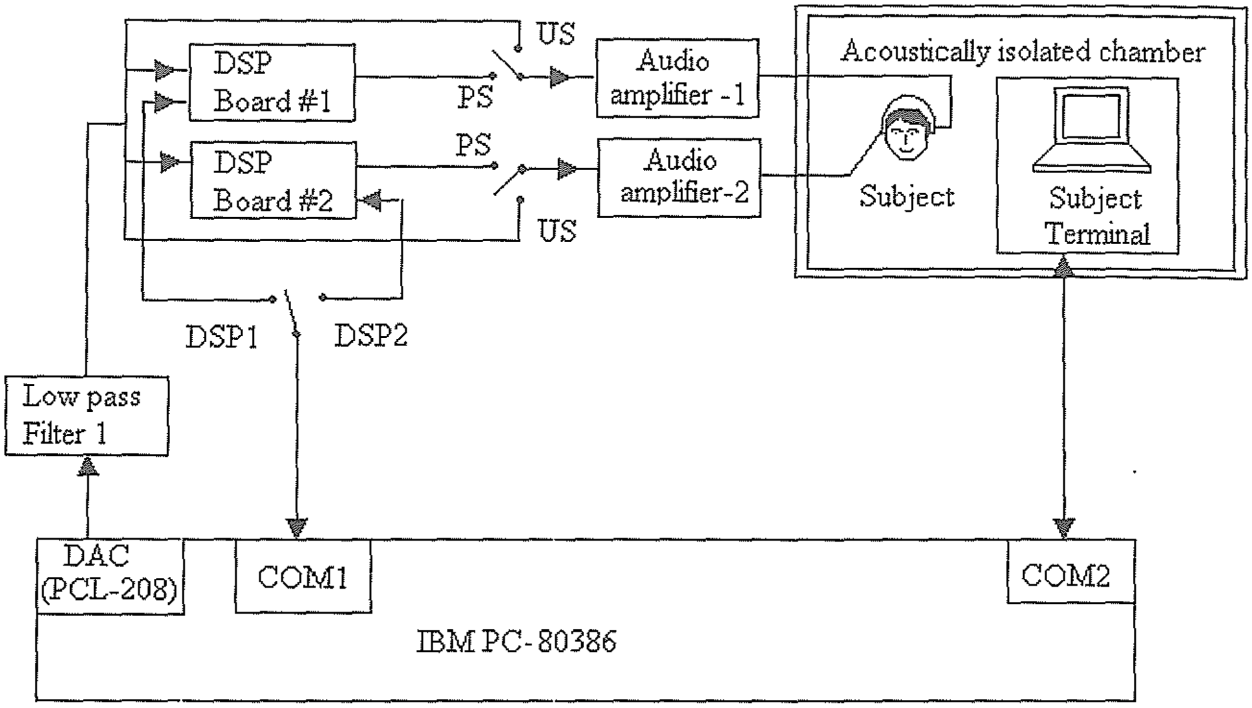


Fig. 2.9) Experimental set-up used in the computerized test administration system for listening tests used in real-line processing. Source: [23].

and resulted in improved recognition of place feature, thus indicating that the scheme reduces the effect of spectral masking. Further shaping of filter bands did not contribute to reduction of spectral masking. On the basis of these listening tests, it has been decided to implement the processing scheme for binaural dichotic presentation by means of 128-coefficient comb filters.

2.6 Scope for further work

It has been established by Chaudhari's work that the splitting of the signal on the basis of critical band filters can improve speech perception for persons with bilateral hearing loss. The objective of this project was to carry out further work towards improved implementation and testing of the scheme. In the processing scheme of Lyregaard [1] and Lunner *et al.* [2], emphasis was on a very efficient realization of the filter. The magnitude response of the filters shows a very smooth roll-off in the transition. In the critical band scheme used by Chaudhari, a 128-coefficient FIR filter has been used and the transitions are relatively sharp. It will be of interest to design the filter in such a way that the sidelobes in the stop band are minimized without sacrificing the sharp transitions of the critical band comb filters.

The spectrographic analysis set-up used for analysis of speech signals could be modified to take advantage of increase in the PC processor speeds and availability of sound card. Also, the dichotic presentation technique could be implemented as a part of multimedia system of PC. The three following chapters present the development work carried out in these three areas.

Chapter 3
SPECTROGRAPHIC ANALYSIS SET-UP

3.1 Introduction

For analysis of speech signals, a spectrographic analysis set-up is required which can display the time-varying spectrum of speech. In a spectrogram, time varying spectral characteristics are displayed as a two-dimensional plot, with time and frequency along x and y axes respectively. The spectral magnitudes as a function of time and frequency are viewed as intensity variations [15] [18]. The digital spectrographic analysis involves short-time Fourier analysis of the acquired signal, conversion of the spectral magnitude to dB scale and display of these magnitudes as a function of time and frequency. This technique is particularly suitable for analysis of nonstationary signals such as speech.

3.2 Digital spectrographic analysis

Spectrograms can be generated by obtaining magnitude spectrum of digitized waveforms by using either a digital filter bank or short-time Fourier transform and displaying time-frequency plots.

The short-time Fourier transform of a sampled waveform is

$$X(n, k) = \sum_{m=0}^{N-1} w(m)\, x(n-m)\, e^{-j2\pi km/N} \quad \text{for } 0 \leq k \leq N\text{-}1 \tag{3.1}$$

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

$$w(m) = 0.54 - 0.46 \cos(2\pi m/(L\text{-}1)) \quad \text{for } 0 \leq m \leq L\text{-}1 \tag{3.2}$$

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

The frequency resolution of the spectrographic analysis with a particular window is equivalent to its bandwidth. For Hamming window, it is $\Delta f = 1.36 f_S / L$, where f_S is the sampling rate. For speech analysis, wideband spectrogram with spectral resolution of 300 Hz is useful in observing the voiced speech as vertical striations and for seeing formant transitions. For unvoiced speech, the vertical striations do not appear and the spectral pattern is much more ragged. On the other hand, narrow band spectrograms with spectral resolution of 45 Hz is useful for observing the fundamental frequency and its harmonics.

3.3 Previous set-up

A PC-based spectrographic analysis package was developed earlier at IIT Bombay by Thomas [19] [20] in the programming language Pascal. The spectrogram was displayed on an area of 500 x 256 pixels using VGA card with 16 simultaneous colors on a monochrome monitor. The digitized data file is created using a data acquisition card. The selected segment of the data file was divided into 500 overlapping time frames. An L-point Hamming window was applied to each frame and a 512-point DFT was then taken. The log magnitude in dB was computed and spectral information for each frame was displayed in the form of 256 vertical pixels. The pixel intensity indicated the spectral magnitude in shades of gray, with white used for minimum magnitude level and black used for maximum magnitude level. The program provided an option for generating a wideband, narrowband or a combined spectrogram to preserve both time and frequency resolution. The combined spectrogram was generated by evaluating the geometric mean of the wideband spectrogram and the narrowband spectrograms. Since all the computational tasks were performed by the PC, the execution speed was slowed down.

To increase the speed of execution, Baragi and Prasad in 1996 developed a spectrograph package, which divided the tasks between the PC and a DSP board based on TMS320C25 processor [21] [22]. In this package, a C program running on the PC handled the user-interface and display, while the DSP board handled the data acquisition and computationally intensive FFT routines. Analog signal conditioning circuit consisted of antialiasing lowpass filter at input and smoothing lowpass filter at the output. The PC performed windowing and pre-emphasis of the data frame, and downloaded it to the DSP board. The DFT size used was 256, and the resulting spectrogram was displayed on an area of 500 x 128 pixels. While FFT of one block is being calculated on the DSP board, the PC uploads 128 samples of the computed FFT of the previous block and computes the log magnitude. However, this program didn't have the feature of combined spectrogram.

D. S. Chaudhari in 1996 introduced the facility for capturing the displayed spectrogram and storing it in the Postscript format, which could be printed on a 600-dpi laser printer [23]. The DSP board performed the data acquisition and FFT computations as before. Fig. 3.1 shows the previous set-up for spectrographic analysis.

As this spectrographic analysis set-up was based on the DSP board, it was not portable to other computers. Also, as the PC could only transfer integer values to and from the DSP board, there was truncation error in the computation of FFT. The Postscript file displayed the spectrogram with a resolution of 16 gray levels, which can be increased to 256 gray levels [24]. Also, the spectrogram displayed on screen had a resolution of 16 levels, which can be increased to 64 gray levels using an SVGA card and compatible monitor. Also, the set-up could not display a monochrome spectrogram on a color screen. Due to the increase in processor speeds, and the easy availability and low cost of sound cards as compared to DSP boards, it was decided to modify the previous set-up which is described in the next section.

3.4 Present set-up

The present set-up utilizes the sound card available on most current PCs for playback and recording of the speech signal. The software consists of two programs. The main program, "Spec99.c", written in C language, handles the spectral analysis and display functions, and invokes the program "audio.cpp" written in Visual C++, which is used for recording and playback of sound. Fig. 3.2 shows the modified set-up. The software prompts the user for digitized input either from the microphone or from a pre-recorded speech file. The data acquisition routine has been developed jointly with A. V. Patil [25]. Spectral analysis is then carried out for selected sample range of data file. The selected segment is divided into 500 overlapping time frames. After pre-emphasis, an L-point

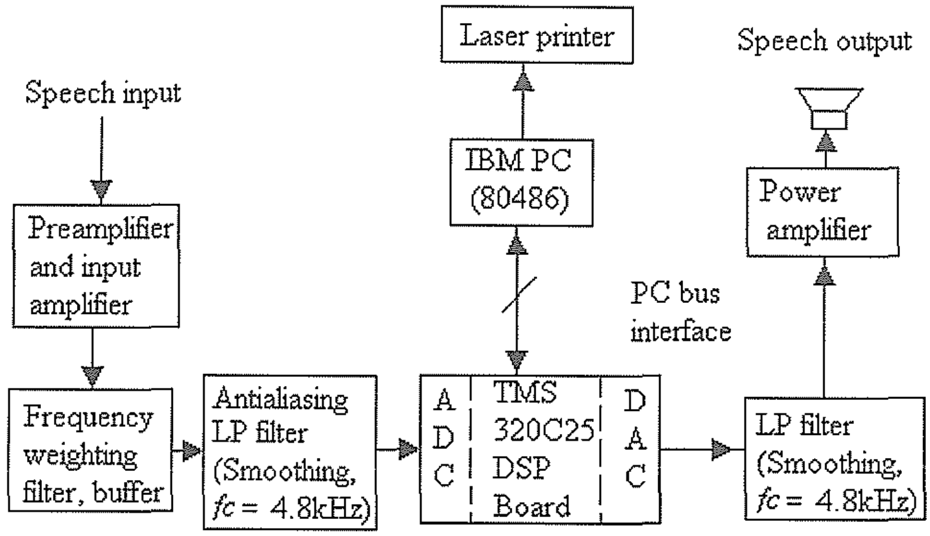


Fig. 3.1) Hardware set-up of the DSP board – based spectrographic analyzer. Source: [23]

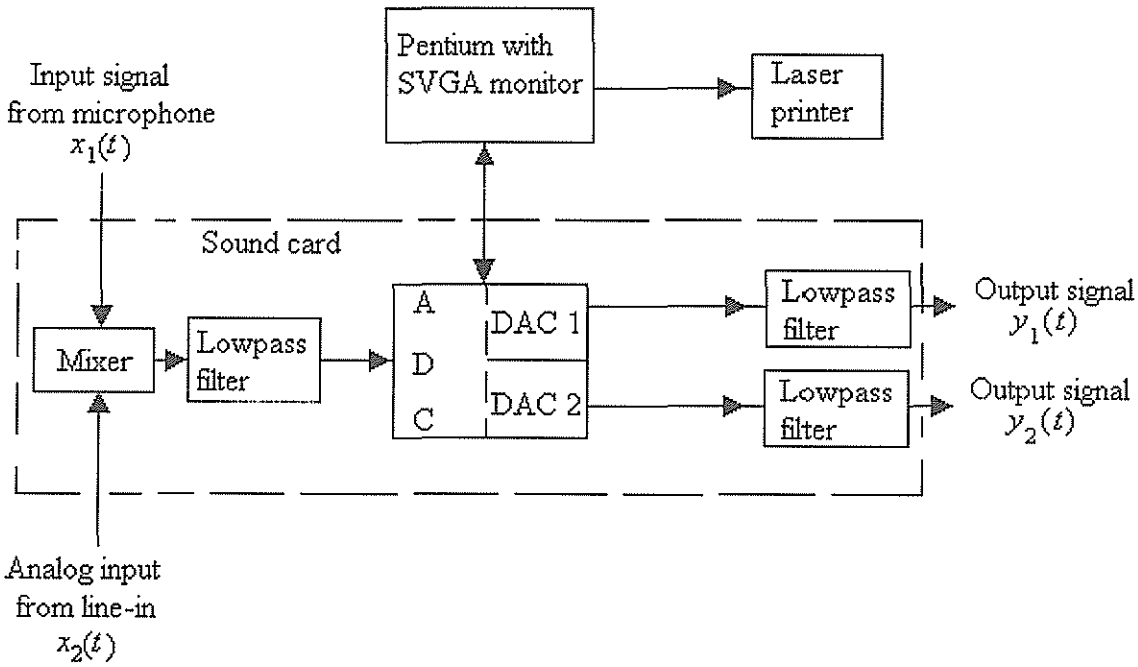


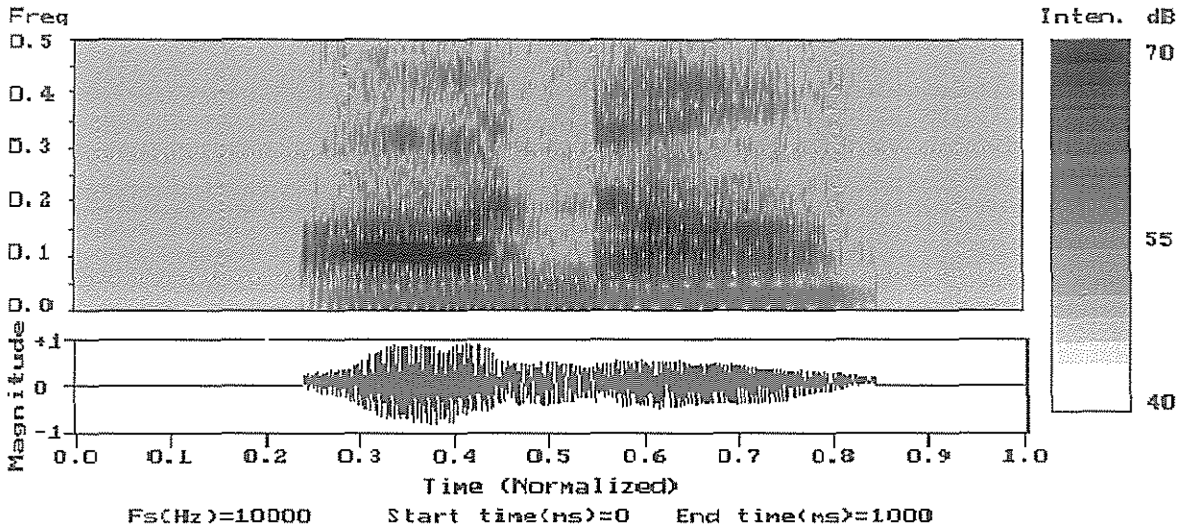
Fig. 3.2) Present set-up of the spectrographic analyzer

Hamming window is applied to each frame. Zero padding is then done so that the magnitude spectrum can be obtained using a 256-point FFT, irrespective of the window length. The spectral values $|X|_{db}(k)$ are scaled to obtain $|X|_{dbN}(k)$, in order to compensate for the variation in the energy of signal for different window lengths, due to zero padding.

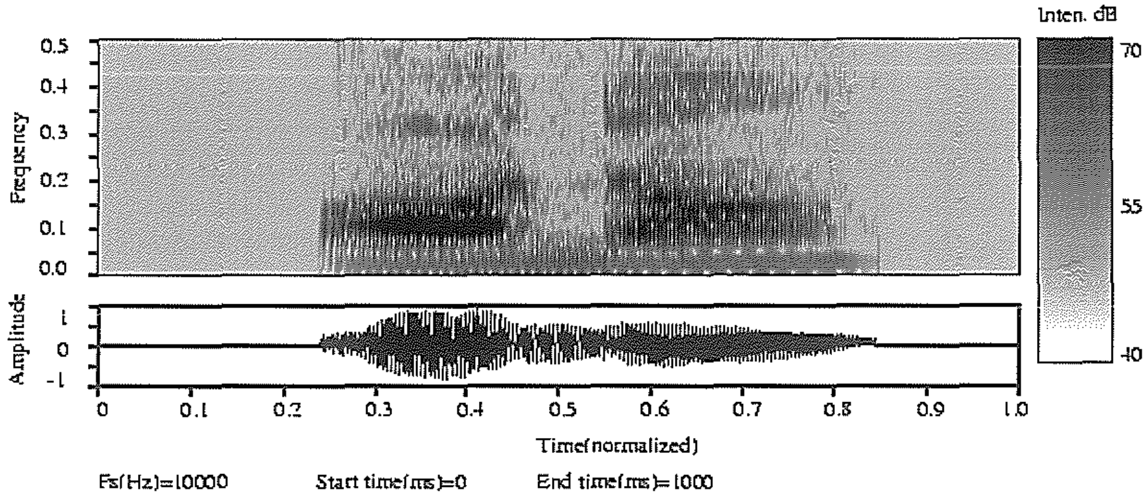
$$|X|_{dbN}(k) = |X|_{db}(k) + W_{dB}(L) \qquad (3.3)$$

where $W_{dB}(L) = 10 \log_{10}(\frac{N}{L})$ (3.4)

The scaling has been done assuming that signal being analyzed has a stationary short time RMS value. The log magnitude is then calculated and the spectrogram is displayed on an area of 500 x 128 pixels with a resolution of 64 gray levels. The spectrogram can be stored in the form of a Postscript file having 256 gray levels of resolution, thus providing a dynamic range of 48 dB. The selected portion of the data can be presented to the left, right or both ears. The software works on a color / monochrome monitor and on any PC with a sound card. Fig. 3.3 shows the spectrograms for the utterance */ana/* using the previous set-up, and with increased resolution using the new set-up.



a)



b)

Fig. 3.3) Wideband spectrograms of speech waveform for utterance */ana/*, a) using the previous set-up and b) using the new set-up

Chapter 4
COMB FILTER OPTIMIZATION

4.1 Introduction

The speech processing for binaural dichotic presentation is done by means of two comb filters which provide signals with complementary spectra. A real-time speech-processing scheme for binaural dichotic presentation was implemented by Chaudhari using two 128-coefficient FIR filters, with comb filter response corresponding to critical bands, on a set-up using two DSP boards [26]. This scheme attempts to reduce the effect of spectral masking due to widening of the critical bands by presenting alternate bands to the two ears. The filter transitions are relatively sharp and achieve spectral splitting. However, to reduce the sidelobes in the stop bands, we need to optimize the frequency response specification of the comb filters.

4.2 Filter optimization

There are two methods widely used for designing FIR filters. One method is to truncate the impulse response of an ideal filter to a finite duration sequence using a window function. The window function should reduce the filter transition overshoot and reduce the sidelobes in stop bands. This is known as windowing technique [27][28]. This approach does not provide adequate control over the parameters of pass band ripple, sidelobes in the stop band, and the width of the transition band. It is suitable for design of prototype filters. However, it cannot be used to realize an arbitrary frequency response, due to difficulties in carrying out the inverse Fourier transform.

Another approach is to determine the finite impulse response which satisfies a frequency response specified at N equidistant frequencies, where N depends on the filter order. The impulse response is determined by taking the IDFT of the specified response, and the impulse response samples directly give the filter coefficients of the FIR filter. This technique is known as frequency sampling method [27][28]. It can be used for designing filters with arbitrary frequency response, and is thus suitable for designing the comb filters based on critical bands. It results in a frequency response which coincides with the desired response at the specified frequency samples, but may have large deviations from it at other frequencies, especially in the transition region. Rabiner *et al.* in 1970 reported an optimization technique, which involves specification of the frequency response in the transition band for reducing the ripples in the response [29]. They reported the optimal set of transition specification for lowpass and bandpass FIR filters for different filter orders and transition widths.

It was decided to carry out similar investigation for reducing the sidelobes in the realization of the critical band based comb filters. This way, we can have an optimum combination of the sidelobes in the stopband and the transition width for a given filter order. Each comb filter involves 18 transitions between passband and stopband, and it is very difficult to find optimal response at transition frequencies. Hence it was decided to find out suitable transition response values for each of the individual critical band filters, using the following steps -

1] Specify the magnitude response of the critical band filter and decide the width of the transition band. Begin with an approximate value of the transition coefficients and find the corresponding FIR filter coefficients.

2] Calculate the frequency response of the FIR filter at a large number of frequencies and find the maximum sidelobe.

3] Readjust the transition coefficients while searching for the minimum sidelobe. This process is carried out for single and double transition samples.

In order to conduct the above procedure, we require a program which allows us to specify the desired frequency response at a set of equidistant frequencies based on the order of the filter. The package should calculate the corresponding filter coefficients, and superimpose the actual response on the desired response, so that we can measure the strength of the sidelobes. The development of this program, as a tool for filter design is given in the next section. Results of the filter design are given in the last section.

4.3 Filter design using frequency sampling technique

A program developed earlier by Preethi Kasthuri in 1997 [30], allows the user to enter the desired frequency response of the filter at N frequency samples either through a graphical interface or from a file, computes the filter coefficients and displays the interpolated response at $2N$ frequency samples. The filter coefficients can be downloaded to a TMS320C50 based board interfaced to the serial port of PC for implementing the filter. The program can be used for filters upto a maximum of 131 coefficients. Starting with the FIR filter design program developed earlier by Preethi Kasthuri, a program "fstfir.c" has been developed to carry out the tasks outlined in the previous section.

The desired filter characteristics are specified through a graphical interface in the frequency domain in terms of magnitude and phase response. From this, the filter coefficients which approximate the desired frequency response specifications are determined. The program offers two options to the user: (1) Specification of magnitude and phase response, and (2) specification of magnitude response of a linear phase filter.

1] *Specification of magnitude and phase response.* The user specifies the magnitude response $|H(e^{j\omega_k})|$ and phase response $\angle H(e^{j\omega_k})$ at equally spaced frequency samples

$$\omega_k = \frac{2\pi k}{M}, \qquad \text{for } k = 0, 1, \ldots, U \qquad (4.1)$$

where M = order of the filter

$$U = \frac{M-1}{2} \qquad \text{for } M \text{ odd}$$

$$\frac{M}{2} \qquad \text{for } M \text{ even}$$

From this, the real and imaginary parts are calculated. The response for the other half of the unit circle is taken as the complex conjugate of the first half so as to obtain a real valued impulse response. Thus,

$$H(k) = |H(e^{j\omega_k})|\ e^{j\angle H(e^{j\omega_k})} \qquad \text{for } 0 \le k \le U \qquad (4.2)$$

$$H^*(M\text{-}k) \qquad \text{for } U \le k \le M-1$$

where $U = \dfrac{M-1}{2}$ for M odd

$\dfrac{M}{2}$ for M even

The impulse response is calculated as the IDFT of $H(k)$ using

$$h(n) = \frac{1}{M}\sum_{k=0}^{M-1} H(e^{j\omega_k})\, e^{j2\pi kn/M} \tag{4.3}$$

2] *Specification of magnitude response of linear phase filter*. The user specifies the magnitude values $|H(e^{j\omega_k})|$ at U equally spaced frequency samples ω_k around half the unit circle.

$$\omega_k = \frac{2\pi k}{M} \qquad \text{for } k = 0, 1, \ldots, U$$

where $U = \dfrac{M-1}{2}$ for M odd

$\dfrac{M}{2} - 1$ for M even

We define a set of real frequency samples

$$G(k) = (-1)^k |H(e^{j\omega_k})| \qquad \text{for } k = 0, 1, \ldots, U \tag{4.4}$$

The impulse response is calculated as

$$h(n) = \frac{1}{M}\left[G(0) + 2\sum_{k=1}^{U} G(k)\cos\frac{2\pi k}{M}\left(n + \frac{1}{2}\right)\right] \tag{4.5}$$

This results in a filter with symmetric unit sample response and linear phase [26] [27]. After $h(n)$, the M-point FIR impulse response, has been calculated, corresponding frequency response $H(e^{j\omega_{\cancel{k}}})$ is given as ~~is calculated at L user specified number frequency samples~~

~~using~~ $$\cancel{H(e^{j\omega_k}) = \sum_{k=0}^{M-1} H(e^{j\omega_k})\, e^{-j2\pi kn/L}} \tag{4.6}$$

$$H(e^{j\omega}) = \sum_{n=0}^{M-1} h(n)\, e^{-j\omega n}$$

And therefore L uniformally spaced samples of the freq response are given as

$$H(e^{j\omega_k}) = \sum_{n=0}^{L-1} h(n)\, e^{-j2\pi kn/L}, \qquad k = 0, 1, \ldots L$$

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

where $\omega_k = \dfrac{2\pi k}{L}$ interpolated L-point

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

$M \leq L \leq 20M$
$20M < L$

4.4 Comb filter design results

The transition bandwidth is defined as the difference between the last sample of the passband and the first sample of the stopband, for high-to-low transition. Alternately, it is the difference between the last sample of the stopband and the first sample of the passband, for low-to-high transition. Thus, the minimum transition bandwidth is one sample. The Fig. 4.1 shows the response specification for a bandpass filter with transition bandwidth of one, two and three samples respectively. The filter design, for the comb filters in the critical band dichotic presentation scheme, uses 128 coefficients and sampling rate of 10 kSa/s. Table 4.1 gives the maximum sidelobe values for all the critical band bandpass filters, for transition bandwidth of one sample. Minimization of the sidelobes was carried out by searching for the optimal value of the frequency samples, for transition bandwidths of (a) two sample (b) three samples at both the low-to-high and high-to-low transitions. The results are shown in Table 4.2 and 4.3 respectively. The sample values in the transition band were found using the filter design package described in the previous section. It can be seen that increase in transition bandwidth can be used for significant decrease in sidelobes. For transition bandwidth of one frequency sample, the maximum sidelobe for the critical band filters range from –21 to –11 dB. The maximum sidelobe range is between –50 to –35 dB for transition bandwidth of two frequency samples, and between –70 to –45 dB for transition bandwidth of three frequency samples. For designing the two comb filters as 128-coefficient FIR filters, the optimal specifcation of individual comb filters, as given in Table 4.2 and 4.3, was taken as a starting point for design optimization.

Table 4.1): Maximum sidelobe values for critical bands with transition bandwidth of one sample. Sampling rate in the design = 10 kSa/s.

Band No	Passband (kHz)	BW, in samples	Maximum Sidelobe (dB)
PB1	0.10-0.20	1	-11.2
PB2	0.20-0.30	1	-11.7
PB3	0.30-0.40	2	-21.1
PB4	0.40-0.51	1	-12.4
PB5	0.51-0.63	2	-21.0
PB6	0.63-0.77	1	-12.6
PB7	0.77-0.92	2	-20.9
PB8	0.92-1.08	2	-21.1
PB9	1.08-1.27	3	-14.5
PB10	1.27-1.48	2	-20.9
PB11	1.48-1.72	4	-19.3
PB12	1.72-2.00	3	-14.5
PB13	2.00-2.32	4	-19.3
PB14	2.32-2.70	5	-15.1
PB15	2.70-3.15	6	-18.4
PB16	3.15-3.70	7	-15.3
PB17	3.70-4.40	9	-15.3
PB18	4.40-5.00	7	-17.9

Table 4.2) : Minimized sidelobe values and transition response values for critical bands with transition bandwidth of 2 samples. Sampling rate in the design = 10 kSa/s.
T = optimal value of the magnitude for the transition samples.

No	Passband (kHz)	BW, in samples	Minimized Sidelobe (dB)	T
PB1	0.10-0.20	1	-39.4	0.430
PB2	0.20-0.30	1	-39.5	0.430
PB3	0.30-0.40	2	-41.7	0.406
PB4	0.40-0.51	1	-41.0	0.430
PB5	0.51-0.63	2	-41.0	0.406
PB6	0.63-0.77	1	-41.1	0.430
PB7	0.77-0.92	2	-40.7	0.394
PB8	0.92-1.08	2	-40.4	0.406
PB9	1.08-1.27	3	-39.4	0.394
PB10	1.27-1.48	2	-40.7	0.406
PB11	1.48-1.72	4	-40.0	0.394
PB12	1.72-2.00	3	-40.3	0.400
PB13	2.00-2.32	4	-45.5	0.103
PB14	2.32-2.70	5	-40.0	0.388
PB15	2.70-3.15	6	-36.0	0.390
PB16	3.15-3.70	7	-38.3	0.382
PB17	3.70-4.40	9	-35.4	0.103
PB18	4.40-5.00	7	-48.9	0.382

Table 4.3) : Minimized sidelobe values and transition response values for critical bands with transition bandwidth of 3 samples. Sampling rate in the design = 10 kSa/s.
T_1, T_2 = optimal value of the magnitude for the transition samples.

No	Frequency Range(kHz)	BW	Minimized sidelobe (dB)	T_1	T_2
PB1	0.10-0.20	1	-60.3	0.097	0.588
PB2	0.20-0.30	1	-60.3	0.097	0.588
PB3	0.30-0.40	2	-48.4	0.103	0.594
PB4	0.40-0.51	1	-61.1	0.097	0.588
PB5	0.51-0.63	2	-47.8	0.103	0.594
PB6	0.63-0.77	1	-61.7	0.097	0.588
PB7	0.77-0.92	2	-47.5	0.103	0.594
PB8	0.92-1.08	2	-47.0	0.103	0.594
PB9	1.08-1.27	3	-56.3	0.109	0.594
PB10	1.27-1.48	2	-46.8	0.103	0.594
PB11	1.48-1.72	4	-50.2	0.103	0.594
PB12	1.72-2.00	3	-59.3	0.109	0.594
PB13	2.00-2.32	4	-49.7	0.103	0.594
PB14	2.32-2.70	5	-61.0	0.103	0.588
PB15	2.70-3.15	6	-52.0	0.103	0.588
PB16	3.15-3.70	7	-48.5	0.103	0.588
PB17	3.70-4.40	9	-44.2	0.103	0.588
PB18	4.40-5.00	7	-67.1	0.103	0.588

In the low frequency range, the critical bands are narrow in terms of frequency samples, so transition bandwidth of two samples is used. In the higher frequencies, the critical bands are wide enough so transition bandwidth of three samples is used. Single transition bandwidth was used upto six critical bands for each comb filter, and double transition bandwidth was used for the remaining three bands. It was found that there were still significant ripples, both in passband and the stopband. To reduce the ripples in the passband, the frequency samples for the individual filters were scaled down. However, this resulted in increased sidelobes. Optimal response for overall realization was then found by iteratively adjusting the frequency samples in the transition band while searching for minimum sidelobes. The process requires on an average 12-15 iterations. The two comb filters have been labeled as "left" and "right" filter. Fig. 4.2 (a) shows the actual response of the left comb filter, with transition width of one frequency sample, superimposed on the desired ideal response. Fig. 4.2 (b) shows the same magnitude response on a dB scale. The ripple in the passband is in 0.01 to 1.40 dB range, while the maximum sidelobe range is -37.92 to -9.95 dB.

Fig. 4.3 shows the frequency response of the right comb filter with transition width of one sample. Passband ripple is in 0.10 to 1.05 dB range and maximum sidelobe range is between –20.1 to –8.21 dB. Fig. 4.4 and 4.5 shows the frequency response of the two comb filters after optimization. For the left filter, the passband ripple range is 0.07 to 0.49 dB and maximum sidelobe range is –22.08 to -40.41 dB. For the right filter, the passband ripple range is 0.06 to 0.57 dB and maximum sidelobe range is –49.28 to -30.51 dB. Fig. 4.6 shows the frequency response of the optimized comb filters superimposed together to illustrate spectral splitting.

The gain of the critcal bands can be adjusted to provide partial compensation for the frequency dependence of the subject's hearing loss. The filter magnitude as a function of frequency is given as

$$A_a(f) = A_c(f) - 3 + 6\,\frac{\alpha(f) - \alpha_{min}}{\alpha_{max} - \alpha_{min}} \tag{4.7}$$

where $A_a(f)$ is the gain in dB for constant gain implementation, $\alpha(f)$ is the interpolated value of the hearing loss in dBHL, and α_{min} and α_{max} are the minimum and maximum

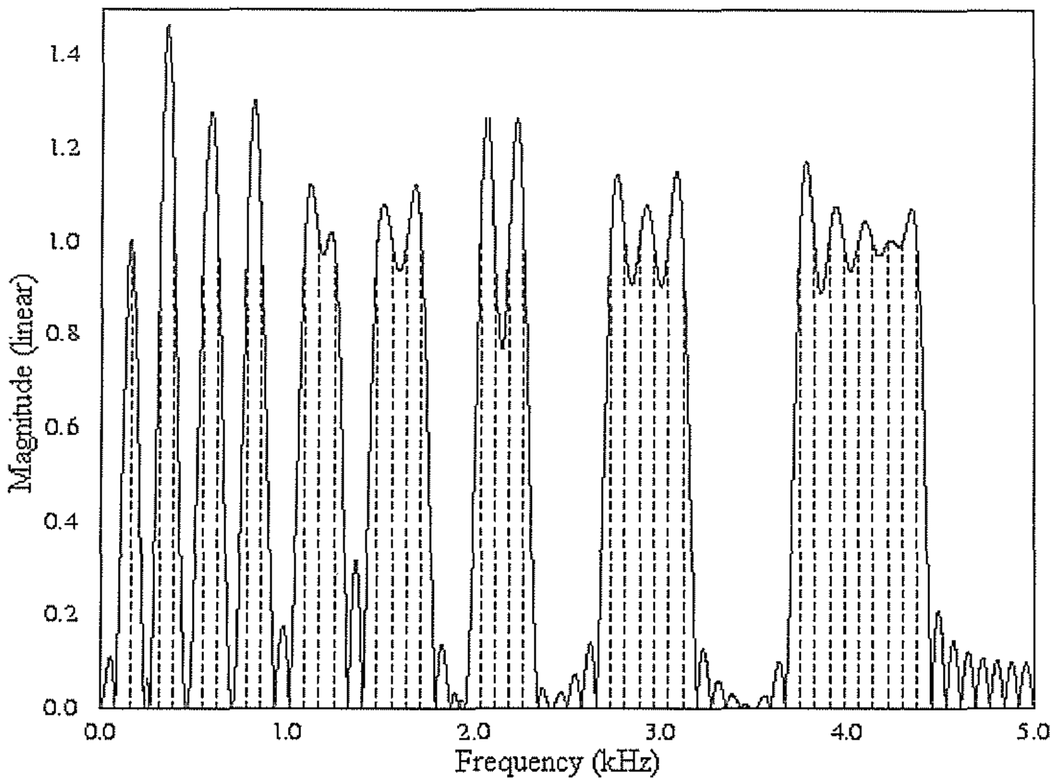


Fig. 4.2 (a) Response of comb filter "left" designed with transition width = 1 sample, on a linear scale. Dotted lines indicate the passband corner frequencies.

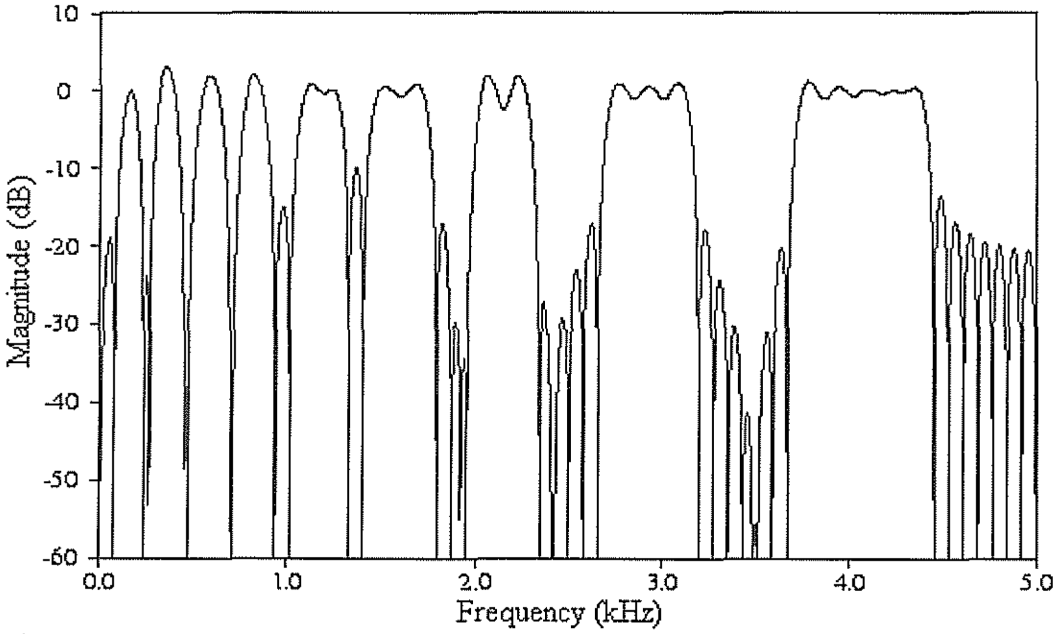


Fig. 4.2 (b) Response of comb filter "left" designed with transition width = 1 sample on a dB scale.

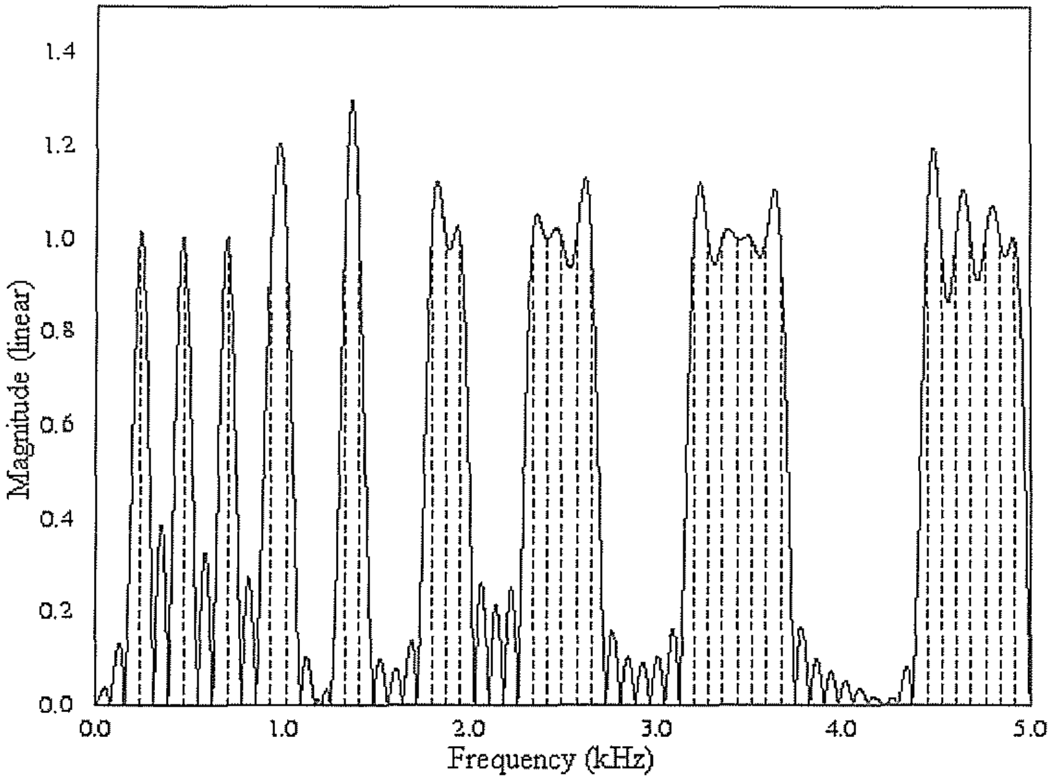


Fig. 4.3(a) Response of comb filter "right", designed with transition width=1 sample, on a linear.scale. Dotted lines are the passband corner frequencies.

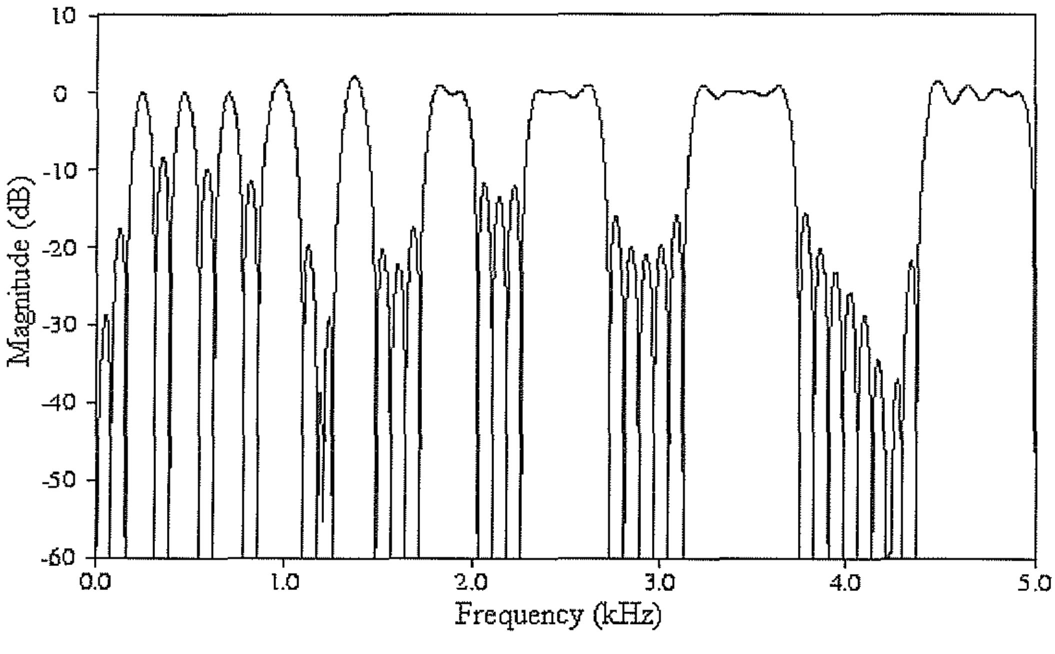


Fig. 4.3(b) Frequency response of comb filter "right" with transition width=1 sample on a dB scale.

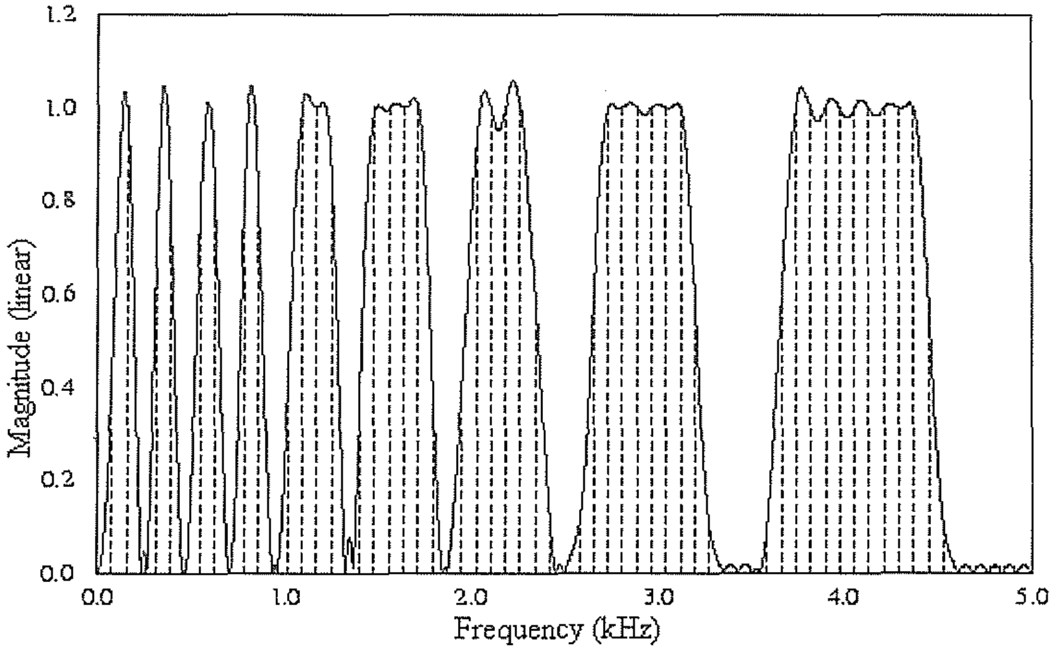


Fig. 4.4 (a) Optimized response of comb filter "left" on a linear scale

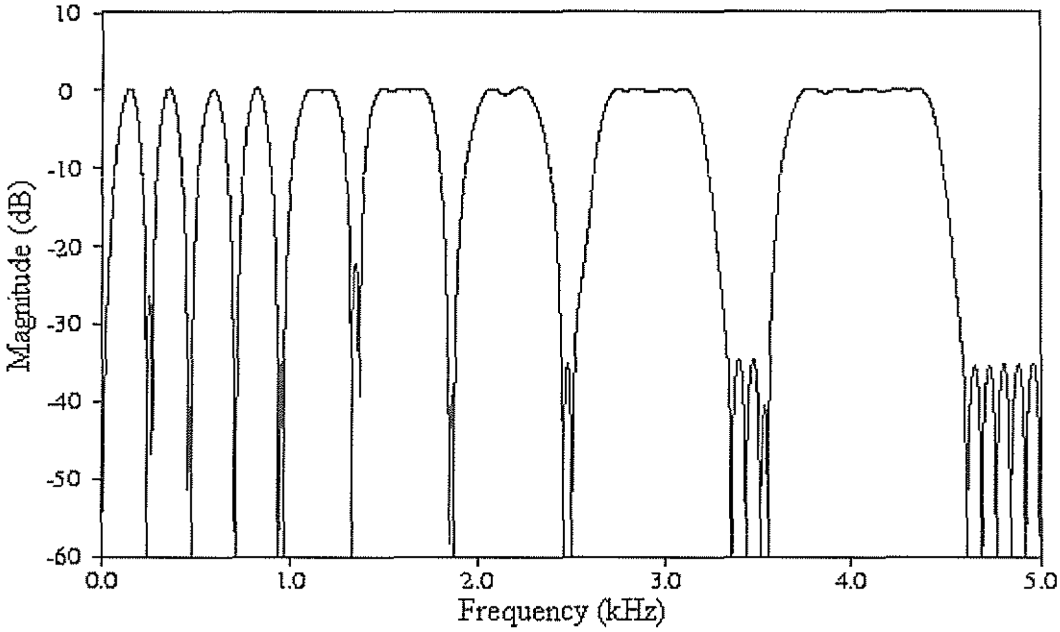


Fig. 4.4(b) Optimized response of comb filter "left" on dB scale

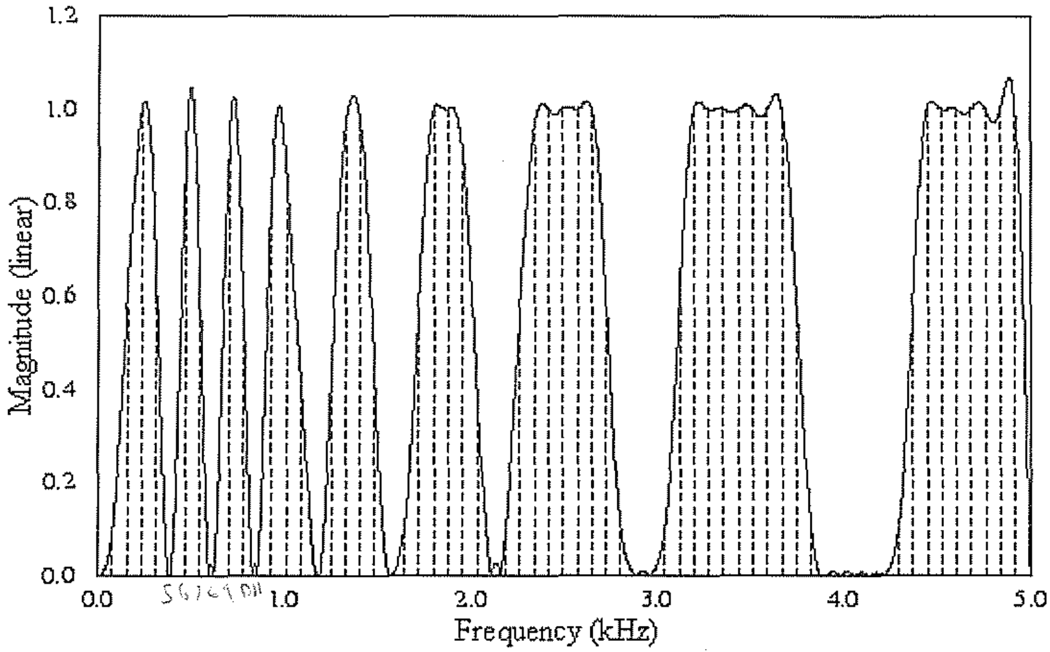


Fig. 4.5(a) Optimized response of comb filter “right” on a linear scale

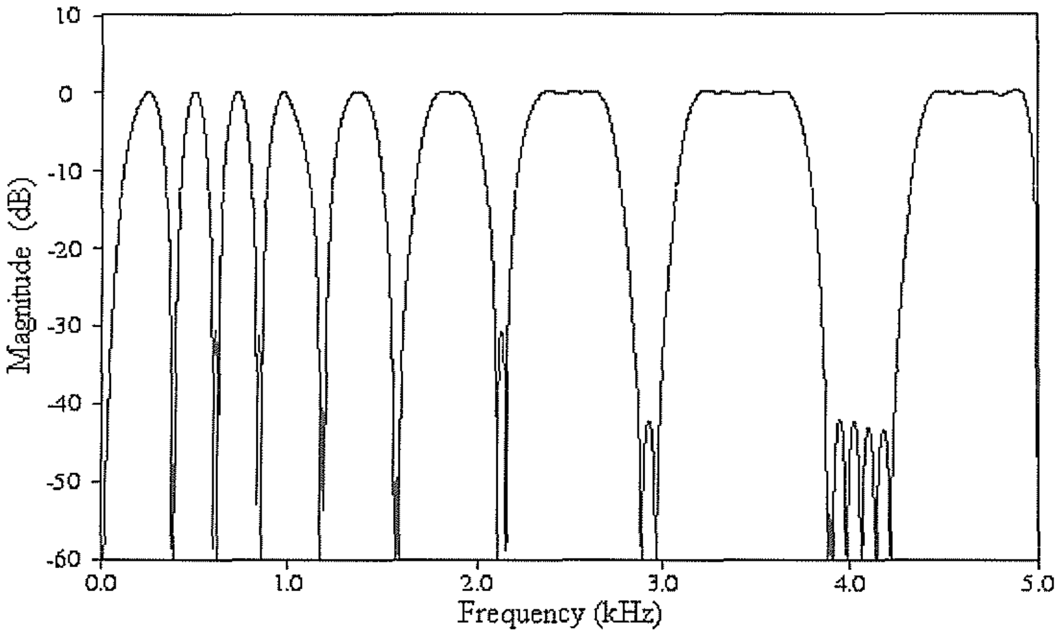


Fig. 4.5(b) Optimized response of comb filter “right” on a dB scale

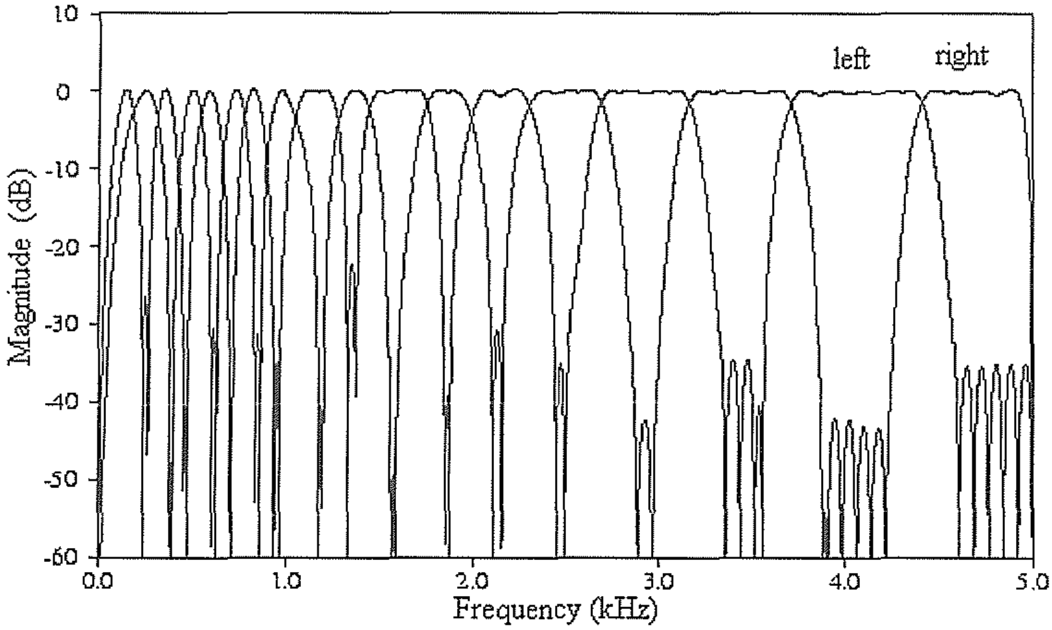


Fig. 4.6) Frequency response of the two optimized comb filters superimposed together to illustrate spectral splitting.

values of the hearing loss over the frequency range of 125 Hz to 5 kHz. Fig. 4.7 shows the relation between the filter gain and the hearing loss.

In this scheme, the filter gains have to be adjusted in accordance with audiogram (a plot of hearing loss as a function of frequency) for the individual subject. The Fig. 4.8 shows the pure-tone audiogram of a subject. A comb filter was designed in which the gain of the filter bands was varied in the range of +3 to –3 dB to match the response of the right ear, according to the equation 4.7. The Fig. 4.9 shows the response obtained using transition width of one sample. The maximum passband ripple is 1.43 dB range and maximum sidelobe range is –11.02 dB. The filter design was optimized using the method described earlier. The Fig. 4.10 shows the optimized response. The maximum passband ripple is 0.48 dB range and maximum sidelobe range is –31.49 dB.

The designed comb filters are used in the implementation of the dichotic presentation technique on multimedia platform, which is described in the following chapter

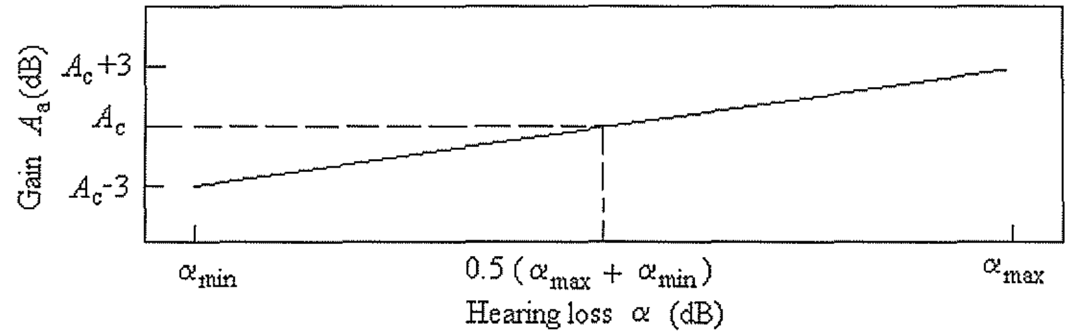


Fig. 4.7) Relation between filter gain and hearing loss

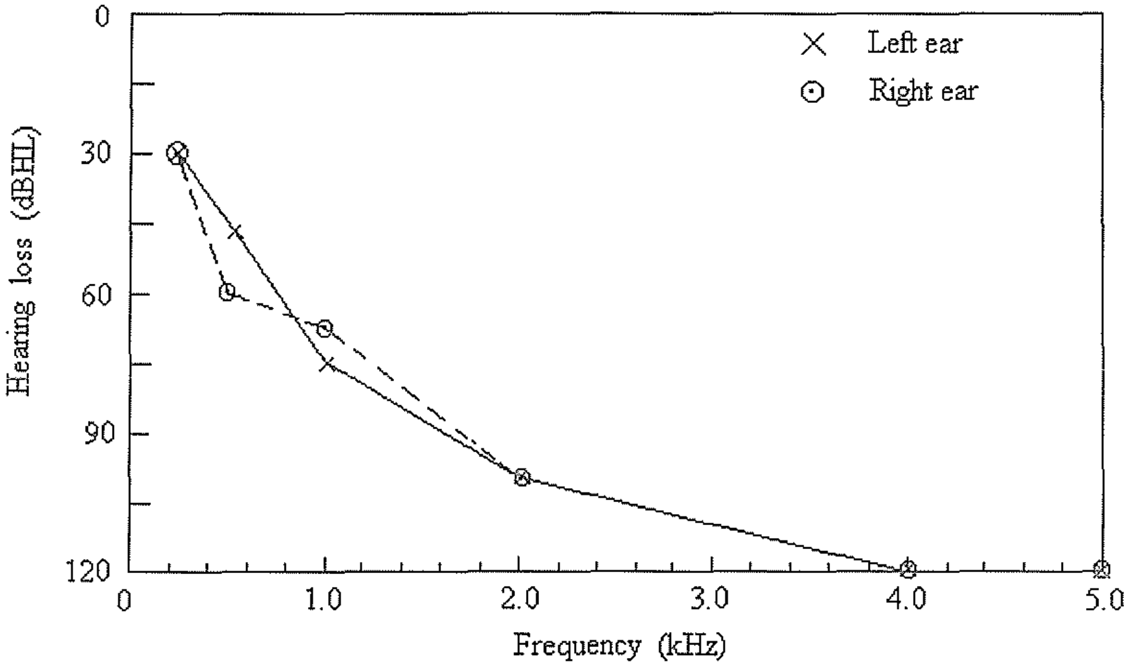


Fig. 4.8) Audiogram of a hearing impaired person

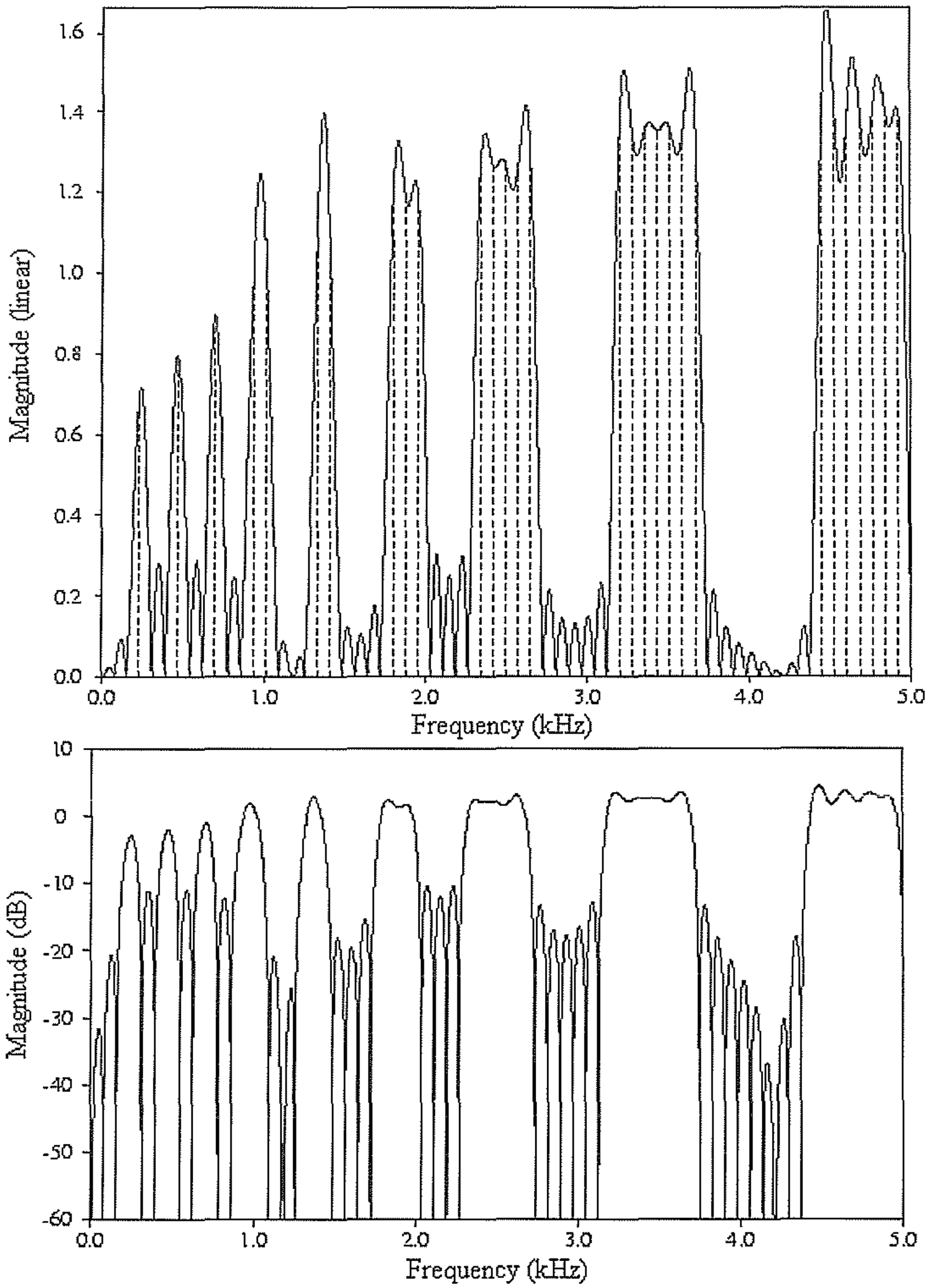
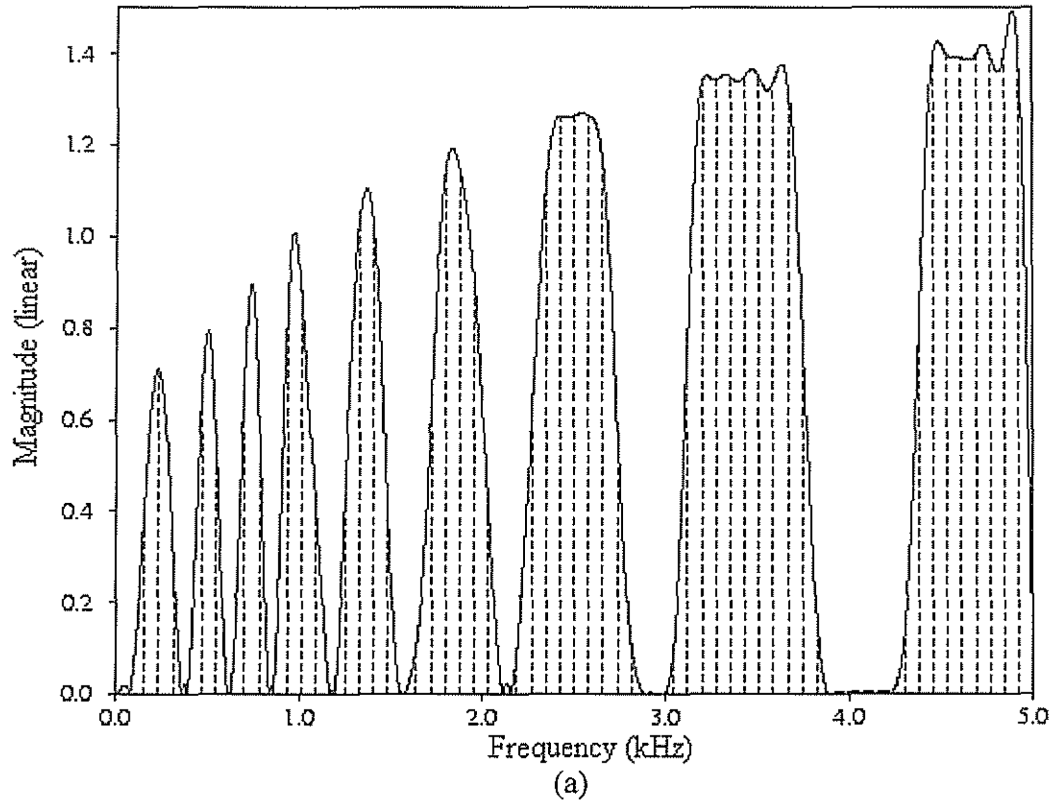


Fig. 4.9) Frequency response of unoptimized comb filter "right" with gain adjustment for compensation of frequency dependent loss a) on a linear scale and b) on a dB scale.



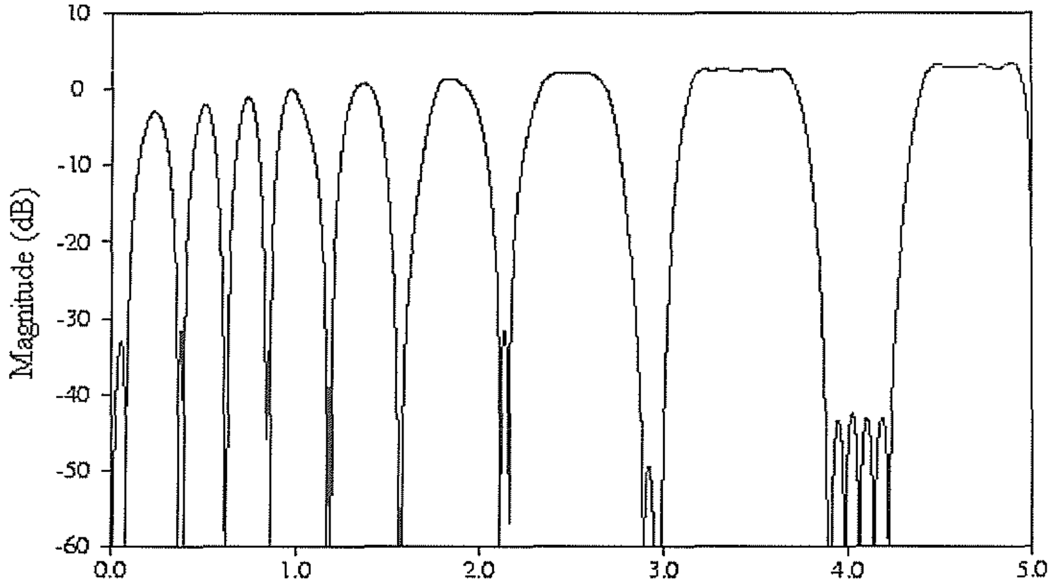


Fig. 4.10) Frequency response of optimized comb filter "right" with gain adjustment for compensation of frequency dependent loss a) on a linear scale and b) on a dB scale.

Chapter 5

DICHOTIC PRESENTATION USING SOUND CARD

5.1 Introduction

A real-time speech-processing scheme for binaural dichotic presentation was implemented by Chaudhari using two 128-coefficient FIR filters, with comb filter response corresponding to critical bands, on a set-up using two DSP boards [26]. In this set-up, a DSP board was used both for signal acquisition, filtering, and for output of the speech signals. Most PCs are equipped with a sound card and speakers as part of multimedia platform, which facilitates speech signal acquisition and two-channel output. This feature can be utilized to implement the dichotic presentation technique for further investigation, without the need for dedicated hardware [31]. Due to the easy availability and low cost of sound cards as compared to DSP boards/data acquisition cards, this signal processing technique can be implemented as a part of the multimedia system of a PC.

The objectives of this implementation are: (i) experimental set-up which can be used for experimenting with various filter bandwidths and gain adjustments for digitized speech, and then to use the optimized coefficients in a DSP board implementation and (ii) a processing system that can be used to enhance the multimedia audio output under adverse listening conditions.

5.2 Present set-up

The present set-up is shown in the Fig. 5.1. The speech signal is acquired using the microphone input of the sound card. The speech signal, which is stored in the form of a data file, is then processed off-line using a C program "dicho.c". The program makes use of FIR filter coefficient files, which may be edited or generated using the filter design program outlined earlier. The program generates two data files corresponding to the two comb filters and combines the two files to prepare a stereo file for binaural output. The speech is output using the line-out of the sound card. The sound card specifications are

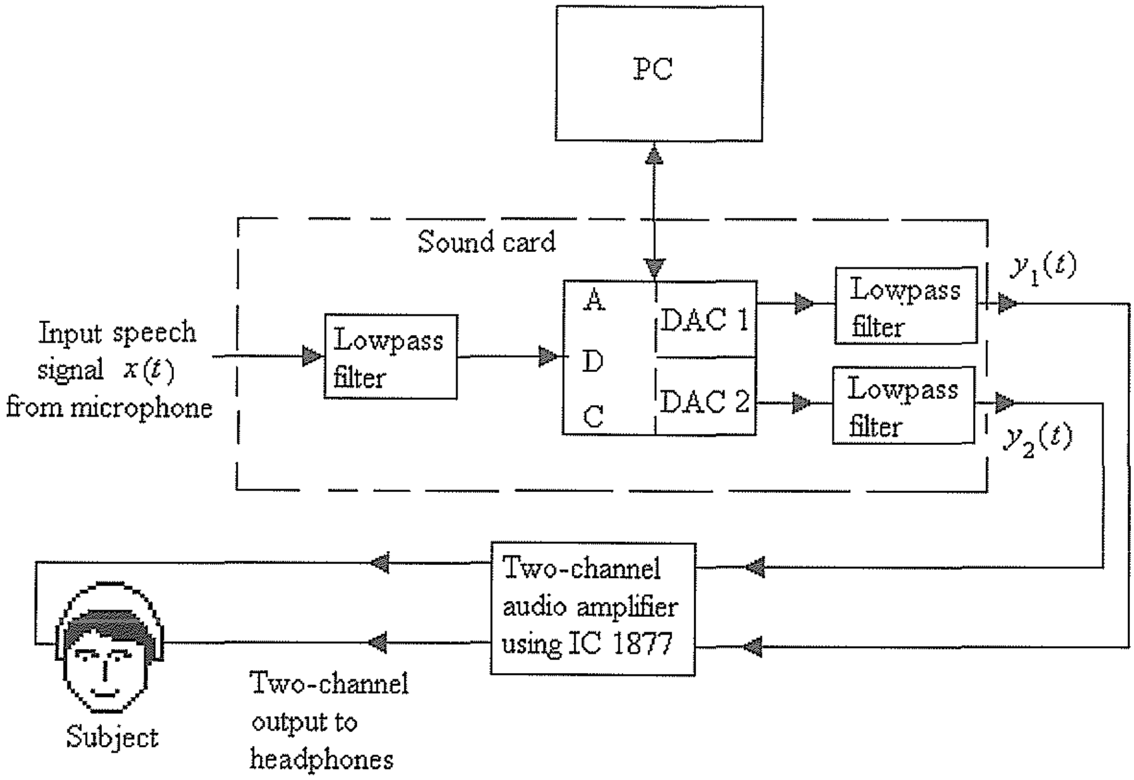
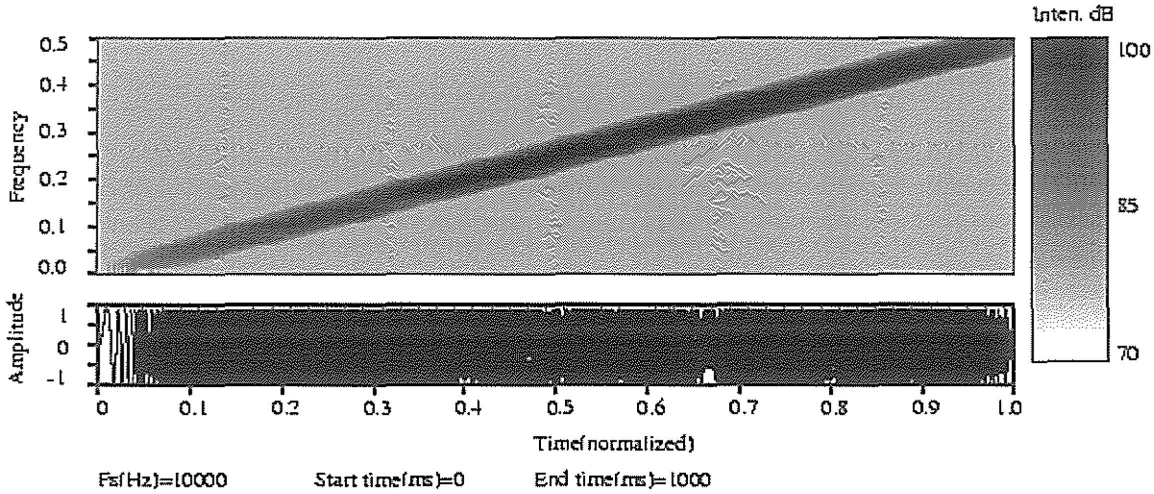


Fig. 5.1) Implementation of dichotic presentation technique using sound card and PC.

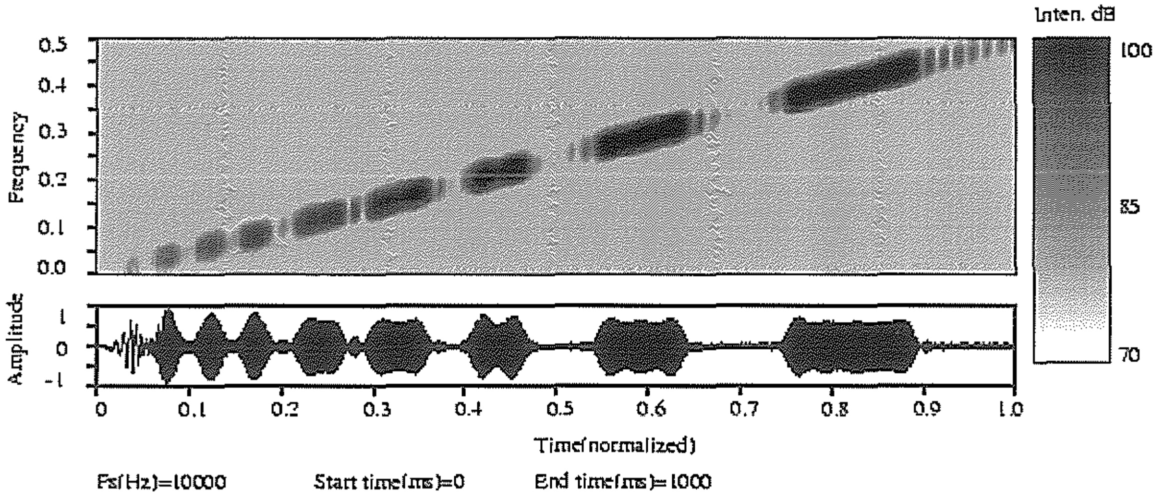
given in Appendix B. The signal can be presented in monaural / binaural, diotic as well as dichotic modes to the subject. This provides a platform for conducting listening tests for various filter bandwidths and gain adjustment of the critical bands. To conduct listening tests, the two line-out outputs of the sound card need to be connected to the headphones. For this purpose, a 2-channel audio amplifier has been designed and built. This is described in Appendix B.

The implementation was tested using the filter coefficients for the comb filters with transition width of one sample (responses given in Fig. 4.2 and 4.3) and using the filter coefficients for the optimized filters (responses given in Fig. 4.4 and 4.5). The implementation has been used for processing test waveform and speech signals, and spectrographic analysis of the processed waveforms was carried out using the program described in Chapter 3. For a sine wave, swept from 0 to 5 kHz in 1 s, the spectrogram for the unprocessed signal and the two comb filter outputs are given in the Fig. 5.2 and 5.3 for the unoptimized and optimized filters respectively. These spectrograms have been taken with 64 gray levels of intensity scale on the display and 256 gray levels on the laser printer outputs. In case of unoptimized filter outputs, effect of sidelobes in the stopband and ripples in the passband are clearly visible. In case of the optimized filters, the swept sine wave segments with frequency components of the different critical bands are clearly separated i.e. sidelobes in the magnitude response are insignificant. Also, the ripples in the passband do not show any significant amplitude variations.

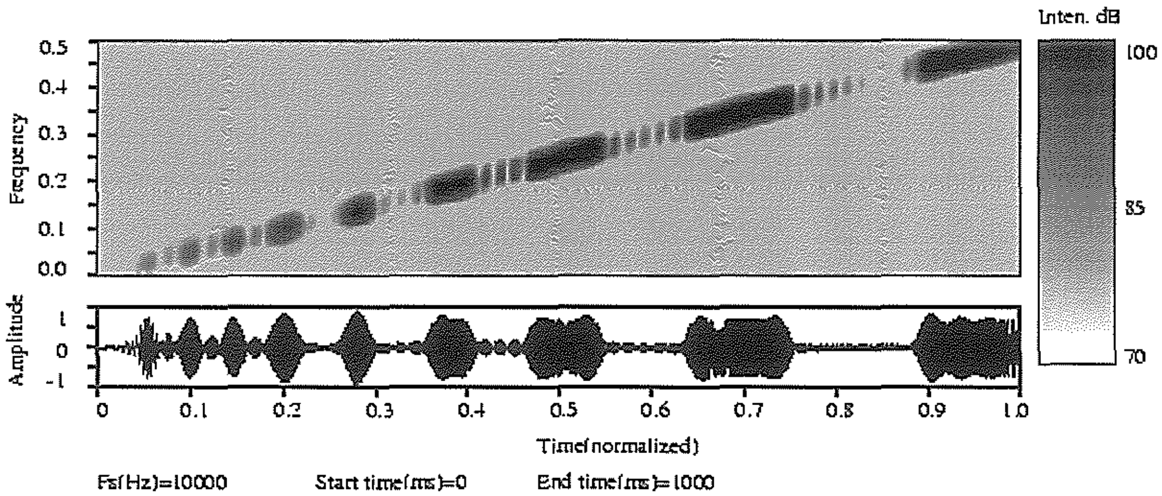
Detailed listening tests for evaluating the improvement due to optimization of the filter design need to be carried out. The comb filters for speech processing have been designed for a sampling frequency of 10 kHz. If they are to be used for filtering data files obtained by sampling at a different frequency, then a sampling rate conversion program would be required.



a)



b)



c)

Fig. 5.2) Wideband spectrograms of a) sine wave swept from 0 to 5000 Hz in 1s and b, c) its filtered versions.using the unoptimized comb filters

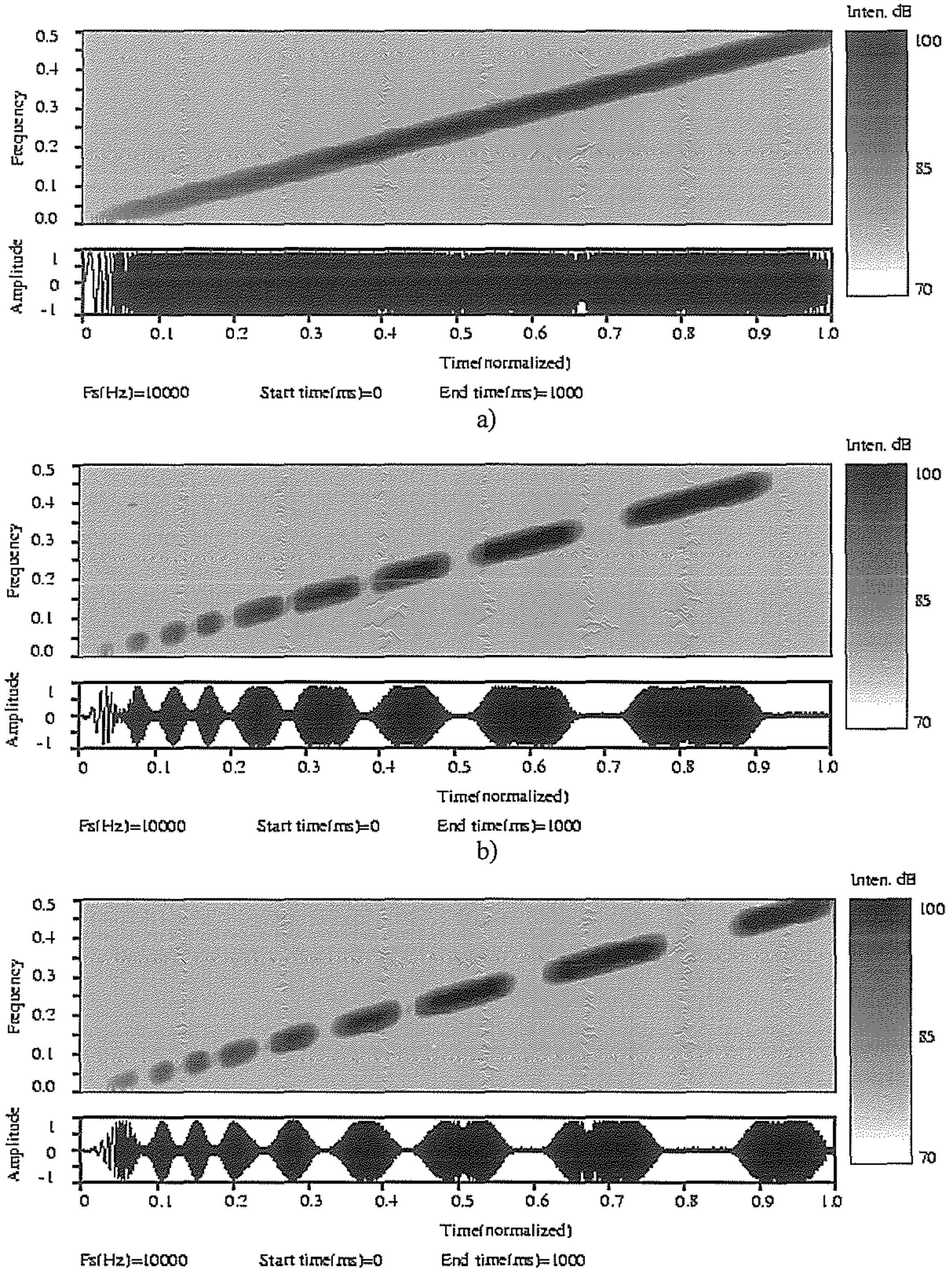


Fig. 5.3) Wideband spectrograms of a) sine wave swept from 0 to 5000 Hz in 1s and b, c) its filtered versions using optimized comb filters.

Chapter 6

SUMMARY AND CONCLUSIONS

6.1 Work done

Sensorineural hearing impairment results in degradation in speech perception due to loss in frequency selectivity as a result of spectral masking. To improve intelligibility, the speech signal can be split spectrally to reduce the effect of spectral masking and presented binaurally. This scheme has been studied earlier by Lyregaard [1], Lunner *et al.* [2] and Chaudhari & Pandey [3] and has yielded favorable results. However, the comb filter implementation by Lyregaard and Lunner *et al.* placed more emphasis on efficiency of realization and spectral splitting was not very effective. The comb filters used by Chaudhari & Pandey [3] provided sharp transitions between passband and stopband. To further reduce the sidelobes in the stopband, we need to optimize the frequency response specification of the filters.

The primary project objective was to design the comb filters for speech processing in such a way that the sidelobes in the stop band are minimized without sacrificing the sharp transitions of the critical band comb filters. For reduction of sidelobes in the comb filters, the frequency specification has been optimized by experimental determination of transition response values. For this purpose, a FIR filter design software based on frequency sampling method was developed, for facilitating the process of finding the optimal filter coefficients. The dichotic presentation technique has been implemented using the PC sound card for signal acquisition and playback. A spectrographic analysis package has been developed which displays the time-varying spectrum of speech.

The spectrographic analysis set-up takes advantage of the increase in the PC processor speeds and the availability of sound card at low cost. The software can work on a PC that has a sound card and a microphone, without the need for any special signal acquisition hardware or DSP peripherals. This set-up can be used by others for analysis of speech and other signals, and has already been used in the Ph.D. research work of

Chaudhari [3]. For conducting listening tests, an audio amplifier circuit was designed to act as a buffer between the sound card of the PC and the headphones.

6.2 Suggestions for future work

The comb filters were optimized by an iterative procedure of adjusting the transition band samples while searching for the mimimum sidelobe. This procedure could be carried out by a program which automatically searches for the optimal transition band samples in filter design. Formal listening tests need to be conducted to verify the effectiveness of the implementation of the scheme. The effect of frequency response shaping of the comb filters as a partial compensation for frequency dependent loss should be investigated. The present scheme for dichotic presentation considers only one aspect of sensorineural hearing loss – spectral masking. The scheme should be extended to investigate the combined effect of temporal and spectral masking. Also, features such as amplitude compression need to be added to compensate for the reduction in the subjects' dynamic range of hearing. It should be seen if this confers additional advantage to the technique.

Appendix A

Design of audio amplifier

The input impedance of the speaker unit of the sound card is 65 kΩ, while that of the headphones is 8Ω. Therefore in order to connect the line-out output of the PC sound card to the headphones, we need to have a buffer audio amplifier. The amplifier IC selected is LM1877, which is a dual power audio amplifier designed to deliver 2W / channel continuous into 8Ω loads. Its significant features are - very low crossover distortion, AC short circuit protection, wide supply voltage range of 6-24 V, and internal thermal shutdown.

Since the LM1877 is internally compensated only for gains greater than 10 in the non-inverting configuration, it is used in the inverting configuration for providing unity gain. The circuit diagram for the two channels of the audio amplifier are shown in the Fig. B.1. The input resistance of the circuit is 100kΩ. Resistors R1 and R2 are used as volume control potentiometers. Diodes D1 and D2 are used for clipping the input signal to an amplitude equal to half the supply voltage, thus protecting the input circuit of the IC. Capacitor C1 is used for providing a stable DC bias of half the supply voltage to the amplifier, and capacitors C4 and C5 provide decoupling to the IC. The present sound card of Creative Technologies Ltd however has the facility for driving non-powered speakers with o/p power of 4W per channel for 4Ω stereo output. So the audio amplifier circuit would be useful if the sound card cannot drive non-powered speakers.

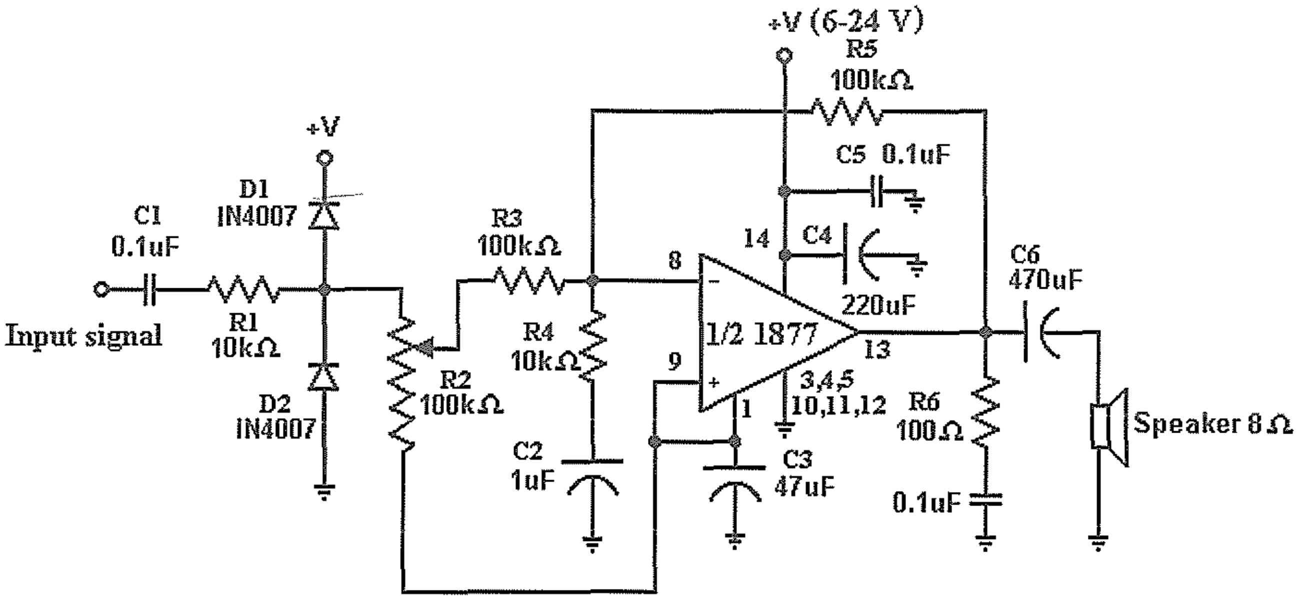


Fig. A.1(a) Circuit diagram of one channel of dual audio amplifier using IC 1877.

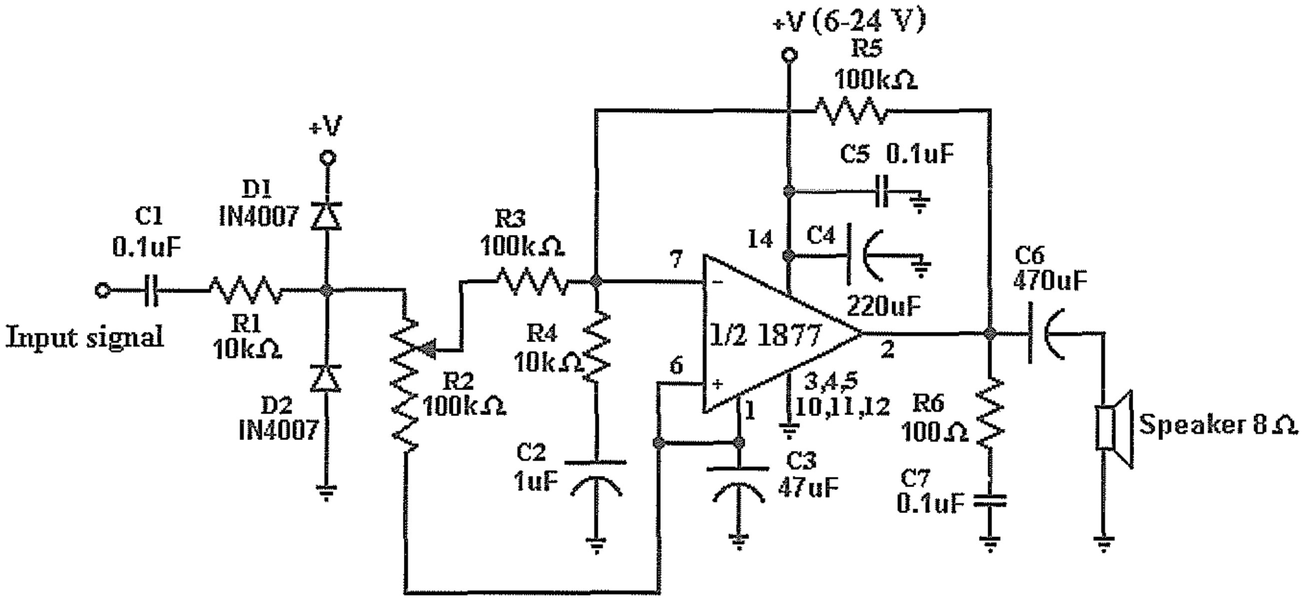


Fig. A.1(b) Circuit diagram of second channel of dual audio amplifier using IC 1877.

Appendix B

Sound card specifications

The project has utilized a sound card manufactured by Creative Technology Ltd.[32] for recording and playback of speech signal . The Fig. B.1 gives a schematic of the sound card connections. The specifications of the sound card are given below:

1] *System requirements*

a) Intel Pentium or AMD K5 90 MHz processor

b) 4 MB RAM

c) VGA or SVGA card

d) 7.5 MB of free hard disk space

e) Windows 95 or Windows 3.1 with MS-DOS 5.0 and a Plug and Play (PnP) configuration manager.

2] *Analog input port characteristics* (established by measurement)

Analog input is a sum obtained through an internal mixer from two analog inputs:

a) Microphone (Mic) powered by 2.5 V supply from sound card

b) Line-in with voltage range of 0-9 Vpp.
Input resistance = 53 kΩ at 1 kHz.
Bandwidth =10 Hz to f_S/2 Hz.

3] *Output*

a) Speaker Out/ Line Out connects non-powered speaker by default with output impedance of 0.767 Ω at 1 kHz with 4 W per channel for 4 Ω stereo output Also connects powered speaker and an external amplifier when built-in amplifier is disabled by changing the jumper settings on the sound card.

4] *Digitization*	a) For analog input : Sampling rate = 11025, 22050, or 44100 Sa/sec Number of quantization bits = 8 or 16. b) For analog output : Sampling rate = adjustable to a value in the range of 5000 Hz to 44100 Hz Number of quantization bits = 8 or 16

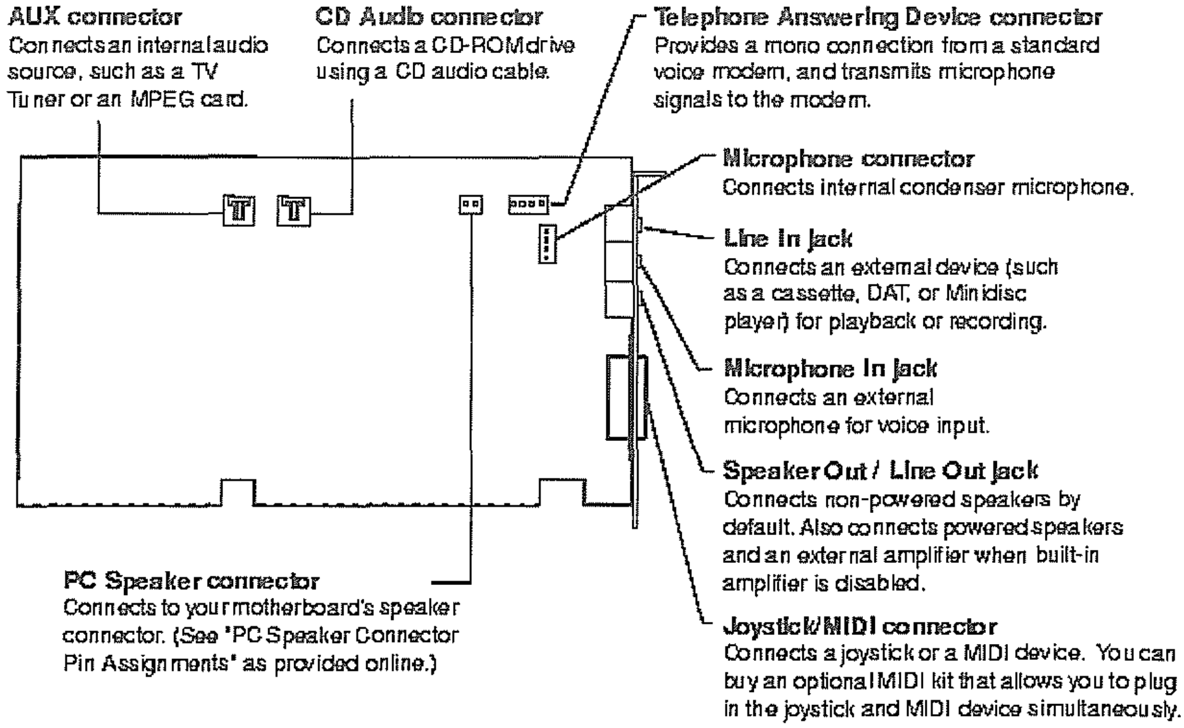


Fig. B.1: Jacks and connections of the sound card. Source: [31]

REFERENCES

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

[2] T. Lunner, S. Arlinger, & J. Hellgren, "8-channel digital filter bank for hearing aid use: preliminary results in monaural, diotic and dichotic modes", *Scand. Audiol. Suppl.*, 38, pp 75-81, 1993.

[3] D.S. Chaudhari & P.C. Pandey, "Dichotic presentation of speech signal using critical filter bank for bilateral sensorinueral hearing impairment", *Proc.16th International Congress on Acoustics (ICA)*, Seattle, Washington, USA, 1998, pp 213-214.

[4] H. Levitt, J.M. Pickett, & R.A. Houde, *Sensory Aids for the Hearing Impaired*, New York: IEEE Press, 1994.

[5] B. C. J. Moore, *An Introduction to the Psychology of Hearing*, London: Academic, 1997.

[6] CHABA, "Speech perception aids for the hearing-impaired people: Current status and needed research", *J. Acoust. Soc. Am.,* 90, pp 637-685, 1991.

[7] H. K Dunn, & S. D. White, "Statistical measurements on conversational speech", *J. Acoust. Soc. Am.,* 11, pp 278-288, 1940

[8] P. J. Fitzgibbons & F. L Wightman, "Gap detection in normal and hearing-impaired listeners", *J. Acoust. Soc. Am.,* 72 (3), pp 761-765, 1982.

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

[10] C. W. Turner, M. P. Ribb, "Audibility and recognition of stop consonants in normal and hearing-impaired subjects", *J. Acoust. Soc. Am.,* 81, pp 1566-1573, 1987.

[11] R. C. Bilger & M. D. Wang, "Consonant confusions in noise: a study of perceptual features", *J. Acoust. Soc. Am.,* 54, pp 1248-1266, 1973.

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

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

[14] A. Syrdal, R. Bennett, & S. Greenspan, *Applied Speech Technology*, Florida: CRC Press, 1995.

[15] D. O' Shaughnessy, *Speech Communication: Human and Machine*, Reading, Massachusetts: Addison Wesley, 1980.

[16] S. Mithal, "Splitting of speech signal for binaural presentation", *B. Tech project report*, Department of Electrical Engineering, IIT Bombay, 1996.

[17] E. Zwicker, "Subdivision of audible frequency range into critical bands", *J. Acoust. Soc. Am.*, 33, pp 248, 1961.

[18] A. V. Oppenhiem, "Speech spectrogram using FFT", *IEEE Spectrum,* 7, pp 57-62, 1970.

[19] T. G. Thomas, "Experimental evaluation of improvement in speech perception with consonantal intensity and duration modification", *Ph.D. thesis*, Department of Electrical Engineering, IIT Bombay, 1996.

[20] T. G. Thomas, P. C. Pandey, & S. D. Agashe, "A PC – based multiresolution spectrograph", *J. IETE (India)*, 40, pp 105-108, 1994.

[21] B. N. Ashok Baragi,. "A speech training aid for the deaf", *M.Tech. dissertation*, Department of Electrical Engineering, IIT Bombay, 1996.

[22] V. V. S. R. Prasad, "Speech processing for single channel hearing aid", *M. Tech dissertation*, Department of Electrical Engineering, IIT Bombay, 1996.

[23] D. S. Chaudhari, "Dichotic presentation for improving speech perception by persons with bilateral sensorinueral hearing impairment", *Ph.D thesis*, School of Biomedical Engineering, IIT Bombay, 2000.

[24] Adobe Systems Inc, *Postscript Language Reference Manual*, Reading, Massachusetts: Addison-Wesley, 1988.

[25] A. V. Patil, "A Glottal pitch extractor", *M.Tech. dissertation*, Department of Electrical Engineering, IIT Bombay, 2000.

[26] D. S. Chaudhari & P. C. Pandey, "Real time speech processing for dichotic presentation in binaural hearing aids", *Proc. National Conference on Biomedical Engineering (NCBME)*, Manipal, Karnataka., 1998, pp II-45.

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

[28] L. R. Rabiner, & B. Gold, *Theory and Application of Digital Signal Processing*, New Delhi: Prentice Hall of India, 1988.

[29] L. R. Rabiner, B. Gold, & Mc Gonegal, "An approach to the approximation problem for nonrecursive digital filters", *IEEE Transactions on Audio and Electroacoustics*, 18, pp 83-106, 1970.

[30] P. Kasthuri, "Real time digital filter with tunable frequency response using TMS320C50 processor", *Project report*, Signal Processing and Instrumentation Lab, Department of Electrical Engineering, I.I.T Bombay, 1997.

[31] D. S. Chaudhari & P. C. Pandey, "Binaural audio in multimedia systems to improve auditory perception for the hearing impaired", *Proc. Fifth International Symposium on Signal Processing and its Applications (ISSPA)*, Brisbane, Australia, 1999.

[32] Creative Technology Ltd, *Sound Blaster Manual*, 1997