# SPEECH PROCESSING

# FOR

# SINGLE CHANNEL SENSORY AID

Dissertation

Submitted in partial fulfillment of the requirements for the degree of

Master of Technology

by

V. V. Siva Rama Prasad
Roll no. 94307021

under the guidance of
Dr. P. C. Pandey
Prof. T. Anjaneyulu

Department of Electrical Engineering
Indian Institute of Technology, Bombay
January 1996

INDIAN INSTITUTE OF TECHNOLOGY, BOMBAY

DISSERTATION APPROVAL SHEET

Dissertation entitled "**Speech Processing for Single Channel Sensory Aid**" by V. V. Siva Rama Prasad, is approved for the award of the degree of Master of Technology in Electrical Engineering.

| | | | |
|---|---|---|---|
| Guide | : | PCPandey | (Dr. P. C. Pandey) |
| Co-Guide | : | [illegible] | (Prof. T. Anjaneyulu) |
| Internal Examiner | : | Vikram M Gadre | (Dr. V. M. Gadre) |
| External Examiner | : | [illegible] | (Sri ~~S. K.~~ R. S. Sardesai) |
| Chairman | : | [illegible] 1/2/96 | (Prof. B. P. Kashyap) |

**V.V.Siva Rama Prasad : Speech Processing for Single Channel Sensory Aid**, M.Tech Dissertation, Department of Electrical Engineering, I.I.T. Bombay, January 1996.

## ABSTRACT

The aim of the project is to develop a speech processor for single channel sensory aid. This aid can be useful as a hearing aid to the deaf, who are suffering from sensori-neural hearing loss.

The speech processing scheme of this speech processor is based on frequency lowering principle. This scheme presents distinct cues about intonation and rhythm of the speech, and about the fricative sounds. This distinction has been achieved by separating the residual hearing frequency region of deaf, in two bands and providing the cues in these two bands. Only one of the two cues is presented at a time. This scheme processes the speech signal using two channels. Channel 1 estimates the pitch, voicing and generates a periodic or random waveform based on voiced or unvoiced speech, respectively. Channel 2 maps the the high frequency components to the higher band of residual hearing region of the deaf. This scheme has been implemented in real-time using a Digital Signal Processing (DSP) card based on TMS320C25 DSP processor. Pitch is estimated by computing the short time autocorrelation. High frequency bursts are mapped using a pulse repetition rate mapping algorithm. An earlier implementation was tested and problems were solved by modifying the scheme as well as the implementation. This modified scheme has a reference to the formant energy for deciding whether to present either channel 2 output or Channel 1 output. This modified implementation is tested thoroughly in modular and overall methods. Listening tests have been conducted using vowel-consonant-vowel syllables, on five normal hearing subjects. Results show that this scheme helps in providing cues about the features like voicing, manner, and place of articulation.

## ACKNOWLEDGEMENT

I sincerely thank my guides Dr. P. C. Pandey and Prof. T. Anjaneyulu for their guidance, encouragement and valuable suggestions during the project work.

I thank Mr. Ashok Baragi for the timely technical discussion support. I also thank Mr. Pravin Survase, Mr. Gunendra V. Patil, Mr. Satish Kumar C., and Mr. Ashok Baragi for, being as subjects during the listening tests and also for their help during my stay at IIT Bombay.

I thank shri Apte, and shri Mantode for their help, in standards lab and in instrumentation lab during my project work. I am grateful to the Director, ISTRAC, ISRO, Dept. of Space, Bangalore for granting me extra-ordinary leave for one and half year for studying M.Tech at IIT Bombay. I am extremely grateful to Mr. V. Murali Krishna, Mr. P. Guna Sekhar, and Mr. Subhash Danodia for their help in domestic matters.

V. V. Siva Rama Prasad
(94307021)

## CONTENTS

## LIST OF TABLES

LIST OF FIGURES

# CHAPTER 1

# INTRODUCTION

## 1.1 Problem Overview

Hearing impaired persons experience a lot of problems while communicating with others. They can not fully perceive the speech of the speaker and also the intelligibility of the speech of a deaf person reduces as there is no auditory feedback. This results in severe restrictions in the social interaction, and educational and carrier opportunities for the deaf and therefore results in degraded quality of life [1].

Deaf people try to perceive the speech by observing the lip movements of the speaker, known as "lipreading". The supra-segmental aspects of speech such as stress, number of syllables, and intonation contour are generally not received by lip reading [1]. In case of receiving consonants confusion occurs in detection of voicing or manner of articulation etc. In case of vowels, confusions relate to the similarity of lip shape. So hearing aids can improve the lipreading by supplementing the cues about fundamental frequency, voicing duration, formants in the speech, etc.

Hearing loss can be classified as conductive and sensori-neural loss. Conductive losses can be cured by medical intervention. For the sensori-neural losses sensory aids are to be used. Sensori-neural loss is generally characterized by severe hearing loss at high frequencies. So a hearing aid which can translate the high frequency speech information to the residual low frequency hearing region can be useful in providing the cues about the certain fricatives, to such deaf people. The hearing aid should also be able to provide distinct cues about intonation and rhythm of the speech. These cues should be provided either one at a time to avoid masking.

## 1.2 Project Objective

There is an on-going effort at IIT Bombay to develop a single channel sensory aid (SCSA) for the deaf to take care of the problem mentioned earlier. This aid is based on a frequency lowering method proposed earlier by Pandey [2]and [3]. This method processes the speech signal in two channels and presents either low frequency components of the speech or translated high frequency components of the speech, one at a time to avoid masking. This scheme was implemented with off-line processing by Sapre [4] for developing a speech processor for single channel auditory prosthesis and was subsequently implemented with real-time processing by Shah [5] for developing a sensory aid for the deaf.

The aim of the project is to develop a speech processor for single channel sensory aid. The implementation done by Shah [5] is referred, in this report, to as `single channel sensory aid-1' (SCSA-1). Thorough testing of the SCSA-1 has to be done and the existing problems are to be solved. This testing comprises testing each individual block of the above scheme with proper excitation, storing the excitation and response, and analyzing the response. The complete implementation is to be tested using vowel-consonant-vowel (VCV) normal speech data and the response of the scheme is to be studied, using spectrogrăpic analysis, described in Appendix A, for observing the mapping of the high frequency components of the fricatives. The problems are to be solved by modifying the software and also by modifying the speech processing scheme, if required. The modified software has to be tested. Listening tests are to be conducted using VCV normal speech syllables to evaluate the performance of the scheme in providing the cues about the features like manner and place of articulation, voicing, etc of the speech signal.

### 1.3 Report Outline

Chapter 2 describes various sensory aids for the deaf.. It also describes various techniques being used in electro-acoustical hearing aids and cochlear implant aids.

In chapter 3, a speech processing scheme for single channel cochlear prosthesis has been explained. Two modified forms of this processing scheme, which are used for developing sensory aids based on the use of residual hearing, are also explained.

Chapter 4 describes the speech processing scheme used for developing the speech processor in this dissertation project and its implementation in brief. The problems and modifications of SCSA-1 and the implementation SCSA-2 are discussed. The test details alongwith test results are also presented.

In chapter 5 the listening test setup and listening test procedure have been described. The analysis of the test results is presented and discussed.

Chapter 6 summarizes the work done and presents some suggestions for further work.

Appendix A describes a spectrograph software package.

Appendix B presents the confusion matrices of the listening tests.

Appendix C describes the supporting software programs, which are developed during the testing of the developed speech processor.

# CHAPTER 2

# SENSORY AIDS FOR THE DEAF

## 2.1 Introduction

Hearing loss is broadly divided into two types, conductive loss and sensori-neural loss. Conductive loss results from dysfunction of ear canal or middle ear structures. So less acoustic energy reaches the auditory receptors in the cochlea. In general, this results in a more or less uniform hearing loss as a function of frequency [6]. Sensori-neural loss results in reduced resolving power of the neural receptor mechanism [1]. Usually, the extent of the loss increases with frequency, but the difficulty experienced by the sufferer is not always well predicted from the audiogram [6]. Sensori-neural loss may arise as a result of defects in many parts of the auditory system. The particular difficulties of the sufferer depend upon the part that has got defect.

The hearing impairments caused by a disorder of the cochlea are commonly associated with a loss of frequency selectivity. In normal ear the auditory filter is narrow band around the signal frequency where as in impaired ear it is broader. The major consequence of this is that there will be a greater susceptibility to masking by interfering sounds. Masking can be defined as the process by which the threshold of audibility for one sound is raised by the presence of another (masking) sound [6].

Most conductive losses can be treated through medical intervention, where as sensori-neural hearing loss is not curable by medical intervention. So sensory aids are to be used for such type of deaf persons. Many deaf persons perceive the speech by lipreading. However, the speech information received by them will be less due to the confusions in discriminating the various vowels and consonants. Therefore information available through lipreading needs to be supplemented by the cues about voicing, manner and place of articulation through sensory aids.

Hearing aids are classified as electro-acoustic hearing aids and sensory or speech perception aids. Both the type of aids are explained in this chapter. The speech processing schemes of some of the single channel sensory aids, reported in literature, which uses frequency lowering techniques are also explained in this chapter.

## 2.2 Electro-acoustic Hearing Aids

Acoustic amplification is the method most commonly used to raise the acoustic signal above the hearing threshold of the hearing impaired. But the frequency versus threshold characteristic of impaired persons may not be uniform at all frequencies of the residual hearing region. So while amplifying the weakest frequency signal to above threshold, the amplifier may increase the other frequency components above the loudness discomfort level. To avoid this problem, frequency selective amplification is done so as to best fit the speech signal in to the residual hearing area without causing discomfort.

In modern hearing aids, for protection against excessive amplification, compression limiting or syllabic compression is done [1]. In compression limiting, the hearing aid behaves as a conventional amplifier for signals below the threshold of compression. For the signals above the threshold of compression, the amplifier gain is reduced substantially. In syllabic compression, the parameters of the amplifier are chosen so as to alter the relative intensities of individual speech sounds [1]. The main aim of the syllabic compression is to improve intelligibility in the amplified speech. Most of the information in speech is carried by consonantal sounds. The sound levels of consonants, with the exception of nasals, are much lower than those of the vowels. So intelligibility can be improved by amplifying the weaker consonantal sounds and by reducing amplification for stronger vowel sounds.

**2.3 Sensory Aids**

For highly impaired deaf persons, electro-acoustic hearing aids may not help much. So sensory or speech perception aids can be used to provide the cues for supplementing the lipreading. These sensory aids can be used for stimulation of the auditory nerve or for presentation through alternative sensory modalities like visual aids and vibro tactile aids. These sensory aids can be classified as single channel and multi channel based on the number of channels that are exciting or stimulating. In this section, cochlear prosthesis and substitutional aids will be described briefly.

**2.3.1 Cochlear prostheses**

For the deaf with intact auditory nerve, the nerve fibers in cochlea can be electrically stimulated. The stimulation has a dynamic range of 2-15 dB. This cochlear prosthesis has been successfully implemented in post lingual persons with total sensori-neural hearing impairment. For prelingual deaf people improvement is not much. But both cases requires extensive training for better auditory and speech skills.

In extra-cochlear prosthesis, a stimulating electrode is placed near the entrance of the cochlear duct. In the simplest cochlear prosthesis, the stimulating electrode is surgically inserted about 6-15 mm into the cochlea. The tip of the electrode is placed close to the auditory nerve cells. In multi-channel prostheses, a number of electrodes (generally 6-22) are placed at different distances in the cochlear duct and these electrodes are indiadually stimulated. These devices may employ a frequency coding scheme [1].

The stimulation pulse rate and electrode position are most important parameters. In a strategy proposed by Tong et. al.[7] the fundamental frequency of the acoustic speech signal was converted to electric pulse rate and the second formant frequency to electrode position.

It has been reported that cochlear prostheses have helped some subjects in improved recognition of manner of articulation, when auditory stimulation was supplemented by visual cues.

### 2.3.2 Substitutional aids

In this approach, speech information is presented via the senses of vision or touch. In Visual aids, cues like pitch information, formant information, vocal tract shape, and energy content in the speech signal are visually displayed. The tactile aids can be used to vibrate the certain sensitive parts of the body to provide cues to supplement the lipreading.

## 2.4 Single Channel Sensory Aids

Sensory aids can be classified as single channel and multi channel based on the number of channels that are exciting or stimulating the deaf. One of the most common property of sensori-neural hearing impairment is that hearing loss is greater at high frequencies. So the audio spectrum of the speech can be translated downward (to lower frequencies), where it is audible to the hearing impaired and presented to the hearing impaired. This is known as frequency lowering principle. A few single channel sensory aids which are using the frequency lowering techniques are explained in next subsections. These are based on modulation, distortion, vocoder, and frequency division techniques [9].

### 2.4.1 Modulation technique

Johansson [10] implemented a sensory aid using modulation technique. The block diagram of the scheme is given in Fig. 2.1 (a). The Speech input is processed in two channels. In channel 1 the speech signal is passed through a compressor. In the second channel the speech input is passed through high pass filter, modulator, low pass filter, and compressor. The carrier frequency of the modulator is in 3-5 kHz range. So the frequency components above 4 kHz are transposed to 0-1.5 kHz band. The outputs of both the channels are added and passed to the ear phones. Johansson reported some advantages, but the deaf children had to be trained intensively. Ling [11] conducted experiments using Johansson technique, and has reported that the discrimination skills with linear amplification were either equal to or better than those with transposition.

### 2.4.2 Distortion technique

Johansson proposed a distortion technique also [9]. The block diagram of this is shown in Fig. 2.1 (b). The frequency components in the range 4 to 5 kHz are passed through a distortion network which is working in nonlinear region. The low frequency range distortion products are added to the normal speech and passed to the head phones. No systematic spectrum differences between different fricatives are observed in the output.

### 2.4.3 Vocoder technique

Priminov [9] proposed a vocoder system. The block diagram of this is shown in Fig. 2.1 (c). This system divides the speech signal into a number of bands and the energies of these bands are modulated over pure tones or noise bands in the hearing range of the subjects. According to Risberg [9] the number of channels in the vocoder should depend on the number of fricatives in language e.g., for four fricatives /f, ʃ, t, s/ there should be 3 or 4 channels. In the first channel, the speech is band pass filtered and presented. The second channel has band pass filter in 0.7 -2 kHz range, a detector, a low pass filter. The low pass filter output is modulated with a noise source. The third channel is same as the second one but with band pass filter in the range 2-7 kHz. The outputs of the three channels are added and given to earphones.

### 2.4.4 Frequency division technique

Frequency division method was proposed by Guttman and Nelson [12]. The block diagram is shown in Fig. 2.1 (d). This aid was to make clear distinction between fricatives /s/ and /ʃ /. The speech signal is passed through a spectrum shaper to enhance the spectral difference between the two fricatives. The speech signal is passed through voicing detection circuit or frication separator, parallelly. Based on this circuit's output low frequency signal corresponding to fricatives is generated using zero crossing frequency divider. This low frequency information is added to the normal speech. This device can be used as a speech training aid.

## CHAPTER 3

## SINGLE CHANNEL SPEECH PROCESSOR SCSA-1

### 3.1 Introduction

One of the most common property of sensori-neural hearing impairment is that the hearing loss is greater at high frequencies. The hearing thresholds are also high, over the low frequency range, which decreases the dynamic range of the loudness. This residual hearing frequency region should be used economically for providing the cues about voicing, and manner of articulation, etc. The cues about the intonation and rhythm should be distinct, in frequencies, from the cues about the high frequency components. To avoid the masking, both the cues should not be presented simultaneously. A speech processing scheme which can accommodate the above requirements has been proposed earlier by Pandey [2] and [3]. A modified form of this scheme has been implemented with off-line processing by Sapre [4] in his M.Tech dissertation, a speech processor for single channel auditory prosthesis. This has been implemented in real-time, with minor modifications, by Shah [5] in his M.Tech dissertation. These three schemes are explained briefly in this chapter. The implementation done by Shah is referred to as SCSA-1 in this report. The implementation of SCSA-1 is also explained briefly in this chapter.

### 3.2. Single Channel Speech Processing Scheme

This scheme has been proposed earlier by Pandey [2] and [3], to provide cues about certain high frequency fricatives and stops like /s ʃ z t/ alongwith low frequency periodicity and rhythm information in the limited PRR and dynamic range available with single channel cochlear prosthesis.

The block diagram of the scheme is shown in Fig.3.1. Thespeech input is processed in two channels. Channel 1 processes low frequency components while channel 2 processes high frequency components. The BPF-1 bandpass filters the input speech signal. The rhythm of the speech is obtained by the envelope detector. The pitch pulses are obtained either from pitch extractor or by center clipping and peak limiting the filtered signal. The available pulse repetition rate (PRR) range is divided into two bands, lower and upper bands. The expected pitch range is mapped onto lower band by a mapping scheme. The amplitude envelope is compressed to suit the available dynamic range and is superimposed on the mapped pulses to obtain channel 1 output.

In channel 2, the high frequency fricatives are filtered by the BPF-2. This filtered signal is center clipped and infinite peak limited to obtain random pulses. These pulses are mapped onto the available upper PRR band using a mapping scheme. The compressed amplitude envelope is superimposed on these random pulses. The comparator compares the output of envelope detectors, of both channels and then switches the output of the stronger channel to stimulator.

A speech processor, shown in Fig.3.2, was built, earlier, in hardware by Pandey [2] and [3], by incorporating the principle of the scheme. The BPF-1 bandwidth is 75-250 Hz, while BPF-2 bandwidth is 3-6 kHz. If channel-2 envelope detector output is more than the reference level, channel-2 output is switched to stimulator. If it is less, channel-1 output is switched. It was tested for extra-cochlear implant and the results indicated that high frequency information is perceived distinctly from the periodic low frequency pulses.

### 3.3 Development of Speech Processor SCSA-1

A modified form of the speech processing scheme, explained in earlier section, was used by Sapre [4] for developing a speech processor for single channel auditory prosthesis. The block diagram of the scheme is shown in Fig.3.3. A low pass filter, with cutoff frequency of 900 Hz, is used in channel-1. A high pass filter, with cutoff frequency of 1 kHz, is used in channel-2. The pitch estimator is estimating pitch and also detecting the voicing. Waveform generators have been incorporated in each channel. The major modification, of the scheme explained in earlier section, is that the voicing information is used for presenting output of either channel-1 or channel-2. So the output is periodic during voiced segments and is aperiodic during unvoiced segments of speech.

This scheme was implemented in software, for processing the speech in off-line mode. The sampling rate is 10 k Sa/s. The frame length is 100 samples. For mapping the random pulses of channel 2, a pulse repetition rate mapping algorithm has been developed by Sapre [4]. Informal listening tests were conducted. Results showed that this scheme helps in presenting the features like place, manner of articulation and voicing.

A modified form of the scheme, discussed in this section, has been used by Shah [5] in his M.Tech dissertation. The block diagram of this scheme is shown in Fig.3.4. The major modification is that an additional branch is included in channel-1 for generating a random waveform which preserves the randomness of the low pass filtered unvoiced speech. This branch presents the random waveform, when the input speech is unvoiced and the average magnitude of channel-2 is less than the reference value. Such a condition can exist during the transition from voiced to unvoiced and vice versa. This scheme has been implemented in real-time by shah. This implementation is referred to as SCSA-1. The scheme as well as its implementation are expalined briefly in this section.

### 3.3.1 Speech processing scheme of SCSA-1

The block diagram of the scheme is shown in Fig. 3.4. In channel 1, amplitude estimator estimates the average magnitude of the low pass filtered speech signal. Zero crossing detector outputs a pulse whenever its input crosses zero level. Pitch estimator estimates the pitch and it decides whether the input speech signal is voiced or unvoiced. If it is voiced, square wave generator generates periodic square waves using the timing information from pitch estimator, i.e., with pitch period, and amplitude information from amplitude estimator. If the speech signal is unvoiced, random wave generator 1 generates random pulses using timing information from zero crossing detector and amplitude information from amplitude estimator.

Channel 2 translates or maps the high frequency fricative information to upper part of the residual band. The amplitude estimator estimates the average magnitude of the high pass filtered speech. The zero crossing detector output is a train of pulses corresponding to input zero crossings. Pulse repetition rate (PRR) mapper maps this train of pulses to a train of low frequency pulses. Random wave generator 2 generates random pulses using the timing information from the PRR mapper and amplitude information from the amplitude estimator. The output low pass filter smoothens the output of waveform generators.

If the input speech is voiced, periodic output of square wave generator is presented. If it is unvoiced, the output of channel 2 amplitude estimator is compared with a threshold to present random pulses of either channel 1 or channel 2.

This scheme has been implemented in real-time using a PC add-on DSP card based on TMS320C25 DSP processor. The sampling rate is 10 k Sa/s. Rectangular window, of length 128 samples, is used. So the block length is 128 samples. Listening tests have been conducted, in off-line mode, using 12 VCV (vowel consonant vowel) normal speech sounds. In off-line mode, the normal speech data is transferred to the shared memory, which is on the DSP board, by the PC using the PC bus interface. This data is processed on DSP board and processed data is transferred to PC and is stored in a file. These processed files are presented during the listening tests.

#### 3.3.2 Software implementation of SCSA-1

The speech processing scheme of SCSA-1 was implemented in assembly language of the DSP chip TMS320C25. The low pass filter is second order Butterworth filter and its cutoff frequency is 500 Hz. After center clipping the low pass filtered speech, short time auto correlation is being computed for estimating the pitch and for voicing detection. The parameters of the pitch estimator are, the threshold clipping level is 42% of the max. value, the minimum peak is 7.8% of R0, and minimum R0 is 37. The high pass filter is second order Butterworth filter with cutoff frequency 2500 Hz. The high frequency components of fricative sounds are being mapped to low frequency using

Pulse Repetition Rate (PRR) mapping algorithm, which is discussed in detail in next section. The output low pass filter is second order Butterworth filter with cutoff frequency of 1000 Hz.

### 3.3.3 Pulse repetition rate (PRR) mapper of SCSA-1

The PRR mapping algorithm was developed by Sapre [4]. The output of the zero crossing detector, of channel-2, is divided by 4 to reduce its frequency, and then the mapping algorithm is applied to this train of pulses. Basically this alogrithm is mapping the input pulse period, time period between the consecutive input pulses, to the pulse period which corresponds to the PRR of the upper band of the residual hearing region. A linear function is being used for this pulse period mapping. This has been implemented as y = m x + c, where x is the average input pulse period, y is the output pulse period, and the m and c are determined as per the input and output PRR ranges.

The pulse repetition rate mapping algorithm is described below. The flow chart of the algorithm is shown in Fig. 3.5. The timing diagram is shown in Fig. 3.6. As shown, the $c_{in}$ is the counter for measuring the input pulse period, 'azct' is the average zero crossing time, 'opc' is the average input pulse period, and $c_{out}$ is the counter for output pulse period. Whenever there is a pulse at the input, the count in $c_{in}$ used for calculating 'azct' and the counter $c_{in}$ is reset, for measuring the next pulse period and is incremented at each next sample instant till the arrival of next pulse. The $c_{out}$ is calculated using the linear function, whenever its values reaches zero, and a pulse is passed to the mapper output. The $c_{out}$ is decremented by one and zero level is passed to the mapper output, at each next sample. When the $c_{out}$ reaches zero, again the procedure repeats.

Algorithm :

1) $i=0;$

2) $if\ x(i)=0\ then\ c_{in}=c_{in}+1;\ goto\ step\ (4)\ else$

3) $azct=\dfrac{(a_0 * azct+c_{in})}{a_0+1}\ ;\ c_{in}=1.$

4) $if\ c_{out}>0,\ c_{out}=c_{out}-1\ ;\ y(i)=0;\ goto\ step(6)\ ;\ else$

5) $O_{opc}=\dfrac{b_0 \cdot azct + c_{in}}{c_0}\ ;\ c_{out}=(f_m \cdot O_{opc}+f_c);\ y(i)=1.$

6) $i=i+1\ ;\ if\ i<128\ goto\ step\ (2)\ else\ end.$

In SCSA-1, The input and output PRR ranges are 3k-7k and 300-600, respectively. The input zero crossing rate (ZCR) is scaled down by a factor 4. So now input ZCR is 750 to 1750 per 10000 samples.

PRR mapping input zero crossing time range is

(10000/1750 to 10000/750) = (5.71 to 13.3)

PRR mapping output zero crossing time range is

(10000/600 to 10000/300) = (16.67 to 33.33)

A linear mapping is used as below

slope = (33.33 - 16.67 ) / (13.33 - 5.71)

y = slope * (x - 5.71) + 16.67

then y = 2.1854 x + 4.1857

So m = 2.1854 and c = 4.1857 .

The remaining parameters of PRR mapper are $a_0$, $b_0$, and $c_0$, and their values are $a_0$ =2, $b_0$ = 1, $c_0$= 1.9.

# CHAPTER 4

## SPEECH PROCESSOR FOR SINGLE CHANNEL SENSORY AID SCSA-2

### 4.1 Introduction

A speech processing scheme has been proposed earlier by Pandey [2] and [3], which can provide cues about the certain high frequency fricatives and stops like /s, ʃ, z, t/ alongwith low frequency periodicity and rhythm information in the limited PRR and dynamic range available with single channel cochlear prosthesis. The modified forms of that scheme have been used by Sapre [4] and Shah [5] in their M.Tech dissertations, for sensory aids based on the use of residual hearing. The speech processing scheme and the implementation have been described in the previous chapter. The implementation done by Shah is referred to as SCSA-1. This SCSA-1 has to be tested throughly, and the problems should be solved by modifying the software and by modifying the speech processing scheme, if required. This modified scheme should be implemented and tested throughly, and listening tests should be conducted for evaluating the performance of the scheme in providing the cues about the features like voicing, manner, and place of articulation, etc.

SCSA-1 was tested thoroughly. The problems were rectified by modifying the processing scheme as well as the implementation. The processing scheme of SCSA-1 has been modified for mapping certain fricatives, which are having considerable high frequency energy, e.g. /z/. The modified implementation is referred to as SCSA-2. SCSA-2 has been tested thoroughly, first modularly and as a system. Listening tests were conducted on five normal hearing subjects.

In this chapter the testing of SCSA-1, the speech processing scheme of SCSA-2, the implementation details of SCSA-2, and testing of SCSA-2 are explained. The listening tests are described in the next chapter.

### 4.2 Testing of SCSA-1

While testing the SCSA-1 a few problems were found. The problems and their solution are explained in this section.

### 4.2.1 Low pass filter

This low pass filter was tested with sine wave input. It was saturating at 14 V amplitude. This was due to the high scaling up factor ($2^{16}$) of the filter coefficients. This problem was solved by reducing the scaling factor to $2^{12}$.

### 4.2.2 Window length and pitch estimation

The window length of SCSA-1 is 128 samples. With this window length the pitch estimation is proper when SCSA-1 was tested with sine wave, square wave, and triangular wave inputs with various frequencies. But when tested with normal speech vowel consonant vowel (VCV) files, the variation in estimated pitch is abnormally high. If the clipping level threshold and voiced/unvoiced detection thresholds are adjusted for one speech file, e.g. ATA file, for other files the pitch estimation is not proper. Also the voiced/unvoiced thresholds could not be fixed at a particular value. Therefore the window length was increased to 256 samples. With this window length, thepitch estimation and the voiced/unvoiced detection was proper. Also for sine, square, and triangular waveform inputs proper pitch estimation was observed.

### 4.2.3 Square wave generator

When the channel-1 software was tested with sine wave input, is was observed that occasionally the polarity of the output of square wave generator 1 is same at the block boundaries, causing constant amplitude with same polarity for one pitch period. This problem has been rectified by developing another square wave generator module (function).

### 4.2.4 Pulse repetition rate (PRR) mapper

When the PRR mapper was tested with sine wave (input) it was observed that the division (dividing factor is 4) of the input PRR is taking place but it is not being mapped to the desired output PRR range (300-600). This problem was due to incorrect implementation of the PRR algorithm. So the implementation was modified. Modification details are as follows,

The step 3 of the algorithm, explained in section 3.3.3, is implemented as follows,

$$azct = a_1 \, . \, azct + a_2 \, . \, cin, \quad \text{where } a_1 = 0.6667 \text{ and } a_2 = 0.3333.$$

One more observation was that the output PRR was varying in 4 discrete PRR levels for the input PRR range (5k-7k). This was due to the dividing factor 4, of the output of ZCR2. So this dividing factor has been modified to 8, to have more discrete PRR levels in output.

The input PRR and output PRR ranges of SCSA-1 are 3k-7k and 300-600 respectively. As the frequency range of the frication bursts of the fricatives is greater than 3 kHz, the PRR mapper input and output ranges of interest are small. So they are modified to 5k-9k and 500-1050 respectively.

### 4.2.5 Output low pass filter

The output low pass filter of SCSA-1 implementation was behaving as band pass filter, when it was tested with sine wave of different frequencies. This was due to the improper implementation. This problem was rectified by scaling up the coefficients with an appropriate scaling factor.

After all the above modifications, the speech processing scheme is tested with normal speech VCV files and the processed output was stored in files. The frequency spectrum of the processed files was analyzed using spectrograph, explained in Appendix A, software. The processed output spectrum was having considerable energy up to 2000 Hz. So to limit the processed energy to low frequencies the order of the output low pass filter has been increased to four.

### 4.2.6 Execution of the scheme

In SCSA-1 all the program modules of the scheme are not being executed in every block duration. If the input speech is voiced, the required modules of the channel 1 are being executed and if it is unvoiced, based on the output of the comparator either channel 1 or channel 2 blocks are being executed.

In SCSA-2, all the program modules are executed irrespective of the voicing of the input speech signal.

### 4.3 Speech Processing Scheme of SCSA-2

There are two drawbacks in the processing scheme of SCSA-1. One of them is, if the input speech signal is unvoiced, the channel 2 amplitude estimator output is being compared with a threshold amplitude for deciding either to present output of RWG1 or RWG2. As this threshold amplitude can vary with signal levels, this method is not suitable in dynamic conditions. The other drawback is that this scheme is mapping only unvoiced sounds. It is not mapping the voiced fricatives which are having considerable amount of high frequency energy, e.g. /z/. These drawbacks are solved by modifying the scheme. This modified speech processing scheme is shown in Fig. 4.1. The formant energy of the speech sounds is being used as reference for presenting either channel-1 output or channel-2 mapped output.

As shown in Fig. 4.1, The speech signal is processed in three channels. Channel-1 processes low frequency components. Amplitude estimator 1 estimates the average magnitude of the low pass filtered speech. The pitch estimator estimates the pitch and also determines the voicing of the input speech signal. Square wave generator (SWG) generates periodic waveform with 50% duty cycle, with period equal to the pitch period. Zero crossing detector-1 outputs a pulse whenever its input reverses the polarity. The random wave generator-1 generates random waveform and also preserves the PRR of the low pass filtered speech. The estimated amplitude is superimposed on the output of waveform generators to preserve the rhythm of the input speech.

Channel-2 processes the high frequency components. This channel maps the high frequency fricative components to upper band of the residual hearing frequency region of the deaf. (The residual hearing frequency region, of the deaf is divided in two bands, lower and upper bands.) Amplitude estimator-2 estimates the average magnitude of the high pass filtered speech. ZCD2 (zero crossing detector-2) outputs a pulse whenever its input reverses its polarity. The PRR (pulse repetition rate) mapper maps, the input train of pulses, onto a train of pulses whose PRR is corresponding to the upper band of the residual hearing frequency region. RWG2 (random wave generator-2) generates random waveform using the timing information from this mapped train of pulses.

The third channel, filters the formant energy, in the 500-2300 Hz frequency band, and Amplitude estimator-3 estimates the average magnitude of this bandpass filtered speech.

The output of channel-1 is either periodic square wave or random waveform depending on the voiced speech or unvoiced speech, respectively. The output of AE3 is compared with five times the output of AE2. If the latter is stronger, output of RWG2 (after superimposing the estimated amplitude of channel 2) is passed to the OLPF (output low pass filter). If the former is stronger, output of channel 1 (after superimposing the estimated amplitude of channel 1) is passed to the OLPF. The OLPF smoothens its input.

One important observation is that this modified scheme helps in distinguishing the vowels /a/ and /i/. In case of /i/, the average magnitude of bandpass filtered speech and high pass filtered speech are equal. So channel-2 mapped energy will be presented.

This speech processing scheme has been implemented in real-time and this implementation is referred to as SCSA-2.

## 4.4 Hardware Setup of SCSA-2

The hardware setup of SCSA-2 is shown in Fig 4.2. The speech signal from an electret microphone is amplified and filtered by the input signal conditioning hardware. The PC add-on DSP board is used for acquiring and processing the speech data. The processed output is smoothened and amplified by the output signal conditioning hardware. A PC is used for downloading the DSP software to DSP board and also for displaying the speech parameters (read from the shared memory space on DSP board). The input and output signal conditioning hardware were built earlier by Shah [5] during his M.Tech dissertation. The individual blocks of the hardware setup are explained briefly in this section.

### 4.4.1 Input signal conditioning hardware

The microphone (electret type) generates a voltage which is typically tens of mV. This is amplified by the preamplifier of the input signal conditioning hardware to +/- 10 V to make full use of the dynamic range of the ADC. The input signal conditioning hardware also has an anti-aliasing filter, which is active $7^{th}$ order elliptic low pass filter. This filter has passband upto 4.6 kHz with a passband ripple of 0.3 dB, and the stopband starts at 5 kHz with a minimum attenuation of 40 dB.

### 4.4.2 TMS320C25 based DSP board [13]

The TMS320C25 based DSP PC add-on board, (DSP25, Dynalog Micro Systems, Bombay, India), is used for acquiring and processing the speech signal. This board operates at a clock frequency of 40 MHz. It has got an on-board ADC, DAC, and a timer. This board is interfaced with the PC using PC bus. The memory (both program and data) on this board can be accessed by both the PC and the TMS320C25 chip (shared memory).

### 4.4.3 Output signal conditioning hardware

The smoothing filter of the output signal conditioning hardware smoothens the processed speech of the DSP board. This filter is a $7^{th}$ order active elliptic low pass filter. The specifications of this filter are same as that of anti-aliasing filter. The output signal conditioning hardware also contains a power amplifier. The power amplifier is class B push-pull type. The amplified signal is passed to TDH39 head phones or to a bone vibrator.

#### 4.4.4 IBM PC/AT

A PC is used for downloading the DSP software to the program memory of the DSP board and also for displaying the speech parameters (which are read by the PC from DSP board using PC bus I/F).

### 4.5 Software Implementation of SCSA-2

The block diagram of the speech processing scheme is shown in Fig.4.1. This scheme has been implemented in assembly language of the DSP chip TMS320C25. A rectangular window with length 256 samples is used.

This software has been implemented for operation in off-line and real-time modes. In off-line mode, data is taken from speech files. A block of 256 samples data is transferred to the shared memory, which is on the DSP board, using the PC bus interface. This data is processed by the DSP processor. The processed data is again transferred to the PC and is stored in a file. In real-time mode, the speech signal is connected to the ADC port of the DSP card, as shown in Fig.4.2. The on-board timer interrupts the C25 processor at a rate of 10 k interrupts/s. With each interrupt a sample is acquired from the ADC port and stored in a temporary input buffer on DSP card, and a sample from the temporary output buffer is transferred to DAC port. After acquiring 256 samples, this data block is transferred to the input buffer and the processed data, corresponding to earlier data block, is sent to the temporary output buffer. The data block present in the input buffer is processed while acquiring the next data block.

The drawback of the off-line processing is that the VCV unprocessed files are loaded on the DSP board using the PC Bus interface. So the analog to digital conversion of the speech data is not taking place. So the noise effects of the quantizing and sampling of ADC can not be observed during the presentation.

Two programs have been developed for each mode. One program processes the speech signal using the complete scheme while the other one has Channel 1 processing software alone. This is done for observing the merits of the complete scheme over those of Channel-1 alone. These programs, for the real-time mode, are `rt.asm' and `rt1.asm'. For the off-line mode, `nrt.asm' and `nrt1.asm' are developed. In off-line mode, for transferring the data between the PC and DSP board a `C' program has been developed. It is `off.c '. In real-time, for downloading the DSP software and also for displaying the speech parameters a software module `chan.c' has been developed. The listings of the above programs have been given in a seperate volume, as appendix to this dissertation [14].

A detailed description of the software implementation of the scheme is given in next sub-sections.

### 4.5.1 Low pass filter

The low pass filter is third order Butterworth filter with cutoff frequency, $f_c$ of 300 Hz. Its normalized transfer function [15] is

$$H(s) = \frac{1}{(s^2+s+1)(s+1)} \tag{4.1}$$

The above equation is denormalized by substituting, $s = \frac{s}{\Omega_c}$ where,

$$\Omega_c = \tan\left(\frac{\omega_c}{2}\right) \quad where \ \omega_c = \frac{2\Pi f_c}{f_s}$$

where $f_s$ is sampling rate in samples/s

$f_c$ is cutoff frequency in Hz

$\omega_c$ is normalized cutoff frequency in rad/sample

$\Omega_c$ is analog cutoff frequency in rad

Then Eq. (4.1) is translated to discrete time domain using bilinear transformation [15], i.e., by substituting $s = \frac{z-1}{z+1}$. From the bilinear transformed equation the filter coefficients can be obtained.

### 4.5.2 Amplitude estimator-1

The absolute values of the output buffer of low pass filter, are passed to amplitude estimation buffer of amplitude estimator 1. This buffer is divided into four equal segments. For each segment, average amplitude is calculated. These four average amplitudes will be used during output waveform generation of channel 1.

### 4.5.3 Pitch estimator

The speech sounds can be classified as voiced and unvoiced. Voiced sounds are produced by forcing air through the glottis with the tension of the vocal cords adjusted so that they vibrate in a relaxation oscillation, thereby producing quasiperiodic pulses of air which excite the vocal tract [16]. This vibration frequency of vocal cords is called pitch or fundamental frequency $F_0$ .

Pitch estimator determines the pitch period and also determines whether the speech signal is voiced or unvoiced. Pitch estimation is implemented using short time autocorrelation method. The block diagram of pitch estimatior is shown in Fig.4.3. The low pass filtered speech signal is center clipped and then its short time autocorrelation is calculated.

The center clipping is being done to remove the unwanted peaks in autocorrelation function, which are due to the vocal tract response. The buffer of amplitude estimator-1 is divided into three equal segments. The maximum amplitudes of first and third segments are found. Clipping level or clipping threshold is set as a 67% of the minimum of the two maximum amplitudes found earlier. For samples above threshold, the center clipper output is equal to input minus clipping threshold while for samples below the clipping level, the output is zero.

The computation of short time autocorrelation function has been done using the following equation [16].

$$r_n(k) = \sum_{m=0}^{N-1-k} [x(n+m)\cdot w(m)]\cdot[x(n+m+k)\cdot w(k+m)]$$

where $r_n(k)$ is kth autocorrelation lag at sample, n

$x(\cdot)$ is input sample sequence

$w(\cdot)$ is window coefficient sequence

$N$ is window length.

As the rectangular window is being used, $w(i) = 1 \quad 0 \le i \le N-1$ . The autocorrelation sequence attains a maximum at samples displaced by the period of the input sequence. Using this property, pitch period is being estimated.

If the peak of the autocorrelation function is more than 20% of r (0) then it is voiced otherwise it is classified as unvoiced. The minimum r (0) is 10. If it is less than 10, the speech signal is unvoiced.

### 4.5.4 Square wave generator

This generate square waves, with period equal to the estimated pitch period, with 50% duty cycle. The amplitude of the square wave is the average magnitude as estimated by the amplitude estimator. The polarity of the last output sample, in a block, is stored for maintaining the continuity at the block boundaries.

### 4.5.5 Zero crossing detector 1

This module detects the zero crossings of its input and outputs a unit impulse, whenever there is zero-crossing at its input.

### 4.5.6 Random wave generator 1

Random wave generator generates random waveform using timing information from zero crossing detector 1 and amplitude information from amplitude estimator 1. The output of this generator reverses polarity whenever there is a pulse at its input. So this maintains the zero crossing rate of the low pass filtered sequence.

### 4.5.7 High pass filter

The purpose of the high pass filter (HPF) is to filter the high frequency fricative information from the input speech signal. The cutoff frequency, $f_c$ of this high pass filter is 3000 Hz. A fourth order Butterworth HPF has been implemented. Its normalized transfer function is

$$H(s)=\frac{1}{s^4+2.613s^3+3.414\,s^2+2.613s+1} \qquad (4.2)$$

The above equation is denormalized by substituting, $s=\frac{\Omega_c}{s}$, where

$$\Omega_c=\tan\left(\frac{\omega_c}{2}\right) \; where \; \omega_c=\frac{2\Pi f_c}{f_s}$$

where $f_s$ is sampling rate in samples/s

$f_c$ is cutoff frequency in Hz

$\omega_c$ is normalized cutoff frequency in rad/sample

$\Omega_c$ is analog cutoff frequency in rad

Then eq. 4.2 is translated to discrete time domain using bilinear transformation, i.e., by substituting $s=\frac{z-1}{z+1}$. So the coefficients of the filter can be obtained after the bilinear transformation.

### 4.5.8 Pulse repetition rate (PRR) mapper

The PRR mapping algorithm was developed by Sapre [4]. The output of the zero crossing detector, of channel 2, is divided by 8 to reduce its frequency, and then the mapping algorithm is applied to this train of pulses. Basically this alogrithm is mapping the input pulse period, time period between the consecutive input pulses, to the pulse period which corresponds to the PRR of the upper band of the residual hearing region. A linear function is being used for this pulse period mapping. This has been implemented as y = m x + c, where x is the average input pulse period, y is the output pulse period, and the m and c are determined as per the input and output PRR ranges.

The pulse repetition rate mapping algorithm is described below. This algorithm is same as that explained in section 3.3.3. Only the step 5 is changed. Like earlier, the '$C_{in}$' is the counter for measuring the input pulse period, 'azct' is the average zero crossing time, and '$C_{out}$' is the counter for output pulse period. Whenever there is a pulse at the input, the count in '$C_{in}$' is used for calculating 'azct' and the counter '$C_{in}$' is reset, for measuring the next pulse period and is incremented at each next sample instant until the arrival of next pulse. The $C_{out}$ is calculated using the linear function, whenever its values reaches zero, and a pulse is passed to the mapper output. The $C_{out}$ is decremented by one and zero level is passed to the mapper output at each next sample. When the $C_{out}$ reaches zero, again the procedure repeats.

Algorithm:

1) $i=0;$

2) $if\ x(i)=0\ then\ c_{in}=c_{in}+1;\ goto\ step\ (4)\ else$

3) $azct=\dfrac{(a_0 * azct+c_{in})}{a_0+1}\ ;\ c_{in}=1.$

4) $if\ c_{out}>0,\ c_{out}=c_{out}-1\ ;\ y(i)=0;\ goto\ step(6)\ ;\ else$

5) $c_{out}=m\cdot azct+c\ ;\ y(i)=1.$

6) $i=i+1\ ;\ if\ i<256\ goto\ step\ (2)\ else\ end.$

While implementing the step 3 of the algorithm variables a1 and a2 are used, where a1 = $a_0/(a_0+1)$ and a2 = $1/(a_0+1)$. The input PRR 5k-9k is to be mapped onto 500-1050 pulses/s. The input zero crossing rate (ZCR) is scaled down by a factor 8. So now ZCR is 625 to 1125 per 10000 samples.

PRR mapping input zero crossing time range is

(10000/1125 to 10000/625) = ( 8.889 to 16)

PRR mapping output zero crossing time range is

(10000/1050 to 10000/500) = ( 9.524 to 20)

slope = ( 20-9.524 ) / ( 16- 8.889) = 1.4732
y = slope * (x - 16) + 20
then y = 1.4732 x - 3.57
hence m = 1.4732 and c = -3.57.

For the above PRR ranges, the parameters values are $a_0$ =2, m = 1.4732 and c = -3.57.

### 4.5.9 Bandpass filter

The purpose of the bandpass filter is to filter the formant energy of the input speech. Its lower and higher cutoff frequencies are 500 Hz and 2300 Hz respectively. A fourth order Butterworth filter has been implemented. Its normalized transfer function is

$$H(s)=\frac{1}{s^2+\sqrt{2}\,s+1} \tag{4.3}$$

The above Eq. is denormalized by substituting

$$s=\frac{s^2+\Omega_0^2}{\Omega\, s}$$

where $\Omega_0^2 = \Omega_1 \cdot \Omega_2 \;\; and \;\; \Omega = \Omega_2 - \Omega_1$ ,

$$\Omega_1 = \tan\left(\frac{\omega_1}{2}\right), \; where \;\; \omega_1 = \frac{2 \cdot \Pi f_1}{f_s},$$

$$\Omega_2 = \tan\left(\frac{\omega_2}{2}\right), \; where \;\; \omega_2 = \frac{2 \cdot \Pi f_2}{f_s},$$

where $f_s$ is sampling rate in samples/s
$f_1$ is lower cutoff frequency in Hz
$f_2$ is higher cutoff frequency in Hz
$\omega_1$ is normalized lower cutoff frequency in rad/sample
$\omega_2$ is normalized upper cutoff frequency in rad/sample
$\Omega_1$ is analog lower cutoff frequency in rad
$\Omega_2$ is analog upper cutoff frequency in rad

after denormalizing the eq. (4.3), it is translated to discrete time domain using bilinear transformation i.e. by substituting $s=\frac{z-1}{z+1}$. So the coefficients of the filter can be obtained after the bilinear transformation.

The random wave generator 2, amplitude estimator 3 and zero crossing detector 2 are similar to that of channel 1.

### 4.5.10 Output low pass filter

The output low pass filter function is to smoothen the waveform generators' output. It is a fourth order Butterworth filter. Its cutoff frequency, $f_c$ is 1000 Hz. Its normalized transfer function is

$$H(s)=\frac{1}{s^4+2.613s^3+3.414\,s^2+2.613s+1} \qquad (4.4)$$

The Eq. 4.4 can be denormalized and can be translated to discrete time domain using bilinear transformation, as like as in the case of input low pass filter, discussed in section 4.5.1. So the filter coefficients can be obtained from the transformed equation.

## 4.6 Testing of SCSA-2

SCSA-2 has been tested in two methods, modular testing and overall testing. In modular testing each software module is tested with appropriate stimuli. In overall testing, the SCSA-2 is tested in off-line processing and in real-time processing modes, as explained in section 4.5, using normal speech VCV files. It is observed that Channel-1 (alone) real-time processing is taking 14 ms per block, whereas complete scheme real-time processing is taking 19.6 ms per block. A brief discussion of the two methods of testing and their results follows.

### 4.6.1 Modular testing

The various modules of SCSA-2 have been tested. The testing details are as follows.

#### 4.6.1.1 Input low pass filter

The input low pass filter was tested with sinusoidal signal of 20 $v_{pp}$ amplitude with different frequencies. Its frequency response characteristic is given in Fig.4.4. Its 3 dB cutoff frequency is 300 Hz.

#### 4.6.1.2 Pitch estimator

The pitch estimator was tested extensively using sinusoidal, square, and triangular waveform inputs with different frequencies, and also in various signal-to-noise ratio (SNR) conditions. It is also tested with synthesized vowels /a/ and /i/ with various pitch frequencies, and also in various signal to noise ratio conditions.

Table 4.1 (a) presents the test results of sine wave input, of 10 $v_{pp}$ amplitude. It is observed that without noise the pitch estimation is well within standard deviation of

0.9 for all the frequencies. The noise has deteriorated the pitch estimation at 80 Hz, and 120 Hz to small extent. For other frequencies noise effect was not observed in pitch estimation.

Table 4.1 (b) describes the test results of square wave input,of 10 $v_{pp}$ amplitude. Without noise, the pitch estimation is correct with zero standard deviation for all frequencies except for 150 Hz. Noise has affected the pitch estimation of 80 Hz, 100 Hz, and 120 Hz to a small extent. For 150 Hz, 200 Hz, and 250 Hz the noise effect was not at all observed.

Table 4.1 (c) describes the test results of triangular wave input, of 10 $v_{pp}$ amplitude. Without noise, the pitch estimation is well within standard deviation of 0.98 for all the frequencies. The noise effect was not there for 250 Hz. 6 dB SNR has very much deteriorated the pitch estimation, having standard deviation up to 2.84 for 100 Hz.

Table 4.1 (d) describes the test results of the synthesized sound /a/ as the input to the pitch estimator. Its amplitude is 10 $v_{pp}$ . The various. pitch frequencies are 100 Hz, 125 Hz, and 200 Hz. The noise effect was not at all observed for all the frequencies.

Table 4.1.(e) describes the test results of the synthesized sound /i/ as the input to the pitch estimator. Its amplitude is 10 $v_{pp}$ . The various. pitch frequencies are 100 Hz, and 200 Hz. The noise effect was not at all observed for 200 Hz. For 100 Hz signal, with 9 dB SNR and 6 dB SNR, pitch estimation has got affected to some extent.

#### 4.6.1.3 Square wave generator

This was tested with sine wave input, for different frequencies. It has been observed that the square wave generator output is a square wave with 50% duty cycle and its frequency is same as that of the sinusoidal input.

#### 4.6.1.4 Zero crossing detector

This was tested with sine wave input, for different frequencies. It has been observed that the output pulse rate is double the frequency of the input sinusoidal waveform.

### 4.6.1.5 Random wave generator 1

This was tested in cascade with zero crossing detector. So sine wave of different frequencies is given to zero crossing detector input. Its output was fed to random wave generator-1 (RWG1). The output of RWG1 is observed as square wave and its frequency is equal to that of the sine wave input.

### 4.6.1.6 High pass filter

The high pass filter was tested with sinusoidal signal of 20 $V_{pp}$ amplitude with different frequencies. Its frequency response characteristic is given in Fig.4.5. It has been observed that its 3 dB cutoff frequency is 3000 Hz.

### 4.6.1.7 Pulse repetition rate mapper

The pulse repetition rate mapper is tested with sinusoidal input of different frequencies ranging from 2500 to 4500 Hz. The test results are given in Table 4.2. Its transfer characteristic is given in Fig. 4.6. It has been observed that the average zero crossing time, azct, is not able to converge to the period of the input pulses. This is due to the fixed point implementation of the algorithm.

The PRR mapper is also excited with white noise and its mapped output is stored in a file. A sinusoidal tone, of 4 kHz frequency, is also given as input to the mapper and its output is stored in a file. Both the PRR mapper responses, response to white noise and response to sinusoidal tone, are compared by their spectrograms, explained in Appendix A, by listening them through headphones, and also by computing their autocorrelation. The response to the 4 kHz sinusoidal tone is a tone of 500 Hz whereas the response to the white noise is having various frequency components.

### 4.6.1.8 Bandpass filter

The bandpass filter is tested with sinusoidal signal of 20 $V_{pp}$ amplitude with different frequencies. Its frequency response characteristic is given in Fig.4.7. It has been observed that its lower and upper 3 dB cutoff frequencies are 500 Hz and 2300 Hz, respectively.

#### 4.6.1.9 Output low pass filter

The output low pass filter is tested with sinusoidal signal of 20 $V_{pp}$ amplitude with different frequencies. Its frequency response characteristic is given in Fig.4.8. It has been observed that its 3 dB cutoff frequency is 1000 Hz.

### 4.6.2 Overall Testing

Overall testing of the speech processing scheme is done in two ways, off-line and real-time. For this purpose 13 normal speech VCV files, of sounds */apa, aba, ata, ada, aka, aga, ama, ana, asa, aza, afa, ava, aʃ a/*, are being used as stimuli. In the above normal speech files the classification of the consonants is, the sounds */p, t, k/* are unvoiced stops, */b, d, g/* are voiced stops, */m, n/* are nasals, */s,ʃ, f/* are unvoiced fricatives, and */z, v/* are voiced fricatives. As the main interest of the project is to map the fricative information to the upper part of the residual hearing band, the mapping can be observed in the processed files using the spectrograph, explained in Appendix A. In case of real-time processing speech signal from the microphone also can be given as input to the system. A brief description of this overall testing follows.

#### 4.6.2.1 Testing using off-line processing

The off-line processing method and the software (alongwith the name of the related programs) are explained in section 4.5. The stimuli (speech files) can be processed with complete speech processing scheme and also with channel-1 processing alone. Using these two methods, the 13 normal speech VCV files were processed and their processed responses were stored in files. These processed responses were analyzed using the spectrograph, explained in Appendix A. In the case of fricatives */s, z, ʃ/* mapping was observed, while in case of */f, v/* mapping was not observed.

The spectrogram of */asa/* is shown in Fig.4.9 (a), while the spectrograms of the processed /asa/ by channel 1 alone and by complete scheme are shown in Fig.4.9 (b) and (c) respectively. In Fig.4.9 (a), the formants can be observed in the /a/ portion and frication can be observed in /s/ portion. For the /a/ portion, the speech signal is detected as voiced and the periodic square wave is passed to the output as shown in Fig.4.9 (b) and (c). The various parallel strips are higher harmonics of the fundamental frequency of the square wave. The high frequency components of the frication part are mapped as shown in Fig.4.9 (c) whereas in Fig. 4.9 (b) no mapping is observed.

The spectrogram of the */aʃ a/* is shown in Fig. 4.10 (a), and that of its processed output of channel 1 alone and that of complete scheme are shown in Fig.4.10 (b) and (c), respectively. By comparing the spectrograms, it can be observed that during the frication, for few blocks mapping can be observed, while for few blocks channel 1 output is passed to the output. This latter one is because of the less frication energy in high freqencies during that period.

The spectrogram of the */aza/* and that of its processed output are shown in Fig.4.11 (a), and Fig. 4.11 (b) and (c), respectively. Mapping can be observed during the frication portion in Fig. 4.11 (c).

#### 4.6.2.2 Testing using real-time processing

The setup for real-time processing scheme is shown in Fig.4.2. The speech processing scheme can be tested using speech input through microphone, as the input to the preamplifier. Other way is using another PC, with plug in data acquisition card [17]. So the speech files can be transmitted from other PC and its analog output can be the input to preamplifier. The real-time processing method and the software (alongwith the name of the related programs) are explained in section 4.5.

The speech processing scheme was tested in real-time using normal speech VCV i.e */ ata, apa, aka, ada, aba, aga, ama, ana, afa, asa, aza, ava, aʃ a/* utterances of male speaker. The distinct cues were observed for voiced sounds and also for fricatives */s, z, ʃ/*. Listening tests also have been conducted using the real-time processing, which will be discussed in next chapter.

# CHAPTER 5

# LISTENING TESTS USING SCSA-2

## 5.1 Introduction

Listening tests have been conducted for evaluating the performance of the speech processing scheme in providing the cues about the intonation and rhythm of the speech, and also about certain high frequency fricatives /s, ʃ, z/. The listening tests can be conducted by using either the real-time processing or the off-line processing. The listening test setup for both the processing methods, the listening test procedure, and the analysis of the results are presented in this chapter.

## 5.2 Listening Test Setup

The listening test setup for both the processing methods, i.e. real-time and off-line, are explained in this subsection.

### 5.2.1 Setup for listening tests with real-time processing

The setup for conducting listening tests with real-time processing is shown in Fig.5.1. The PC is used for presenting the stimuli to subject, for storing the responses of the subject, and also for analyzing the responses. Samples of the speech signal file are output at 10 k Sa/s through the D/A port of the data acquisition card PCL-208 (from Dynalog Micro Systems, Bombay) [17].The D/A output is connected to the input of the signal conditioning hardware (explained in section 4.4.1). The filtered signal is applied to the ADC port of the DSP board, on-which the speech processing software is running. The DSP board and the Data acquisition card are installed on the same PC. This signal is sampled using the on-board timer of the DSP board and is processed in real-time and output through its D/A port. This processed analog signal is conditioned using the output signal conditioning hardware (explained in section 4.4.3). The output of power amplifier is passed to the TDH39 type headphones. The subject will be wearing the headphones and will be responding to the stimuli using the remote terminal, which is connected to the PC through RS232-C interface. The remote terminal is another PC with the communication software "kermit" running on it.

### 5.2.2 Setup for listening tests with off-line processing

The setup is shown in Fig.5.2. The DSP card is not being used while conducting the listening tests. The VCV normal speech files are processed in off-line using off-line processing, as explained in section 4.5. The processed speech data are stored in files. The PC is used for stimulating the subject, for storing the responses of the subject, and also for analyzing the responses. The processed speech data file is passed to the data acquisition card (PC add-on card) [17] using the PC Bus interface. The data acquisition card sends the speech data to the DAC port at a rate of 10 k sa/s. The DAC output of the data acquisition card is connected to the output signal conditioning hardware (explained in section 4.4.3). The output of power amplifier is connected to the TDH39 type headphones. The subject will be wearing the headphones and will be responding to the stimuli using the remote terminal, which is connected to the PC through RS232-C interface.

## 5.3 Listening Test Procedure And Test Results

The test material consisted of 12 VCV syllables with consonants */p, t, k, b, d, g, m, n, s, f, z, v/* and vowel */a/*. These syllables are digitized records of a male speaker. The main elements of the listening test are presentation of the set of processed VCV syllables to the subjects, and recording the subject responses. The subject has to identify the presented syllable by hitting the corresponding key on the terminal.

For presentation of stimuli and recording the responses of the subject a "computeried test administrator" program "CTA.c", has been developed earlier by Thomas [18]. This software has been modified as per the requirements and renamed as "Test.c".In one test session, each of the 12 VCV syllables, is presented 5 times in a randomized order with certain uniformity constraints [18].

Before starting a session, the DSP software is to be downloaded to the DSP board using the "Chan.c" program, then the session should be started using the "Test.c" program. The stimulation level should be adjusted to the comfortable listening level of the subject, using the potentiometer of the power amplifier.

In a session, the subject can listen to each stimulus item separately, any number of times. After becoming familiar with the stimuli, the subject proceeds to the test session. This test session can be with or without feedback. The PC records the responses from the subject during the test session, and forms a confusion matrix. In the confusion matrix, the rows correspond to the stimuli while the columns correspond to the responses. The diagonal entries correspond to correct responses and off-diagonal entries correspond to errors. Based on the number of correct responses the computer gives a score for each test session.

The test sessions should be conducted consecutively. When the latest 3 scores are within ±5%, the tests can be terminated. The confusion matrices of these 3 sessions can be added. For adding the confusion matrices a program cummat.c has been developed by Thomas [18]. This program has been modified as per the requirements and renamed as "cummat2.c".

## 5.4 Analysis of Listening Test Results

The confusion matrices are given either as stimulus-response frequencies or the probabilities estimated from the measured frequencies [19].

Let X be the set of n stimuli $\{x_1, x_2, \ldots, x_n\}$ and Y be the set of n responses $\{y_1, y_2, \ldots, y_n\}$. Let N(x ), N(y ), N(x ;y ) are the frequencies of stimulus x, response y, and the stimulus-response pair $(x_i, y_j)$ in a sample of N observations. The probabilities can be estimated as,

$$p(x_i, y_j) = \frac{N(x_i, y_j)}{N}$$

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

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

As it is difficult to study the error patterns in the confusion matrix, it is needed to reorganize the data. Two commonly used methods are calculation of recognition scores, and relative information transmission.

Recognition or articulation score, $R_s$ is the probability of the correct responses which is given by

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

Although this recognition score is useful, it does not provide the information on the distribution of the errors. If some of the stimuli are having a common feature, the stimuli can be grouped as per that feature and the resulting new matrix will give the recognition score for the transmission of this feature.

Information transmission analysis as used by Miller and Nicely [19] provides a measure of covariance between stimuli and responses. This method uses mean logarithmic probability (MLP) measure of information.

Let $I_s(x)$ and $I_s(y)$ are the information measures of the input stimulus, x and output response, y respectively. They are given by

$$I_s(x) = MLP(x) = -\sum_i p(x_i) \cdot \log_2 [p(x_i)] \quad bits,$$

$$I_s(y) = MLP(y) = -\sum_j p(y_j) \cdot \log_2 [p(y_j)] \quad bits,$$

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

$$I(x,y) = MLP(x) + MLP(y) - MLP(x,y)$$

$$= -\sum_{i,j} p(x_i, y_j) \log_2 \frac{p(x_i)\, p(y_j)}{p(x_i, y_j)}$$

The relative transmission from x to y is given by

$$I_{tr}(x,y) = \frac{I(x,y)}{I_s(x)} \quad where \quad I_s(x) \geq I(x,y) \geq 0.$$

The relative information transmission measure can also be applied to to the matrices that are derived from the original confusion matrix by grouping the stimuli in accordance with certain features.

### 5.5 Analysis Results

The listening tests were conducted on five normal hearing subjects PRS, ASH, PRV, PTL, and SAT. The listening tests were conducted with real-time processing using the complete scheme and also using channel-1 alone. The confusion matrices of the subjects, for both the cases i.e. complete scheme and channel-1 alone, are given in Appendix B. For four subjects, the recognition score of the complete scheme show improvement of around 6% when compared with that of the channel-1 alone. For one subject, there is almost no improvement in recognition score for both the cases. For all the subjects the most confusion was in recognizing the sounds /ama/ and /ana/. The presence of mapping had helped in recognizing the fricative /z/. This can be concluded form the matrices. In the case of the testing using channel-1 alone, there was confusion among /aza/ and /ava/ which has reduced the recognition score.

These confusion matrices should be analyzed for evaluating the performance of the speech processing scheme in providing the cues about the features like voicing, manner and place of articulation, etc., which can be supplemented to lipreading.

For analyzing the confusion matrices a program, "info.c", has been developed by Thomas [18]. This program is modified as per the requirements and renamed as "info2.c". This program expects a file infogr.dat, which has the grouping information of the stimuli as per various features. The 12 stimuli are grouped according to the features voicing, manner, and place of articulation. The classification table is shown in table 5.1.

The confusion matrices of the listening tests, using complete scheme, of the five normal subjects are analyzed using the program info2.c. Table 5.2 shows the summary of the listening test results. The recognitiom scores of voicing for both the processors are almost same for all the subjects. The ch1 processor is transmitting more relative information of voicing in case of three subjects. So in case of voicing, there is not much improvement due to channel-2. The recognition scores of manner, for both the processors are almost same for three subjects. Ch 1&2 is transmitting more relative information, about the manner of articulation, for three subjects, than that of ch 1. So ch2 is helping in moderate improvement in perceiving the manner. In case of the feature, place, the ch 1&2 is providing more relative information and also achieving high recognition scores compared to that of channel 1, for all the subjects. Overall, the recognition scores of ch 1&2 are more than that of channel 1, by nearly 6% (for one subject SAT, there is no improvement due to channel-2) and also the relative information transmitted by ch 1&2 is more than that of channel 1 alone.

# CHAPTER 6

## SUMMARY, CONCLUSIONS, AND SUGGESTIONS

The conductive hearing loss can be cured by medical intervention whereas for sensori-neural hearing loss sensory aids are to be used for improvement in hearing. The hearing aids can be classified as electro-acoustic aids and sensory aids. The sensory aids present the information by electrical stimulation of the auditory nerve or through alternative sensory modalities. These sensory aids can be classified as single channel and multi-channel depending upon the number of channels that are stimulating. In literature four types single channel sensory aids are reported. These aids are based on the frequency lowering techniques. These aids are presenting the cues about high frequencies and the normal speech simultaneously. So the cues can be masked by the speech.

A single channel speech processing scheme was proposed earlier by Pandey [2] and [3] which can provide cues about certain high frequency fricatives and also cues about intonation and rhythm of the input speech. Both the cues are separated in frequencies and these frequencies are chosen so as to accommodate the residual hearing frequency region of the deaf. The cues are provided one at a time to avoid masking. The modified forms of this scheme are used by Sapre [4] and Shah [5] for development of speech processor SCSA-1. Sapre [4] has implemented the scheme in off-line mode, in software, while Shah [5] has implemented in real-time, in software, using the DSP processor TMS320C25.

The aim of the project is to develop a speech processor. The implementation SCSA-1 has been tested thoroughly, i.e all the modules were tested independently and also as a system. During the testing a few problems were observed and these problems were solved by modifying the software of SCSA-1 and also by modifying the speech processing scheme of SCSA-1. This modified implementation is referred to as SCSA-2. The input and output PRR ranges are modified to 5k-9k and 500-1050, respectively. The SCSA-2 has been tested thoroughly and the results are presented. For testing the system, VCV normal speech sounds of male speaker are used. The mapped response of the speech processor for these sounds were analyzed using the spectrograph software, explained in Appendix A. Mapping was observed for the sounds */asa, aʃa, aza/*. Listening tests were conducted, with real-time processing, using SCSA-2 on five normal hearing subjects. These tests were conducted in two cases, using the complete scheme and using channel-1 alone. This has been done to compare the performance of the overall scheme over the channel-1 alone. The results show that this scheme provides more cues about the features manner, and place of articulation of the speech signal compared to that of channel-1 alone. In case of cues about voicing, there is no improvement due to channel-2.

Suggestions for the further work are as follows. Still more number of listening test sessions should be conducted with more normal hearing subjects for quantifying the performance. Listening tests should also be conducted on subjects with known sensori-neural loss with adequate residual hearing. The test sessions of channel-1 alone and channel-1&2 should be conducted in random sequence to avoid biasing. The modification, the decision criterion whether to present the mapped output of channel-2 or the output of channel-1, of the speech processing scheme has been done, based on the speech data of five male speakers. So this criterion should be tested for more number of speakers. Amplitude mapping has to be done for accommodating the dynamic loudness range of the residual hearing region of the deaf. The speech tests should also be conducted with other than VCV syllables. If this scheme is found effective, improvements in the present implementation may be done. The scheme should be modified to accommodate the cues about the features of female and child speech. For this purpose the speech data of children and female speakers should be analyzed and a criterion should be developed to recognize the speaker whether it is male, female or child. Based on this recognition the respective mapping algorithms should be executed for providing the cues of the speech signal.

Table 4.1 (a) Pitch estimator test results, testing with sinusoidal input, 10 $V_{pp}$

| | Frequency (Hz) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 80 | | 100 | | 120 | | 150 | | 200 | | 250 | |
| Signal | mean | std. dev | mean | std. dev | mean | std. dev | mean | std. dev | mean | std. dev | mean | std. dev |
| without noise | 80.3 | 0.4 | 100.5 | 0.5 | 120 | 0.0 | 150.4 | 0.92 | 200 | 0.0 | 250 | 0.0 |
| 12 dB SNR | 80.1 | 0.6 | 100.5 | 0.5 | 120 | 0.7 | 150.2 | 0.98 | 200 | 0.0 | 250 | 0.0 |
| 9 dB SNR | 79.9 | 0.8 | 100.0 | 0.5 | 120 | 0.7 | 149.6 | 0.92 | 200 | 0.0 | 250 | 0.0 |
| 6 dB SNR | 79.9 | 1.1 | 100.0 | 0.6 | 120 | 0.8 | 149.8 | 0.98 | 200 | 0.0 | 250 | 0.0 |

Table 4.1 (b) Pitch estimator test results, testing with square wave input, 10 $V_{pp}$

| | Frequency (Hz) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 80 | | 100 | | 120 | | 150 | | 200 | | 250 | |
| Signal | mean | std. dev | mean | std. dev | mean | std. dev | mean | std. dev | mean | std. dev | mean | std. dev |
| without noise | 80 | 0.0 | 100.0 | 0.0 | 120.0 | 0.0 | 149.4 | 0.8 | 200 | 0.0 | 250 | 0.0 |
| 12 dB SNR | 80 | 0.0 | 100.0 | 0.0 | 119.8 | 0.4 | 149.4 | 0.8 | 200 | 0.0 | 250 | 0.0 |
| 9 dB SNR | 80 | 0.0 | 100.2 | 0.4 | 119.7 | 0.5 | 149.4 | 0.8 | 200 | 0.0 | 250 | 0.0 |
| 6 dB SNR | 80 | 0.0 | 100.1 | 0.5 | 119.7 | 0.5 | 149.8 | 1.0 | 200 | 0.0 | 250 | 0.0 |

Table 4.1 (c) Pitch estimator test results, testing with triangular wave input 10 $V_{pp}$

| | Frequency (Hz) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 80 | | 100 | | 120 | | 150 | | 200 | | 250 | |
| Signal | mean | std. dev | mean | std. dev | mean | std. dev | mean | std. dev | mean | std. dev | mean | std. dev |
| without noise | 80.3 | 0.4 | 100.7 | 0.8 | 120.1 | 0.30 | 150.2 | 0.98 | 200.0 | 0.0 | 250 | 0.0 |
| 12 dB SNR | 80.6 | 0.6 | 100.7 | 0.8 | 120.0 | 0.45 | 150.0 | 1.00 | 200.0 | 0.0 | 250 | 0.0 |
| 9 dB SNR | 80.8 | 1.4 | 100.9 | 1.1 | 120.2 | 1.08 | 149.8 | 0.98 | 200.0 | 0.0 | 250 | 0.0 |
| 6 dB SNR | 81.4 | 2.1 | 101.1 | 2.8 | 120.7 | 1.85 | 149.8 | 1.33 | 200.4 | 1.2 | 250 | 0.0 |

Table 4.1 (d) Pitch estimator test results, testing with synthesized sound /a/ , 10 $V_{pp}$

| | Frequency (Hz) | | | | | |
|---|---|---|---|---|---|---|
| | 100 | | 125 | | 200 | |
| Signal | mean | std. dev | mean | mean | std. dev | mean |
| without noise | 100 | 0.0 | 125 | 0.0 | 200 | 0.0 |
| 12 dB SNR | 100 | 0.0 | 125 | 0.0 | 200 | 0.0 |
| 9 dB SNR | 100 | 0.0 | 125 | 0.0 | 200 | 0.0 |
| 6 dB SNR | 100 | 0.0 | 125 | 0.0 | 200 | 0.0 |

Table 4.1 (e) Pitch estimator test results, testing with synthesized sound /i / , 10 $V_{pp}$

| | Frequency (Hz) | | | |
|---|---|---|---|---|
| | 100 | | 200 | |
| Signal | mean | std. dev | mean | std. dev |
| without noise | 100.0 | 0.0 | 200 | 0.0 |
| 12 dB SNR | 100.0 | 0.0 | 200 | 0.0 |
| 9 dB SNR | 99.9 | 0.3 | 200 | 0.0 |
| 6 dB SNR | 99.8 | 0.4 | 200 | 0.0 |

Table. 4.2 Tranfer Characteristic of PRR mapper

| Input PRR | Output PRR |
|---|---|
| 5000 | 554 |
| 5800 | 624 |
| 6200 | 666 |
| 6800 | 768 |
| 7400 | 834 |
| 8200 | 1000 |
| 9000 | 1110 |

Table 5.1 Grouping of the stimuli

| Feature | Stimuli | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *apa* | *aba* | *ata* | *ada* | *aka* | *aga* | *ama* | *ana* | *asa* | *aza* | *afa* | *ava* |
| Voicing | uv | v | uv | v | uv | v | v | v | uv | v | uv | v |
| Manner | s | s | s | s | s | s | n | n | f | f | f | f |
| Place | fro. | fro. | m | m | b | b | fro. | m | m | m | fro. | fro. |

uv : unvoiced
v : voiced
s : stop
n : nasal

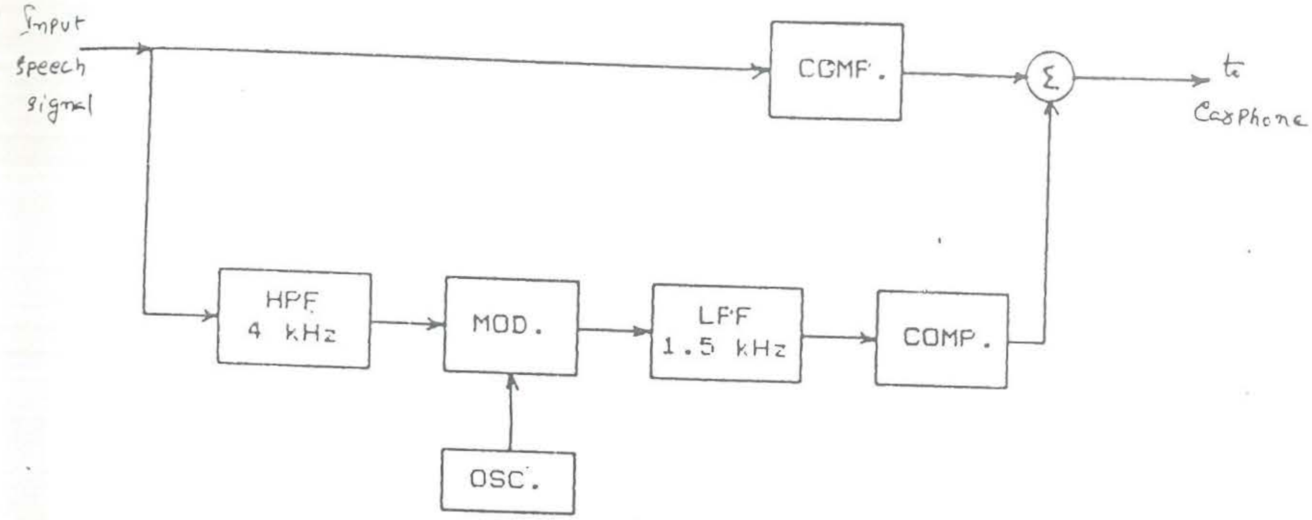
f : fricative
fro. : front
m : middle
b : back

Table 5. 2 Summary of listening test results

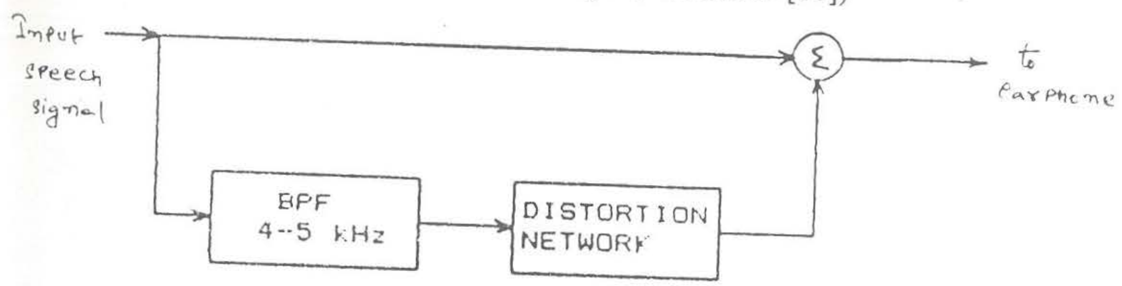
| Subject | Procesor | Recognition score | | | | Relative information transmitted | | | | Mean resp. time s |
|---|---|---|---|---|---|---|---|---|---|---|
| | | VO | MN | PL | TTL | VO | MN | PL | TTL | |
| PRS | ch 1 | 99 | 96 | 86 | 82 | 92 | 83 | 59 | 83 | 1.40 |
| | ch 1&2 | 98 | 99 | 89 | 88 | 84 | 94 | 67 | 89 | 1.71 |
| PRV | ch 1 | 98 | 95 | 78 | 74 | 87 | 81 | 38 | 77 | 3.18 |
| | ch 1&2 | 91 | 96 | 84 | 83 | 55 | 84 | 51 | 83 | 2.88 |
| ASH | ch 1 | 99 | 95 | 81 | 78 | 92 | 79 | 48 | 79 | 3.24 |
| | ch 1&2 | 97 | 91 | 90 | 83 | 82 | 71 | 67 | 82 | 3.39 |
| PTL | ch 1 | 98 | 88 | 82 | 78 | 84 | 66 | 47 | 77 | 3.25 |
| | ch 1&2 | 98 | 96 | 89 | 87 | 90 | 85 | 62 | 86 | 2.38 |
| SAT | ch 1 | 84 | 82 | 71 | 54 | 37 | 48 | 34 | 59 | 3.23 |
| | ch 1&2 | 78 | 78 | 73 | 54 | 22 | 41 | 31 | 56 | 5.91 |

ch 1 : Test using channel-1 software alone
ch 1&2 : Test using channel-1&2 software

VO : Voicing
MN : Manner
PL : Place
TTL : Overall

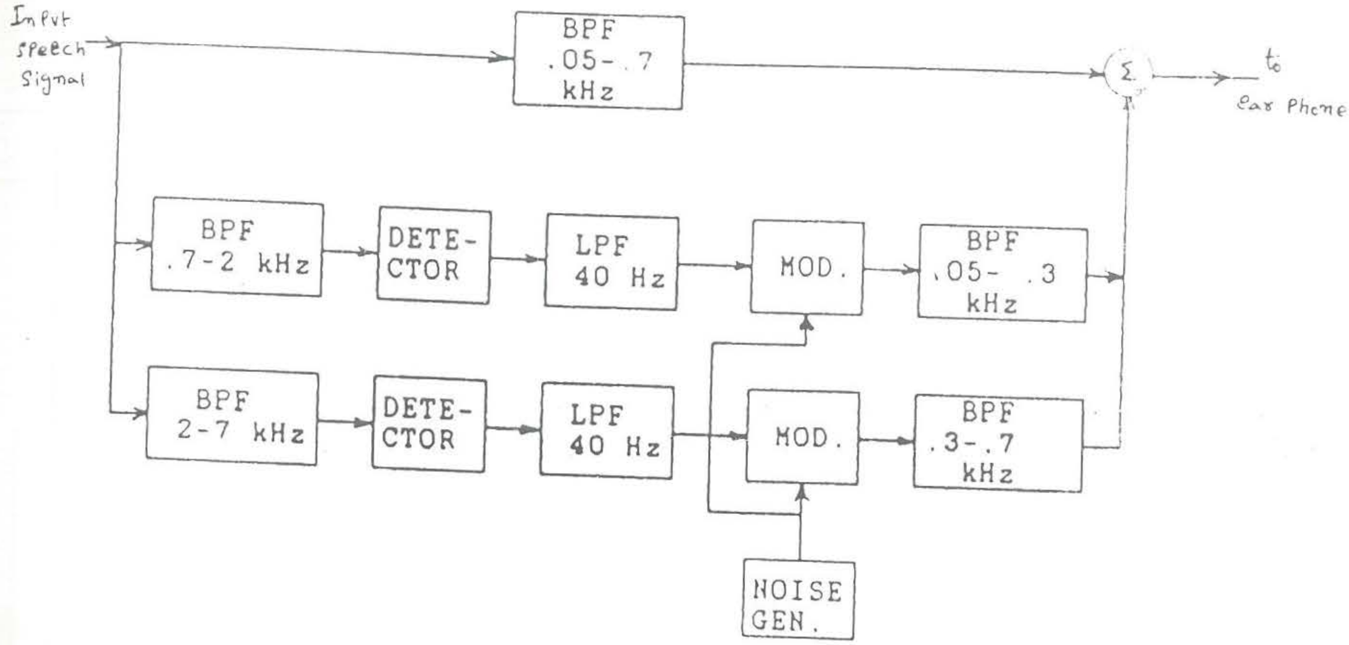



(a) Modulation technique (Johannson [10])

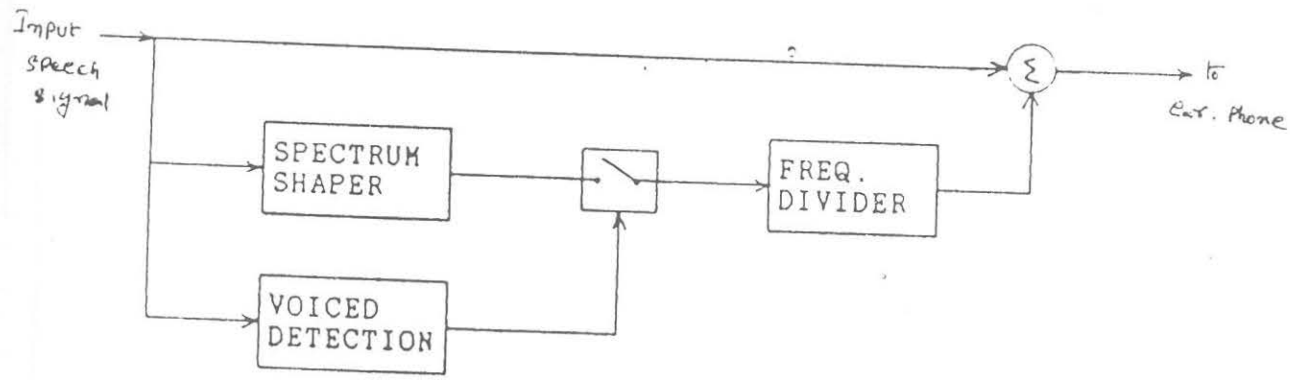



(b) Distortion technique (Johannson [9])

Fig. 2.1 Block diagrams of different frequency transposing hearing aids ( Levitt *et. al.* [8])




(c) Vocoder technique (Priminov [9])




(d) Frequency division technique (Guttman-Nelson [12])

Fig. 2.1 contd. Block diagrams of different frequency transposing hearing aids ( Levitt *et. al.* [8])

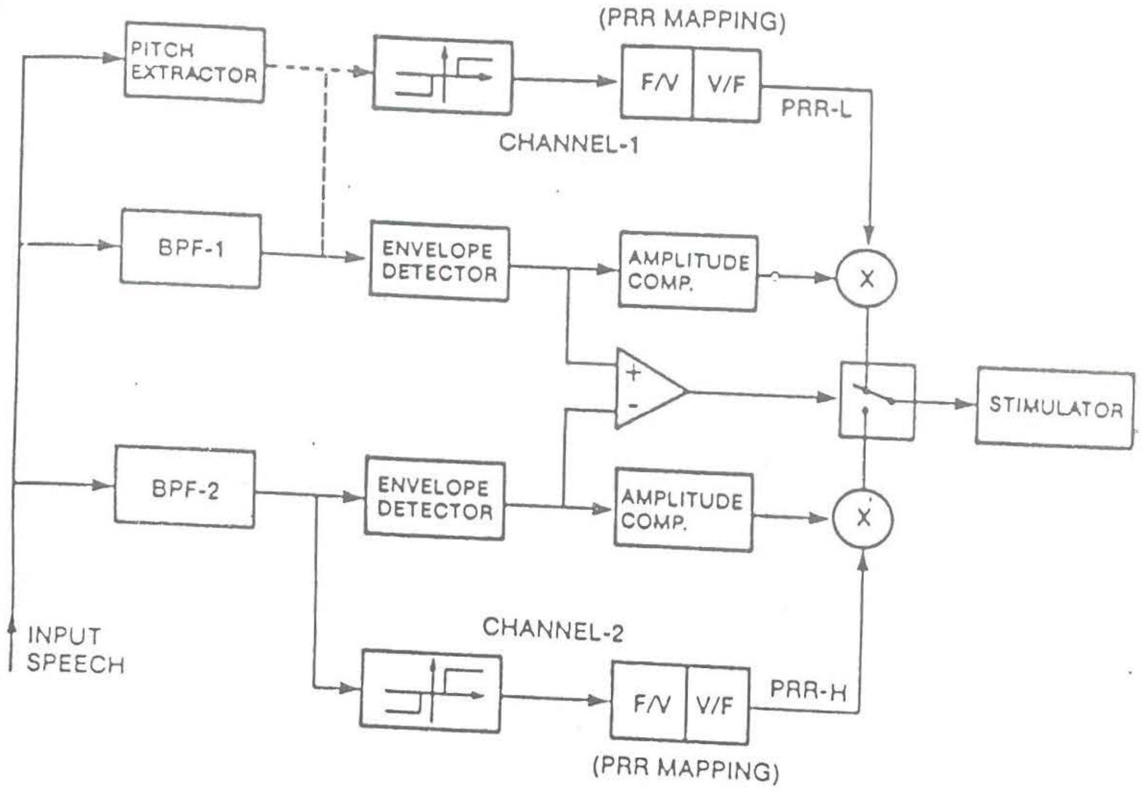



Fig. 3.1 Block diagram of speech processing scheme (Pandey [2] and [3])
BPF: bandpass filter, F/V: frequencyy to volatage converter, V/F: voltage to frequency converter, COMP: compressor, Channel-1: low frequency periodicity (voicing & suprasegmmentals), Channel-2: high frequency noise bursts, PRR: pulse repetition rate

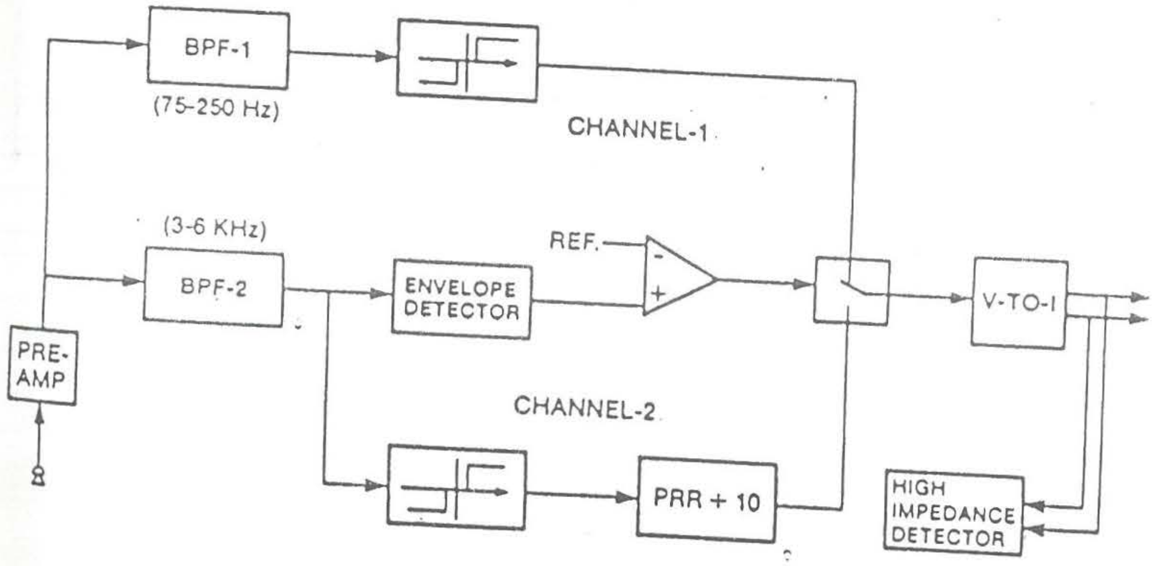



Fig. 3.2 Block diagram of the speech processor and stimulator (pandey [2] and [3])
V-TO-I : voltage to current converter

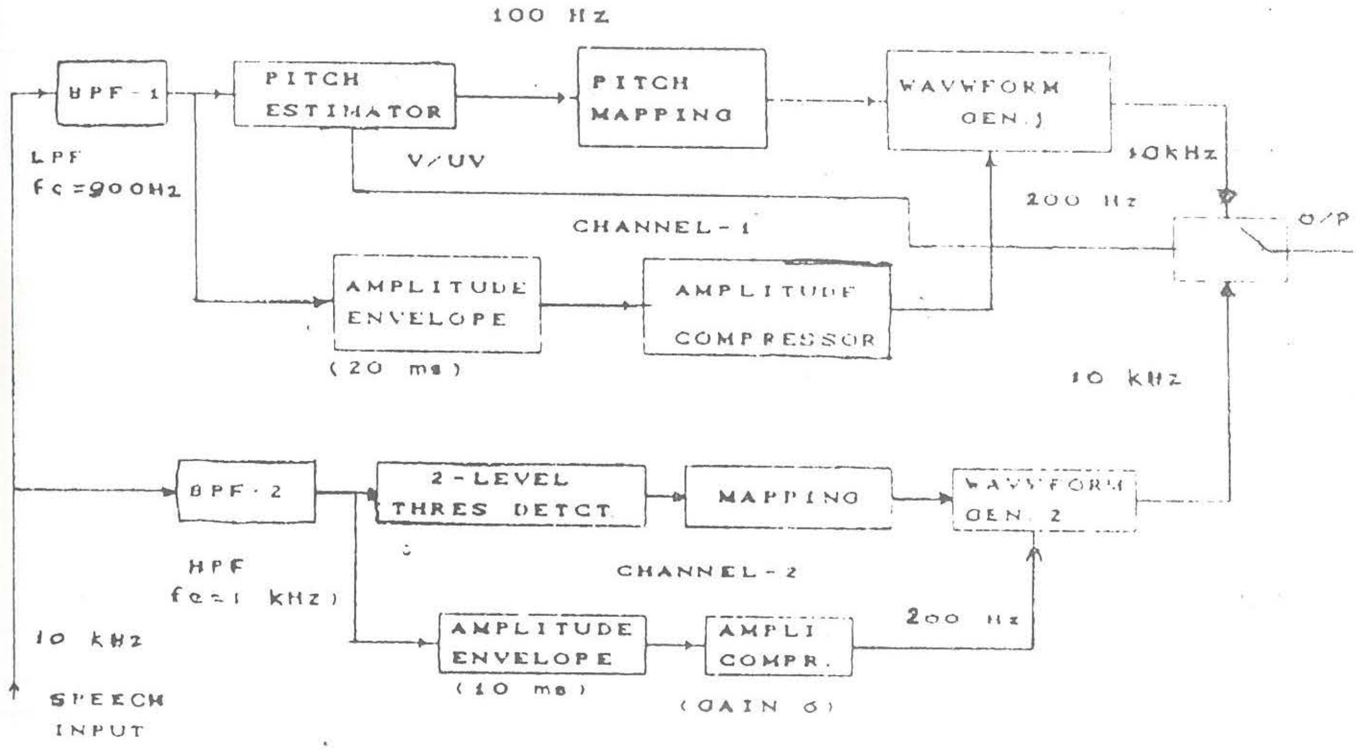



Fig.3.3 Block diagram of the speech processor for single channel auditory prosthesis (Sapre [4])

AMPLI. COMPR: amplitude compressor, BPF: bandpass filter, HPF: high pass filter, LPF: low pass filter, V/UV: voiced/unvoiced signal, CHANNEL-1: low frequency prosodic information, CHANNEL-2: frication inforation, 2-LEVEL THRES. DETCT: two level threshold detector.

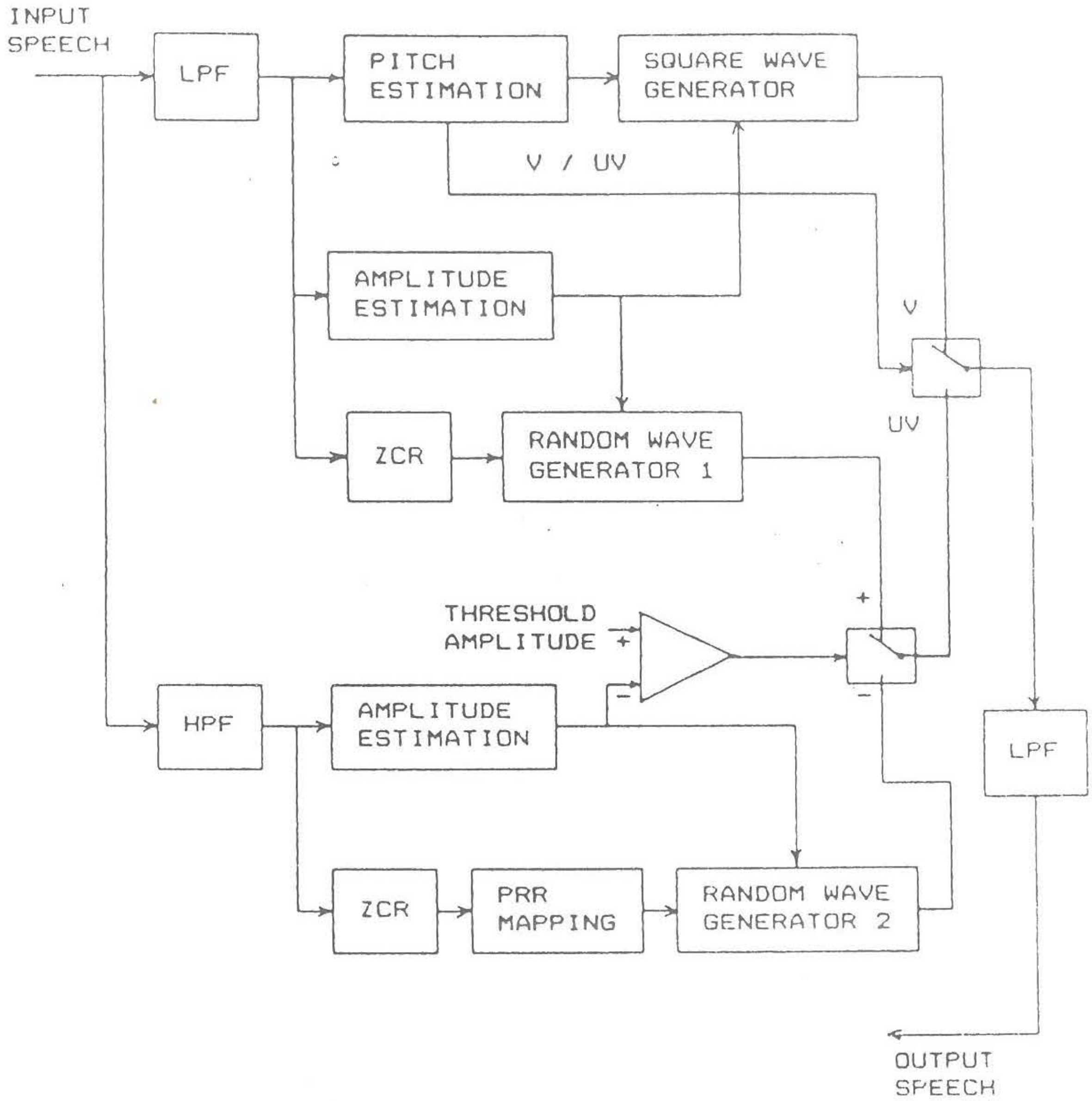



Fig. 3.4 Block diagram of the speech processing scheme of SCSA-1 (Shah [5])

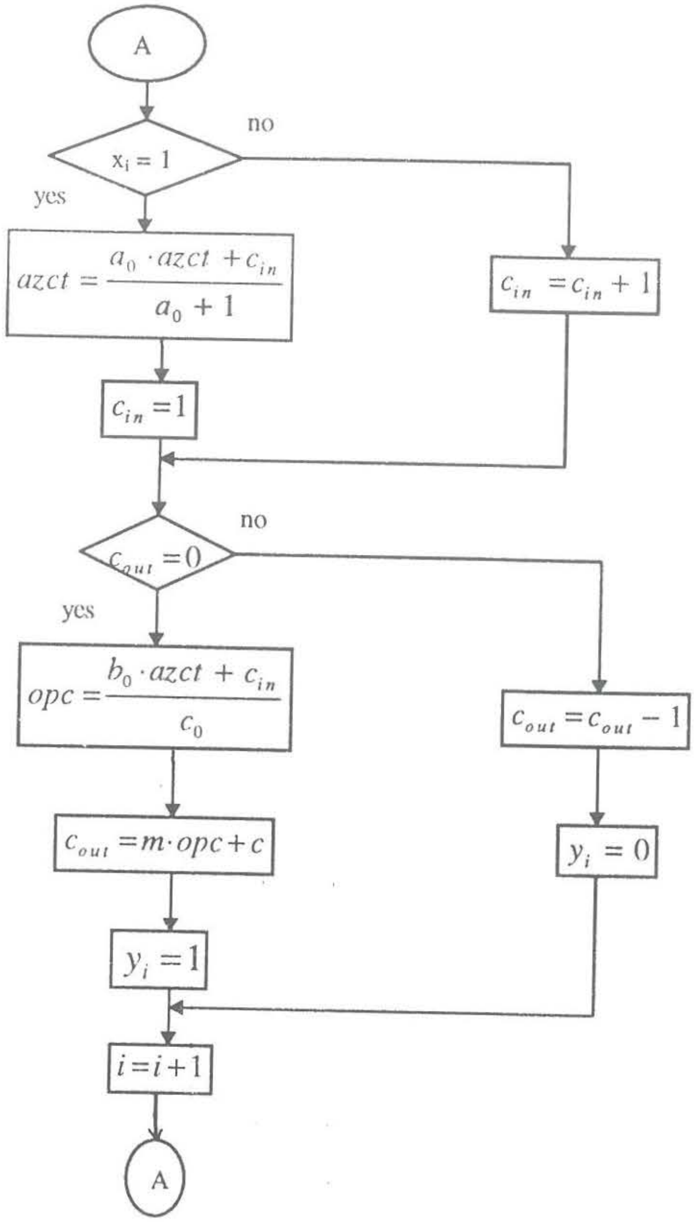



Fig. 3.5 Flow chart of PRR mapping algorithm

$x_i$ : input pulse sequence
$c_{in}$ : input counter
azct : average zero crossing time
$a_0$, $b_0$, $c_0$, m, and c are parameters

$y_i$ : output pulse sequence
$c_{out}$ : output counter
i : sample nummber

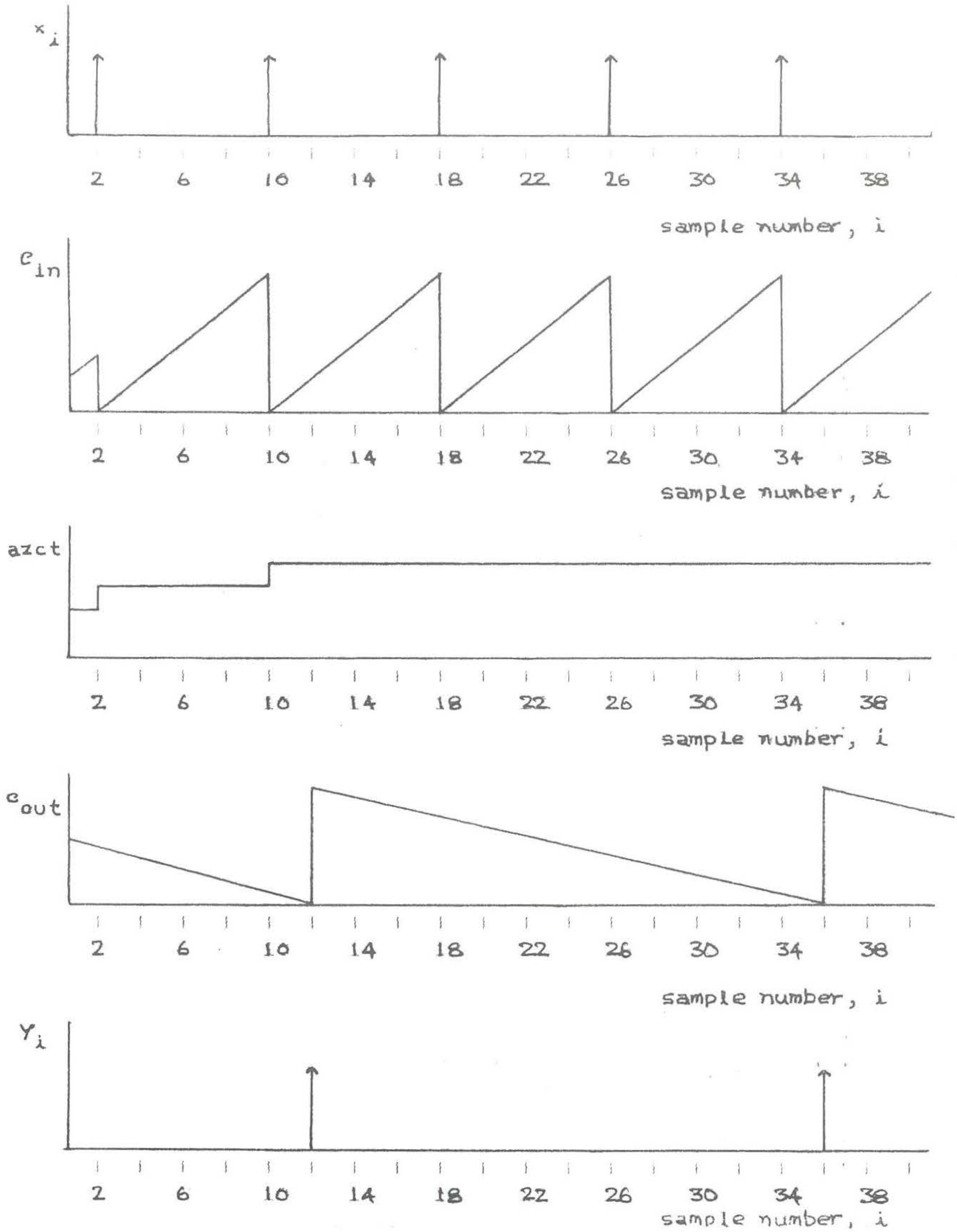



Fig. 3.6 Timing diagram of PRR mapping
(for this figure, constant 5000 input PRR is assued)

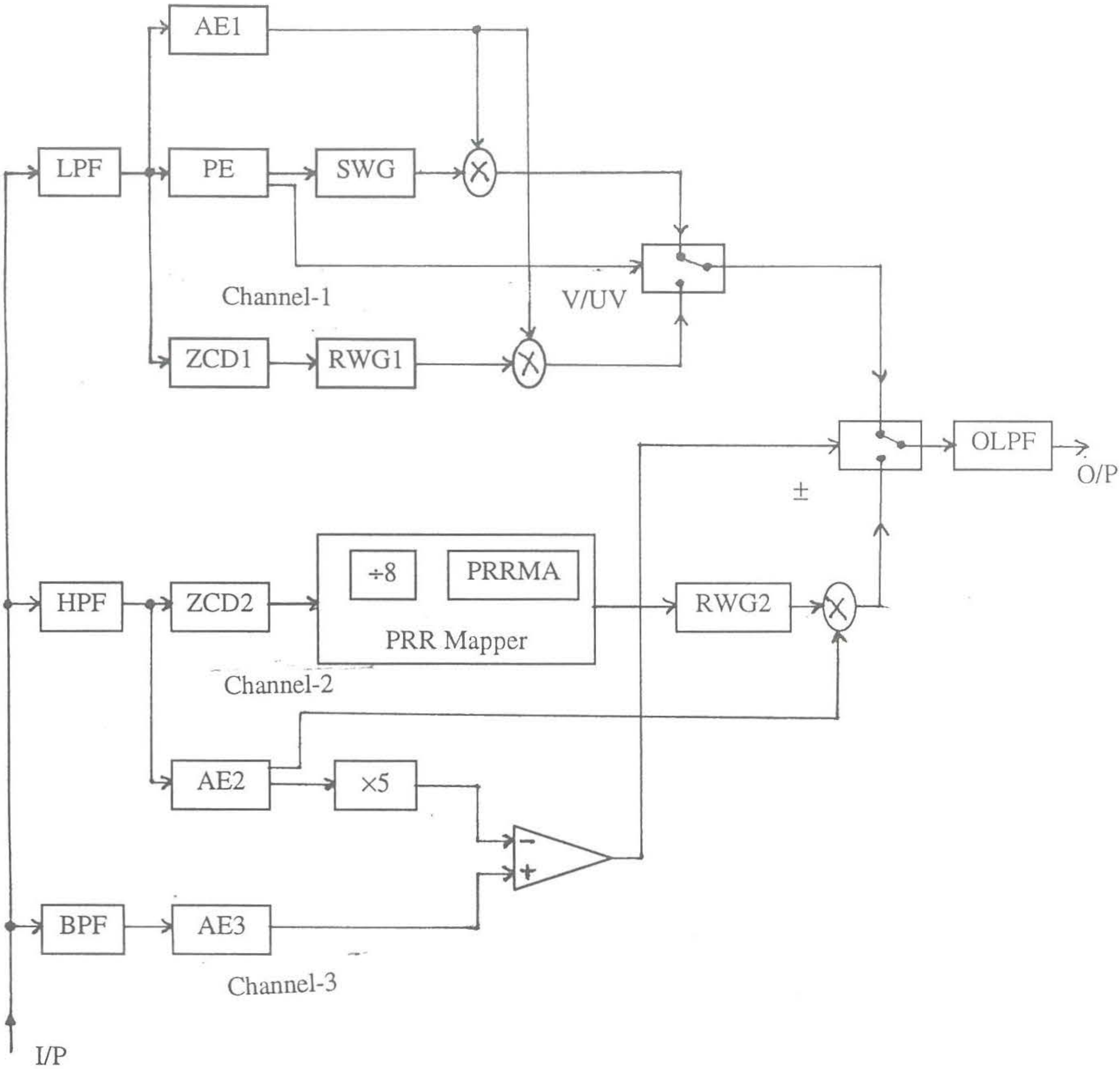



Fig. 4.1 Block diagram of the speech processing scheme of SCSA-2

| | | | |
|---|---|---|---|
| I/P | : Input Speech | O/P | : Processed Output |
| LPF | : Low pass filter | SWG | : Square Wave Generator |
| BPF | : Band Pass Filter | RWG | : Random Wave Generator |
| HPF | : High Pass Filter | PRRMA: | Pulse Repetition Rate Mapping Algorithm |
| PE | : Pitch Estimator | OLPF | : Output Low Pass Filter |
| AE | : Amplitude Estimator | V/UV | : Voiced/Unvoiced |
| ZCD | : Zero Crossing Detector | | |

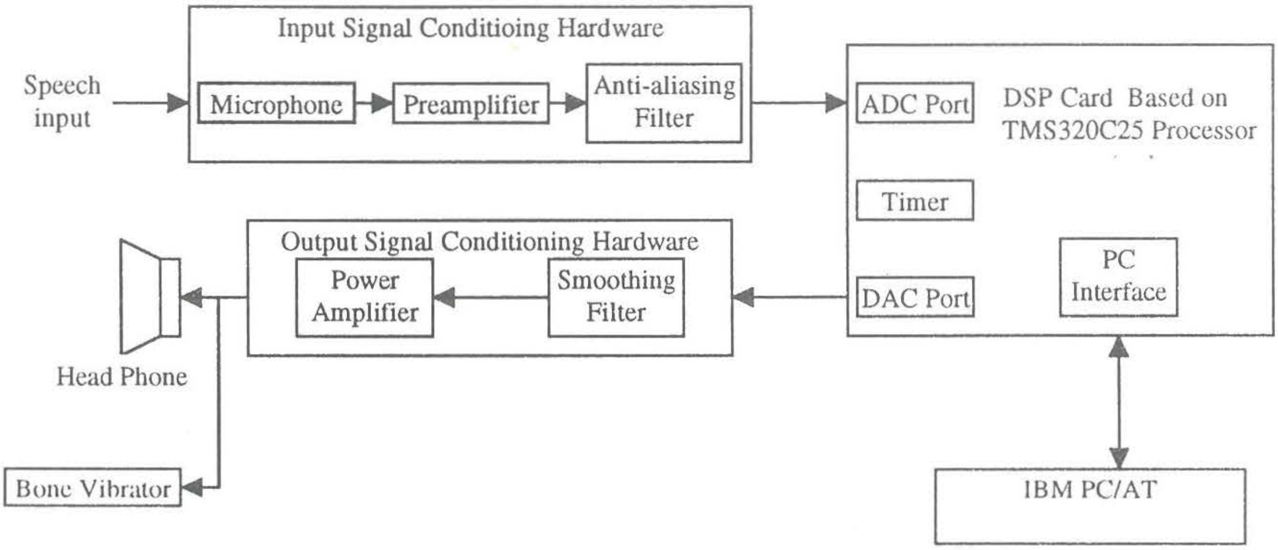



Fig. 4.2 Hardware Setup of SCSA-2

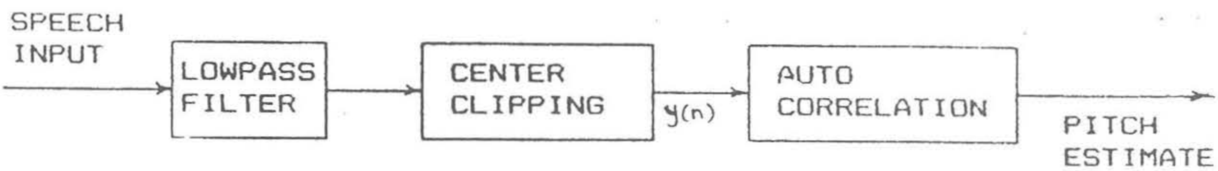



Fig. 4.3 Block diagram of Pitch Estimator

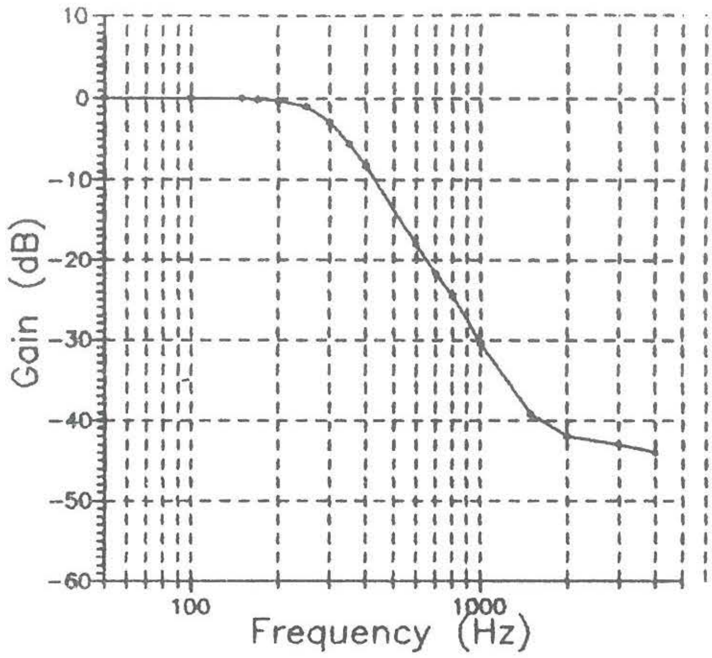



Fig. 4.4 Frequency response of low pass filter

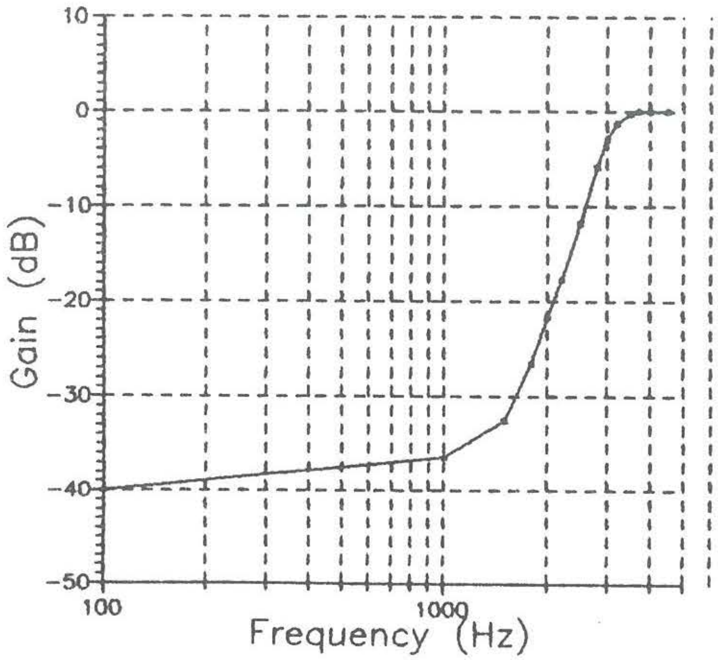



Fig. 4.5 Frequency response of high pass filter

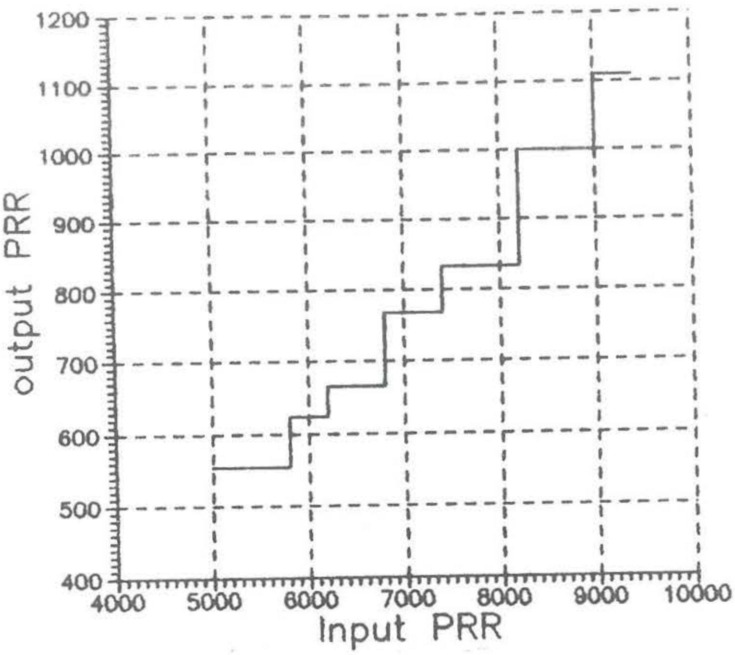



Fig. 4.6 Pulse repetition rate mapper transfer characteristic

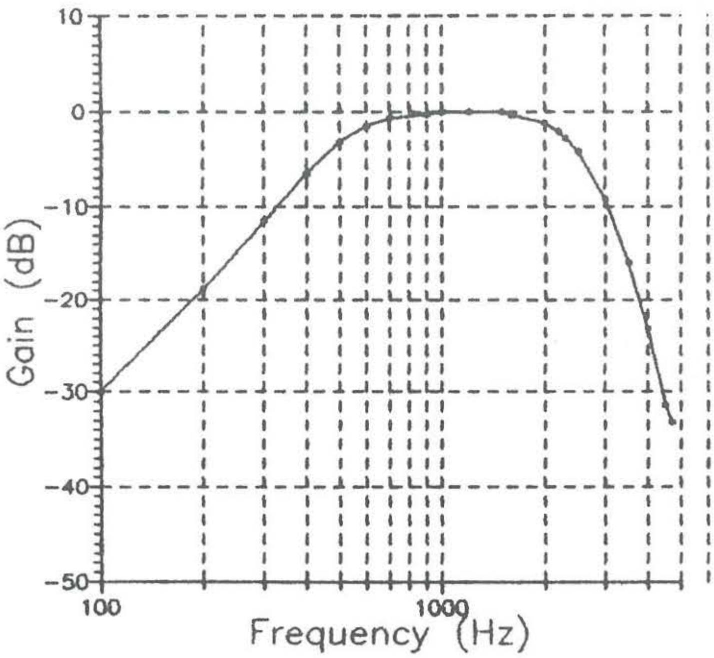



Fig. 4.7 Frequency response of bandpass filter

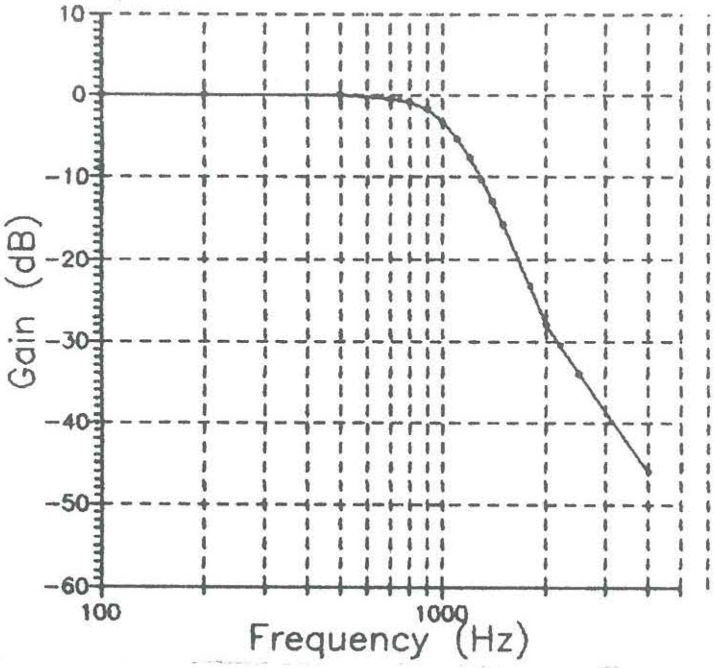
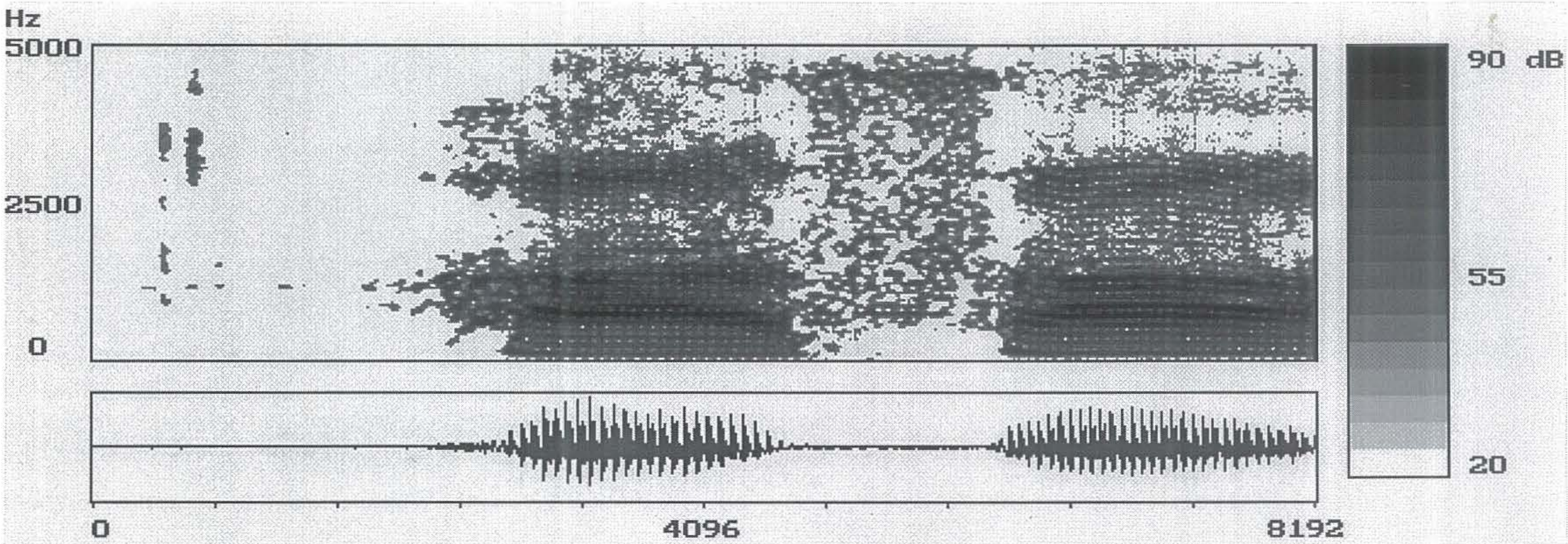


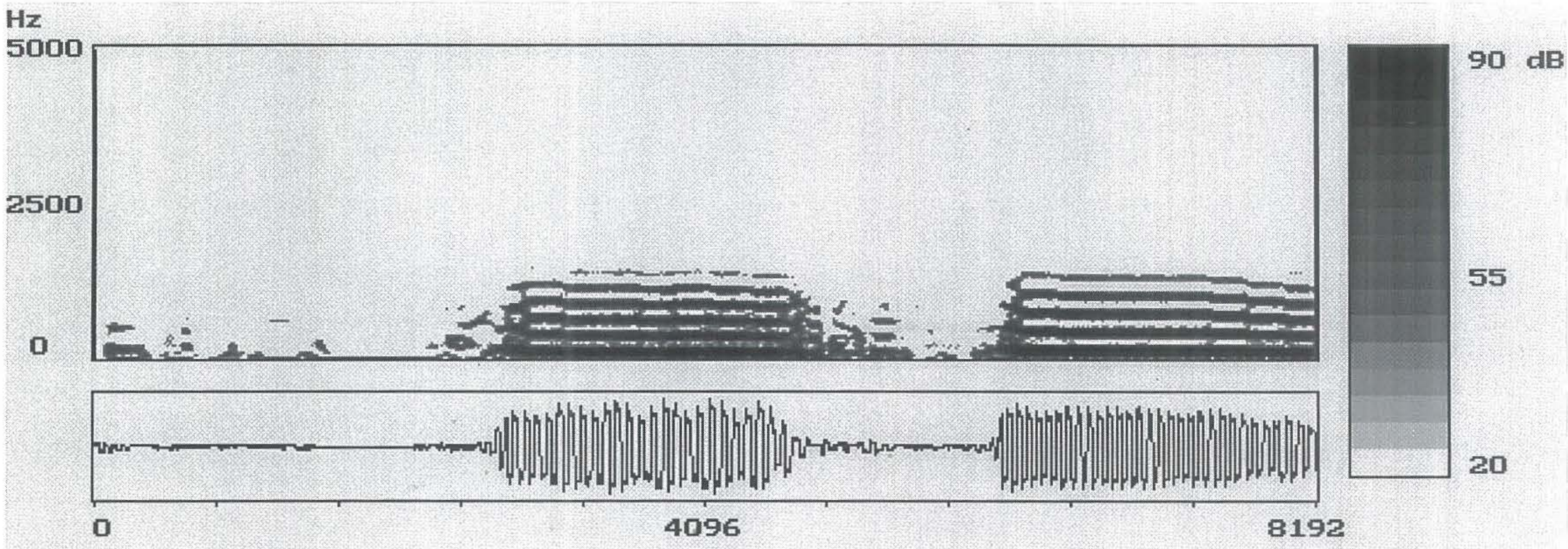

Fig. 4.8 Frequency response of output low pass filter




```
fine movement      : arrow keys
coarse movement    : 2,4,6,8 (arrows with numlock)
measurements       : Enter
cursors off/on     : F1
band analysis      : F2
quit               : Esc
```

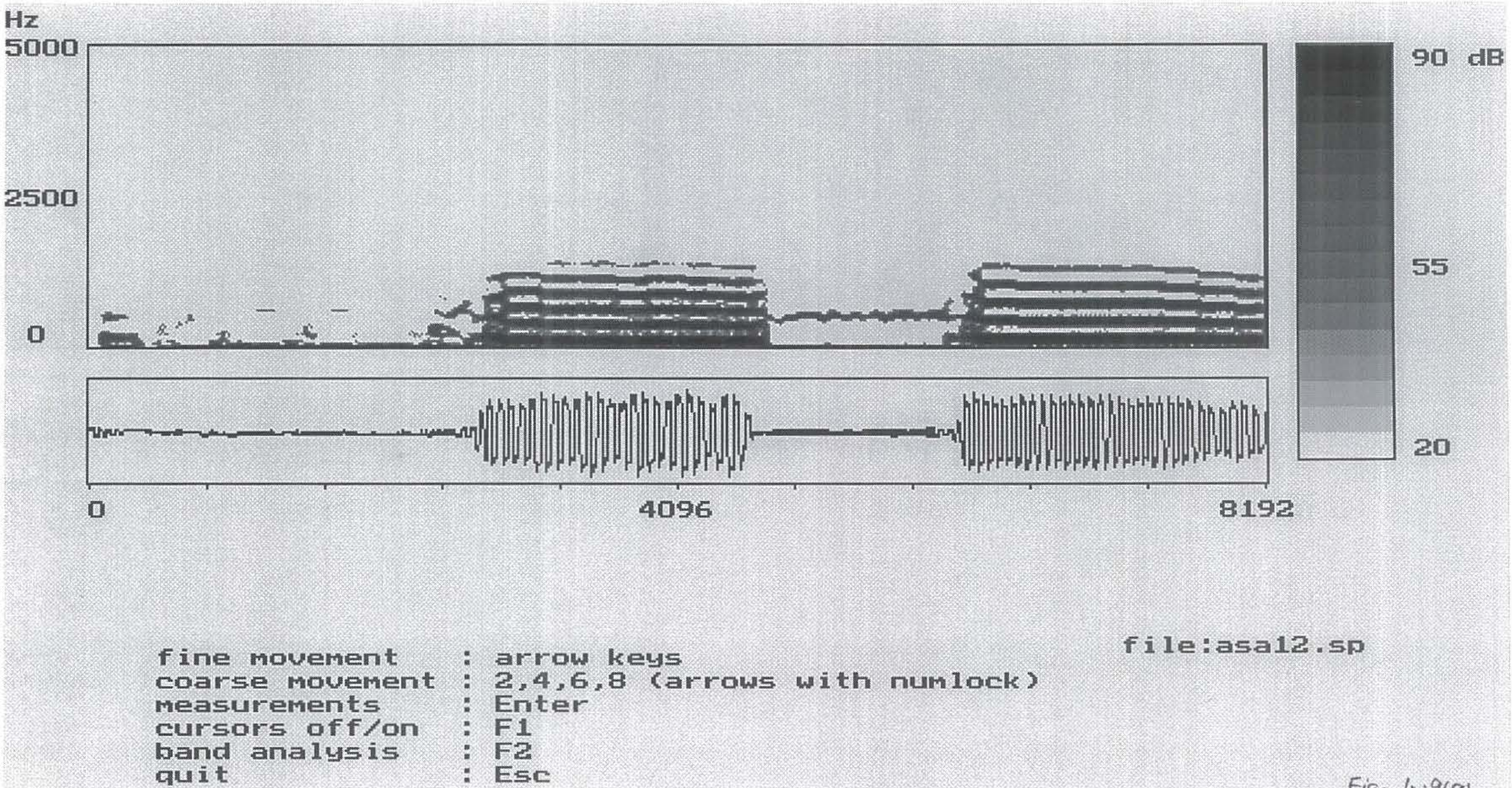
file:asa.spg

Fig. 4.9 (a)



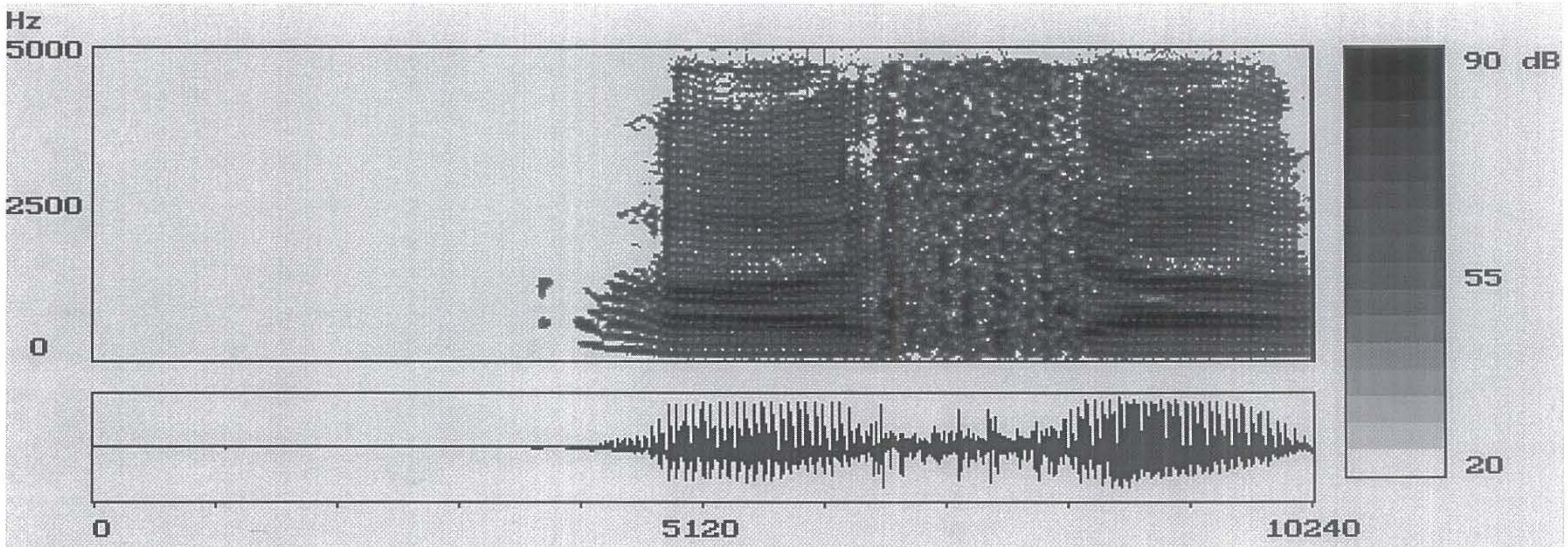

```
fine movement     : arrow keys                          file:asa1.spg
coarse movement   : 2,4,6,8 (arrows with numlock)
measurements      : Enter
cursors off/on    : F1
band analysis     : F2
quit              : Esc
```

Fig. 4.9(b)



Hz
5000
2500
0
90 dB
55
20
0
4096
8192
file:asa12.sp
fine movement : arrow keys
coarse movement : 2,4,6,8 (arrows with numlock)
measurements : Enter
cursors off/on : F1
band analysis : F2
quit : Esc

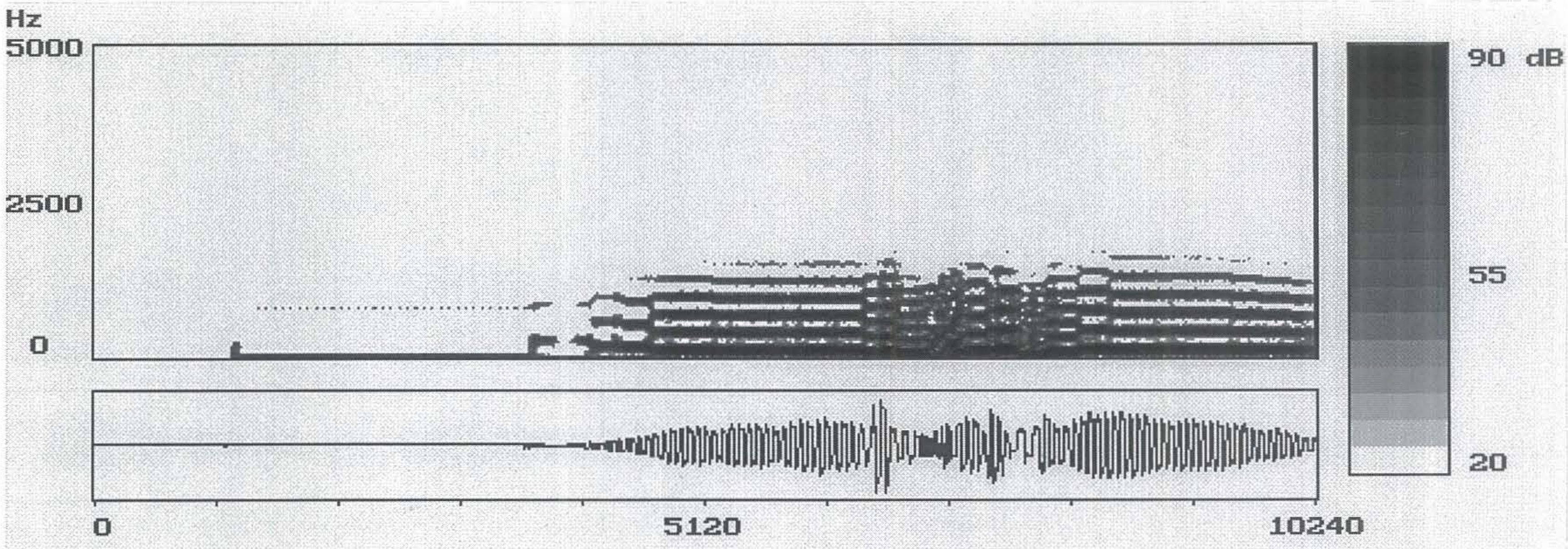

Fig. 4.9(c)




```
fine movement      : arrow keys
coarse movement    : 2,4,6,8 (arrows with numlock)
measurements       : Enter
cursors off/on     : F1
band analysis      : F2
quit               : Esc
```

file:asha.spg

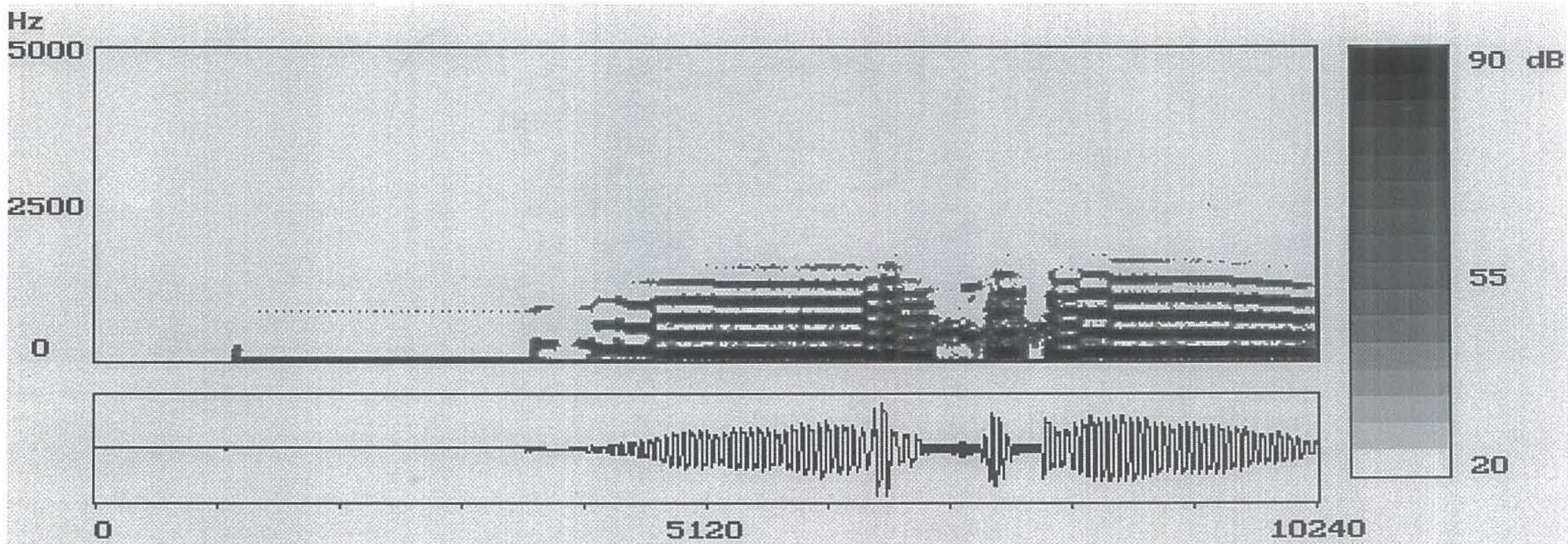
Fig. 4.10 (a)




file:ashal.sp

```
fine movement    : arrow keys
coarse movement  : 2,4,6,8 (arrows with numlock)
measurements     : Enter
cursors off/on   : F1
band analysis    : F2
quit             : Esc
```

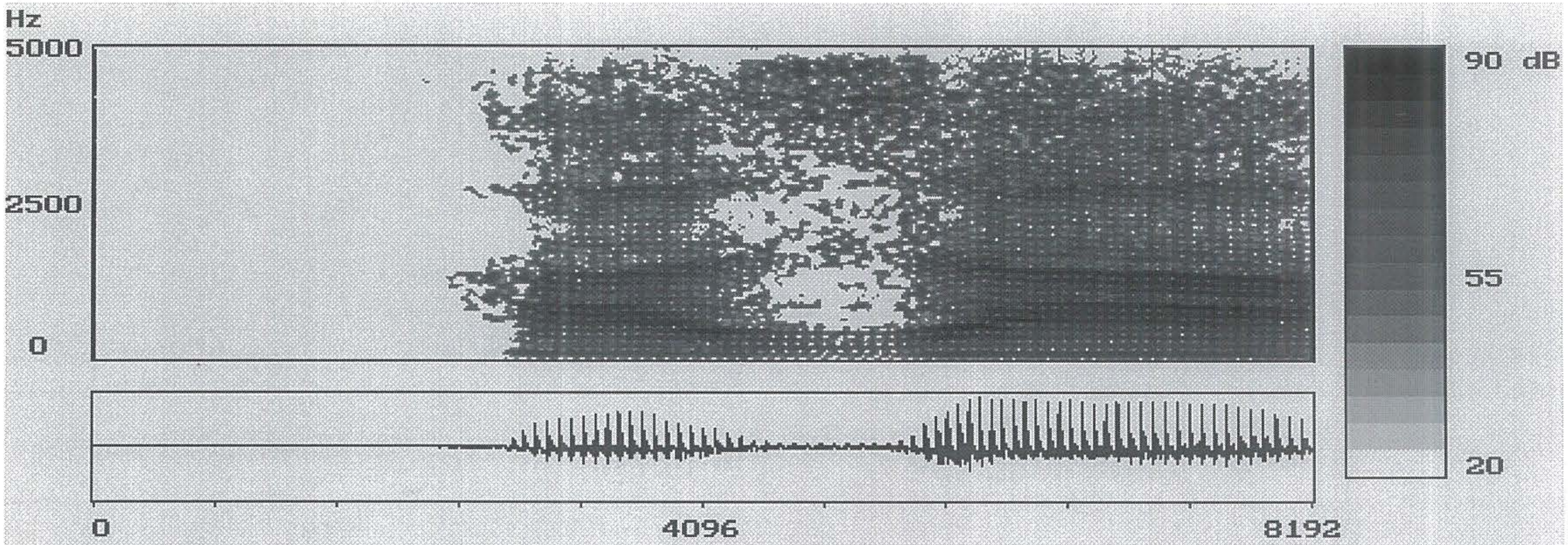
Fig. 4.10(b)




file:asha12.s

```
fine movement     : arrow keys
coarse movement   : 2,4,6,8 (arrows with numlock)
measurements      : Enter
cursors off/on    : F1
band analysis     : F2
quit              : Esc
```

Fig. 4.10(c)




file:aza.spg

fine movement : arrow keys
coarse movement : 2,4,6,8 (arrows with numlock)
measurements : Enter
cursors off/on : F1
band analysis : F2
quit : Esc

Fig. 4. 11(a)

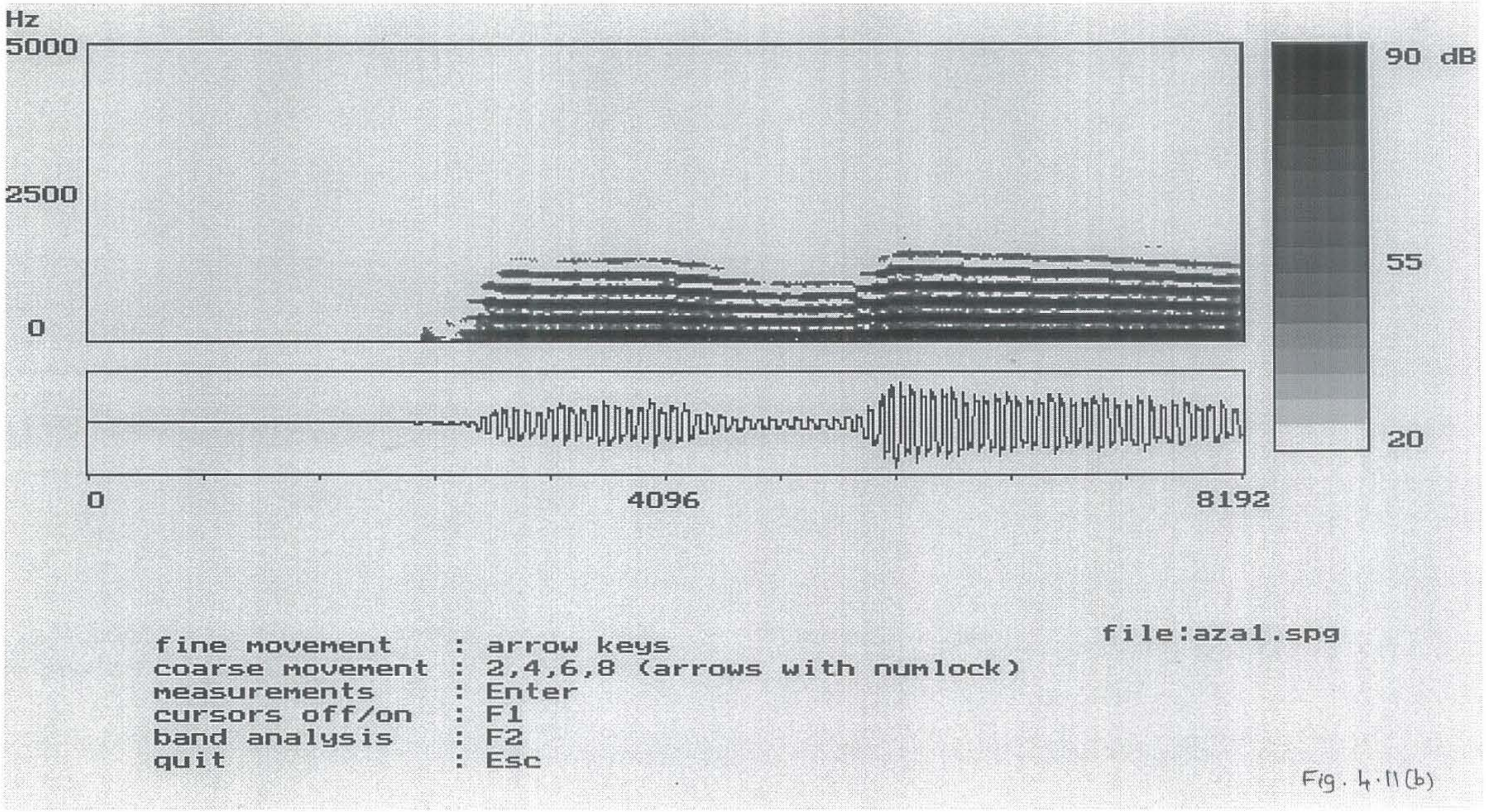

Hz
5000
2500
0
90 dB
55
20
0
4096
8192
file:aza1.spg
fine movement : arrow keys
coarse movement : 2,4,6,8 (arrows with numlock)
measurements : Enter
cursors off/on : F1
band analysis : F2
quit : Esc

Fig. 4.11(b)

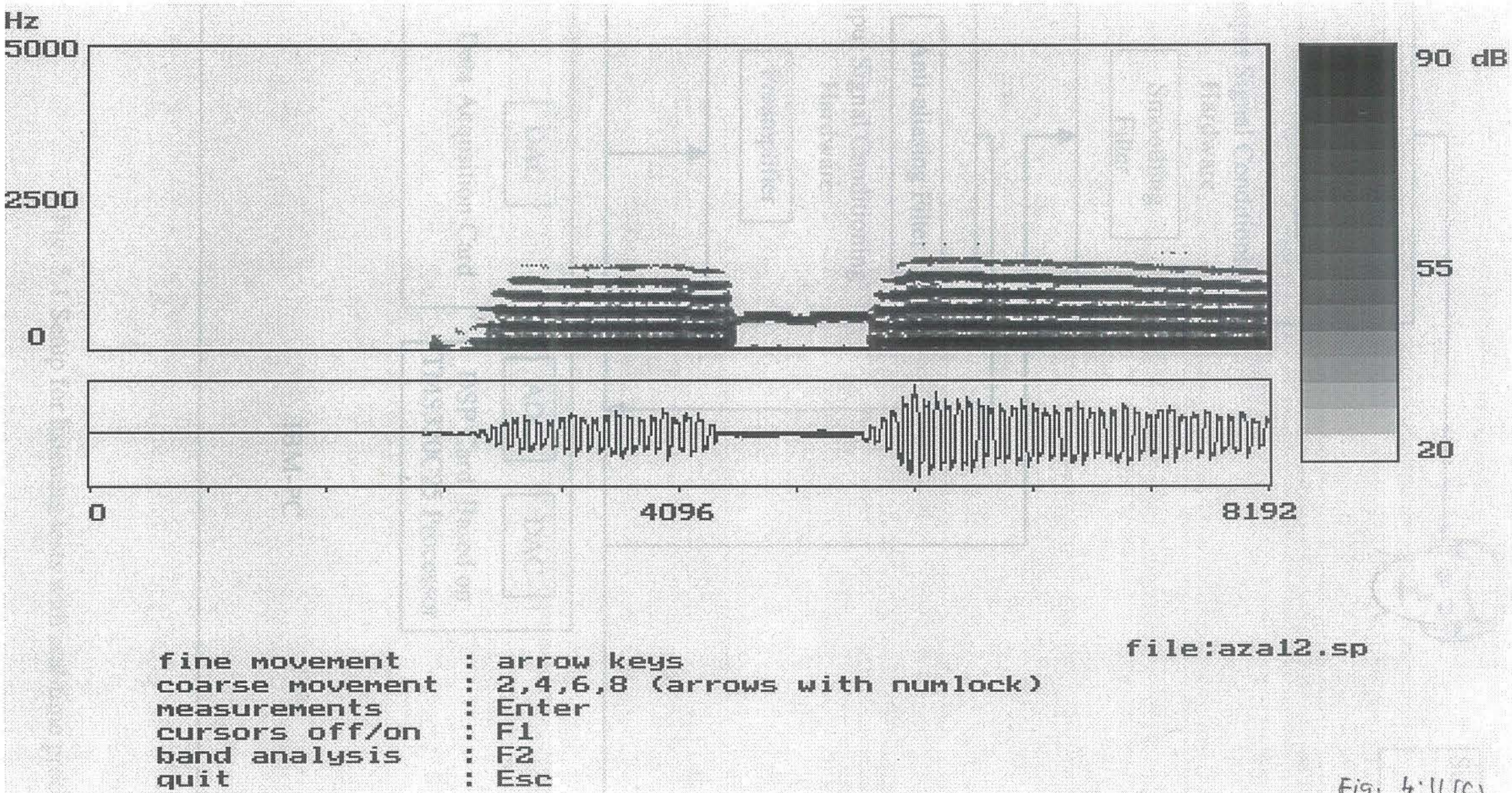

Hz
5000
2500
0
90 dB
55
20
0
4096
8192
file:aza12.sp
fine movement : arrow keys
coarse movement : 2,4,6,8 (arrows with numlock)
measurements : Enter
cursors off/on : F1
band analysis : F2
quit : Esc

Fig. 4·11 (c)

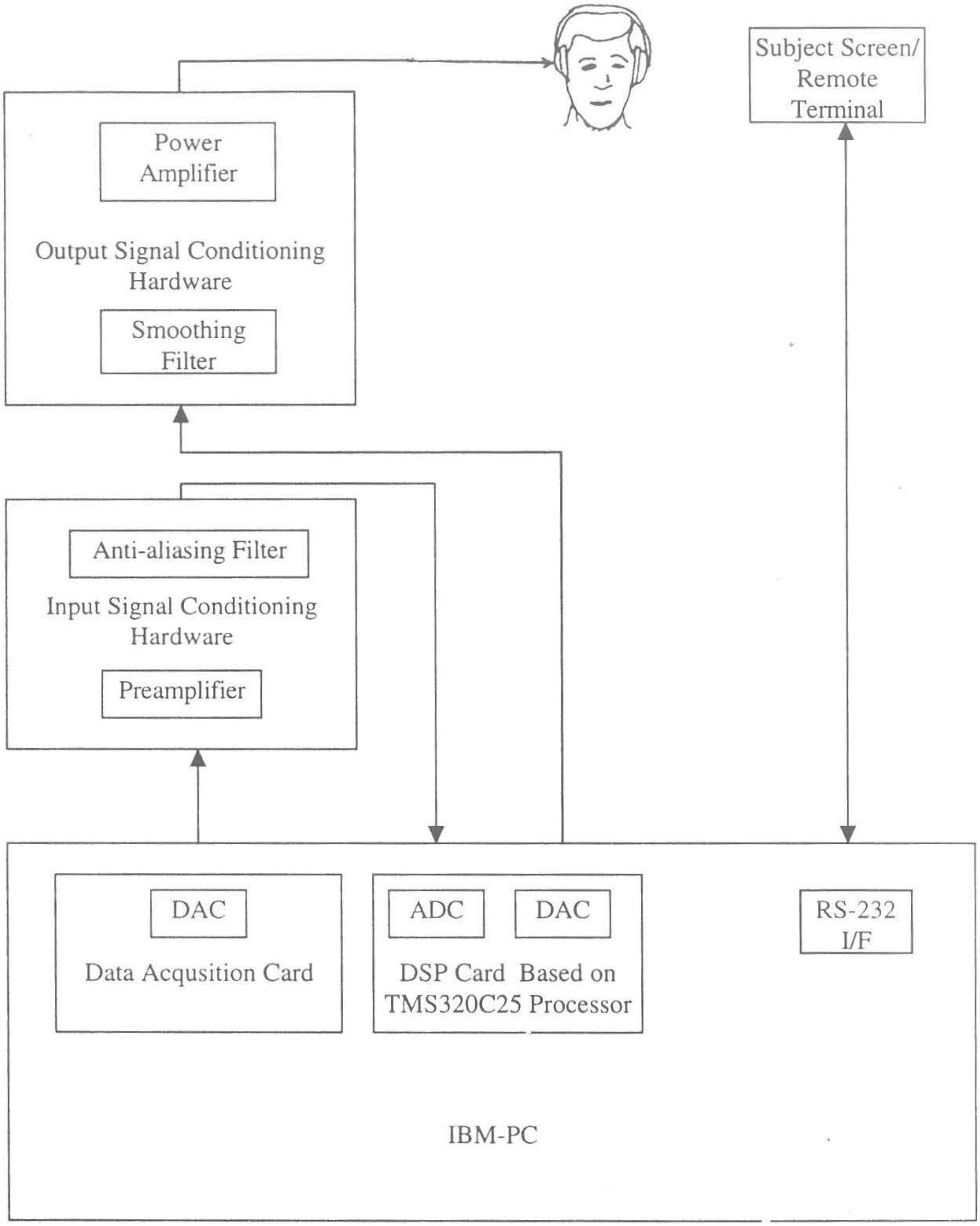



Fig. 5.1 Setup for listening tests with real-time processing

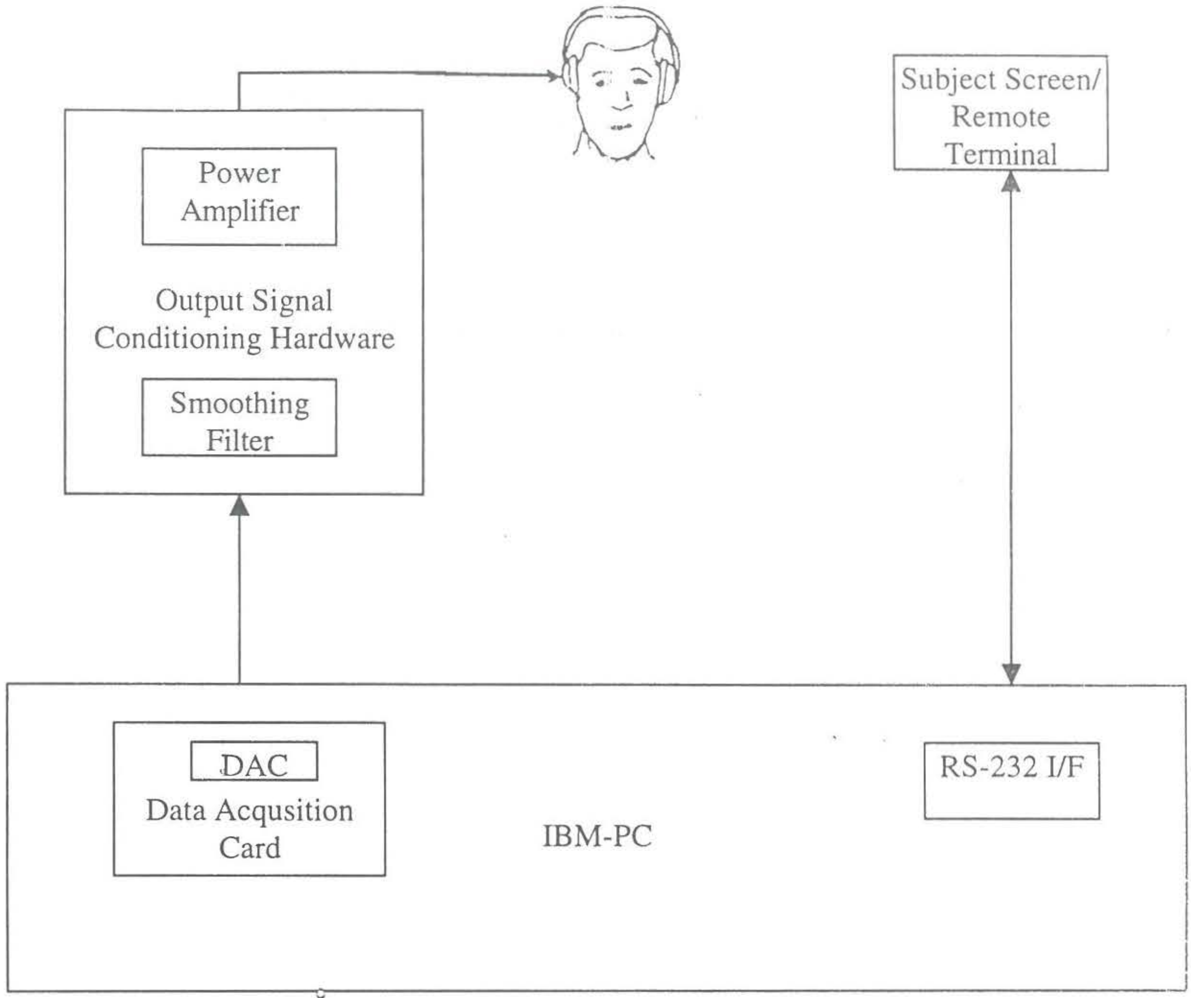



Fig. 5.2 Setup for listening tests with off-line processing

# APPENDIX A

## SPECTROGRAPH

Spectrograph is an instrument that displays the time dependent spectrum of the input speech signal, as frequency versus time. The energy of the frequency components is displayed in various gray levels. The lowest energy level is displayed as white and the darkness of the gray level increases proportional to the energy of the signal. So the highest energy is represented as darkest gray level (black). This two dimensional representation is known as spectrogram.

In one of the methods of generating the spectrogram [20], the spectrogram of a short (2 s) speech utterance is recorded on teledeltos paper using an electromechanical system. In this system, the speech utterance modulates a variable frequency oscillator. This modulated signal is bandpass filtered and the average energy of the output is recorded. The frequency and time resolutions of the spectrogram are dependent on the bandwidth of the bandpass filter (BPF). If is has a wide bandwidth (300 Hz) the spectrogram displays good time resolution and poor frequency resolution. If the BPF has a narrow bandwidth (45 Hz), the spectrogram displays good frequency resolution and poor time resolution. This method of generating the spectrogram takes long time around 10 s and also its dynamic range is low. So this method is not being used for generating the spectrogram. For displaying good resolution in both the time and frequency axes, a combined spectrogram can be generated, as explained later. The spectrogram can also be generated using the computers i.e. by digitizing the speech, by computing its short time fourier transform (STFT), and by displaying the frequency components on console. The STFT can be obtained by computing the discrete fourier transform (DFT) of the speech signal x(n) and is given by

$$x(n,k)=\sum_{m=0}^{N-1} w(m)\cdot x(n-m)\cdot e^{-j\frac{2\Pi m k}{N}}$$

where n represents the discrete time samples, k the discrete frequency, N is the DFT size, and w(n) is the window sequence with length L samples, where L < N. The frequency resolution of the spectrogram is dependent on the type of window, window length L, and the sampling rate $f_s$. For Hamming window, the frequeny resoution $f_{rs}$ is given by

$$f_{rs} = \frac{2\cdot f_s}{1.5\cdot L}$$

where the factor 1.5 is due to the Hamming window [3]. So for generating wideband (300 Hz) spectrogram, for the speech signal which is sampled at 10k Sa/s, the L can be chosen as 43 samples while for narrowband (45 Hz) spectrogram, L is equal to 289 samples. The DFT size N has to be maintained constant for obtaining same number of spectral samples for different values of N.

As per Cohen [21], a simpler method of providing both frequency and time resolutions simultaneously is by the combined spectrogram. The frequency components of the combined spectrogram, $x_{cb}(\ )$ can be obtained as

$$|x_{cb}| = [|x_{wb}| \cdot |x_{nb}|]^{\frac{1}{2}}$$

where $x_{nb}(\ )$ and $x_{wb}(\ )$ are the frequency components of narrowband and wideband spectrograms respectively.

A spectrograph has been developed, using a PC, by Thomas [22] at IIT Bombay. This is a software package which works on PC having an EGA card and an EGA display. This spectrograph displays the spectral componets by computing the DFT of the input speech signal. The DFT size is 512 samples. This package provides an option for generating either a wideband/narrowband or a combined spectrogram. This software has been developed using pascal language. All the tasks, i.e. windowing the input speech signal, obtaining the 512 poing DFT, calculating the log spectrum, and displaying on the video display are being done by the PC. As the process is computationally intensive, it is very slow.

For generating the spectrogram fastly a software package, high speecd spectrograph, has been developed during this M.Tech project alongwith a fellow M.Tech student Ashok Baragi B.N [23]. This spectrograph uses a PC with a VGA card, a video display of 640x480 pixel wide, and a DSP card based on TMS320C25 DSP processor [13]. This DSP card is used to compute the DFT of the input speech signal. This package has two modules, one module will be working on the PC while the other one will be working on the DSP card. The first one is developed in C language while the second one is developed in assembly language of the TMS320C25 DSP processor. The C module displays the input speech file along the time axis. A required segment can be selected using the cursors. So for the selected segment, the spectrogram is generated. The C program requests the dynamic range of the input speech signal, the window length (L) of the Hamming window, and also the scaling factor to scale down the magnitude spectrum (on DSP board) for avoiding the overflow. As mentioned earlier, based on the L (which should be < 256) wideband or narrowband spectrogram can be generated. In the selected speech segment, a block of L samples are Hamming windowed and are padded with N-L zeros, as the DFT size is N (N=256, for this spectrograph) samples. This block is downloaded to the shared data memory, which is on the DSP card using the PC bus interface, and 256 poing FFT (fast fourier transform) is computed on this downloaded data block. After the computation of the FFT, the PC uploads 128 samples of the DFT sequence, as the DFT is symmetric for real valued signals. The log magnitude of this sequence is diplayed on the monitor, as time versus frequency. The high speech has been achievied by using the DSP card for computing the 256 point FFT, and also as the PC displays the log manitude of the previous DFT sequence while the DSP processor is computing the FFT of the present downloaded data block, in parallel. This spectrogram is 500 pixels wide and 128 pixels high and is displayed above the display of the selected time segment of the speech signal. The gray scale of the dynamic range is also displayed on the monitor. After displaying the spectrogram, the magnitude of the various frequency components can be obtained using the cursors. This spectrograph also displays the spectrum of the input speech signal over a window length, at the selected sample. There is a facility to store the spectrogram in a file, and also to acquire the speech signal from the microphone and then to generate its spectrogram. There is also an option for pre-emphasis of the input speech signal. This option is required in the case of the speech signal, as the higher frequency components, otherwise would not be visible beacause of the limited dynamic range of the magnitude scale. The pre-emphasis is obtained by computing the first difference of the input speech signal. This program does not provide an option for generating the combined spectrogram, but it permits the change of resolution by changing the Hamming window length. The operation of this spectrograph is explained in next section.

**Operation :**

The C module is 'spectro.c' while the assembly module is 'fft256.asm'. Upon invoking the spectrograph with the command 'spectro', the C module checks whether the DSP card is present or not. If it is not present, the program terminates with error. The assembled version, of the assembly module 'fft256.mpo' should be present in the current directory for proper operation of the spectrogram. If it is not available, the program terminates with error. The assembly module 'adc3.asm' acquires the input speech data and stores it in a file. So the assembled version 'adc3.mpo' should be present in the current directory, for acquiring the input speech signal. Certain precautions have to be taken while running the spectrogram as the FFT is being calculated on a fixed point DSP processor. The program may otherwise terminate with floating point error or it may give wrong results. The queries asked by the program are explained in this section.

* Acquire data for spectrograph (y/n) :

By pressing the key 'y', the program asks the next queries related to the acqusition of speech data. By pressing the key 'n', the program assumes that the input for the spectrograph is from a file and it asks the related queries.

For the case of the response 'y', the next queries are,

* Sampling rate, greater than 77 sa/sec (in sa/sec) :

The user should respond with the required sampling rate, in the range 77-50000 sa/sec. For the speech signal 10000 sa/sec should be given.

* Output file type (b: bin, t: text) :

The type of the file, in which the acquired speech data will be stored. The format of the stored file is, the first item is the number of samples in the that file and then the speech samples will be stored.

* Output file name :

The user should respond by entering the file name in which the acquired data should be stored.

* No. of samples (multiple of 128) :
  The user should respond by entering the number of samples that are to be acquired. This response may not be a multiple of 128, but the program rounds the the response to multiple of 128.
* Zero intensity level (dB) :
  The user should respond by entering the minimum intensity of the speech signal. The typical value is 20 dB.
* Max. intensity level (dB) :
  The user should respond by entering the maximum intensity of the speech signal. The typical value is 90 dB.

For the case of the response 'n', the next queries are,

* Sampling rate , greater than 77 sa/sec (in sa/sec) :
* File name :
* file type (b: bin, t: text) :
* Zero intensity level (dB) :
* Max. intensity level (dB) :

The above queries are realated to the speech file that is to be processed for generating the spectrogram. So those can be responded by the user accordingly.

After getting the above responses the program plots the acquired speech file or the selected speech data file along the time axis. The user can select a segment by using the cursors. The menu is self explanatory. After selecting the required segement, the program redraws the selected segment and also the gray scale with 16 gray levels. Then the program asks for

* Window length (max : 200, min : 10) :
  The user can respond by entering the required window length of the Hamming window. As explained earlier, this determines the type of spectrogram. For wideband spectrogram of bandwidth 300 Hz, the window length should be 43 (at 10k sa/sec sampling rate) samples. For narrowband spectrogram of bandwidth 84 Hz, the window length should be 159 samples (at the same sampling rate as above).

* Fst_diff (y/n) :
  If the speech signal is to be pre-emphasized, then the user should respond by entering 'y'.

* Mag. scaling factor (2 ** n) :
  This is the scaling down factor of the magnitude of the DFT sequence, for avoiding overflow, on the DSP board. The maximum value is 15. This should be given as per the signal level (count). For a speech file of maximum signal level of count 2000, the typical value of the scale down factor is 13 or 14.

After responding to the above queries, the program resumes for generating the spectrogram. It displays the log magnitude of the 500 DFT sequences and each DFT sequence is displayed as 128 spectral components. After displaying the spectrogram, a menu is displayed below the spectrogram. The various frequency components, at various time instants can be measured by moving the cursors in vertical and horizontal axes. By pressing the return or Enter key the spectrum of the signal, over the window length, will be displayed. The vertical and horizontal cursors can be made off by pressing the 'F1' key. The band analysis can be done by pressing the 'F2' key. The spectrogram can be stored in a file by pressing 'F3' key. The complete screen can be stored by pressing the 'F4' key. By pressing 'Esc' key, the user can quit the spectrogram. Then the program asks the query

* another WB/NB spectrogram on the seletced data segment ? (y/n) :
  The user can generate another spectrogram on the selected data segment by pressing the key 'y'. If the response is 'y', the program resumes with the query of window length. If it is 'n', the program asks

* Resume spectrogram on another data segment ? (y/n) :
  The user can generate the spectrogram of another data segment of the same speech file, by pressing the key 'y'.

**Results :**

Fig. A1 shows the wideband spectrogram of a square wave, whose frequency has a step variation from 500 Hz to 750 Hz and back to 500 Hz. The spectrogram clearly shows the abrupt variation of frequency, as the time resolution is good. The Fig.A2 shows the narrowband spectrogram of the same squarewave. In this case the funadamental frequency and its harmonics can be clearly seen, as the frequency resolution is good. However, the time resoltion has suffered as is seen from the smears at the points of frequency changes.

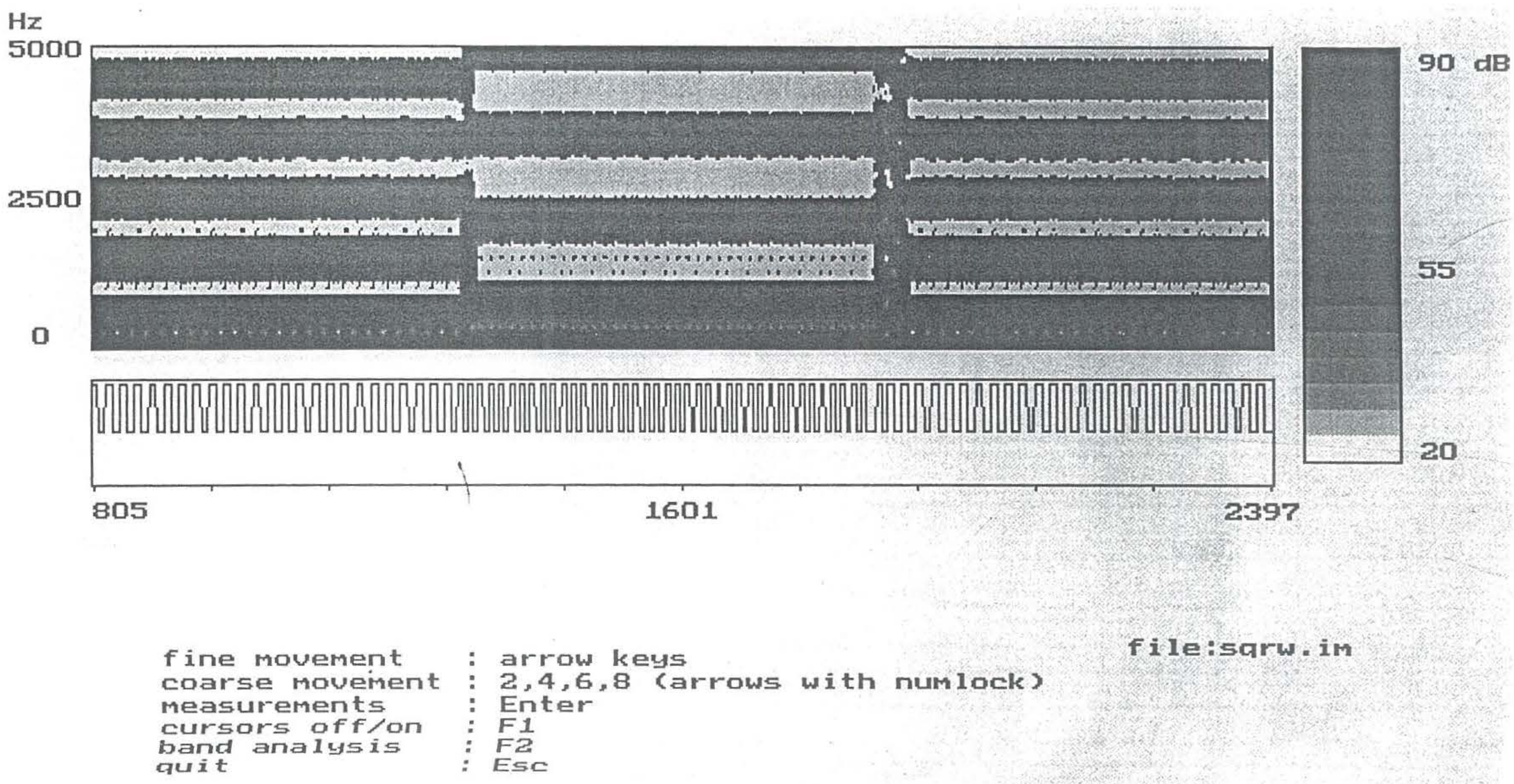



Figure A.1: Wideband spectrogram for a square wave input, with step change in frequeny

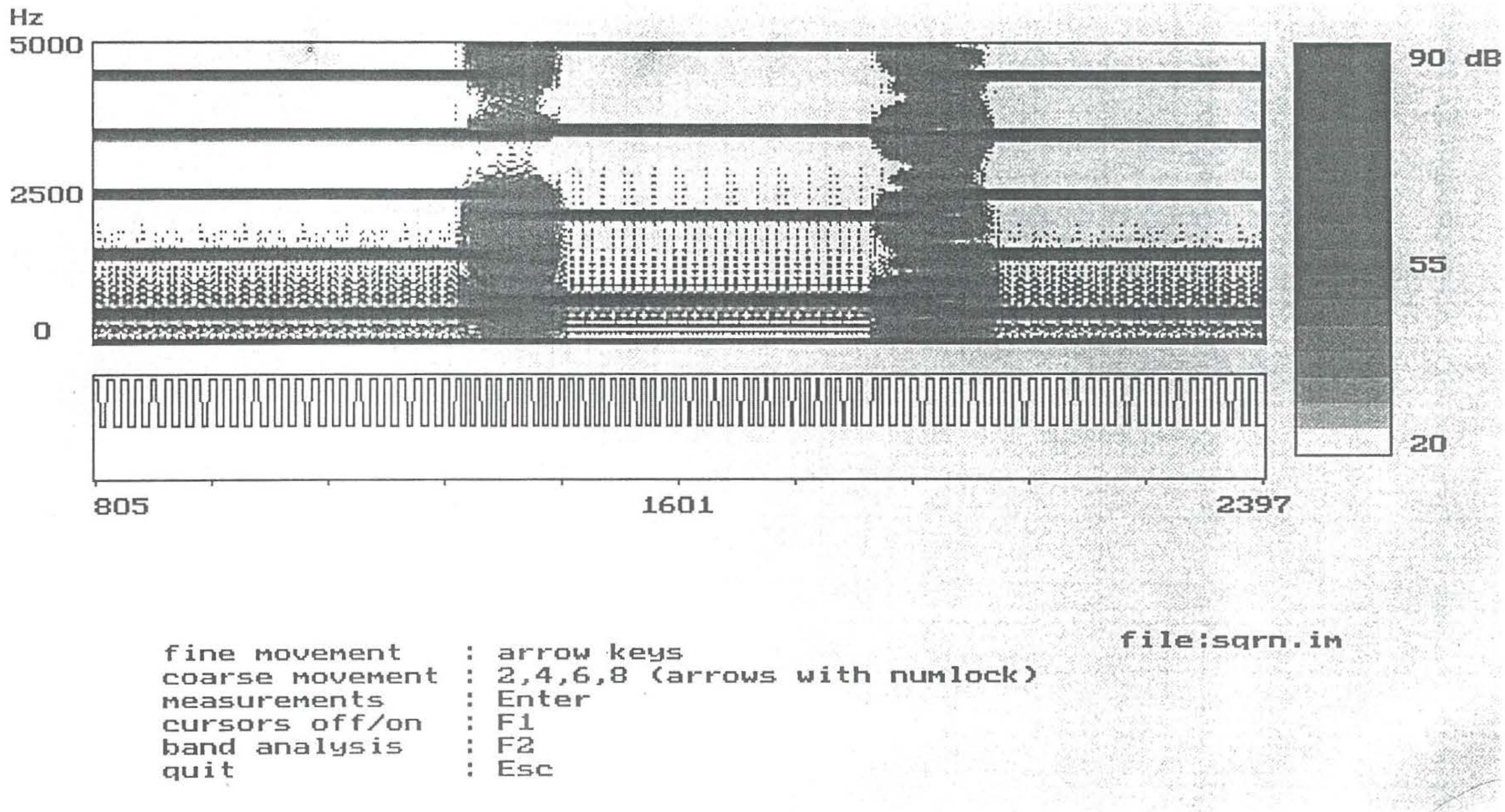



Figure A.2: Narrowband spectrogram for a square wave input, with step change in frequeny

# APPENDIX B

## Table B1. Confusion matrices of the listening tests with the subject PRS

Table B1.1 Test with channel-1 alone

| S/R | apa | aba | ata | ada | aka | aga | ama | ana | asa | aza | afa | ava | + |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| apa | 13 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 15 |
| aba | 0 | 15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 15 |
| ata | 0 | 0 | 13 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 15 |
| ada | 0 | 0 | 0 | 15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 15 |
| aka | 0 | 0 | 1 | 0 | 14 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 15 |
| aga | 0 | 0 | 0 | 0 | 2 | 13 | 0 | 0 | 0 | 0 | 0 | 0 | 15 |
| ama | 0 | 0 | 0 | 0 | 0 | 0 | 10 | 5 | 0 | 0 | 0 | 0 | 15 |
| ana | 0 | 0 | 0 | 0 | 0 | 0 | 6 | 9 | 0 | 0 | 0 | 0 | 15 |
| asa | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 15 | 0 | 0 | 0 | 15 |
| aza | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 7 | 0 | 5 | 15 |
| afa | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 | 0 | 15 |
| ava | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 11 | 15 |
| + | 16 | 18 | 14 | 15 | 18 | 13 | 16 | 14 | 15 | 11 | 14 | 16 | 180 |

Test score : 82%

Table B1.2 Test with complete scheme

| S/R | apa | aba | ata | ada | aka | aga | ama | ana | asa | aza | afa | ava | + |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| apa | 15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 15 |
| aba | 0 | 15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 15 |
| ata | 0 | 0 | 13 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 15 |
| ada | 0 | 0 | 0 | 14 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 15 |
| aka | 0 | 0 | 0 | 0 | 13 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 15 |
| aga | 0 | 0 | 0 | 0 | 0 | 15 | 0 | 0 | 0 | 0 | 0 | 0 | 15 |
| ama | 0 | 0 | 0 | 0 | 0 | 0 | 8 | 7 | 0 | 0 | 0 | 0 | 15 |
| ana | 0 | 0 | 0 | 0 | 0 | 0 | 7 | 7 | 0 | 0 | 0 | 1 | 15 |
| asa | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 15 | 0 | 0 | 0 | 15 |
| aza | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 14 | 1 | 0 | 15 |
| afa | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 15 | 0 | 15 |
| ava | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 14 | 15 |
| + | 15 | 16 | 13 | 14 | 16 | 17 | 15 | 14 | 15 | 14 | 16 | 15 | 180 |

Test score : 88%

**Table B2. Confusion matrices of the listening tests with the subject PRV**

Table B2.1 Test with channel-1 alone

| S/R | apa | aba | ata | ada | aka | aga | ama | ana | asa | aza | afa | ava | + |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| apa | 15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 15 |
| aba | 0 | 14 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 15 |
| ata | 0 | 0 | 5 | 0 | 9 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 15 |
| ada | 0 | 3 | 0 | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 15 |
| aka | 0 | 0 | 4 | 0 | 6 | 2 | 0 | 0 | 1 | 0 | 2 | 0 | 15 |
| aga | 0 | 0 | 1 | 0 | 0 | 11 | 0 | 0 | 0 | 1 | 0 | 0 | 13 |
| ama | 0 | 0 | 0 | 0 | 0 | 0 | 10 | 5 | 0 | 0 | 0 | 0 | 15 |
| ana | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 13 | 0 | 0 | 0 | 0 | 15 |
| asa | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 15 | 0 | 0 | 0 | 15 |
| aza | 2 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 7 | 0 | 5 | 13 |
| afa | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 11 | 0 | 14 |
| ava | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 11 | 15 |
| + | 17 | 17 | 10 | 14 | 16 | 13 | 12 | 18 | 17 | 12 | 13 | 16 | 175 |

Test score : 74%

Table B2.2 Test with complete scheme

| S/R | apa | aba | ata | ada | aka | aga | ama | ana | asa | aza | afa | ava | + |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| apa | 15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 15 |
| aba | 0 | 15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 15 |
| ata | 0 | 0 | 9 | 0 | 4 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 15 |
| ada | 0 | 0 | 0 | 15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 15 |
| aka | 0 | 0 | 1 | 0 | 13 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 15 |
| aga | 0 | 0 | 1 | 0 | 0 | 12 | 0 | 0 | 0 | 0 | 0 | 1 | 14 |
| ama | 0 | 0 | 0 | 0 | 0 | 0 | 13 | 2 | 0 | 0 | 0 | 0 | 15 |
| ana | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 13 | 0 | 0 | 0 | 0 | 15 |
| asa | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 15 | 0 | 0 | 0 | 15 |
| aza | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 8 | 0 | 15 |
| afa | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 10 | 0 | 15 |
| ava | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 13 | 15 |
| + | 17 | 16 | 12 | 15 | 17 | 13 | 15 | 15 | 16 | 11 | 18 | 14 | 179 |

Test score : 83%

**Table B3. Confusion matrices of the listening tests with the subject ASH**

Table B3.1 Test with channel-1 alone

| S/R | apa | aba | ata | ada | aka | aga | ama | ana | asa | aza | afa | ava | + |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| apa | 15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 15 |
| aba | 0 | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 2 | 15 |
| ata | 0 | 0 | 11 | 0 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 15 |
| ada | 0 | 0 | 0 | 14 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 15 |
| aka | 0 | 0 | 3 | 0 | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 15 |
| aga | 0 | 0 | 0 | 0 | 0 | 15 | 0 | 0 | 0 | 0 | 0 | 0 | 15 |
| ama | 0 | 0 | 0 | 0 | 0 | 0 | 12 | 2 | 0 | 0 | 0 | 1 | 15 |
| ana | 0 | 0 | 0 | 0 | 0 | 0 | 10 | 5 | 0 | 0 | 0 | 0 | 15 |
| asa | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 15 | 0 | 0 | 0 | 15 |
| aza | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 11 | 0 | 2 | 15 |
| afa | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 12 | 0 | 15 |
| ava | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 7 | 0 | 6 | 15 |
| + | 16 | 14 | 14 | 14 | 16 | 16 | 23 | 8 | 15 | 21 | 12 | 11 | 180 |

Test score : 78%

Table B3.2 Test with complete scheme

| S/R | apa | aba | ata | ada | aka | aga | ama | ana | asa | aza | afa | ava | + |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| apa | 14 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 15 |
| aba | 0 | 13 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 15 |
| ata | 0 | 0 | 10 | 0 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 15 |
| ada | 0 | 0 | 0 | 14 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 15 |
| aka | 0 | 0 | 4 | 0 | 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 15 |
| aga | 0 | 0 | 0 | 0 | 0 | 15 | 0 | 0 | 0 | 0 | 0 | 0 | 15 |
| ama | 0 | 0 | 0 | 0 | 0 | 0 | 14 | 1 | 0 | 0 | 0 | 0 | 15 |
| ana | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 13 | 0 | 0 | 0 | 0 | 15 |
| asa | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 15 | 0 | 0 | 0 | 15 |
| aza | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 | 1 | 0 | 15 |
| afa | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 10 | 1 | 15 |
| ava | 0 | 5 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 8 | 15 |
| + | 20 | 18 | 14 | 14 | 16 | 16 | 17 | 15 | 15 | 13 | 11 | 11 | 180 |

Test score : 83%

## Table B4. Confusion matrices of the listening tests with the subject PTL

Table B4.1 Test with channel-1 alone

| S/R | apa | aba | ata | ada | aka | aga | ama | ana | asa | aza | afa | ava | + |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| apa | 13 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 15 |
| aba | 0 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 8 | 0 | 0 | 15 |
| ata | 0 | 0 | 12 | 0 | 1 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 15 |
| ada | 0 | 0 | 0 | 15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 15 |
| aka | 0 | 0 | 1 | 0 | 12 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 15 |
| aga | 0 | 0 | 0 | 0 | 0 | 14 | 0 | 0 | 0 | 0 | 1 | 0 | 15 |
| ama | 0 | 0 | 0 | 0 | 0 | 0 | 11 | 4 | 0 | 0 | 0 | 0 | 15 |
| ana | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 11 | 0 | 0 | 0 | 0 | 15 |
| asa | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 13 | 0 | 0 | 0 | 15 |
| aza | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 9 | 1 | 4 | 15 |
| afa | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 13 | 0 | 15 |
| ava | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 1 | 0 | 1 | 0 | 10 | 15 |
| + | 14 | 8 | 14 | 15 | 14 | 18 | 15 | 16 | 16 | 19 | 17 | 14 | 180 |

Test score : 78%

Table B4.2 Test with complete scheme

| S/R | apa | aba | ata | ada | aka | aga | ama | ana | asa | aza | afa | ava | + |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| apa | 14 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 15 |
| aba | 0 | 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 3 | 15 |
| ata | 0 | 0 | 13 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 15 |
| ada | 0 | 0 | 0 | 15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 15 |
| aka | 0 | 0 | 3 | 0 | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 15 |
| aga | 0 | 0 | 0 | 0 | 0 | 15 | 0 | 0 | 0 | 0 | 0 | 0 | 15 |
| ama | 0 | 0 | 0 | 0 | 0 | 0 | 7 | 8 | 0 | 0 | 0 | 0 | 15 |
| ana | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 14 | 0 | 0 | 0 | 0 | 15 |
| asa | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 15 | 0 | 0 | 0 | 15 |
| aza | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 | 3 | 0 | 15 |
| afa | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 14 | 0 | 15 |
| ava | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 14 | 15 |
| + | 14 | 11 | 17 | 15 | 14 | 16 | 8 | 22 | 15 | 13 | 18 | 17 | 180 |

Test score : 87%

**Table B5. Confusion matrices of the listening tests with the subject SAT**

Table B5.1 Test with channel-1 alone

| S/R | apa | aba | ata | ada | aka | aga | ama | ana | asa | aza | afa | ava | + |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| apa | 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 4 | 1 | 15 |
| aba | 1 | 6 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 4 | 1 | 2 | 15 |
| ata | 0 | 0 | 4 | 4 | 1 | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 15 |
| ada | 0 | 0 | 1 | 8 | 0 | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 15 |
| aka | 0 | 0 | 3 | 0 | 9 | 2 | 0 | 0 | 1 | 0 | 0 | 0 | 15 |
| aga | 0 | 0 | 1 | 3 | 6 | 4 | 0 | 0 | 1 | 0 | 0 | 0 | 15 |
| ama | 0 | 0 | 0 | 0 | 0 | 0 | 9 | 5 | 0 | 1 | 0 | 0 | 15 |
| ana | 0 | 3 | 0 | 0 | 0 | 0 | 7 | 5 | 0 | 0 | 0 | 0 | 15 |
| asa | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 14 | 0 | 0 | 0 | 15 |
| aza | 1 | 3 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 7 | 2 | 1 | 15 |
| afa | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 11 | 0 | 15 |
| ava | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 12 | 15 |
| + | 15 | 14 | 9 | 15 | 17 | 18 | 16 | 12 | 16 | 14 | 18 | 16 | 180 |

Test score : 54%

Table B5.2 Test with complete scheme

| S/R | apa | aba | ata | ada | aka | aga | ama | ana | asa | aza | afa | ava | + |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| apa | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 1 | 2 | 15 |
| aba | 1 | 7 | 0 | 0 | 0 | 1 | 2 | 1 | 0 | 0 | 2 | 1 | 15 |
| ata | 0 | 3 | 4 | 0 | 3 | 1 | 0 | 1 | 3 | 0 | 0 | 0 | 15 |
| ada | 0 | 0 | 0 | 14 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 15 |
| aka | 0 | 0 | 5 | 0 | 6 | 3 | 0 | 0 | 0 | 0 | 0 | 1 | 15 |
| aga | 0 | 0 | 0 | 2 | 0 | 12 | 0 | 0 | 0 | 0 | 0 | 1 | 15 |
| ama | 0 | 0 | 0 | 0 | 0 | 0 | 9 | 5 | 0 | 0 | 1 | 0 | 15 |
| ana | 0 | 2 | 0 | 0 | 0 | 0 | 4 | 9 | 0 | 0 | 0 | 0 | 15 |
| asa | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 13 | 0 | 0 | 0 | 15 |
| aza | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 9 | 2 | 0 | 15 |
| afa | 5 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 3 | 4 | 15 |
| ava | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 2 | 6 | 4 | 15 |
| + | 18 | 13 | 9 | 17 | 11 | 18 | 15 | 18 | 16 | 17 | 15 | 13 | 180 |

Test score : 54%

## APPENDIX C

## DEVELOPMENT OF SUPPORTING SOFTWARE

### Introduction :

On the course of testing the SCSA (Single Channel Sensory Aid) there was a need to develop a few programs for executing the test procedures. The various programs developed are spectrograph software, software for acquiring a signal and storing in a file, software for echoing a signal file through DAC port of DSP board, software for scaling a signal file, and software for obtaining various signal to noise ratio (SNR) files. The spectrograph software has been explained in Appendix A and the other programs are explained, in brief, in this Appendix.

### Signal Acquisition Software :

This software was developed to acquire a signal using the PC add-on DSP card based on TMS320C25 DSP processor [13] and storing it in a file. This contains two modules, `C ' and DSP assembly modules. The assembly module samples the input signal at 10k sa/sec. The`C ' module stores the sampled data in a file in text or binary mode. This software was used for acquiring sine, square, triangular waveforms and also noise for testing the SCSA-2 in various signal conditions. The C module is `adc4.c' and the assembly module is `adc4.asm'.

### Signal File Echo Software :

This software was developed to echo a signal file to DAC port of the add-on DSP card [13], so that the signal can be observed on oscilloscope. This software echos the signal file at 10k sa/sec. There are two modules `echo_fil.c' and `echo_fil.asm'. Signal file can be either in text or binary format.

### Signal File Scaling Software :

This software is written for scaling a signal file to the required maximum absolute value. Program name is `scale.c'. This program finds the absolute maximum value of the signal file and finds a scaling factor by using the user given required maximum absolute maximum value. Using this scaling factor it scales the signal file and stores the scaled version in a file `tmp.bin' in binary mode.

**Software for Obtaining Various S/N Ratio Files :**

This is a `C ' module, `snr.c'. This program was used for obtaining signal files with various SNR (signal to noise ratio), while testing the pitch estimator of SCSA-2. This program asks for the required SNR, signal file name, noise file name and the range of signal samples to be corrupted. This program then calculates the average signal power and average noise power in the selected range of samples and hence finds a scaling factor for the selected SNR. Using this scaling factor the noise is scaled and then added to signal samples. These corrupted samples are stored in a file.

## REFERENCES

[1] Working group on communication aids for the hearing impaired, "Speech perception aids for hearing impaired people, current status and needed research," *J. Acoust. Soc. Am.*, vol. 90, pp. 637-685, 1991.

[2] P. C. Pandey, H. Kunov, and S. M. Abel, "A speech processor providing fricative and low-frequency periodicity information for single channel cochlear prosthesis," *Int. Conf. on Acoust. Spch. and Sig. Proc.*, (ICASSP '87, Dallas, Texas, USA), April, 1987.

[3] P. C. Pandey, *Speech Processing for Cochlear Prosthesis*, Ph. D. thesis, Electrical Engineering Department, University of Toronto, 1987.

[4] R. M. Sapre, *A Speech Processor for Single Channel Auditory Prosthesis*, M. Tech. dissertation, Electrical Engineering Department, IIT Bombay, 1992.

[5] N. Shah, *A Sensory Aid for the Deaf,* M.Tech dissertation, Electrical Engineering Department, IIT Bombay, 1995.

[6] B. C. J. Moore, *An Introduction to the Psychology of Hearing*, New York : Acadamic Press, 1982.

[7] Y. C. Tong, G. M. Clark, P. J. Blamey, P. A. Busby, and R. C. Dowell, "Psychological studies for two multiple-channel cochlear implant patients," *J. Acoust. Soc. Am.*, vol. 71, pp. 153-160, 1982.

[8] H. Levitt, J. M. Pickets, and R. A. Houde, *Sensory Aids for the Hearing Impaired*, New York, IEEE Press, 1980.

[9] A. Risberg, "A critical review of work on speech analyzing hearing aids," *IEEE Trans. Audio Electroacoust.* AU-17,pp. 290-297, 1969, reprinted in Levitt et al. [8]

[10] B. Johannson, "The use of transposer for the management of the deaf child," *Int. Audiol*, vol. 5, pp. 362-372, Sep. 1966, reprinted in Levitt et al. [8].

[11] D. Ling, "Speech discrimination by profoundably deaf childrenusing linear and coding amplifiers," *IEEE Trans. Audio Electroacoust.*, AU-17, pp. 298-303, 1969, reprinted in Levitt et al. [8]

[12] N. Guttman and J. R. Nelson, "An instrument that creats some artificial speech spectra for the severely hard of hearing," *Amer. Ann. Deaf.*, vol. 113, pp. 295-302, 1968.[4]

[13] *PCL-DSP25 Digital Signal Processor Card User's Manual*, Dynalog MicroSystems Pvt. Ltd., 1993.

[14] V. V. S. R. Prasad, *Program Listings for the M. Tech. Dissertation.*, " Speech Processing for Single Channel Sensory Aid," Electrical Engineering Department, IIT Bombay, 1996.

[15] E. C. Ifeachor, and B. W. Jervis, *Digital Signal Processing*, Great Britan : Addison -Wesley Publisers Ltd., 1993.

[16] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech signals*, Englewood cliffs, NJ : Prentice-hall 1978.

[17] *PCL-208, Data acquisition Card User's Manual*, Dynalog MicroSystems Pvt. Ltd., 1992.

[18] T. G. Thomas, *Experimental Evaluation of Improvement in Speech Perception with Consonantal Intensity and Duration Modification*, Ph. D. Thesis, Electrical Engineering Department, IIT Bombay, 1995.

[19] G. A. Miller, and P. E. Nicely, " An analysis of perceptual confusions among some English consonants," *J. Acoust. Soc. Am.*, vol. 27, pp. 338- 352, 1955.

[20] R. Koenig, H. K. Dunn, and L. Y. Lacey, " The sound spectrograph," *J. Acoust. Soc. Am.*, Vol. 18, pp. 19-49, 1946.

[21] L. Cohen, " Time-frequency distributions - a review," *Proc. IEEE*, Vol. 77, pp. 941-981, 1989.

[22] T. G. Thomas, P. C. Pandey, and S. D. Agashe, " A PC-Based spectrograph for speech and biomedical signals," *Proc. Intl. Conf. Recent Advances in Biomedical Engineering*, (Hyderabad), Jan. 1994, pp. 6-8.

[23] B. N. Ashok Baragi, *A Speech Training Aid for the Deaf*, M. Tech. dissertation, Electrical Engineering Department, IIT Bombay, 1996.