Th-33

A GLOTTAL PITCH EXTRACTOR

A dissertation

submitted in partial fulfillment of the requirements

for the degree of

Master of Technology

by

Abhijit V. Patil
(98307002)

Guide

Dr. P. C. Pandey

Department of Electrical Engineering

Indian Institute of Technology, Bombay

January 2000

Indian Institute of Technology, Bombay

Dissertation Approval

The thesis entitled "A Glottal Pitch Extractor", by Shri. Abhijit V. Patil (Roll No. 98307002) is approved for award of the degree of **Master of Technology** in Electrical Engineering with specialization in Control & Instrumentation.

Supervisor	______________________	(Prof. P. C. Pandey)
External examiner	______________________	(Dr. G. D. Jindal)
Internal examiner	Mukherji	(Prof. S. Mukherji)
Chairperson	______________________	(Prof. S. N. Malhotra)

Date : 9th February 2000

Abhijit V. Patil: *A Glottal Pitch Extractor*, M. Tech. dissertation, Department of Electrical Engineering, IIT Bombay, Jan. 2000, Guide : Prof. P. C. Pandey.

ABSTRACT

Electroglottography (EGG) is a non-invasive technique for measuring impedance variation across the thyroid cartilage of the larynx. This impedance variation, sensed by a pair of electrodes, provides information about the dynamics of the closure of vocal folds. Signal related to the impedance variation, known as L_x waveform, is useful for estimation of voice pitch, diagnosis of voice disorders, and as a speech training aid for the hearing impaired.

An instrument is designed in which a high frequency (400 kHz), low intensity (~1 mA) current is passed through the central discs (15 mm dia.) of a pair of electrodes held in contact with the skin on both the sides of the thyroid cartilages, and the L_x waveform is extracted. A glottal impedance simulator has been designed for testing the sensitivity and frequency response of the impedance sensor. The impedance sensor is interfaced to the PC based sound/multimedia card, for acquisition, analysis, and display of the L_x waveform. The analysis and display package provides the pitch measurement, pitch histogram, and spectrographic analysis.

ACKNOWLEDGMENT

I would like to express my sincere gratitude to my guide, Dr P.C. Pandey for his invaluable guidance.

I thank Shri L. D. Muntode for his invaluable help during my project work. I would also like to thank all of my friends who encouraged and gave me lot of support for the completion of this project.

IIT Bombay　　　　Abhijit V. Patil
Jan. 2000

CONTENTS

SYMBOLS AND ABBREVIATIONS

Symbols	Explanation
F_x	pitch
L_x	laryngogram waveform
ΔF	bin width
$p(F_x)$	probability density function
T_{total}	time over which analysis of the Lx waveform is carried out
$x(n)$	discretized input samples
$x_{max}(n)$	maximum amplitude of the acquired waveform
$x_{min}(n)$	minimum amplitude of the acquired waveform
$x_{mean}(n)$	mean of maximum and minimum amplitude
f_s	sampling frequency
N	number of samples
$w(m)$	Hamming window

Abbreviations	Explanation
EGG	electroglottograph
GIS	glottal impedance sensor
PC	personal computer
LCD	liquid crystal display
PCMCIA	Personal Computer Memory Card Interface Association
ADC	analog-to-digital converter

LIST OF FIGURES

Chapter 1
INTRODUCTION

1.1 Overview

The main organs related with the generation of sound are lungs, vocal tract, and larynx. The lungs are the source of airflow. The vocal tract is an acoustic enclosure and acts as acoustic filter shaping the spectrum of the generated sound [1]. The source of most speech occurs in the larynx. Inside the larynx, there are two folds of the muscular bundle known as vocal chords. These vocal chords obstruct the airflow from the lungs and produce audible vibrations that make the speech. The mechanism of generation of sound is known as phonation. The vibration consists of three phases namely as contact phase, separation phase, and open phase [2].

Diseases of the larynx, such as acute laryngitis, vocal nodules, and cancer directly affect the mechanism of phonation, and result in changes in patient's speech. If these changes are properly detected and measured, they are useful for diagnosis purposes [3] [4]. Electroglottography is a technique, which can be used to assess the vibration of vocal chords during phonation, from which various diseases can be detected. Electroglottograph is an instrument designed to detect the vibration of vocal chords as a time varying signal, the amplitude variations of this signal represent the amount of contact between the vocal chords [5].

The impedance between the vocal chords is a function of tissue path length. When the vocal chords are open, the tissue path length is maximum and hence the impedance is maximum. When the vocal chords are closed, the tissue path length is minimum and hence the impedance is minimum. Electroglottograph measures this impedance variation. It is used to assess the vibration of vocal chords which is useful for clinical and pathological use [3] [6]. The fundamental frequency of vocal fold vibration is known as pitch and the plot of instantaneous pitch as a function of time is known as F_x plot.

In electroglottography, a pair of electrodes are held in contact with the skin on the neck. High frequency constant current is passed through the electrodes. During phonation, as vocal chords vibrate, the voltage across electrodes changes and we get amplitude modulated waveform. This waveform is then demodulated using demodulator. Thus the instrument gives the impedance variation waveform known as electroglottogram (EGG) signal [6].

1.2 Objectives

A number of electroglottograph instruments have been developed. One such instrument named "Laryngograph" developed in 1974 by the group of Dr. A. J. Fourcin at University College, London has been extensively used for studying pitch variation and laryngeal function assessment in therapy and clinical situations [7].

This project is a continuation of development work done at IIT Bombay 1990 onwards, as part of student projects [9] [10] [11]. The earlier work by Chitnis in 1998 involved development of a low cost battery operated glottal pitch extractor to sense impedance variations across the larynx, and a microcontroller based signal acquisition and LCD graphics display unit which can be interfaced with a PC through RS232 serial interface for downloading the measurement results on to a computer or a plotter [11] [12].

In this project, the aim is:

- to increase the sensitivity of the impedance sensor part of the instrument and reduce the level of the excitation current, to reduce the power requirement for the instrument, and to develop an instrument.
- to develop a glottal impedance simulator for testing the sensitivity and response characteristic of impedance sensor.
- to study various electrode configurations and driving/sensing circuit for reducing the distortion of the waveform, and decreasing the noise pick-up.
- to interface glottal impedance sensor with the PC-based sound/multimedia card for acquisition and analysis of the waveform, in place of special signal acquisition hardware.
- to develop the program for analysis of the EGG waveform.

1.3 Outline of the dissertation

Chapter 2 describes the speech production mechanism, physiology of larynx and the phenomenon of vocal fold vibrations. This chapter also explains the detection of impedance variation between the vocal folds, the development work carried out for glottal pitch extractor in brief and the modifications needed for glottal impedance sensor. Chapter 3 explains the design of glottal impedance sensor, and need for a simulator and its design. Chapter 4 explains the signal acquisition, analysis, and display of glottographic waveform by interfacing impedance variation sensor with sound/multimedia card. Chapter 5 explains the testing of the instruments (glottal impedance sensor and simulator) and test results on different subjects. Chapter 6 provides summary and conclusion along with suggestions for future work.

Chapter 2
ELECTROGLOTTOGRAPHY

2.1 Introduction

This chapter explains the anatomy and physiology of various organs related to speech production, the impedance variation between the two vocal folds during phonation, application of electroglottography, EGG instrumentation and review of work done earlier for glottal impedance sensor, and modifications needed.

Lungs, larynx, and vocal tract are the important organs, which are responsible for the generation of sound. Fig 2.1 shows the various speech organs [13]. The two lungs situated in the chest cavity on either side of the heart are organs of respiration. Lungs are expanded and compressed by movement of diaphragm during breathing. During inspiration, the diaphragm pulls the lower surfaces of the lungs downwards. Then during expiration, the diaphragm simply relaxes and the elastic recoil of the lungs, chest wall, and abdominal structures compresses the lungs. They communicate with the atmosphere through the trachea, which opens into the pharynx. The pharynx is a musculo-membranous tube. It passes air from nose or mouth to the larynx. It also acts as a channel to transport food from mouth to esophagus. Thus the lungs acts as a source of airflow that passes through the larynx and the vocal tract [2].

2.2 Anatomy and physiology of larynx

Larynx acts as an air passage, which carries air from pharynx to the lungs. It is situated in the front of the neck, above trachea. During swallowing, it moves upward so that food passes into the esophagus and not into the lower respiratory passages.

Larynx is made up of seven irregularly shaped cartilages attached to each other by ligaments and membranes. The main cartilages are thyroid cartilage, cricoid cartilage, two arytenoid cartilages and an epiglottis. The structure of the larynx as viewed from behind and front is given in Fig 2.2.

The thyroid cartilage is the most prominent and consists of the two flat pieces of the cartilages fused anteriorly (in the front), forming the laryngeal prominence (Adam's apple). The inlet to the larynx lies in the anterior wall of the pharynx. Continuing from the inlet, the cavity of the larynx expands into a wider vestibule. Two sets of thick, membranous ridges protrude from the laryngeal wall at the lower end of the vestibule. The upper pair, called vestibular folds or false vocal folds, are not involved in the production of sound. The lower pair, the vocal folds or true vocal chords, contains the vocal ligaments. The vocal folds together with the intervening space form the sound producing apparatus known as glottis. Each vocal chord is stretched between the thyroid cartilage anteriorly and an arytenoid cartilage posteriorly [2].

2.3 Vocal fold vibration

The vocal chords are set into vibrations during the voiced segments of speech. The vibration cycle consists of three main phases: contact phase, the separation phase, and the open phase [2]. Vibration of the vocal folds is based on the Bernoulii's principle. The vocal folds vibrate when they are sufficiently elastic and close together, and there is a sufficient difference between the pressure below the glottis and the pressure above the glottis. As the air is exhaled from the lungs, the vocal chords move apart. As the velocity increases in the narrow glottis, the local pressure drops. When sufficient pressure difference exists across the glottis to cause a large airflow, a negative pressure develops in the glottis that forces the vocal folds to close. Glottal closure interrupts the airflow, and a pressure gradient develops across the glottis, eventually building to a point where the vocal folds open again. Thus the vocal folds are set into vibrations.

Variations in the pressure below the glottis and vocal fold elasticity due to the muscles of the chest and larynx, respectively, cause changes in the waveshape of glottal air pulses, including their duration and amplitude, which lead to changes in pitch and loudness. Fig 2.3 Shows the position of the vocal chords in a horizontal sections when they are opened and closed [11] [13].

The fundamental frequency of vocal fold vibrations is known as pitch. It depends upon the length and the tightness of the chords. In adults, the length of the vocal chords is typically 15 mm long in male and about 13 mm long in the female [14]. In speech, the pitch may typically vary over about an octave, while singers may be able to use a two-octave range. Average pitch values are 132 Hz and 223 Hz for males and females, respectively. Children have generally higher pitch [14]. When the force of expired air is more than the normal, the vocal chords vibrate more and the louder is the sound. The main factors, which decide the quality and resonance of the voice are, the shape of the mouth, the position of the tongue and the lips, the facial muscles, and the air sinuses in the bones of the face and the skull [11] [13].

Fig 2.4 shows the representation of the sequence in a vocal fold vibration cycle in a vertical sections [15]. The two folds come into contact due to a wavelike motion starting at the lower end as shown in Fig 2.4 (e) and (f). Fig 2.4 (a) shows the condition of maximum contact between the two folds. Fig 2.4 (b) and (c) represents the separation phase, which is also gradual from the lower surface upwards. Fig 2.4 (d) shows the open phase where the two folds are completely apart [15].

2.4 Significance of pitch

Voice pitch variation pattern conveys information such as the speaker's identity and emotional state. The patterns of pitch variations that occur during a phrase, known as intonation, mark the boundaries of syntactic units and sometimes affect their meaning. It has been determined that pitch and amplitude patterns vary with emotions. The emotions often raise pitch and amplitude levels, and their variability [7].

In English, statements have falling pitch, while sentences meant as questions have a rising pitch. In many other languages, pitch variations during vowel segments known as "tone", also contribute to the meaning of the word [7].

2.5 Impedance variation between the vocal folds

Phonation consists of three phases; open phase, separation phase and the contact phase. The impedance between the two vocal chords is the function of the tissue path length. When the vocal folds are open, the tissue path length is maximum and hence the impedance will be maximum. When the vocal folds are in contact phase, the tissue path length is minimum and hence the impedance will be minimum. Hence when the vocal chords vibrate, the impedance between the two chords changes. If we sense this impedance variation, we get information about vibration of vocal chords [16].

Electroglottography is a technique, which senses the impedance variation between the two vocal folds as a time varying signal.

Electroglottograph is an instrument, which senses the impedance variation between the two vocal folds. The impedance variation gives information about the contact phase, open phase and the separation phase; which is very important for laryngeal function assessment. Fig 2.5 shows a model of the impedance variation waveform. This figure labels critical points on the model waveform, which is a straight-line approximation for the waveform. At all the key points, the positions of the vocal chords are shown starting from the closed phase to the open phase [16].

2.6 Application of electroglottography

2.6.1 Laryngeal function assessment

Electroglottography (EGG) is useful for the assessment of laryngeal function, as it is possible to detect the physical condition of the vocal chords using EGG signal. The output from the electroglottograph, proportional to impedance variation, is known as electroglottogram or laryngogram, and is often referred to as L_x waveform [7].

The waveforms for normal voice, breathy voice and creaky voice [7] are shown in Fig 2.6. For normal voice the waveform is shown in (a). When the vocal chords are fully apart, a peak is obtained. For breathy voice, the waveform is shown in (b). Airflow required is larger than that needed for normal voice. So vocal chord contact does not necessarily occur during vibration. The waveform shows small and well-defined closure peaks, which are more symmetrical than for normal voicing. For creaky voice, the waveform is shown in (c). From the analysis of normal voice waveform by Fourcin [7], following conclusions can be drawn,

- Uniform waveform of normal peaks is associated with uniform acoustic output.
- Long closure duration is associated with well-defined formants (resonance of the vocal tract filter).
- Sharply defined contact implies good acoustic excitation of vocal chords.
- Regular periodical contact gives well-defined pitch.

In pathology, an indication of the nature and degree of the disorder can be made by interpreting the L_x waveform with the reference to the above points.

A plot of instantaneous pitch frequency as a function of time is known as F_x waveform. This is considered useful for observing perturbation in the pitch, and also for detecting unvoiced segments [7].

In practical situations, the L_x waveform is acquired for a long period usually 1 to 2 min [7]. The F_x histogram is a plot of probability density [17] [18] of vocal fold vibration frequencies observed over this time. The F_x histogram can be plotted as either a single period or a multiple (often, triple) period histogram. The frequency range of interest is divided into a number of equal bins. For obtaining single period F_x histogram, each bin is incremented when a pitch value falls within its frequency range. The content of each bin divided by the total number of pitch periods is a measure of the probability density, on the basis of count of pitch periods. Let the bin width be ΔF, and probability density function be $p(F_x)$. The probability for the bin starting at frequency F is given as

$$P(F;\, F + \Delta F) = \int_{F}^{F+\Delta F} p(F_x)\,dF_x$$

$$\approx p(F + \Delta F/2).\Delta F \qquad \text{(eqn. 2.6.1)}$$

The probability values are plotted using either log scale (0.1% to 100%) or linear scale (0% to 100%). The probability values can be scaled with respect to the mean pitch period of the bin to get the F_x histogram on the basis of the relative time for which the pitch value lies in a bin. The probability for bin starting at frequency F relative to time is given as

$$P_t\,(F;\, F + \Delta F) = P(F;\, F + \Delta F)\frac{1}{F + \Delta F/2}\;\frac{1}{T_{total}} \qquad \text{(eqn. 2.6.2)}$$

where, T_{total} is the total time over which the analysis is carried out.

In triple period histogram, count in a bin is incremented only if three successive pitch values fall within its frequency range [7]. By incrementing each bin only when, for example, three successive vocal fold periods correspond to the bin range, a measure of vibrational regularity is obtained from both height and width of the distribution [7].

Analysis of F_x histogram was carried out by Fourcin et al. [7]. They estimated the F_x histogram on the basis of relative number of pitch periods plotted on the log scale.

- Healthy larynx characteristically has sharp edges to its frequency distribution and well defined frequencies of vibrations. Fig 2.7 shows the frequency distribution for normal male voice.
- For older people, lower frequencies have a more gradual distribution.
- For a younger person, the low frequency edge distribution is abrupt and higher frequency periods are gradually reduced in probability.
- Smoking increases the irregularity of the vocal fold vibrations.

F_x histogram, therefore is a useful research tool and also has applications in clinical practice [7].

2.6.2 Use of EGG in speech analysis

EGG waveform finds wide application in the area of speech processing, mainly in the following.

- *Voiced–unvoiced classification*: Algorithms for classification of a speech segment into voiced, unvoiced, mixed, and silent have taken on new levels of accuracy and simplicity when EGG signal is used as an aid in the decision making process. For just a voiced-unvoiced decision, using the EGG signal alone is sufficient since the EGG is ideally zero during unvoiced regions and periodic and nonzero during voiced regions [16].
- *Pitch estimation*: Using the EGG, the pitch estimation becomes simpler and accurate, since usually the pitch value is based on either zero crossings or the distance between the minima in the differentiated EGG [16].
- *Synthesis application*: The EGG-aided analysis can play an important role in obtaining parameters for synthesis of high quality speech. This is because the naturalness and intelligibility of synthesized speech are influenced by factors such as accuracy in the vocal tract modeling, voiced-unvoiced classifications, pitch detection, and the nature of excitation used [16].

2.6.3 Improving lip-reading

Lip-reading is one of the alternatives for profoundly deaf person, to understand a conversion. Lip–reading, sometimes becomes very difficult for differentiation of certain words, which have identical lip movements. Some of the phoneme pairs which create the discrimination problem are /g/ and /k/, /b/ and /p/, and /v/ and /f/. Words such as 'vat' and 'fat', which have identical lip movements, are very difficult to distinguish. The only difference between 'vat' and 'fat' is that the vocal folds vibrate throughout the /v/ and not during /f/. It has been shown that when pitch is provided externally, lip-readers are able to comprehend speech much better [8]. It has been shown that if the voice pitch is provided externally, normal listeners are able to lip-read a person reading continuous text, at a rate two and half times faster when they are lip-reading without the provision of pitch [8]. The provision of pitch greatly improves lip-reading ability in those who have acquired an auditory knowledge of speech before becoming deaf [8].

2.7 Development of EGG instrumentation

A lot of research relating to the electroglottography was carried out by the group headed by Dr. A. J. Fourcin, using the instrument developed by them at University College of London. The instrument developed was called 'Laryngograph' [15][19]. The instrument has been currently marketed by Kay-Elemetrics Corp. There have been a number of other such instruments, eg. 'Electroglottograph type EG 830' and 'Portable Electroglottograph type EG 80' from F-J Electronics A/S [7] [11] [19].

Development work for an electroglottography instrument has been carried out at IIT Bombay as part of student projects by Bhagwat (1990), Sriram (1991), Thajudin (1994), Mahajan (1995), Chitnis (1998) [9] [10] [11]. The circuits developed from Bhagwat (1990) to Mahajan (1995) had noise problems and impedance change of less than 1 Ω could not be detected properly. Chitnis in 1998 was successful in developing an glottal impedance sensor, which could detect impedance change of less than 1 Ω. He also developed a stand-alone data acquisition and display unit for acquiring impedance variation waveform and analyzing it.

The impedance variation in the glottis due to vocal fold vibration can be detected using a scheme shown in the Fig. 2.8 [16]. A pair of electrodes is held in contact with the skin on both the sides of the thyroid cartilage. An RF voltage source is applied using a transformer. This high frequency RF signal is given to a pair of electrode. The impedance Z_g between the electrodes and the transformed impedance Z_l of the output circuit (AM detector) form a voltage divider. When the vocal folds vibrates, Z_g varies and the resulting output gets amplitude modulated [16].

$$V_g = V_s \frac{Z_l}{Z_g + Z_l}$$

An AM detector is used for demodulating the amplitude modulated waveform. The output waveform is related to the impedance variation. The transformer in the circuit helps in isolation and in rejection of common mode pick-up.

Glottal pitch extractor developed by Chitnis in 1998 consists of a transformerless glottal impedance sensor, a signal acquisition and display unit, and a program for pitch extraction and display. For sensing the impedance variation, a pair of electrodes is held in contact with the skin on both sides of the thyroid cartilage. A high frequency current (frequency in the range of 300 kHz, current in the range of 3 mA) is passed through these two electrodes. The voltage level between the electrodes is typically of the order of 0.5 V. When the vocal chords vibrate, the impedance between the two vocal chords changes, and voltage across the two electrodes changes in accordance with the impedance between the vocal chords. Thus we get an amplitude modulated waveform. An equivalent circuit of the impedance detector is shown in Fig. 2.9. The output voltage is given as, $V_g = Z_g I_s$. This signal is then amplified and given to demodulator circuit. Demodulator circuit consists of a full wave precision rectifier and low pass filter. The demodulated output is given to an amplifier which further gives an impedance variation waveform known as Electroglottogram (EGG), Laryngogram, or L_x waveform. The base impedance across the thyroid cartilage is approximately 500 Ω and change in impedance is less than 1 Ω [11] [16].

Fig. 2.10 shown as block diagram of the instrument set-up. The Lx waveform is acquired and processed by signal acquisition and display unit. It consists of a dedicated ADC system using a microcontroller for digitizing the waveform obtained from glottal impedance sensor. The digitized signal can be displayed on the graphics display and stored into the data memory. An RS-232 serial interface has been provided for downloading the digitized samples to a computer or a plotter [11]. The unit has a limited amount of data memory and hence cannot be used for storing large amount of data, especially for capturing the impedance variation waveform for a longer duration. Hence, it cannot be directly used for calculating the pitch and the F_x histogram, which requires large amount of data capturing and processing. Thus in order to calculate the pitch and the F_x histogram, a signal acquisition and pitch extraction software is developed on a notebook PC for interfacing DAQ-700 card from National Instruments as a signal acquisition system. The card can be interfaced to the PC through a PCMCIA (Personal Computer Memory Card Interface Association) port.

2.8 Modifications for glottal pitch extractor

The glottal pitch extractor developed by Chitnis in 1998 can be modified, so as to increase the sensitivity of the instrument and reduce the level of excitation current. The excitation current passed through the electrodes is 3 mA which can be further reduced. The various electrode configurations (explained in section 3.3) and driving and sensing circuit have to be studied for reducing the distortion of the waveform, and decreasing the noise pick up. The full wave rectifier in the glottal impedance sensor by Chitnis in 1998 does not show a linear response to the input amplitude. Problem has been traced to the speed limitations of the diodes. The problem has to be corrected. The amplifier stage of the glottal impedance sensor consists of the cascade of second order bandpass filters with 3-dB lower cut-off frequency of 10 Hz and upper cut-off frequency of 10 kHz. The frequency response of the amplifier deteriorates due to the cascade of two bandpass filters. The frequency response of the output stage of the glottal impedance sensor can be improved by replacing the cascade of second order bandpass filters with second order Butterworth high pass (1-dB cut-off frequency of 80 Hz) and low pass (1-dB cut-off frequency of 1 kHz) filter followed by a second order bandpass filter (3-dB cut-off high and low pass cut-off frequency of 10 Hz and 30 kHz respectively). A PCB layout of the glottal impedance sensor has to be designed taking utmost care to avoid the possibility of noise pick-up, etc.

The glottal impedance sensor unit requires ± 8-12V supply, and at present it operates from two 9V batteries. It is desirable to operate the instrument by using single power supply of 9V. For this a high efficiency (COSEL ZUW3) 12V to ±12V, 130 mA DC-DC converter is proposed to be used [20]. For testing and calibration of the EGG instrument, and to study the effect of common-mode pick-up, etc; a glottal impedance simulator needs to be designed and built.

For acquisition and analysis of the L_x waveform, the glottal impedance sensor needs to be interfaced with the PC-based sound/multimedia card in place of special signal acquisition hardware. This will make the use of instrument relatively inexpensive.

The modified design of glottal pitch extractor is given in Chapter 3. The signal acquisition and display is described in Chapter 4.

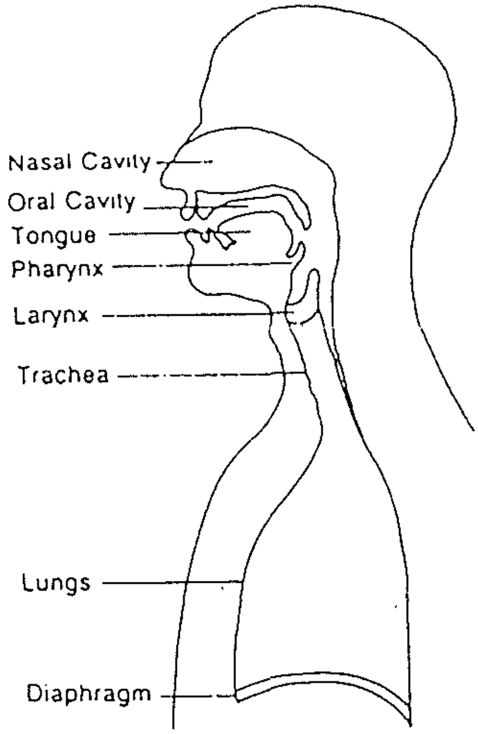


Fig. 2.1 The speech organs. Source : [13]

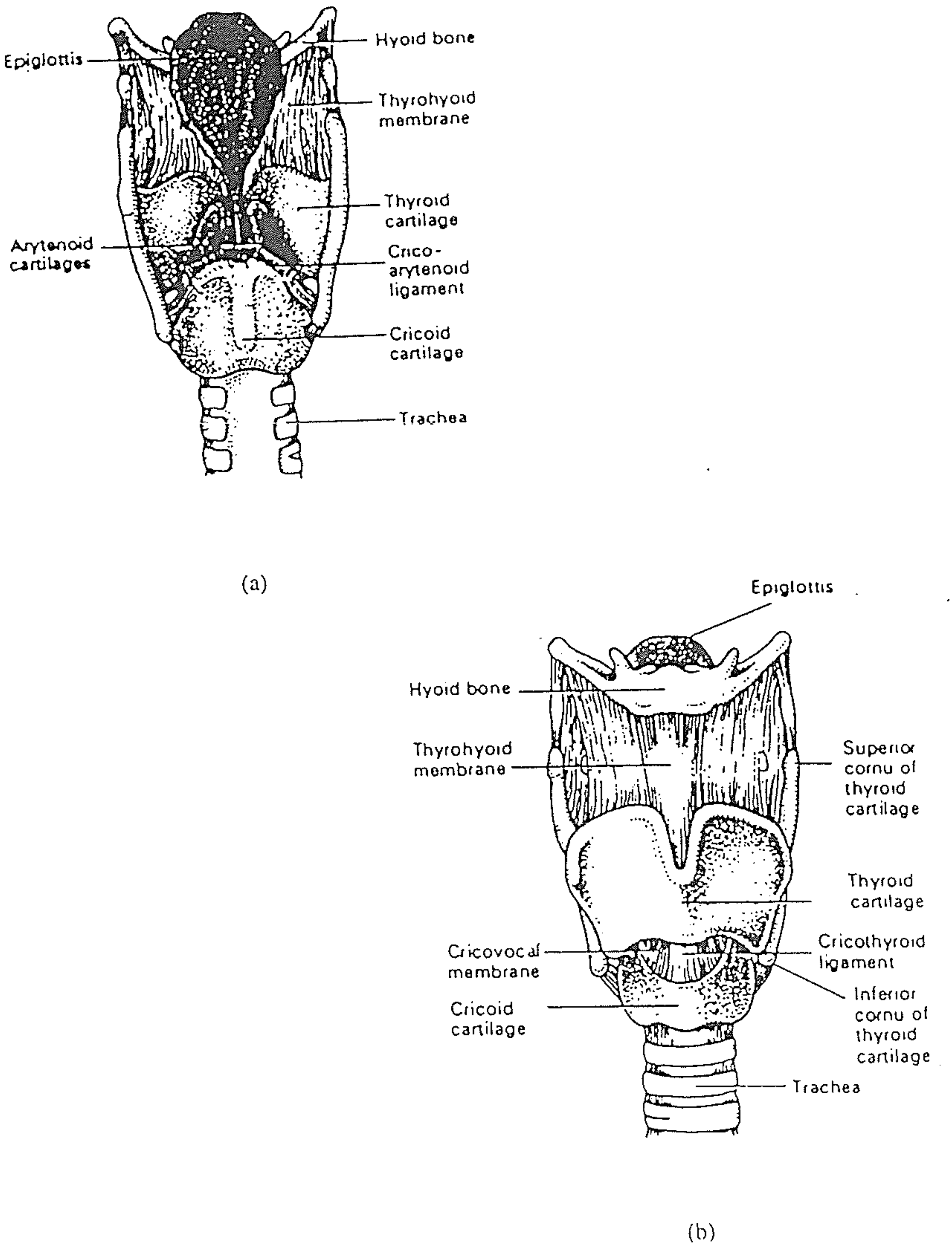


Fig. 2.2 Structure of larynx (a) viewed from behind, (b) viewed form front. Source : [11]

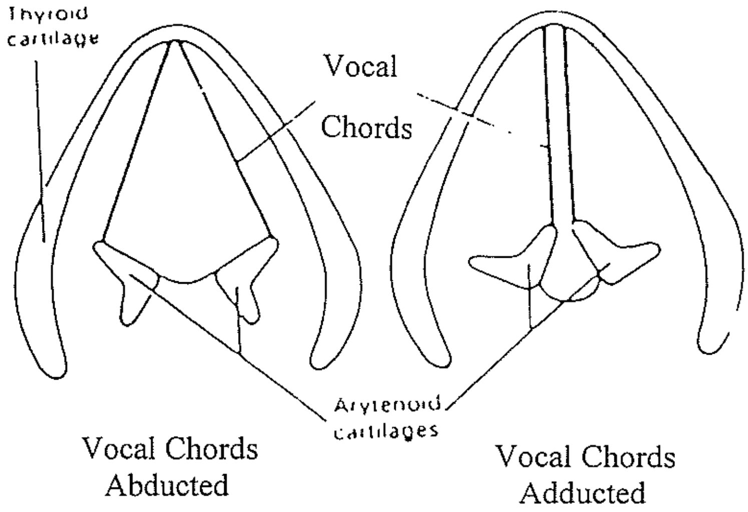


Fig. 2.3 Positions of the vocal chords in a horizontal section. Source : [14]

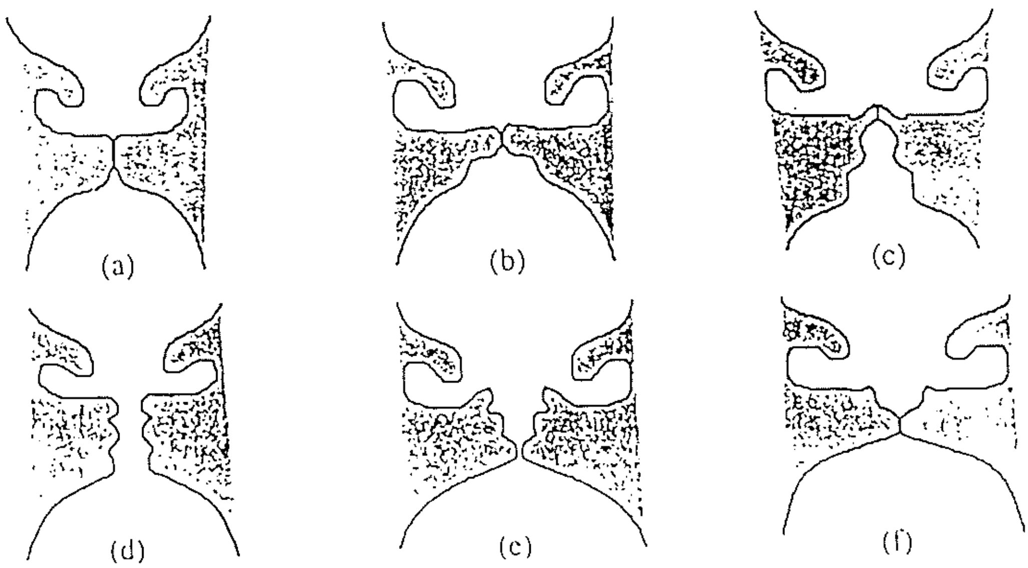
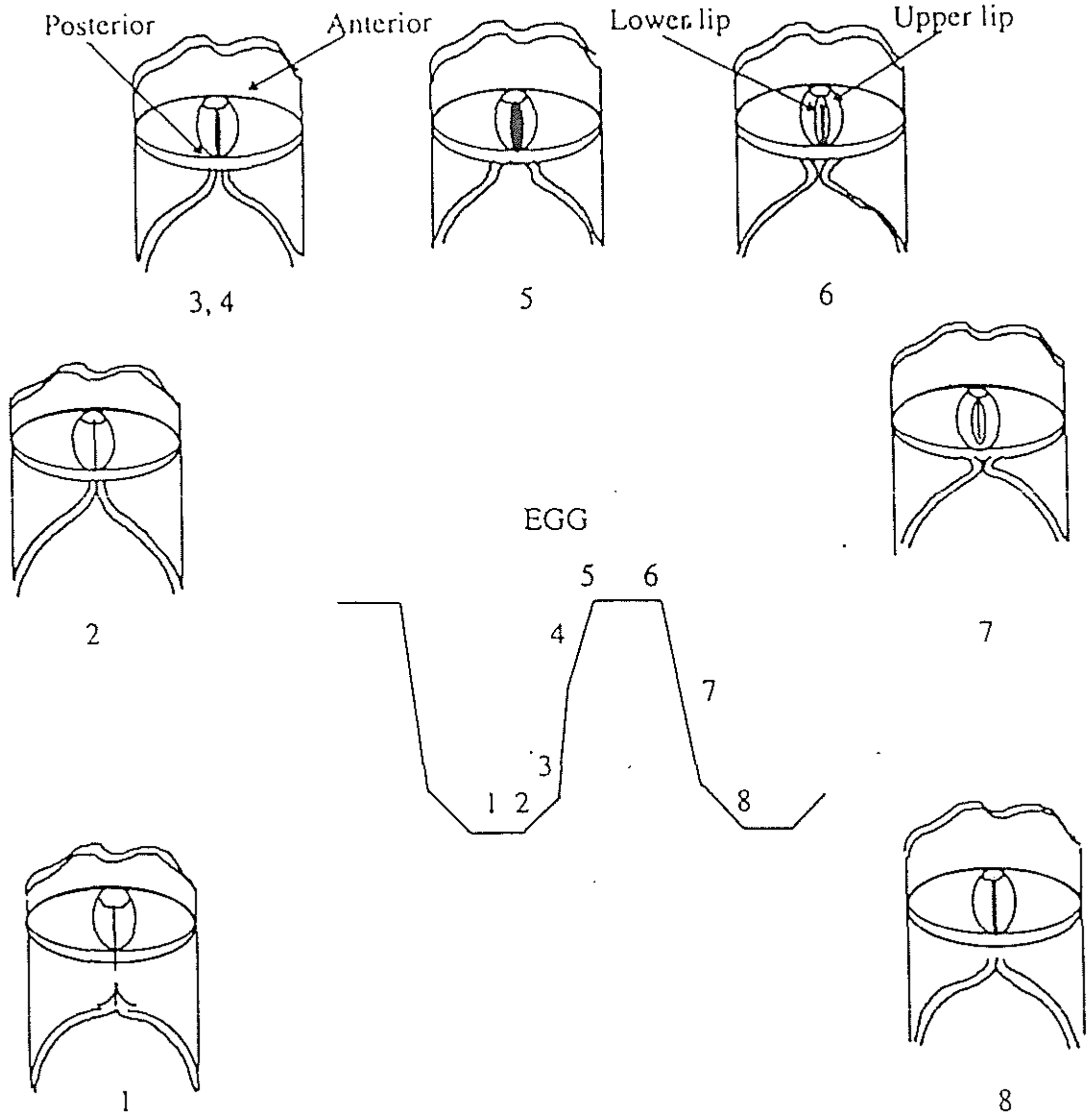


Fig. 2.4 The vocal fold vibration sequence in a vertical section. Source : [15]



1 Vocal folds maximally closed.
2 Folds separating along lower margin toward upper margin.
3 Folds separating along upper margin.
4 Folds continuing to open.
5 Folds apart, no lateral contact.
6 Folds starting to close, starts at the lower margin from posterior to anterior but still no contact.
7 Folds making first lateral contact along lower margin at anterior.
8 Folds completing closure with increasing lateral contact.

Fig. 2.5 Model of EGG waveforms with corresponding drawings of vocal fold configurations. Source: [16]

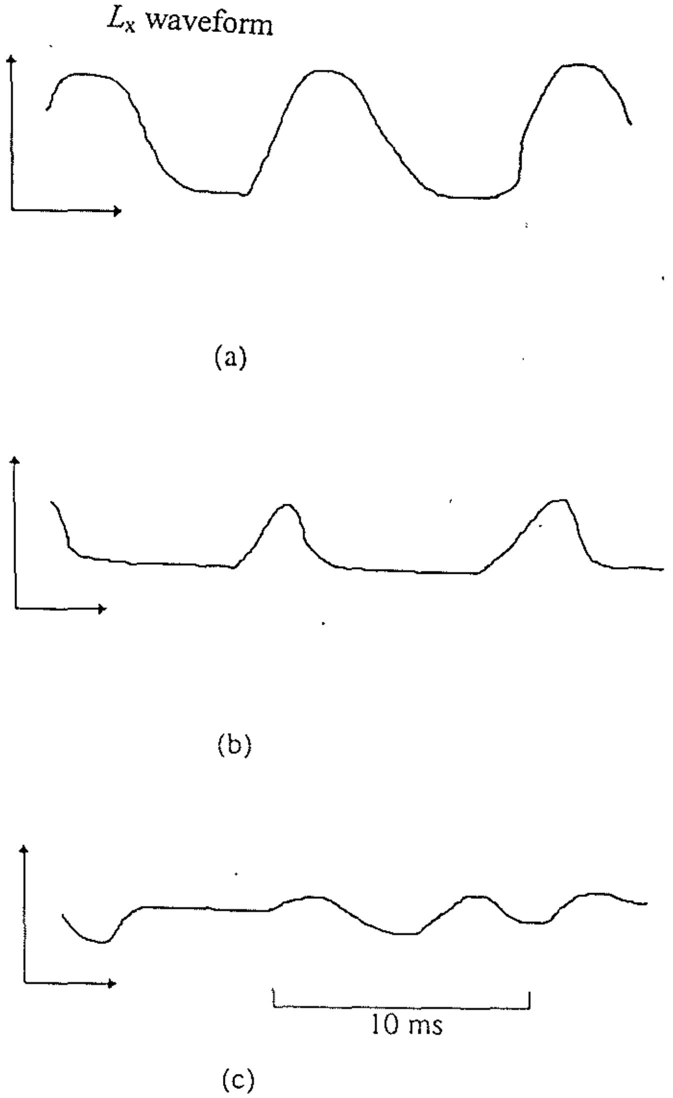


Fig. 2.6 Waveform for characteristic voicing. (a) L_x waveform for normal voice, (b) L_x waveform for breathy voice, (c) L_x waveform for creaky voice. Y-axis indicates the impedance variation. Source : [7]

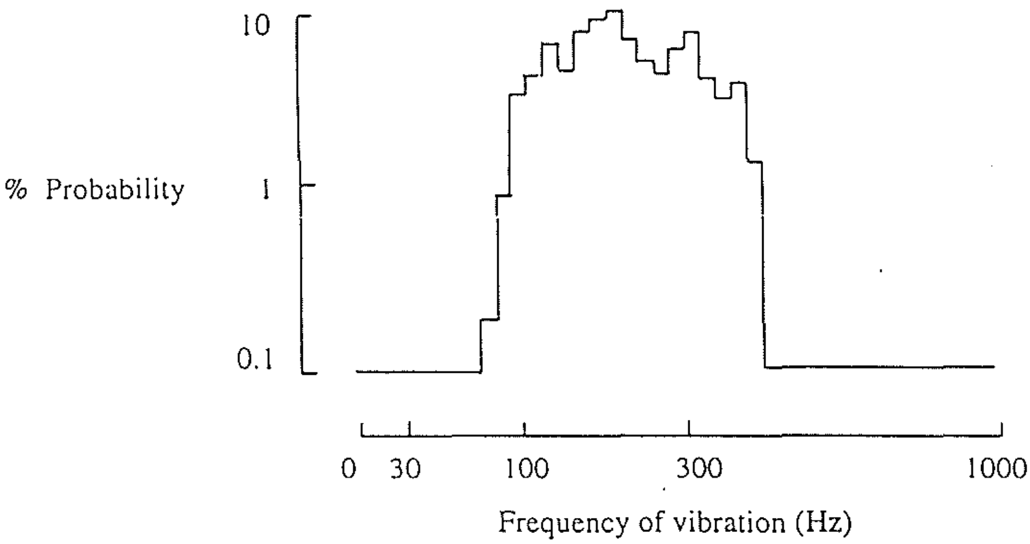


Fig. 2.7 F_x histogram for normal male voice. Source : [7]

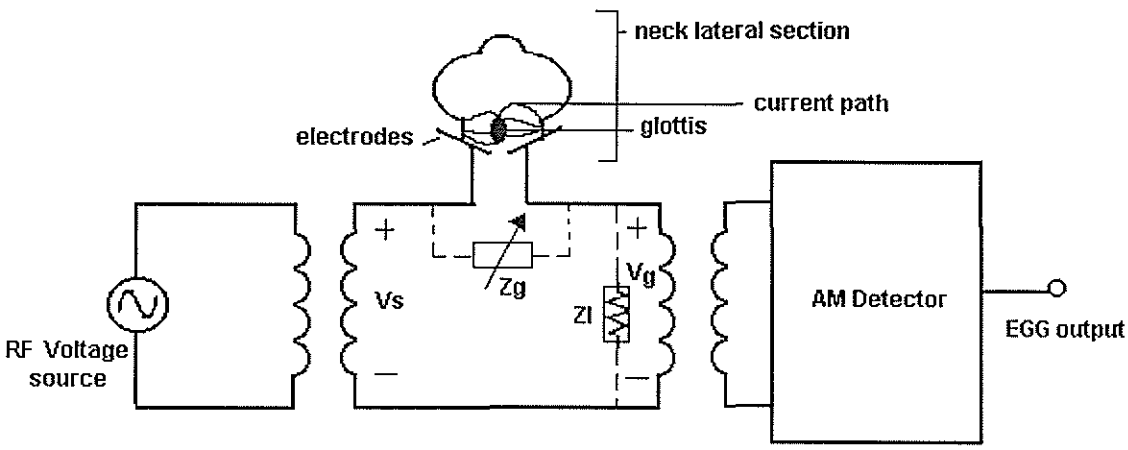


Fig. 2.8 Block diagram of an EGG instrument using RF source. Adapted : [1]

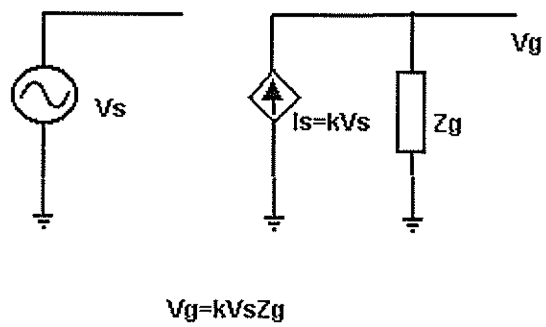


Fig. 2.9 Equivalent circuit of the impedance detector used by Chitnis [11].

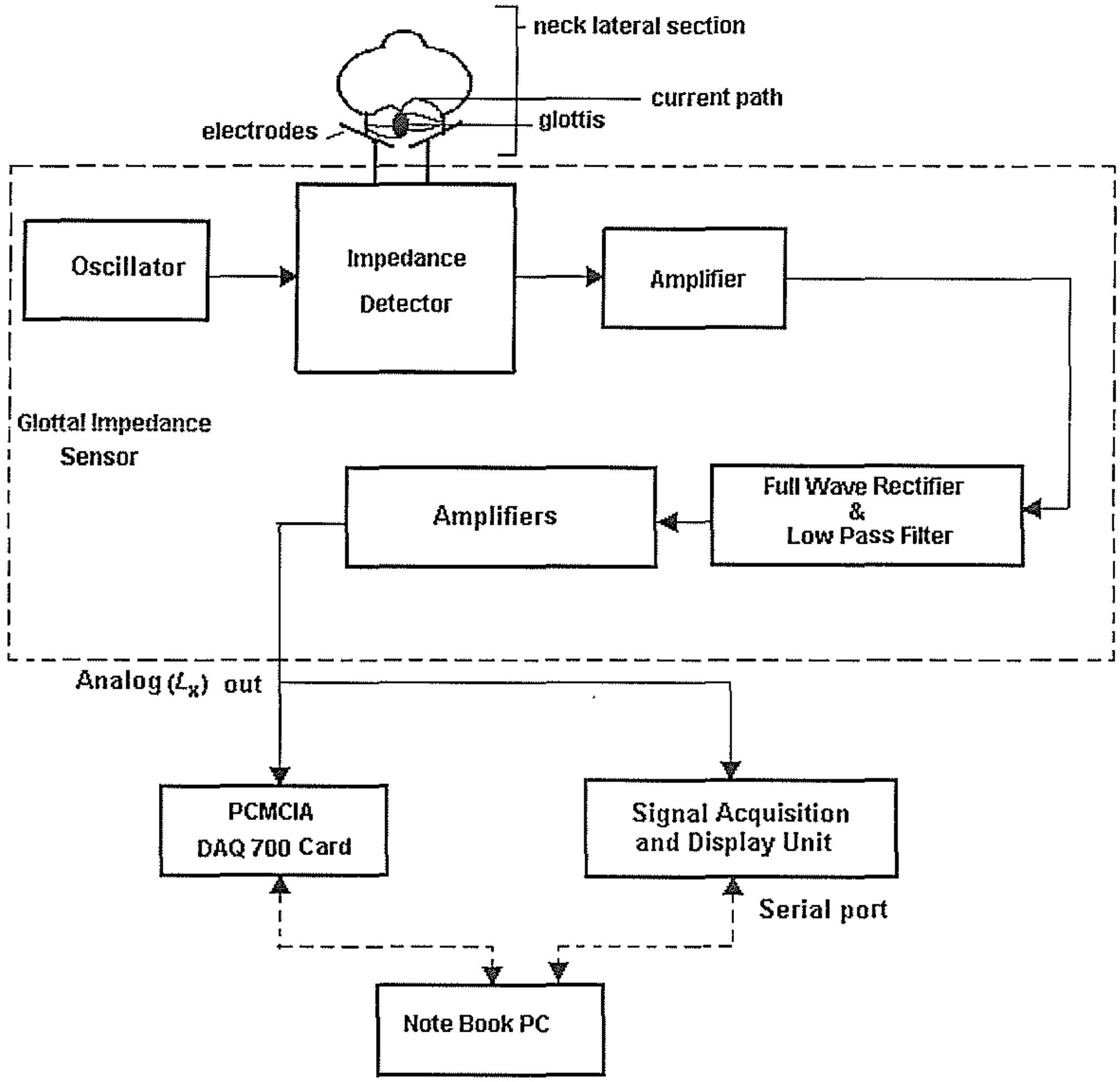


Fig. 2.10 Block diagram of glottal pitch extractor developed by Chitnis in 1998. Adapted : [11]

Chapter 3
INSTRUMENT HARDWARE

3.1 Introduction

The development of the instrument hardware involved the design and assembly of the circuits of glottal impedance sensor, glottal impedance simulator to facilitate testing and calibration, and use of the appropriate electrode configurations. This chapter provides a description of these in the following sections.

3.2 Glottal impedance sensor

The electroglottograph instrument is basically the circuit for sensing the variation in the electrical impedance of the glottis, caused by the opening/closing of the vocal chords and it consists of the following blocks:

- Oscillator
- Impedance detector
- Demodulator and amplifiers

The block diagram of glottal impedance sensor is shown in Fig.3.1. A pair of electrodes is held in contact with the skin on both sides of the thyroid cartilage. A high frequency current (frequency 400 kHz and current 0 to 1 mA) is passed through these two electrodes. The two electrodes are connected across the feedback path of a V/I converter circuits. Hence constant current flows through the electrodes. When the vocal chords vibrate, the impedance between the two vocal chords changes. As current is constant, voltage across the two electrodes changes in accordance with the impedance between the vocal chords and we get an amplitude modulated waveform. This signal is then amplified and given to a demodulator circuit. Demodulator circuit consists of a precision full-wave rectifier and low pass filter that suppresses the high frequency carrier. The demodulated output is given to an amplifier which consists of a cascade of second order Butterworth high pass and low pass filter, and bandpass filter which further gives an impedance variation waveform known as electroglottogram or laryngogram, or L_x waveform. The circuits of these blocks are explained in the following subsections.

3.2.1 Oscillator

A Wien-bridge oscillator circuit shown in Fig 3.2 is used for generating 400 kHz, 5 Vp-p sinusoidal signal. It is built around op-amp U1. The amplitude of the oscillator is stabilized using the circuit around JFET TR1 and zener diode D2. If the negative half cycle of the output voltage exceeds the zener breakdown, excess voltage is filtered in RC network which provides V_{GS} bias for JFET TR1. Thus channel resistance of JFET increases, reducing the effective gain of the amplifier circuit and hence its output amplitude. Thus amplitude of the oscillator circuit is stabilized. The frequency of oscillation is given as

$$f_o = 1/(2\pi RC)$$

where $R = R_{29} = R_{30}$, $C = C_2 = C_3$. With the values shown in Fig. 3.2,

$$f_o = 408 \text{ kHz.}$$

This generated waveform, which acts as a carrier signal is given to V-I converter to detect glottal impedance variation through a voltage follower circuit. The output Vb of voltage follower U5 can be adjusted from 0 to 2.5 Vp-p.

3.2.2.1 Impedance detector

The impedance detector circuit is shown in Fig 3.3a. The oscillator output Va is given to the attenuator buffer formed of resistors R24 and P2, and op-amp U5. The buffer output Vb can be adjusted in 0-2.5 Vp-p range to set the current level to be injected. The current source for current injection is the V/I converter formed using op-amp U2 and resistor R8 and R9. The injection current I_e is given as

$$I_e = V_b/R_9$$

With R9 = 2.2 kΩ, the current can be set in the range of 0-1 mA peak-to-peak. The electrodes placed across the thyroid cartilage are put across terminals E1-E2, i.e. in the feedback path of the op-amp U2. Capacitors C5 and C7 are used to prevent the passage of any dc current through the electrodes. These capacitors should have negligible leakage, and their ac impedance at excitation frequency should be negligible as compared to 500

Ω. Resistor R8 has been put in the feedback path, to limit the dc gain of the circuit and to restrict the dc errors. Its resistance should be large compared to the impedance between the electrodes.

One of the electrodes, E1, is at virtual ground. Therefore, the voltage across the electrodes is given by the output voltage

$$V_c = I_e Z_g$$

where Z_g is the glottal impedance. Due to variation in Z_g, Vc is an amplitude modulated voltage.

3.2.2.2 Contact impedance indicator

The contact impedance indicator circuit is used to verify the contact at the skin-electrode interface, and operation status of V/I converter. The circuit diagram is shown in Fig.3.3b. Two indicator circuits are used for this purpose. Each circuit compares the amplitude level of one half cycle of the sinusoidal waveform, with the rectified (corresponding polarity) average of input voltage Vb. When electrodes are not properly connected, the level in both positive and negative half cycles will become high and both LEDs are turned ON. When both LEDs are turned OFF, V/I converter is operating in active region. V/I converter may go into saturation due to increased input bias current or other dc errors. If either of the LED is ON, it indicates that corresponding half cycle is in saturation.

3.2.3 Demodulator

The output Vc from V/I converter circuit (Fig. 3.3) is connected to demodulator circuit, as shown in Fig. 3.4. Voltage Vc is filtered by the high pass filter formed using op-amp U12, and C12, R26, C13 and R25. The filter is designed for a cut-off frequency of 50 kHz. The high pass filter cutoff frequency is given as

$$f_h = 1/(2\pi R_{26} C_{12})$$

and with the values shown in Fig. 3.5, f_h = 48.2 kHz

The gain of this amplifier can be adjusted by varying the resistance P3 from 1 to 10 kΩ.

$$A = 1 + R_{p3}/R_{25}$$

The role of C13 is to limit the gain of the circuit for internal dc errors. The output voltage of this amplifier varies from 0 to 4 Vp-p.

The demodulator circuit is a rectifier detector consisting of full-wave precision rectifier followed by a first order low pass filter. The rectified signal is low pass filtered to remove the 400 kHz carrier signal (A synchronous detector may have resulted in better signal-to-noise ratio. However in this case the modulation index is very low (<<0.01) and we do not expect interference around the carrier frequency).

The circuit consists of a precision full-wave rectifier using op-amp U9 and U11. The op-amps, LF 356 and diodes OA 85 have been tested for proper rectification at 400 kHz. Condition for balanced amplification of the two cycles of the signal is,

$R_{35} = R_{33}, R_{37} = 2R_{P4},$

The low pass filter cutoff frequency is given as

$f_l = 1/(2\pi R_{39}C_{15})$

and the component values have been selected for f_l = 13 kHz. The cut-off frequency should be selected to eliminate the carrier, but not to introduce any significant error in the L_x waveform.

3.2.4 Amplifier

The amplifier consists of three stages as shown in Fig. 3.5. The output Vd of the demodulator is passed through cascade of second order Butterworth high pass and low pass filter designed by using op-amps U10 and U3 respectively, and a bandpass filter designed by using op-amp U13. The 3-dB cut-off frequencies of second order Butterworth high pass and low pass filter are 50 Hz and 2.5 kHz respectively. The bandpass filter has a high pass cut-off of 10 Hz and low pass cut-off of 22 kHz. The cut-off frequencies have been selected to eliminate low frequency drift and residual of the carrier frequency, without introducing much distortion in the glottal waveform.

$$A_1(s) = \frac{A_{01}s^2}{s^2 + s(\omega_{01}/Q_1) + \omega_{01}^2} \tag{3.1}$$

where, A_{01} is the high-frequency gain, ω_{01} is the *pole frequency,* and Q_1 is the selectivity of the second order response [22].

For Butterworth response, selectivity $Q = 0.707$. The high pass filter stage in Fig. 3.5 is Sallen-Key high pass filter with inverting amplifier configuration [21], and its transfer function $A_1(s)$is given as

$$A_1(s) = \frac{V_{d1}}{V_d} = -\frac{R_{42}s^2}{s^2(R_{42}+2R_{41}) + s\frac{R_{41}}{C}(\frac{1}{R_{45}} + \frac{3}{C(R_{41} \| R_{44})}) + \frac{R_{41}}{C^2 R_{45}(R_{41} \| R_{44})}} \quad (3.2)$$

where, $C = C_{16} = C_{17} = C_{18}$. Compairing Eqns 3.1 and 3.2, we get

$$A_{01} = -\frac{R_{42}}{R_{42}+2R_{41}} = -0.916$$

$$Q_1 = \sqrt{(R_{41}+R_{44})R_{45}R_{44}}\,(\frac{R_{42}+2R_{41}}{R_{41}R_{44}+3(R_{41}+R_{44})}) = 0.7068$$

$$f_h = \omega_{02}/2\pi = \frac{1}{2\pi}[\frac{R_{41}+R_{44}}{C^2 R_{45}R_{44}(R_{42}+2R_{41})}]^{1/2} = 50 \text{ Hz}$$

where, f_h is 3-dB cut-off frequency for Sallen-Key high pass filter.

The transfer function of a second order low pass filter with non-inverting amplifier [22] is given as,

$$A_2(s) = \frac{A_{02}\omega_{02}^2}{s^2 + s(\omega_{02}/Q_2) + \omega_{02}^2} \quad (3.3)$$

where, A_{02} is the dc gain of the amplifier, ω_{02} is the pole frequency, and Q_2 is the selectivity of the second order response.

$$A_{02} = a_{02}/\omega_{02}^2.$$

For Butterworth filter, $Q = 0.707$. The low pass filter stage in Fig. 3.5 is the second order Sallen-Key low pass filter with non-inverting amplifier, and its transfer function $A_2(s)$ is given as [22]

$$A_2(s) = \frac{V_{d2}}{V_{d1}} = \frac{1 + R_{50}/R_{48}}{s^2 R_{47}(R_{49} + R_{46})C_{21}C_{22} + s[C_{21}(R_{47}+R_{46}+R_{49}) + R_{47}C_{22}(1-A_{02})] + 1} \quad (3.4)$$

comparing Eqns 3.4 and 3.5, we get

$A_{02} = 1 + R_{50}/R_{48} = 2$

$$f_l = \omega_{02}/2\pi = \frac{1}{2\pi\sqrt{R_{47}(R_{46}+R_{49})C_{21}C_{22}}} = 2.4 \text{ kHz}$$

where, f_l is 3-dB cut-off frequency of Sallen-Key low pass filter of Fig. 3.5.

$$Q_2 = \frac{\sqrt{R_{47}(R_{49}+R_{46})C_{21}C_{22}}}{R_{47}C_{22}(1-A_{02})+C_{21}(R_{47}+R_{48}+R_{46})} = 0.707$$

To amplify the signal further and eliminate high frequency noise, the output of Butterworth filter is fed to a cascade of first order low pass and high pass filter, designed by using op-amp U13. The transfer function of bandpass filter shown in Fig. 3.5 is as follows,

$$A(s) = V_e/V_{d2} = -\frac{R_{53}}{R_{52}}\frac{\tau_1 s}{(1+\tau_1 s)(1+\tau_2 s)}$$

where, $\tau_1 = R_{52}C_{19}$ and $\tau_2 = R_{53}C_{20}$. Therefore, 3-dB cut-off frequencies are $f_h = 1/(2\pi\tau_1)$ = 10 Hz and $f_l = 1/(2\pi\tau_2) = 22$ kHz. The pass band gain $A_{03} = -R_{53}/R_{52} = -220$.

Thus, the overall frequency response has an overall passband gain of 403 and passband from 50 Hz to 2.4 kHz.

3.3 Electrodes used

In electroglottography, two or four electrode arrangement is used. In the 2-electrode arrangement, two disc or plate electrodes are used. Each electrode is used for both current injection and voltage sensing, as shown in Fig. 3.6. Since, the excitation frequency is high, electrode gel for good contact is not needed. In the 4-electrode arrangement, as shown in Fig. 3.7, there are two discs, each with a centre conducting disc, surrounded by a ring. These can be connected in different configurations for current injection and voltage sensing, as shown in Fig. 3.8. In case of 4-electrode arrangement, electrode gel must not be used, because it is likely to provide short between the centre disc and the surrounding ring. For using these electrodes, the contact surface should be cleaned dry.

In the configuration shown in Fig. 3.8a, the current is passed through the central discs and the modulated voltage is sensed from the ring surrounding the central discs. In the second arrangement shown in Fig. 3.8b, the current is passed through the ring electrodes and the modulated voltage is sensed from the central discs. However, both the methods do not give any significant advantage over the 2-electrode configuration, in extracting impedance variation across the glottis, and in reducing common mode interference.

In the configuration as shown in Fig. 3.8c, the current is passed through the central disc and the modulated voltage is sensed across the same disc. The rings surrounding the central discs are shorted to the ground. In this configuration, the common mode pick-up is reduced. However, part of the injected current flows through the surface tissue to the ring electrode and ground, hence the sensitivity of the instrument is reduced. In another configuration as shown in Fig. 3.8d, the current is passed through the central discs and the modulated voltage is sensed across the same discs. In this configuration, the two rings are shorted, without being connected to the ground, and thus these act as current director rings [23].

Another possible configuration is shown in Fig. 3.8e. It consists of a buffer for driving the individual guard rings. The potential developed at the central discs is used for driving the guard ring of each electrode separately. This reduces the common mode pick-up and directs the flow of the excitation current into the larynx and avoids the leakage of the current from the source across the skin of the neck, resulting in enhanced sensitivity. However, it is to be noted that there will be additional current flow between the outer rings, drawn from the buffer amplifiers. The use of the rings driven by buffered voltage is also known as "guard ring" [23].

The choice of electrode configuration is very important. The sensitivity of the circuit depends upon the electrode configuration used. If the electrodes without guard rings are used, then the current finds an alternate path across the skin of the neck and reaches other electrodes. Thus very small amount of current passes through the larynx and hence the sensitivity decreases. It was decided to use the configuration of Fig. 3.8e.

One terminal of the load in the voltage-to-current converter circuit of Fig. 3.3a is at virtual ground. Hence the output terminals E1, E2 of this circuit can be connected to the electrodes in such a manner that no buffer amplifier are needed, as shown in Fig. 3.9. CE1 is connected to E1, which is at virtual ground, and RE1 is connected to ground. Thus there is no potential difference between CE1 and RE1. Since E2 is the voltage output terminal of the op-amp, there is no need to buffer it, and it is connected to CE2 and RE2 both. The current to be injected $I_e = V_b/R_9$ flows between CE2 and CE1. Voltage across them is sensed as voltage Vc. It is to be noted that there will be some current flowing between RE2 and RE1. This current also will be supplied by terminal E2, but will not affect I_e and V_c [11] [23].

The electrodes are to be placed properly across the neck for detection of the impedance variation across the glottis without applying gels. The application of gel will short the ring electrode with the central electrode. Thus, maximum current will pass across the neck surface instead of passing through the larynx.

The electrodes are constructed on a glass epoxy printed circuit board as shown in Fig 3.10. The PCB is double sided with plated-through-holes for establishing electrical connection between the two sides. The lead wires are soldered on the other side of the electrode. The electrodes are gold plated in order to avoid corrosion due to electrolysis. These electrodes were constructed earlier by Chitnis [11].

3.4 Glottal impedance simulator

For calibration of the glottal impedance sensor and testing of various electrode configurations for reducing noise, there is a need for glottal impedance simulator. For obtaining the frequency response and calibration of the glottal impedance sensor, it has been decided to simulate the impedance variation as step changes, corresponding to the vocal chord being fully open, and fully closed. The impedance across the thyroid cartilage is approximately 500 Ω and the change in the impedance due to vocal fold vibration is less than 1 Ω [6] [15].

This simulator should simulate various fixed impedances, and the variations in the impedances due to glottal vibrations, in the current path. At the frequency of current injection, all the impedance are predominantly resistive. A model of the glottal impedance between the electrodes is given in Fig. 3.11. R_z represents tissue resistance in the current path, with vocal chords fully open. When the switch S is closed, R_y is connected in parallel to R_z, and this corresponds to the condition of vocal chords in full contact. Resistors R_{x1}, R_{x2}, R_{x3} and R_{x4} connected in series represents the resistance across the outer surface of the neck. R_{z1} and R_{z2} represent the resistance in the path of the current flow from the surface to the inner tissue. The ground point between R_{x2} and R_{x3} represents the common mode pick-up point.

A scheme for realizing the model of Fig. 3.11 is shown in Fig. 3.12. The switch S is an analog switch controlled by the glottal pulses V_{gl}. The resistance R_y has been split into two resistances.

$$R_y = R_{y1} + R_{y2}$$

To ensure proper operations of the switch S, the signal voltage at its terminal must be referred to the same ground on V_{gl}. For this purpose, R_z has also been split in two resistances.

$$R_z = R_{z01} + R_{z02}$$

$$R_{z01} = R_{z02}$$

With switch "off" (when V_{gl} = low),

$$R_{eq.}(Soff) = R_x \mid\mid (R_{z1} + R_{z2} + R_{z3})$$

and with switch "on" (when V_{gl} = high),

$$R_{eq.}(Son) = R_x \mid\mid (R_{z1} + R_{z2} + R_z \mid\mid R_y)$$

where, $R_x = R_{x1} + R_{x2} + R_{x3} + R_{x4}$

$$R_y = R_{y1} + R_{y2}$$

$$R_z = R_{z01} + R_{z02}$$

The impedance simulator has been built using the circuit as shown in Fig. 3.13. An astable multivibrator formed using U2 produces ± 3.5 Vp-p square wave output with a frequency range of 100 Hz to 650 Hz. The output of astable multivibrator is given to comparator formed using U1 to obtain ± 4.5 Vp-p square wave output. The circuit is operated by a split power supply designed using op-amp U3 with 9 V DC input, and ± 4.5 V DC output. The switch S of Fig. 3.8 has been realized by using four switches in CD 4066 connected in parallel.

Comparing the components of Fig. 3.12 with circuit components of Fig. 3.13, we get

$$R_{x1} = R_{67}, R_{x2} = R_{68}, R_{x3} = R_{69}, R_{x4} = R_{70},$$

$$R_{z1} = R_{66}, R_{z2} = R_{71},$$

$$R_{z01} = R_{64}, R_{z02} = R_{65},$$

$$R_{y1} = R_{62} + R_{p4}, R_y = R_{63}$$

With the values shown in Fig. 3.13

$$R_{eq.}(Soff) = 591.10\Omega$$

$R_{eq.}(Son) = 590.10\Omega$

Therefore, $R\ (Glottis\ open) = 591.10\ \Omega$

$R\ (Glottis\ closed) = 590.10\ \Omega$

$\Delta R/R = 0.00169$

The testing of the various blocks of glottal impedance sensor and the glottal impedance simulator is explained in Chapter 5. Chapter 4 explains the acquisition and display of the L_x waveform from the glottal impedance sensor.

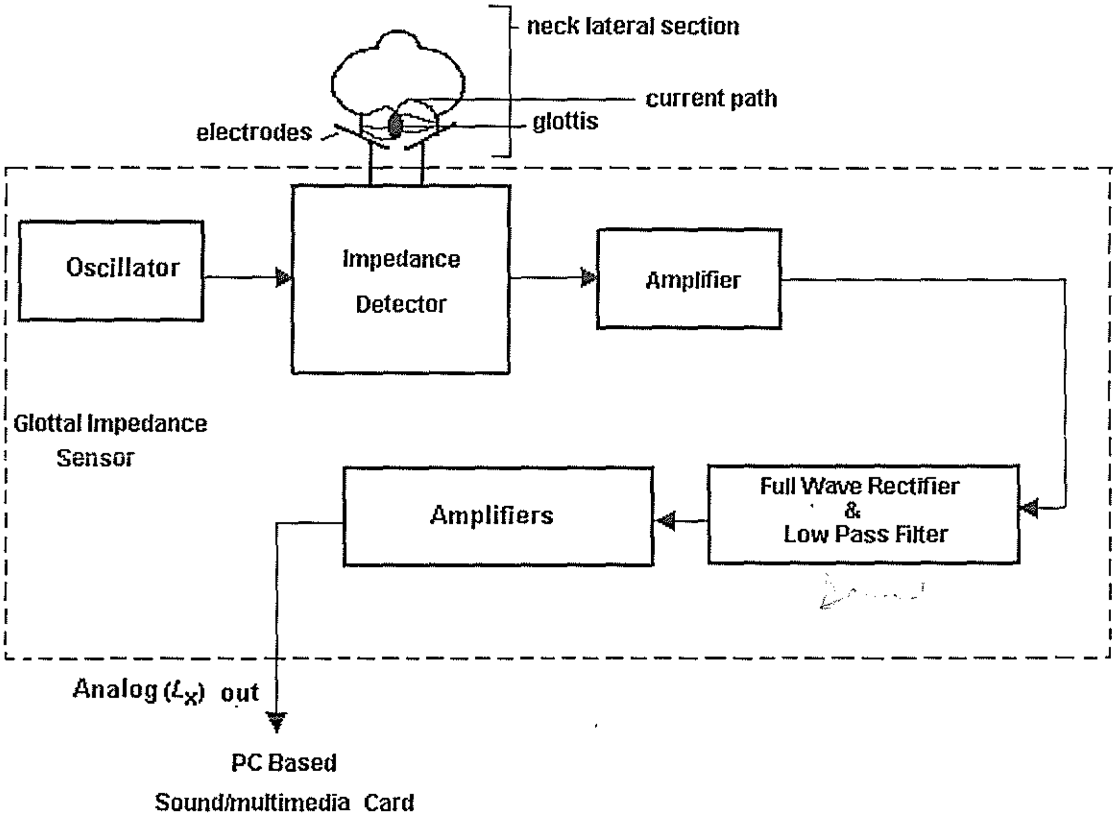


Fig. 3.1 Block diagram of the hardware

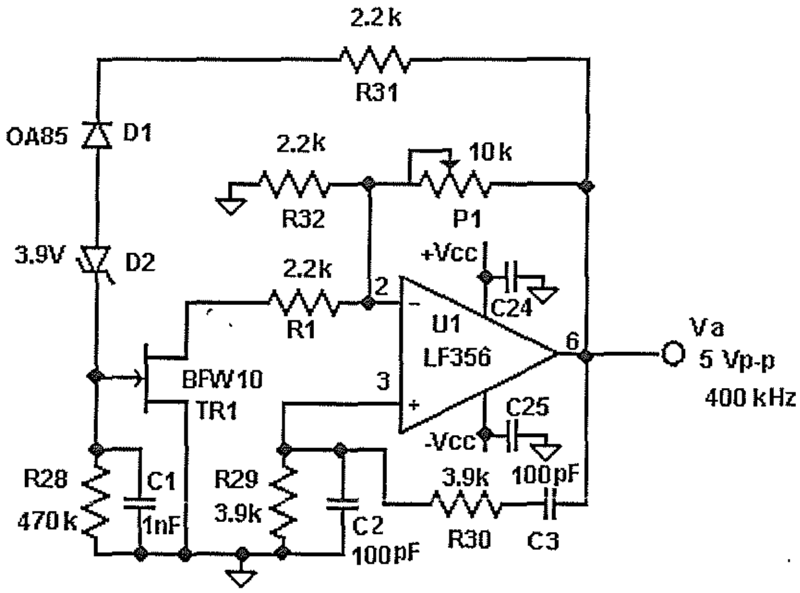


Fig. 3.2 Wien Bridge oscillator

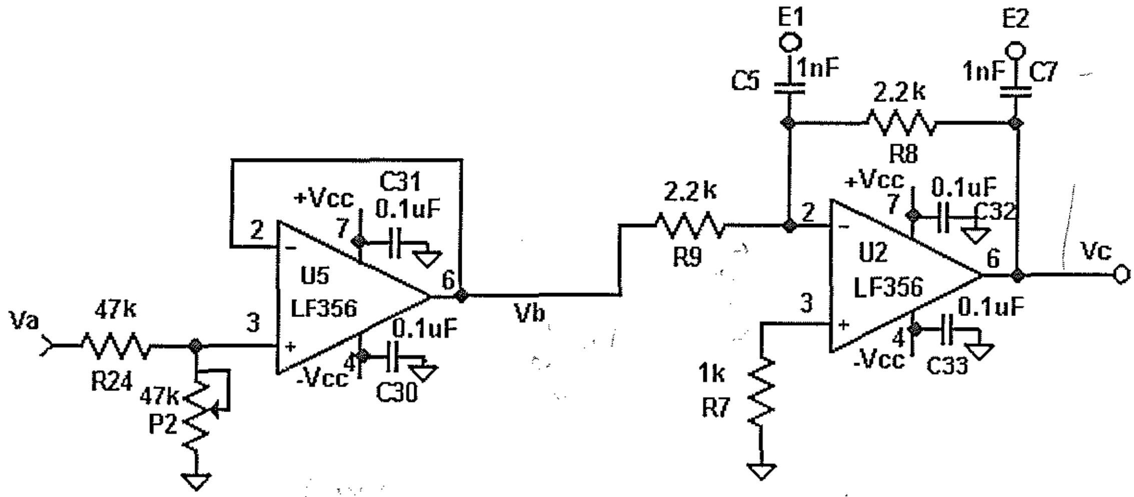


Fig. 3.3a Circuit diagram of impedance detector. Va is the output of the oscillator circuit in Fig. 3.2

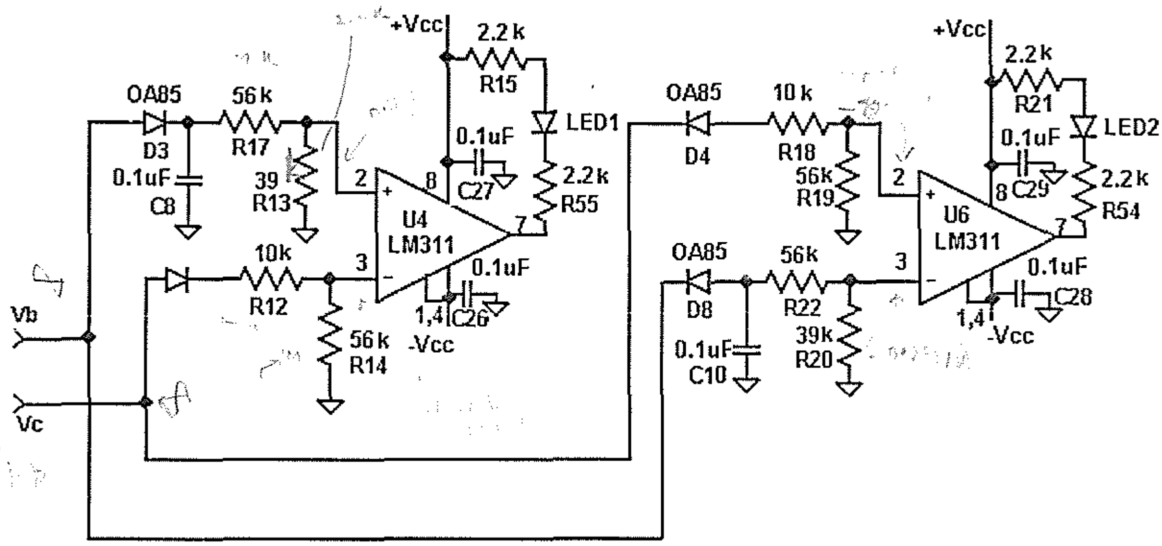


Fig. 3.3b Circuit diagram of high impedance indicator

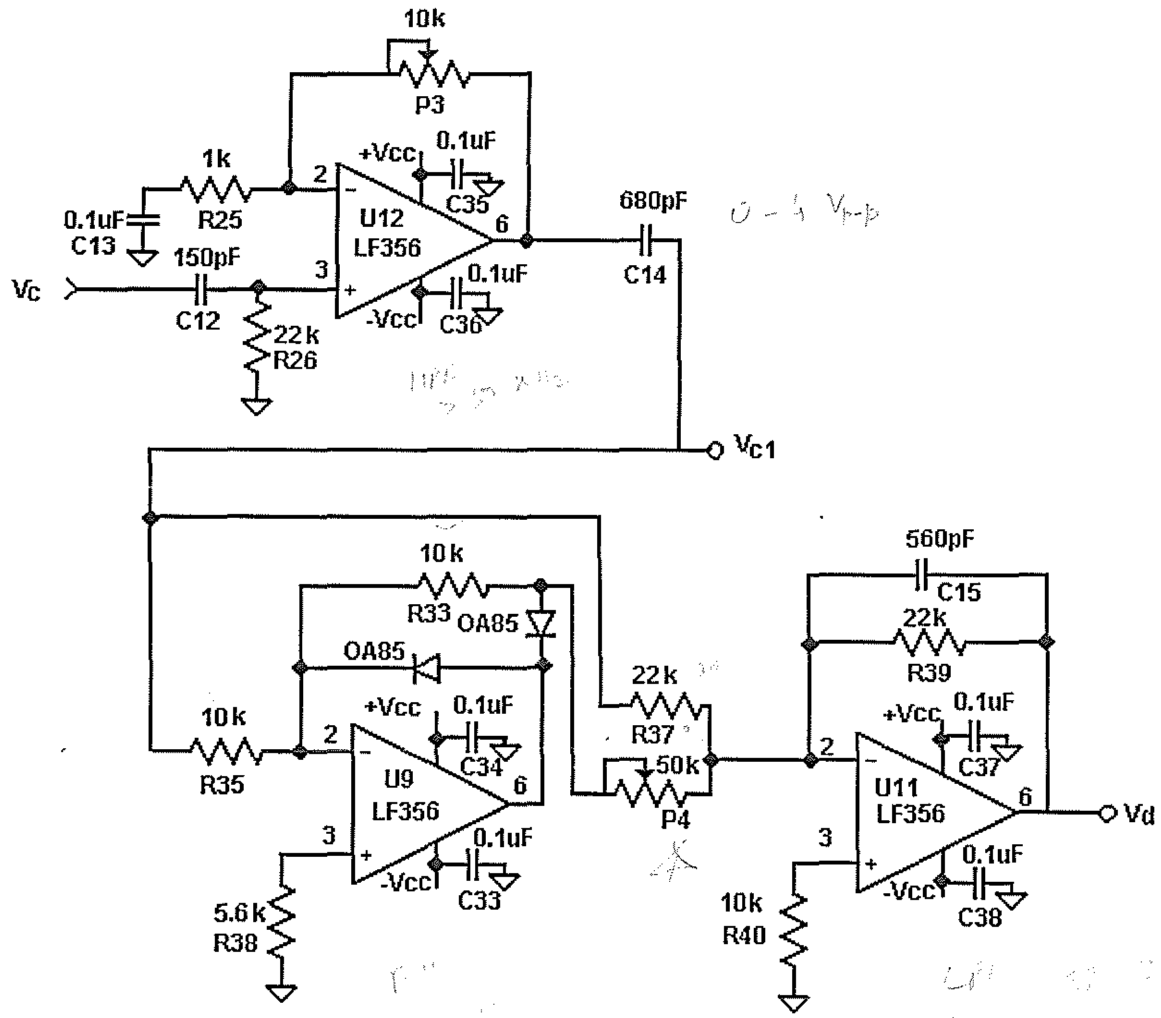


Fig. 3.4 Circuit diagram of demodulator. Vc is the output of impedance detector of circuit in Fig. 3.3a

check.
Should be ≈ 11 k

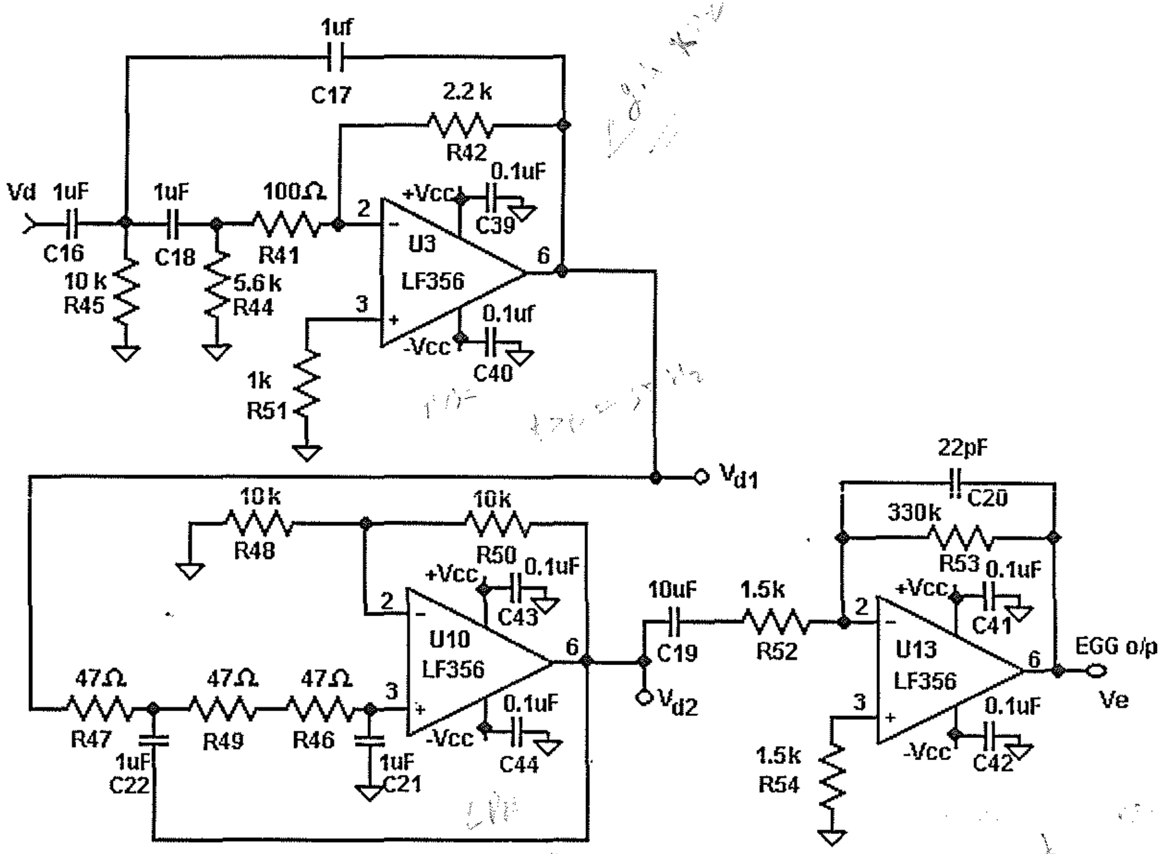


Fig. 3.5 Circuit diagram of amplifiers. Vd is the output of the demodulator circuit in Fig. 3.4

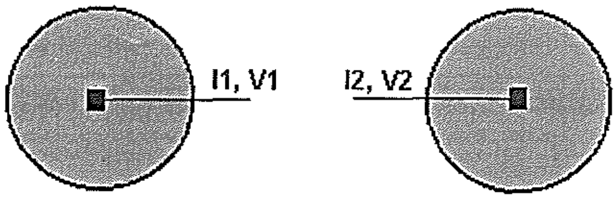


Fig. 3.6 2-electrode arrangement

Fig. 3.7 4-electrode arrangement

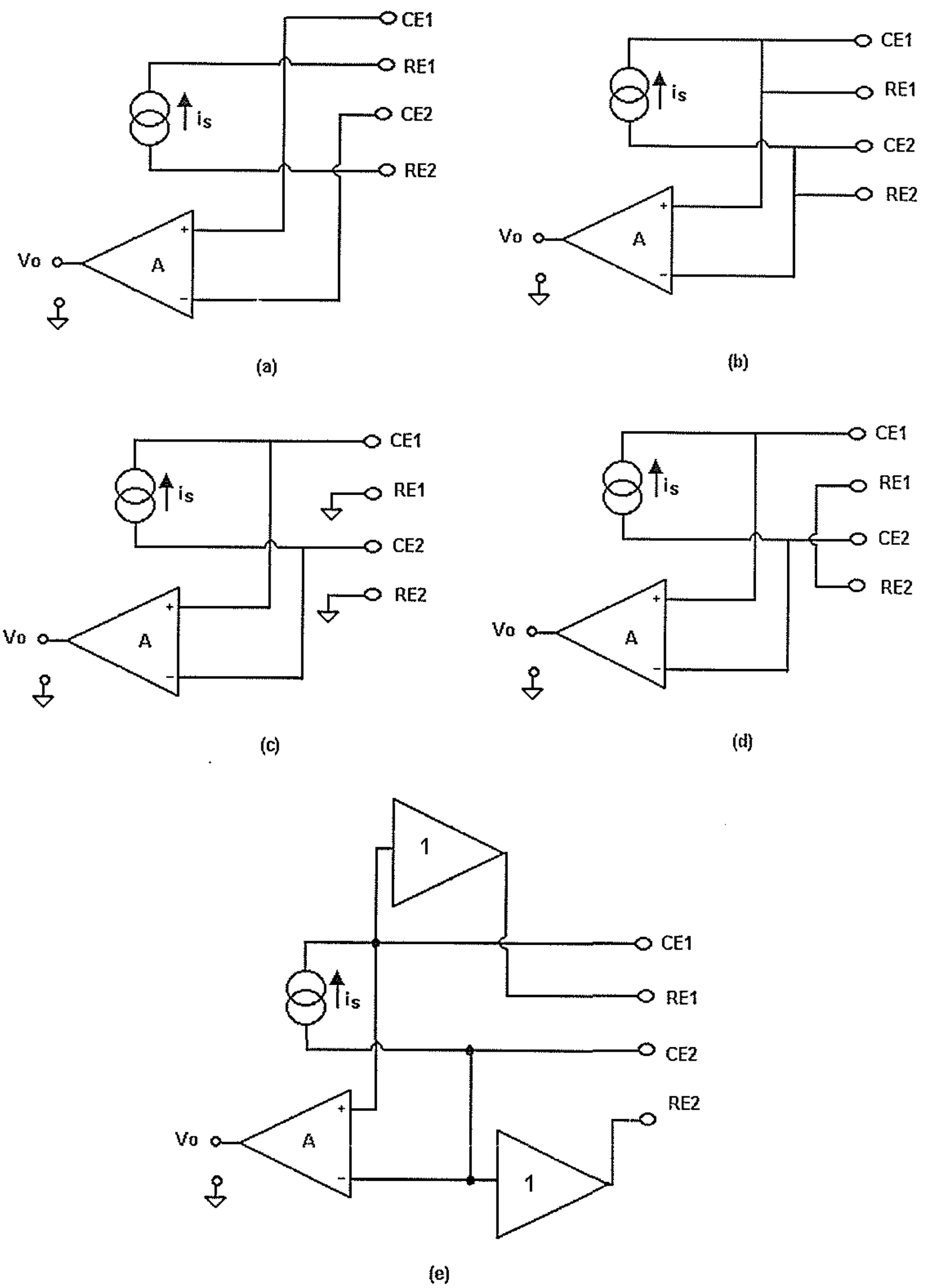


Fig. 3.8a-e Various 4-electrode configurations for EGG

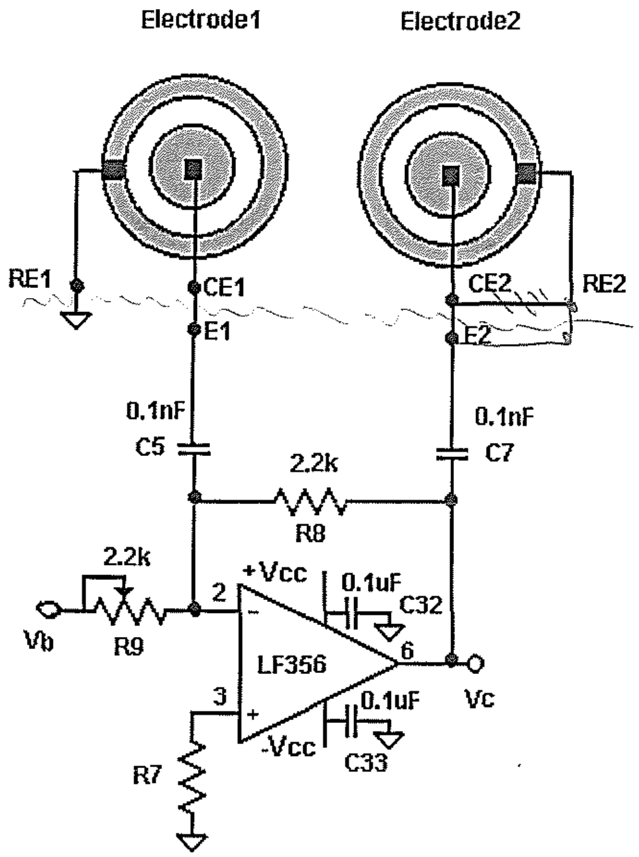


Fig. 3.9 Connections of electrodes to the V/I converter of Fig. 3.3a to provide driven guard rings

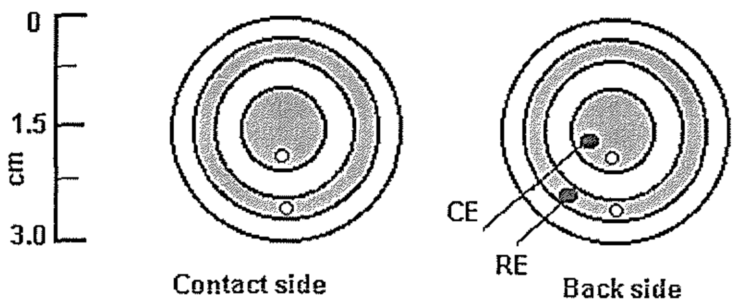


Fig. 3.10 Construction of electrodes. Source [11]

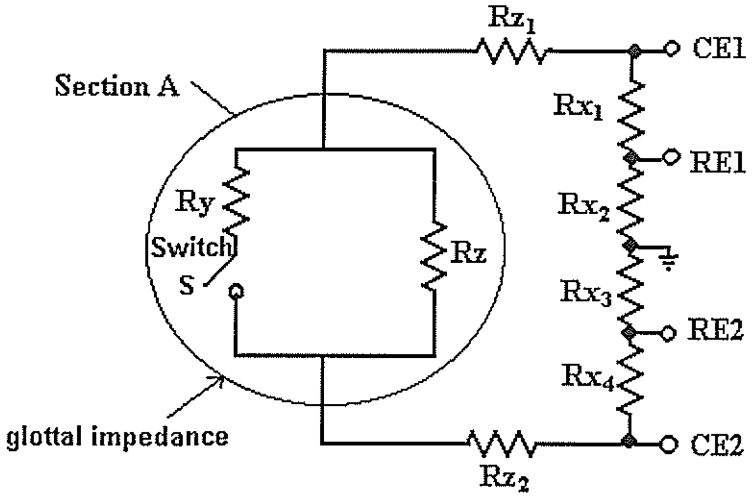


Fig. 3.11 A model of the impedance between the two electrodes

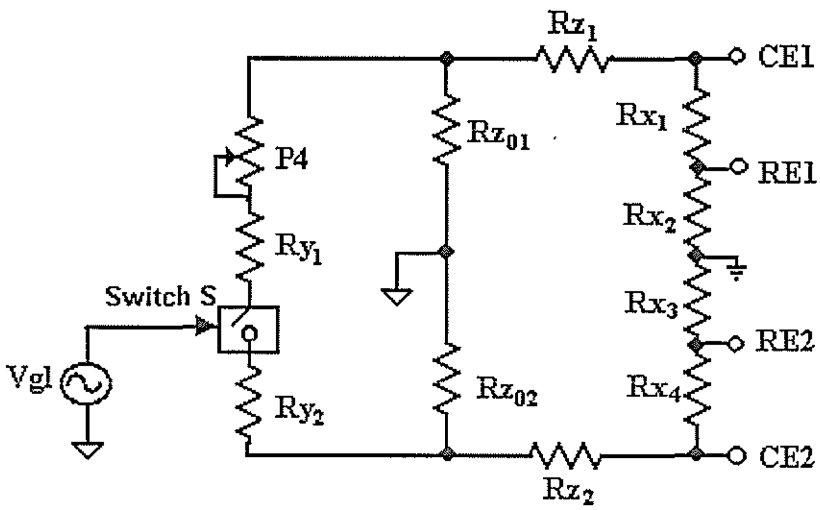


Fig. 3.12 Impedance simulation for the model in Fig. 3.11

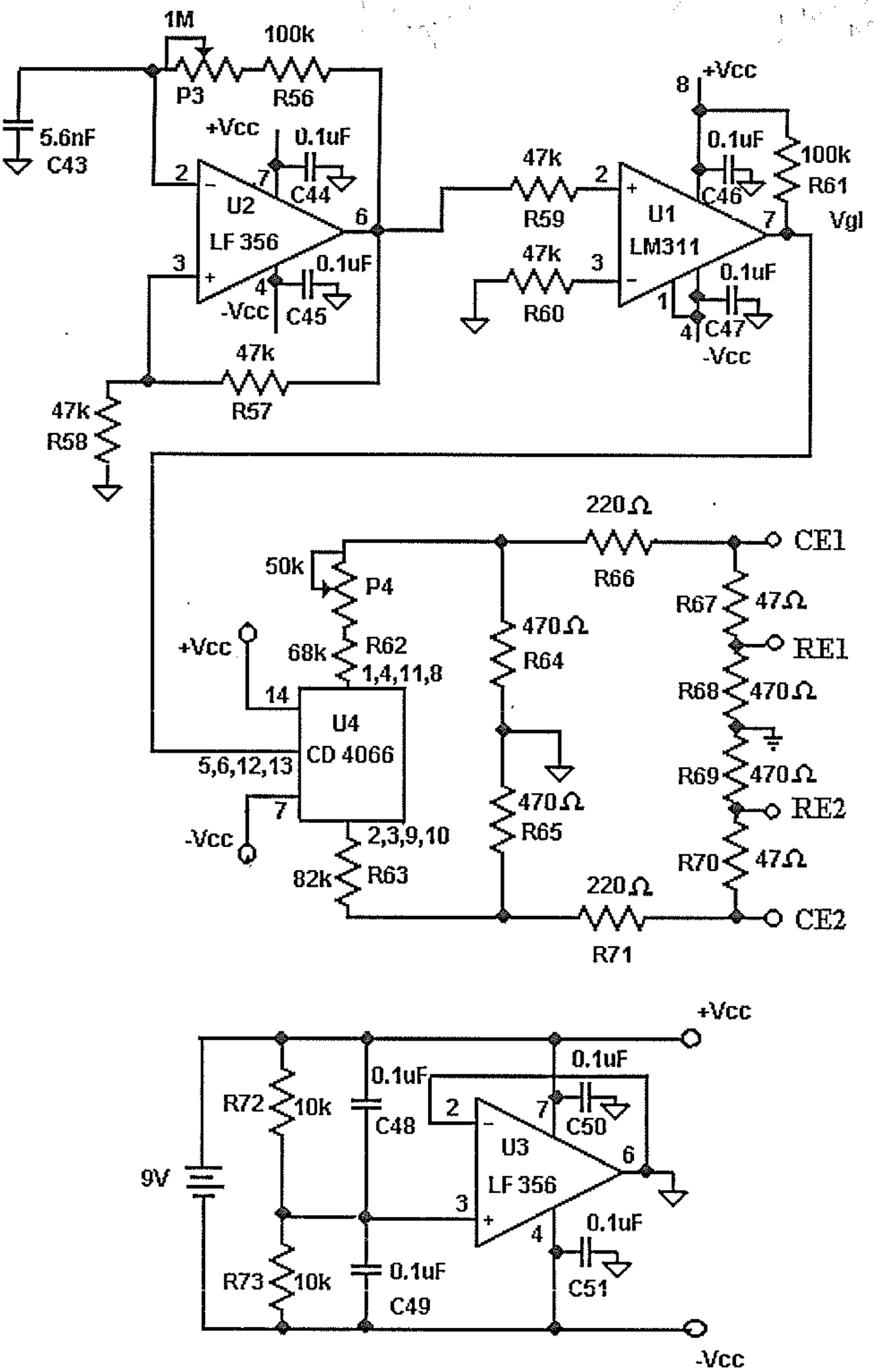


Fig. 3.13 Circuit of glottal impedance simulator

Chapter 4
SIGNAL ACQUISITION AND DISPLAY

4.1 Introduction

Signal acquisition, analysis, and display system has been developed around a PC and multimedia/sound card. The analysis includes obtaining F_x values from the input L_x waveform, F_x histogram, and spectrographic analysis of the input waveform.

4.2 Signal acquisition

The L_x waveform (output signal of the glottal impedance sensor) is acquired using a PC based sound/multimedia card for analysis and display purpose. The specifications of sound/multimedia card are given in the Appendix B.

The amplitude of glottal impedance sensor varies from 0.5 Vp-p to 1 Vp-p. For acquiring the waveform from glottal impedance sensor, a program "sigac99" in "C" has been written jointly with Ratanpal [24]. The program "sigac99" is invoked by the program "hist99" for pitch F_x and F_x histogram calculations. The signal from the glottal impedance sensor can be recorded for a maximum period of 3 min., since the file pointers in "C" program cannot handle more than 10 MB of data file. The L_x waveform can be sampled at the sampling rate of 11025 or 22050, or 44100 Sa/s. The waveform can be stored in the form of binary (8 bit or 16 bit data) or text format [24] [25] [26].

4.3 Pitch calculation and histogram plotting

The algorithm developed for calculating the pitch uses period-by-period analysis for detecting the pitch periods. The algorithm uses a data file of the samples acquired from the input EGG or L_x waveform for calculating the instantaneous pitch periods. The pitch periods are estimated from successive zero crossings, which are detected using the Schmitt trigger method as follows.

The input L_x waveform samples is $x(n)$. The mean $x_{mean}(n)$ of maximum $x_{max}(n)$ and the minimum $x_{min}(n)$ amplitude of the acquired sequence $x(n)$ from the glottal impedance sensor is calculated. This $x_{mean}(n)$ is taken as threshold.

One bit quantization of the sequence $x\ (n)$ is done

$$y(n) = 1 \qquad \text{if } x(n) > x_{mean}(n) + 0.01\, x_{max}(n)$$
$$= -1 \qquad \text{if } x(n) < x_{mean}(n) - 0.01\, x_{min}(n)$$
$$= y(n-1) \qquad \text{otherwise}$$

The 1-bit quantized output values is stored in a file, and then processed again. We calculate the number of samples N between the switching from one quantization level to another for obtaining the instantaneous pitch periods. The instantaneous pitch periods are calculated as

$$F_x = f_s/2N$$

where f_s = sampling rate. Every time the pitch value F_x is calculated, it is stored in a file for obtaining the F_x histogram and the F_x plot.

The F_x plot is obtained by plotting the instantaneous pitch values v/s time. The pitch axis is normalized from minimum pitch as 0% to maximum pitch 100%.

The F_x histogram can be plotted as either a single period or a multiple (often, triple) period histogram. The frequency range of interest is divided into a number of equal bins. In this implementation, the frequency range is taken to be 0 to 1 kHz, and the number of bins is set to 50. Thus, each bin interval is 20 Hz. For a single period F_x histogram, each bin is incremented when a pitch value falls within its frequency range. The content of each bin divided by the total number of pitch values is a measure of the probability density. The Probability values are plotted on the log scale (0.1% to 100%) or linear scale (0% to 100%). In triple period histogram, each bin is incremented only if three successive pitch values fall within its frequency range [7]. The probability values can be scaled with respect to the mean pitch period of the bin to get F_x histogram on the basis of the relative time for which the pitch value lies in a bin using eqn.2.6.2 in section 2.6.1.

The program is called "hist99", and it invokes the program "sigac99" for signal acquiring (as described in the previous section) and the program "spec99" for spectrographic analysis, as described in the following sections.

4.4 Spectrographic analysis

For analysis of glottographic waveform, a spectrographic analysis software is required which can display the time-varying spectrum of glottographic signal. Spectrographic analysis is useful for determining formants present in the speech waveform, voiced-unvoiced classification, fundamental frequency estimation, and speech synthesis. In a spectrogram, time varying spectral characteristics are displayed as a two-dimensional plot, with time and frequency along x and y axes respectively. The spectral magnitudes as a function of time and frequency are viewed as intensity variations [13] [24]. The digital spectrographic analysis involves short-time Fourier analysis of the acquired signal, conversion of the spectral magnitude to dB scale and display of these magnitudes as a function of time and frequency. This technique is particularly suitable for analysis of nonstationary signals such as speech. L_X waveform being a non-stationary signal, spectrographic analysis is likely to be a careful tool for its study.

4.4.1 Digital spectrographic analysis

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

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

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

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

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

The frequency resolution of the spectrographic analysis with a particular window is equivalent to its bandwidth. For Hamming window, it is $\Delta f = 1.36 f_s / L$, where f_s is the sampling rate. For speech analysis, wideband spectrogram with spectral resolution of 300

Hz is useful in observing the voiced speech as vertical striations and for seeing formant transitions. For unvoiced speech, the vertical striations do not appear and the spectral pattern is much more ragged. On the other hand, narrow band spectrograms with spectral resolution of 45 Hz is useful for observing the fundamental frequency and its harmonics.

4.4.2 Previous setup

A PC-based spectrographic analysis package was developed earlier at IIT Bombay by Thomas [28] in the programming language Pascal. The spectrogram was displayed on an area of 500 x 256 pixels using VGA card with 16 simultaneous colors on a monochrome monitor. The digitized data file is created using a data acquisition card. The selected segment of the data file was divided into 500 overlapping time frames. An L-point Hamming window was applied to each frame and a 512-point DFT was then taken. The log magnitude in dB was computed and spectral information for each frame was displayed in the form of 256 vertical pixels. The pixel intensity indicated the spectral magnitude in shades of grey, with white used for minimum magnitude level and black used for maximum magnitude level. The program provided an option for generating a wideband, narrowband or a combined spectrogram to preserve both time and frequency resolution. The combined spectrogram was generated by evaluating the geometric mean of the wideband spectrogram and the narrowband spectrogram [28]. Since all the computational tasks were performed by the PC, the execution speed was slowed down.

To increase the speed of execution, Baragi and Prasad in 1996 developed a spectrograph package, which divided the tasks between the PC and a DSP board based on TMS 320C25 processor [28] [29]. In this package, a C program running on the PC handled the user-interface and display, while the DSP board handled the data acquisition and computationally intensive FFT routines. Analog signal conditioning circuit consisted of antialiasing low pass filter at input and smoothing low pass filter at the output. The PC performed windowing and pre-emphasis of the data frame, and downloaded it to the DSP board. The DFT size used was 256, and the resulting spectrogram was displayed on an area of 500 x 128 pixels. While FFT of one block is being calculated on the DSP board, the PC uploads 128 samples of the computed FFT of the previous block and computes the log magnitude. However, this program didn't have the feature of combined spectrogram.

Chaudhari in 1996 introduced the facility for capturing the displayed spectrogram and storing it in the Postscript format, which could be printed on a 600 dpi laser printer [30]. The DSP board performed the data acquisition and FFT computations as before. The Fig 4.1 shows the previous setup for spectrographic analysis.

As this spectrographic analysis setup was based on the DSP board, it was not portable to other computers. Also, as the PC could only transfer integer values to and from the DSP board, there was truncation error in the computation of FFT. The Postscript file displayed the spectrogram with a resolution of 16 gray levels, which can be increased to 256 gray levels [31]. Also, the setup could not display a monochrome spectrogram on a color screen. Due to the increase in processor speeds and the easy availability and low cost of sound cards as compared to DSP boards, it was decided to modify the previous setup which is described in the next section. This development work, described in the next subsection, has been carried out jointly with Ratanpal [24], in which the spectral analysis and display are primarily his contribution.

4.4.3 Present setup

The present setup utilizes the sound card available on most current PCs for playback and recording of the glottographic signal. The program, written partly in C language and partly in Visual C++, prompts the user for digitized input either from the microphone or from a pre-recorded signal file [24]. Spectral analysis is then carried out for selected sample range of data file. The selected segment is divided into 500 overlapping time frames. After pre-emphasis, an L-point Hamming window is applied to each frame. Zero padding is then done so that the magnitude spectrum can be obtained using a 256-point FFT, irrespective of the window length. The spectral samples are scaled to compensate for the variation in the energy of the signal for different window lenghts, due to zero padding. The log magnitude is then calculated and the spectrogram is displayed on an area of 500 x 128 pixels with a resolution of 64 gray levels. The spectrogram can be stored in the form of a Postscript file having 256 gray levels of resolution. The selected portion of the data can be presented to the left, right or both ears. The software works on a color / monochrome monitor and on any PC with a sound card. The Fig 4.2 shows the modified setup. The Fig 4.3 shows the spectrogram of the utterance */ana/* using the previous setup and the new setup. The program is called "spec99" and is invoked by "hist99". Fig. 4.4 shows the spectrogram of sine wave swept from 200 Hz to 5000 Hz for 1.8 seconds at 11025 Sa/sec. Fig. 4.5 shows the spectrographic analysis of the speech waveform acquired by using microphone for the utterance "*…atma amar hai…*". The spectrogram for the speech signal show the pitch periods and formants structure (resonance of vocal tract filter).

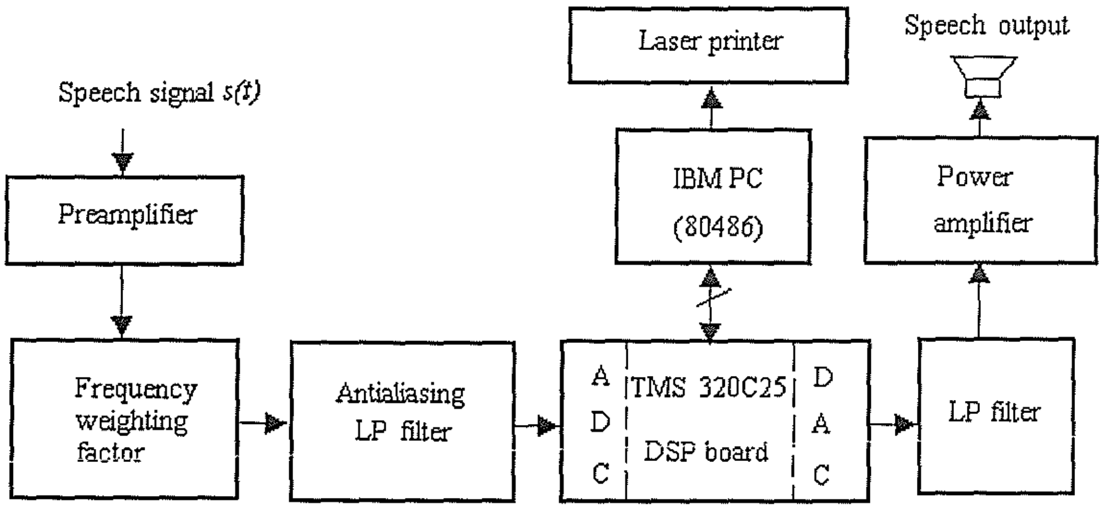


Fig. 4.1 Hardware setup of the DSP board- based spectrographic analyzer. Source [30]

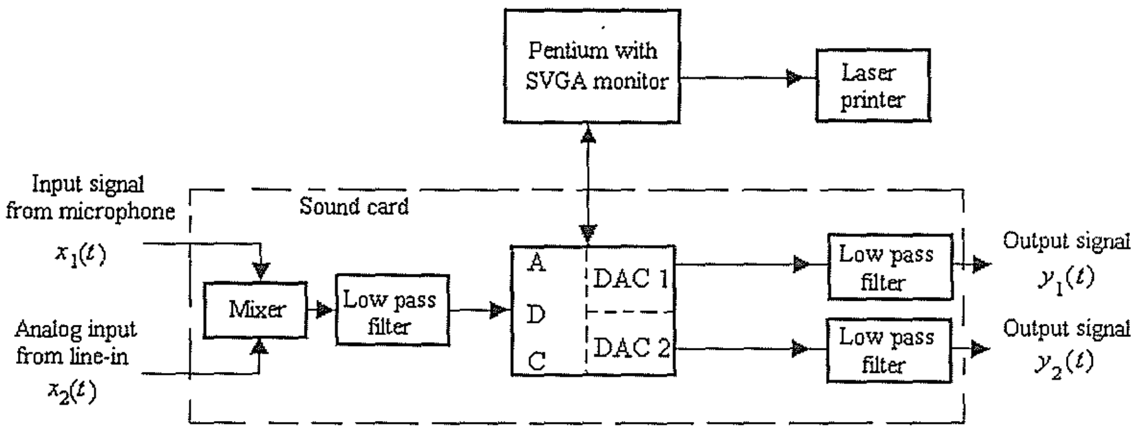
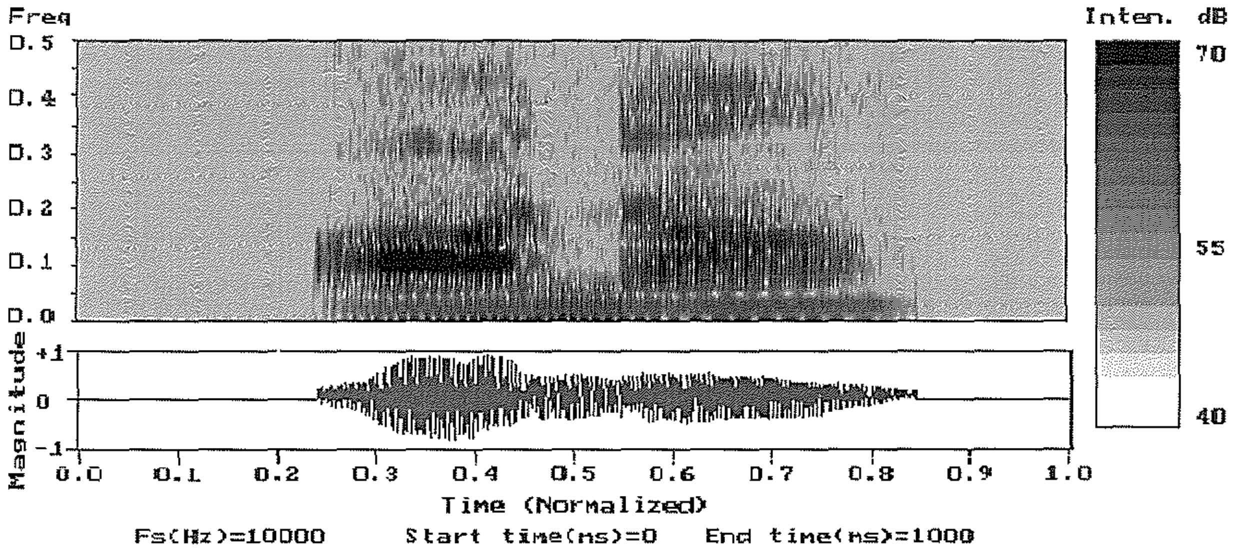
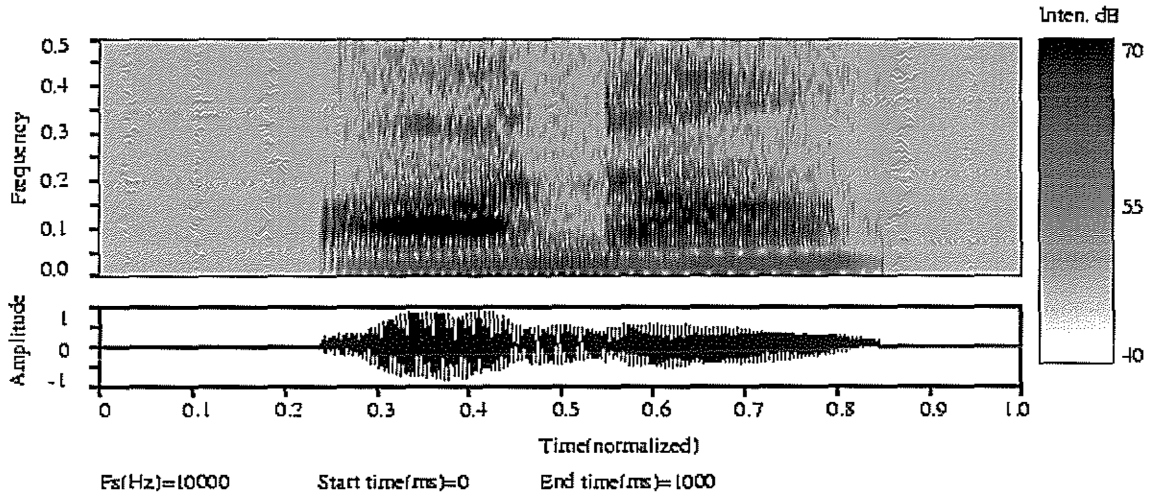


Fig. 4.2 Present setup of the spectrographic analyzer



a)



b)

Fig.4.3 Wideband spectrogram of speech waveform for utterance /ana/, using a) previous setup and b) using the new setup

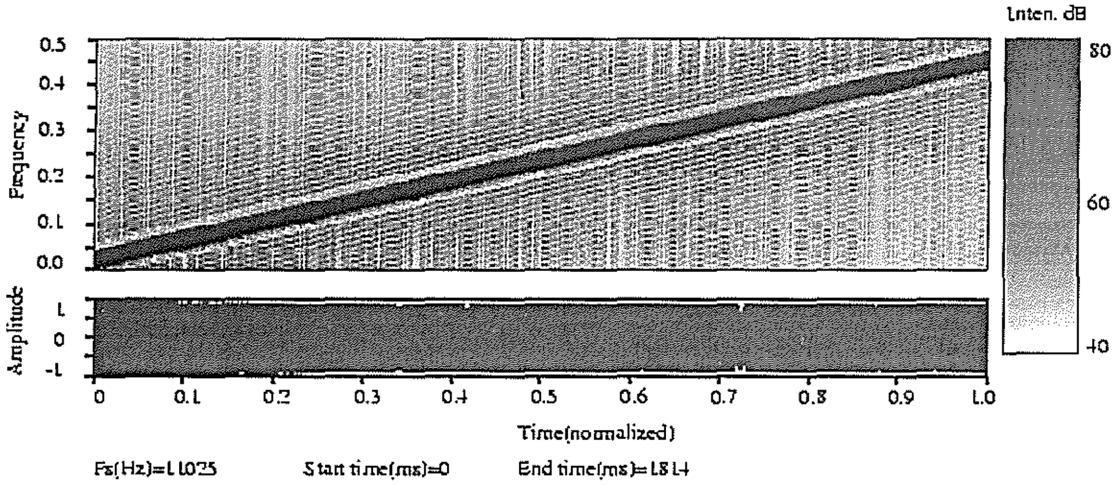


Fig. 4.4 Wideband spectrogram for the sine-wave swept linearly from 200 Hz to 5000 Hz for 1.8 seconds

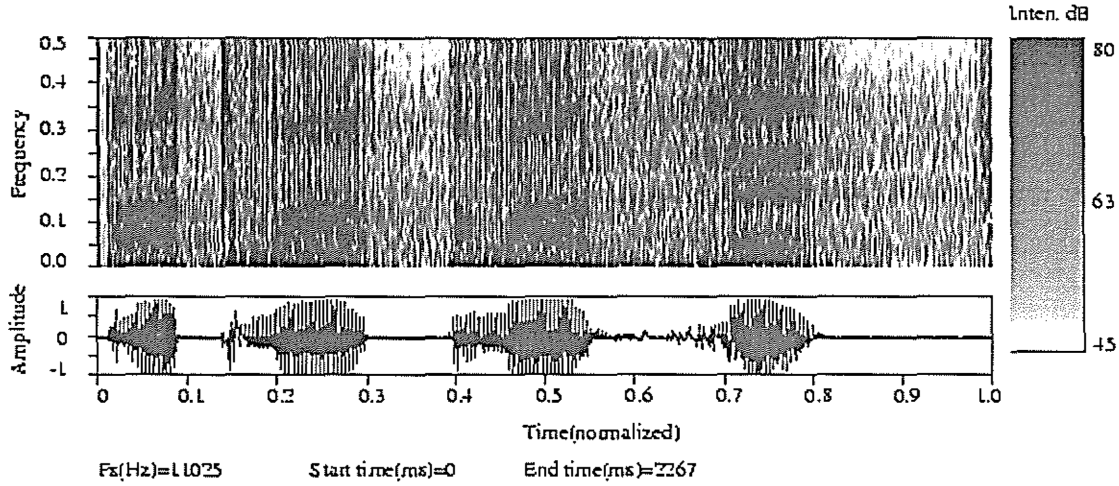


Fig. 4.5 Wideband spectrogram and the speech waveform $s(t)$ for the utterance " *atma amar hai* ..."

Chapter 5

CIRCUIT ASSEMBLY AND SYSTEM TESTING

5.1 Introduction

This chapter explains the assembly and testing of glottal impedance sensor and glottal impedance simulator, and signal acquisition and analysis of laryngograph L_x waveform from male and female subjects.

5.2 PCB design

The EGG circuit and the glottal impedance simulator circuit have been assembled on two separate PCB's and assembled as independent instruments. Both the PCBs are double sided with plated through holes (PTH). The glottal impedance sensor has been built on a PCB of size 11cm × 12.5cm. It consists of Wien bridge oscillator, impedance detector, demodulator and filter circuits. The glottal impedance simulator is built on a PCB of size 6cm × 7.5cm. It consists of split power supply, astable multivibrator and CMOS switch. The PCB layout for both the instruments has been designed with special consideration for reducing the power supply noise by providing tight coupling between the supply and ground, stabilizing the ground by providing a large ground plane and reducing the noise pick-up, by coupling the signal lines with ground track/plane. The entry points of the dc supply on the PCB are decoupled by 220µF/63V electrolytic capacitors in parallel with 0.1µF ceramic capacitor. Positive and negative supply for each IC is decoupled by 0.1µF ceramic capacitor placed as close as possible to the IC. Thick copper planes of supply and ground are provided on opposite sides of the PCB. This gives the effect of the distributed capacitance across the PCB, thus improving the decoupling effect. Care is taken to minimize the length of the supply path for ICs. The electrode terminals E1 and E2 shown in Fig. 3.3 are connected to the PCB by shielded coaxial cable.

Both instruments are battery operated. This helps to reduce the electromagnetic pickups, and takes care of safety considerations. The EGG instrument is powered by ±9V dual supply, obtained by using two 9V batteries. The glottal impedance simulator is operating from a single 9V battery with an internal split power supply circuit. The circuit schematics and PCB layouts are shown in Appendix C.

5.3 Assembly

The EGG instrument and the glottal impedance simulator have been assembled in two separate cabinets. The cabinets are made of 3 mm acrylic sheet. The PCB of glottal impedance sensor has been assembled in a cabinet with all the switches, controls, indicators and connectors for electrode on the front side and the o/p connector (i.e. L_x waveform) on the backside of the panel. The PCB of glottal impedance simulator has been assembled with all the switches, controls and connectors for electrodes on the front side of the panel only. The box has been designed, fabricated and assembled in such a way that the PCB, and the controls and accessories are fixed on the bottom portion of the box, for easy access to all the circuit parts during testing, once the top cover is removed. All the controls, switches and indicators are labeled. The front and the back panel of glottal impedance sensor and simulator are shown in Appendix D1. The cabinet design for glottal impedance sensor and simulator is shown in Appendix D2. The list of components of glottal impedance sensor and glottal impedance simulator is shown in Appendix E1 and E2 respectively. The testing of various blocks of electroglottograph instrument and glottal impedance simulator are explained in subsequent sections.

5.4 Testing of glottal impedance sensor

The electroglottograph instrument is basically the circuit for sensing the variation in the electrical impedance of the glottis, caused by the opening/closing of the vocal chords and it consists of the following blocks:

- Oscillator
- Impedance detector
- Demodulator and amplifiers

5.4.1 Oscillator

The Wien bridge oscillator with amplitude stabilization is shown in Fig. 3.2. It was decided to use a carrier frequency 400 kHz. With the values shown in Fig. 3.2 the oscillator frequency is 408 kHz. In actual testing of the oscillator circuit, the oscillating frequency is found out to be 390 kHz.

The amplitude of the oscillator is stabilized using the circuit around JFET TR1 and zener diode D2. The variation of output voltage of the oscillator due to change in supply voltage is shown in Fig. 5.1. The figure shows that the output of the oscillator remains constant (4.5 V p-p) for power supply voltage fluctuation from ±6 V to ±12 V.

5.4.2 Demodulator and amplifiers

The demodulator consists of precision full-wave rectifier and an averaging circuit. The circuit was tested for linear response to the input amplitude. The response of the demodulator circuit is shown in Fig. 5.2.

The output Vd of the demodulator is given to the cascade of the amplifiers. The amplifiers consists of cascade of second order Butterworth high pass filter and low pass filter U3 filter, and bandpass filter. The second order Butterworth high pass and low pass filter is designed with 3 dB cut-off frequency of 50 Hz and 2.5 kHz respectively. The bandpass filter following the Butterworth filter is designed for 3 dB cutoff frequencies of 10 Hz (high pass) and 22 kHz (low pass). The three stages are designed for passband gain of 0.916, 2, 220 respectively, with an overall gain of 403 or 52.10 dB

The frequency response of the amplifier is shown in Fig. 5.3. The response of the amplifier circuit is taken on the oscilloscope. At frequencies outside the passband, the output amplitude decreases, and once it becomes comparable to the noise and hum, it cannot be reliably measured. Hence the response at frequencies when the gain is 20 dB below the passband gain, could not be correctly measured.

5.5 Testing of glottal impedance simulator

For calibration of the glottal impedance sensor and testing of various electrode configurations for reducing noise, a glottal impedance simulator has been designed. The impedance across the thyroid cartilage is approximately 500 Ω and change in impedance due to vocal fold vibrations is less than 1 Ω [16] [30]. The equations for selecting the values of resistors is explained in section in Section 3.3. The glottal impedance simulator is designed so as to give a change in impedance of 1 Ω. In the actual test of the glottal impedance simulator, the change in impedance is of the order of 0.9 Ω.

$R_{eq.}(Soff) = 591.10\ \Omega$, and $R_{eq.}(Son) = 590.10\ \Omega$, $\Delta R = 1.0\ \Omega$

In actual test it was found that $R_{eq.}(Soff) = 590.73\ \Omega$, $R_{eq.}(Son) = 589.83\Omega$, $\Delta R = 0.9\ \Omega$

Hence the glottal impedance simulator provides a change of 0.9 Ω over a basal impedance of 590.70 Ω.

Therefore,

$R(Glottis\ open) = 590.73\ \Omega$

$R(Glottis\ closed) = 589.83\ \Omega$

$\Delta R/R = 0.0015236$

The current to be injected was set at 1 mA. The gain of the demodulator input stage was set at the highest value where we do not get appreciable noise. When the glottal impedance simulator is connected across the sensor, we get 1 Vp-p, corresponding to step change of 0.9 Ω in the simulator. This gain setting is marked on the instrument dial as a calibrated point. With these injection current and gain setting, the electrodes are connected across the thyroid cartilage of the subjects. The output variation is from 0 to 4 Vp-p.

5.6 Testing of electrode configurations

In electroglottography, two or four electrode arrangement is used. The detailed explanation of various electrode configurations is given in Section 3.3. The various possible connections for 4-electrode arrangement are shown in Fig. 3.8. These arrangement were tested using glottal impedance simulator, and it was observed that maximum sensitivity is attained when the voltages at the two electrodes are buffered and used for actively driving the respective guard rigs. The arrangement for this configuration is shown in Fig. 3.9.

The electrodes are to be placed properly across the neck for detection of the impedance variation across the glottis without applying gels.

5.7 Testing of signal acquisition and display software

Signal acquisition, analysis and display system has been developed around a PC and sound/multimedia card. The analysis includes obtaining F_x values from the input L_x waveform, F_x histogram, and the spectrographic analysis of the input waveform. For acquiring the L_x waveform from the glottal impedance sensor, a program "sigac99" in "C" has been written. The program "sigac99" is invoked by the program "hist99" for F_x and F_x histogram calculations.

The signal from the glottal impedance sensor can be acquired for a maximum period of 3 min., since the file pointers in "C" program cannot handle more than 10 MB of data file. The L_x waveform can be sampled at the sampling rate of 11025 or 22050, or 44100 Sa/s. The waveform can be stored in the form of binary (8 bit or 16 bit data) or text format.

The spectrographic analysis setup has been extensively tested, and some examples of spectrographic analysis have been shown earlier in Fig. 4.3 to 4.5.

The algorithm developed for calculating the pitch uses period-by-period analysis for detecting the pitch periods. The pitch periods are estimated from successive zero crossings, which are detected using the Schmitt trigger method explained in section 4.3. Every time the pitch value F_x is calculated, it is stored in a file for obtaining the F_x histogram and the F_x plot.

The analysis software has been tested by generating data file for sine wave, whose frequency is swept linearly from 300 Hz to 700 Hz. The Pitch F_x and F_x histogram is calculated for the swept sine wave. Fig. 5.4 shows the pitch and F_x histogram (single period) of the swept sine wave. The F_x histogram for swept sine-wave is shown on log scale with percentage probability as related to time and frequency. The F_x histogram shows that in higher frequency range (450 Hz-700 Hz) some of the cycles are not detected, hence the bin is empty. For swept sine-wave, F_x histogram rises linearly when

% probability is plotted as related to frequency, and remains more or less constant when % probability is plotted as related to time.

The output from glottal impedance sensor with simulator connected across the terminal E1 and E2 of Fig. 3.3, is acquired from sound/multimedia card. The pitch F_x and single period F_x histogram (related to frequency) is shown in Fig. 5.5. The frequency from glottal impedance simulator has been set to 250 Hz.

5.8 Acquiring L_x waveform from subjects

The glottal pitch extraction has been done by recording L_x waveform from glottal impedance sensor. Recordings are done for subjects with normal vocal chords. The recordings obtained are given in Fig. 5.6 to 5.9. Here L_x indicates the laryngogram waveform, F_x indicates the pitch, and %P indicates the percentage probability in F_x histogram.

The L_x waveform , pitch F_x, and F_x histogram for sustained vowels */a/*, */i/*, and */u/* for a male and a female speaker are shown in Fig. 5.6 to 5.7. The maximum value of pitch lies in the frequency bin 100-125 Hz for male speaker AVP, and 200-220 Hz for female speaker DSJ.

For studying pitch F_x and F_x histogram over a long duration recording, L_x waveform was recorded for 1 minute duration, with a male subject (author himself) reading the beginning part of Chapter 5 of this dissertation. The single and triple period histogram (related to time) , pitch F_x, and laryngogram L_x waveform is shown in Fig. 5.8. Fig. 5.8a shows the single period F_x histogram with % probability related to time and it can be inferred that maximum value of pitch lies in the bin 80-120 Hz. Fig. 5.8b shows the triple period F_x histogram with % probability related to time. For triple period F_x histogram most of the pitch values lies in the bin 100-120 Hz. This is also indicative of vibrational regularity of vocal folds, since in triple period histogram, the bin is incremented only when three successive values lies in the particular bin.

Similar recording was done for female speaker DSJ for 1 minute and its single and triple period histogram, pitch F_x, and laryngogram L_x waveform is shown in Fig. 5.9. Fig. 5.9a shows that for a single period F_x histogram (related to time), maximum value of pitch lies in the bin 200-220 Hz. Fig. 5.9a also shows some peaks for bins apart from

200-220 Hz. Fig. 5.9b shows the triple period F_x histogram (related to time). The Fig. 5.9b shows that for female speaker DSJ, most of the pitch values lie in the bin 180-220 Hz.

The recording and spectrographic analysis was done for Hindi sentence "...*atma amar hai*..." by the author (AVP) for the speech signal and L_x waveform. Since two channel recording is not possible using a sound card; two PC/sound card set-up were used for simultaneous recording speech $s(t)$ and L_x waveform. The two waveforms along with the spectrograms are shown in Fig. 5.10. the spectrogram for the speech signal show the pitch periods and formant structure (resonance of vocal tract filter). The spectrogram for the L_x waveform shows the vertical striations in the voiced segments and there are no formants (moving resonance frequencies).

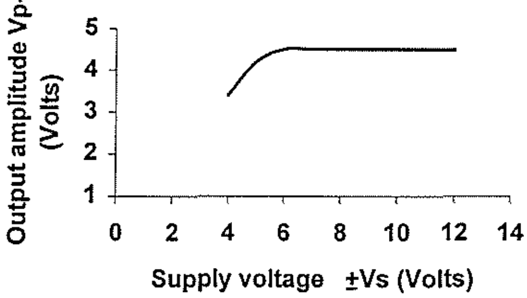


Fig. 5.1 Oscillator output as a function of supply voltage

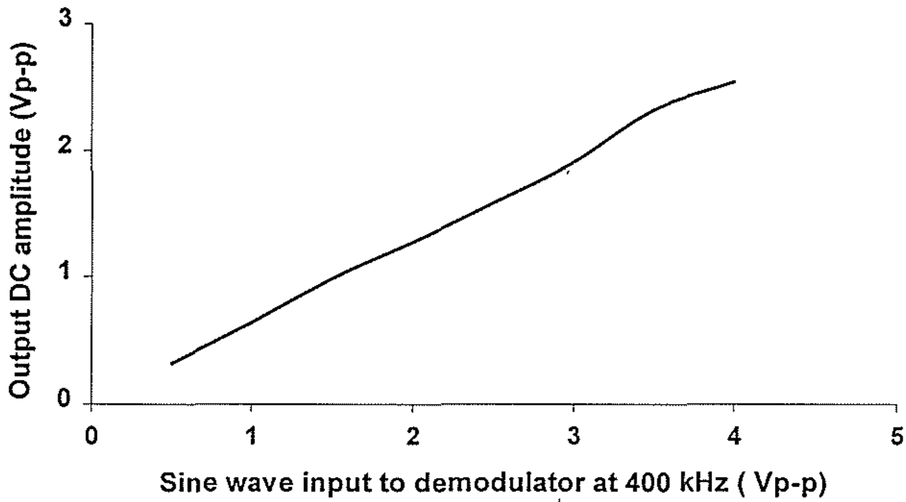


Fig. 5.2 Input output characteristic of the demodulator circuit

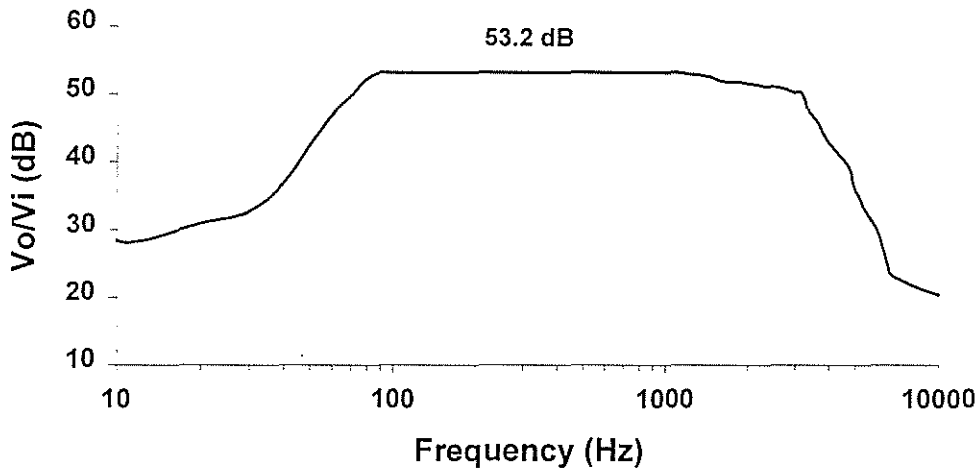


Fig. 5.3 Magnitude response of amplifier circuit in Fig. 3.5

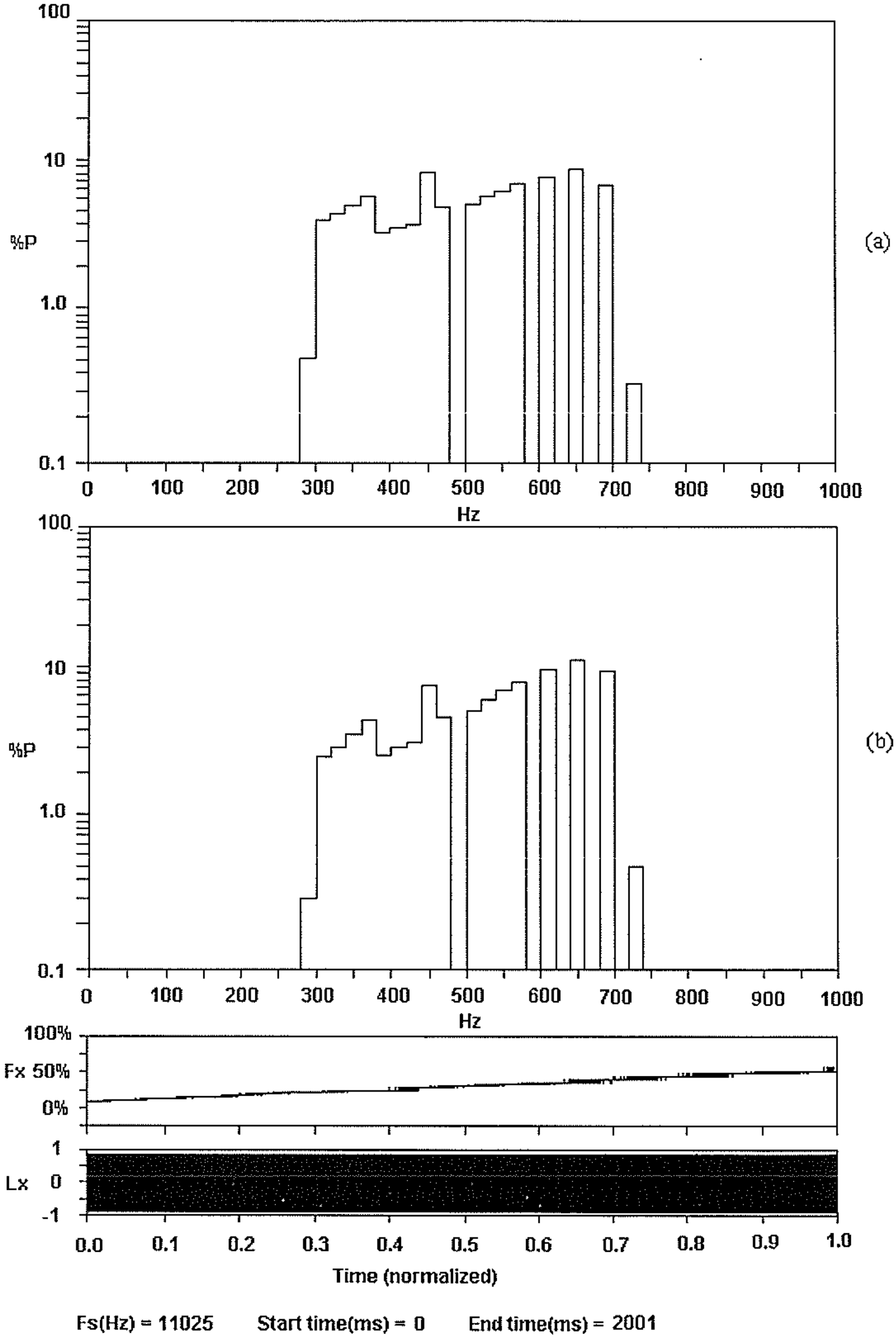


Fig. 5.4 Plot of pitch F_x and single period F_x histogram (a) related to time and (b) related to frequency, for sine-wave swept linearly from 300 Hz to 700 Hz over 2 seconds

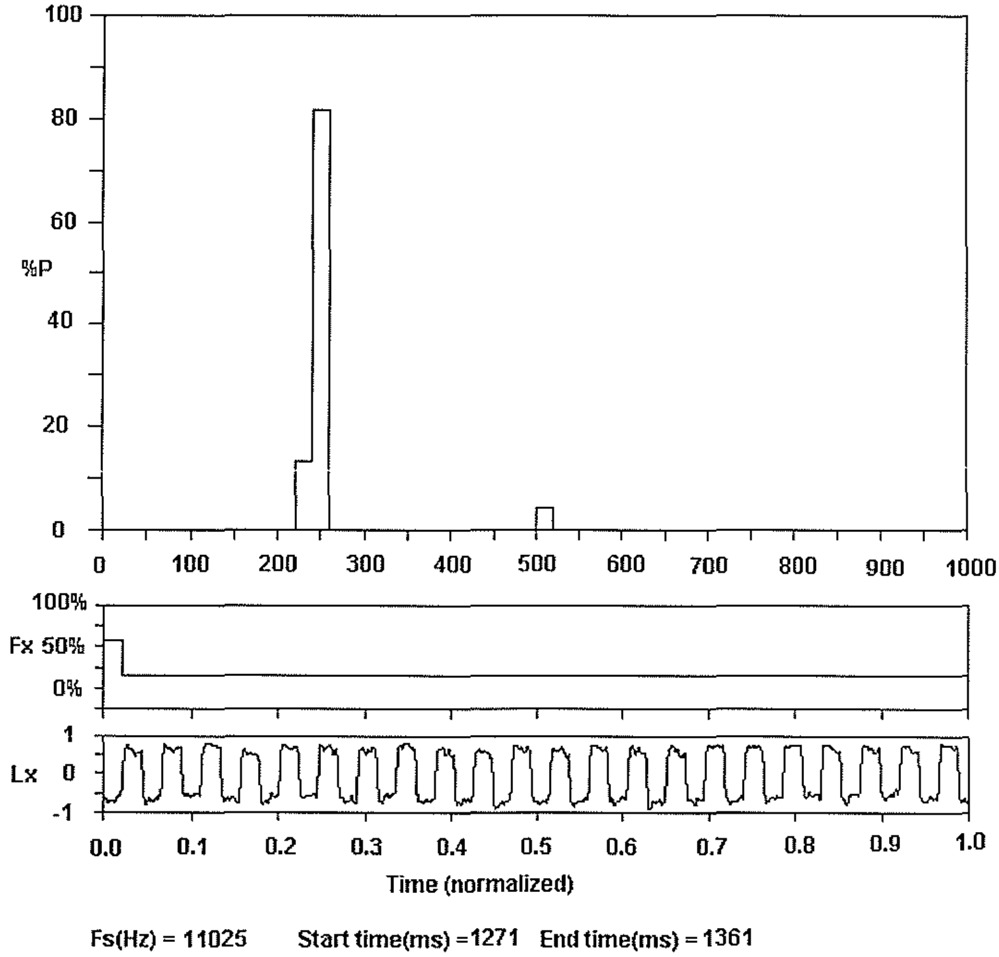


Fig. 5.5 Plot of pitch F_x and single period F_x histogram (related to frequency) for square-wave generated from glottal impedance simulator for 2 seconds

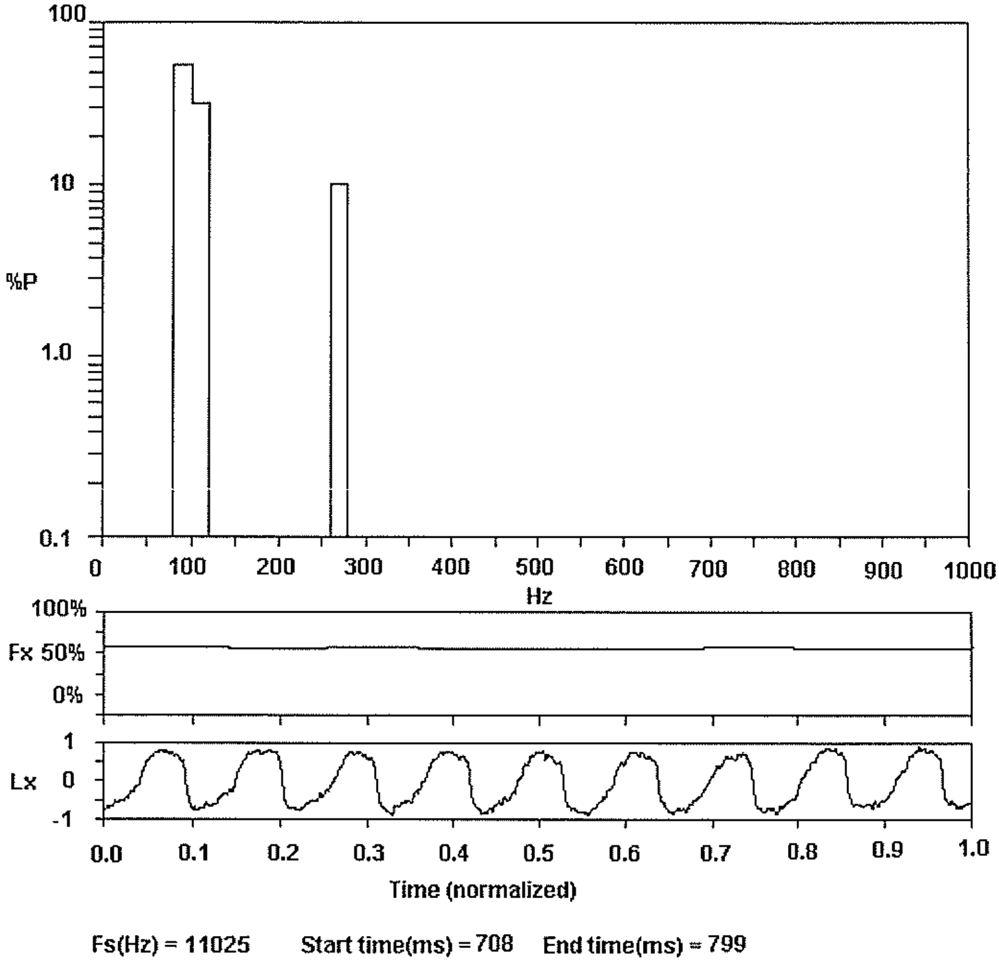


Fig. 5.6a Plot of L_x waveform, pitch F_x, and single period F_x histogram (related to frequency) of male speaker AVP for sustained vowel /a/

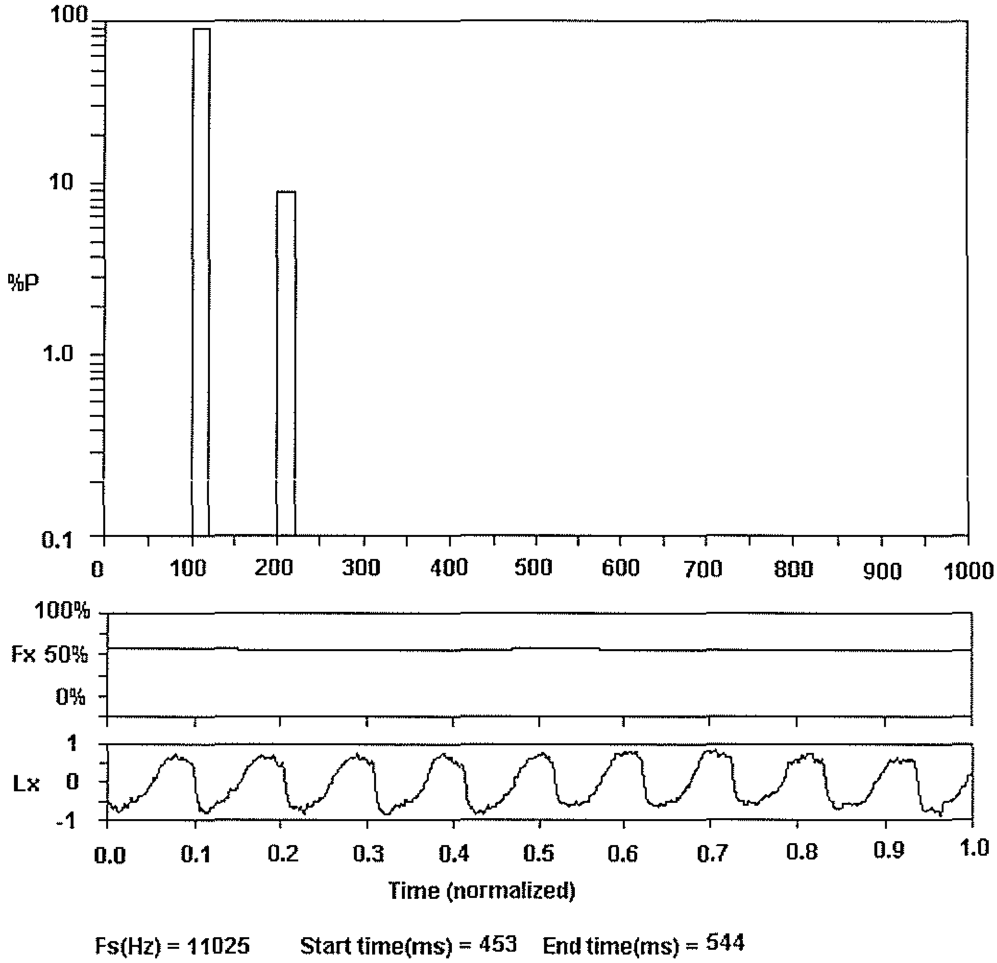


Fig. 5.6b Plot of L_x waveform, pitch F_x, and single period F_x histogram (related to frequency) of male speaker AVP for sustained vowel /*i*/

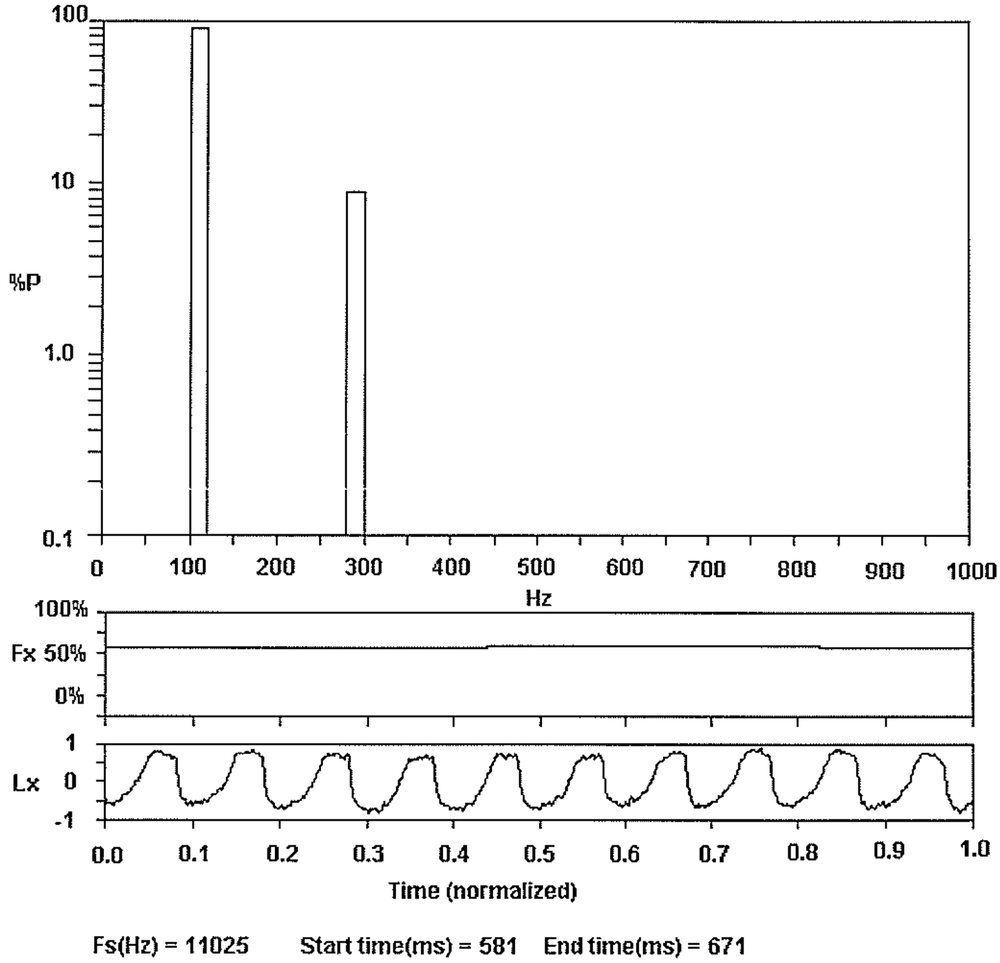


Fig. 5.6c Plot of L_x waveform, pitch F_x, and single period F_x histogram (related to frequency) of male speaker AVP for sustained vowel /*u*/

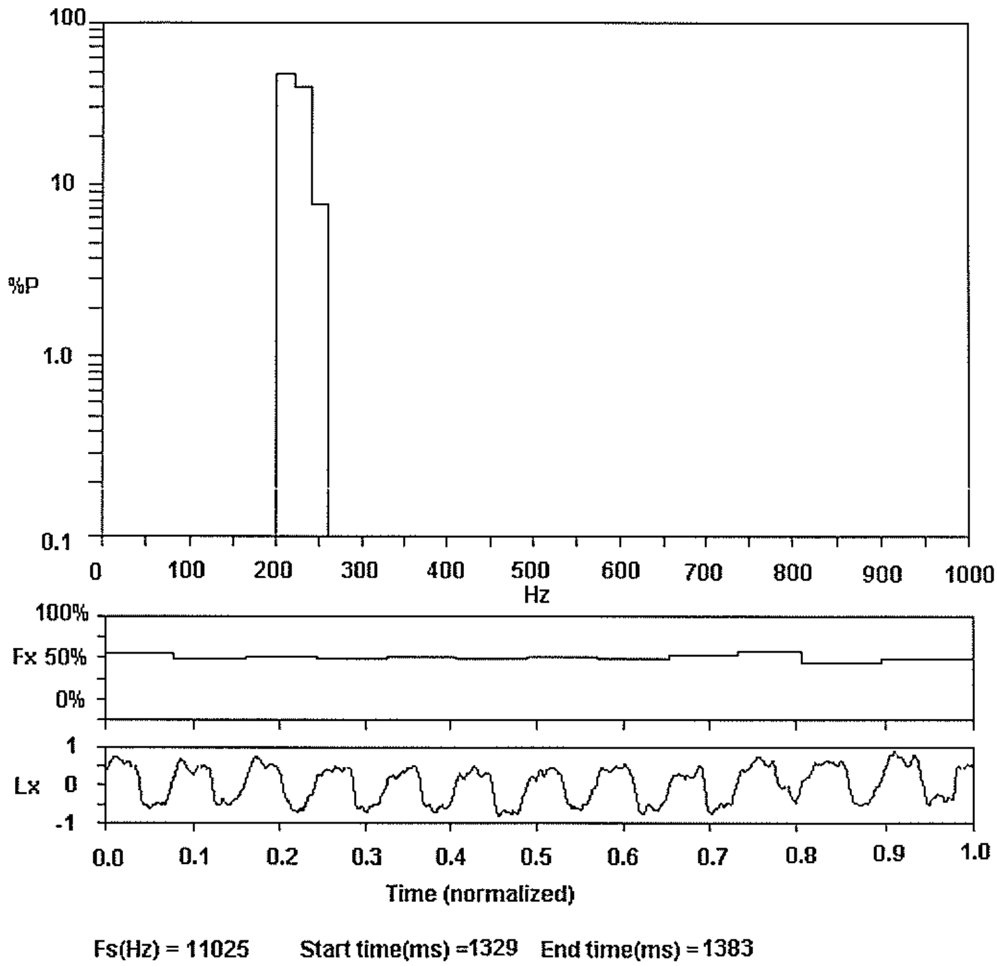


Fig. 5.7a Plot of L_x waveform, pitch F_x, and single period F_x histogram (related to frequency) of female speaker DSJ for sustained vowel */a/*

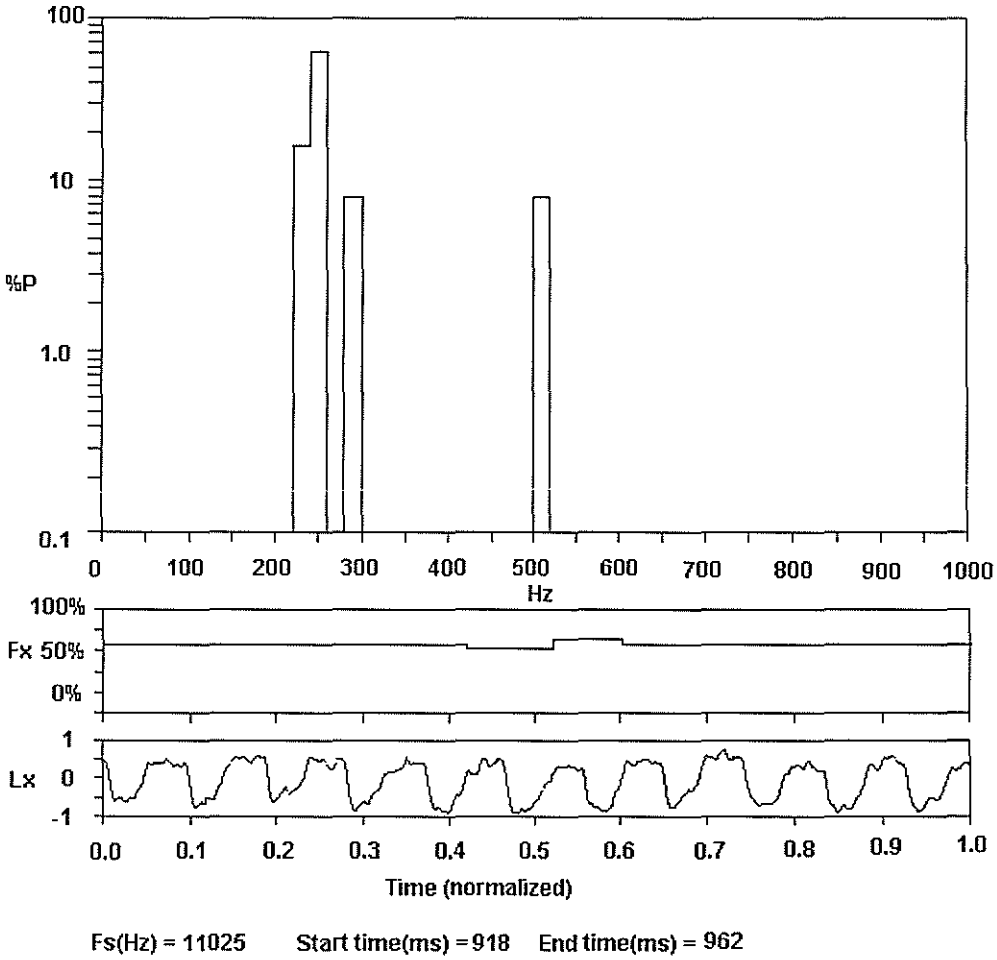


Fig. 5.7b Plot of L_x waveform, pitch F_x, and single period F_x histogram (related to frequency) of female speaker DSJ for sustained vowel /*i*/

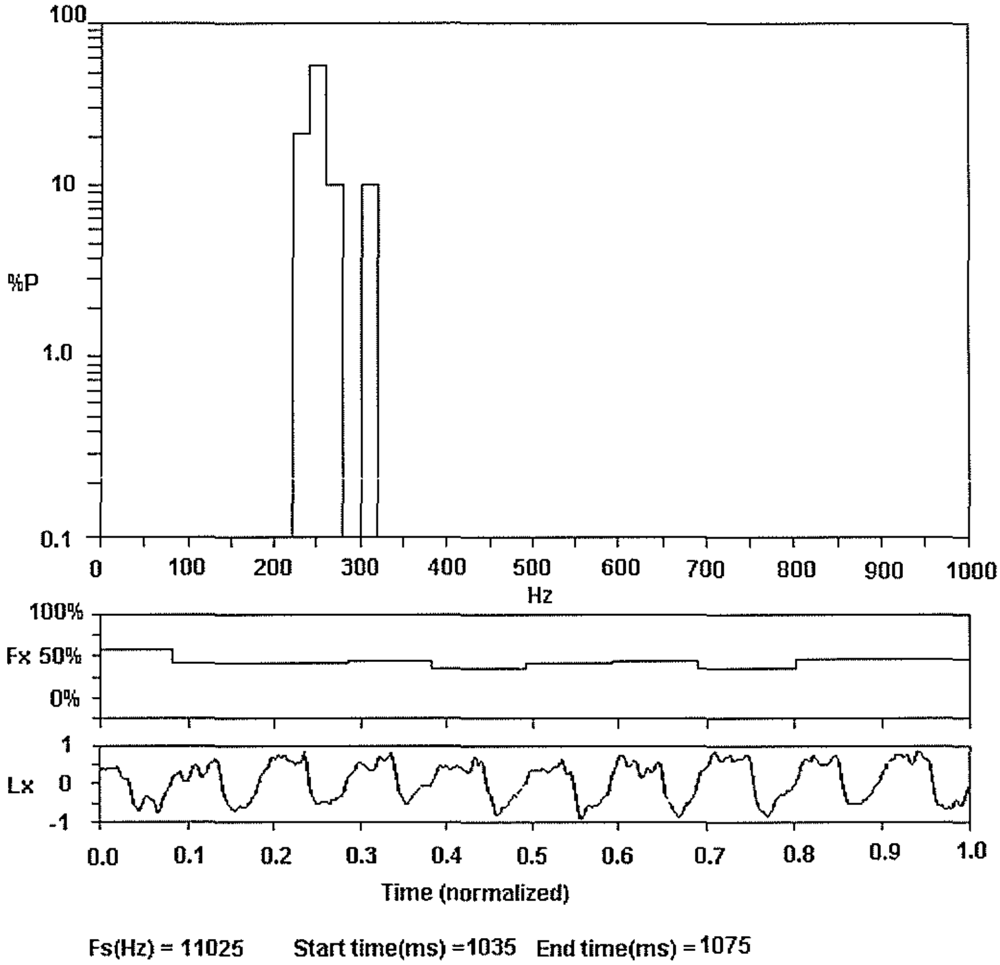


Fig. 5.7c Plot of L_x waveform, pitch F_x, and single period F_x histogram (related to frequency) of female speaker DSJ for sustained vowel */u/*

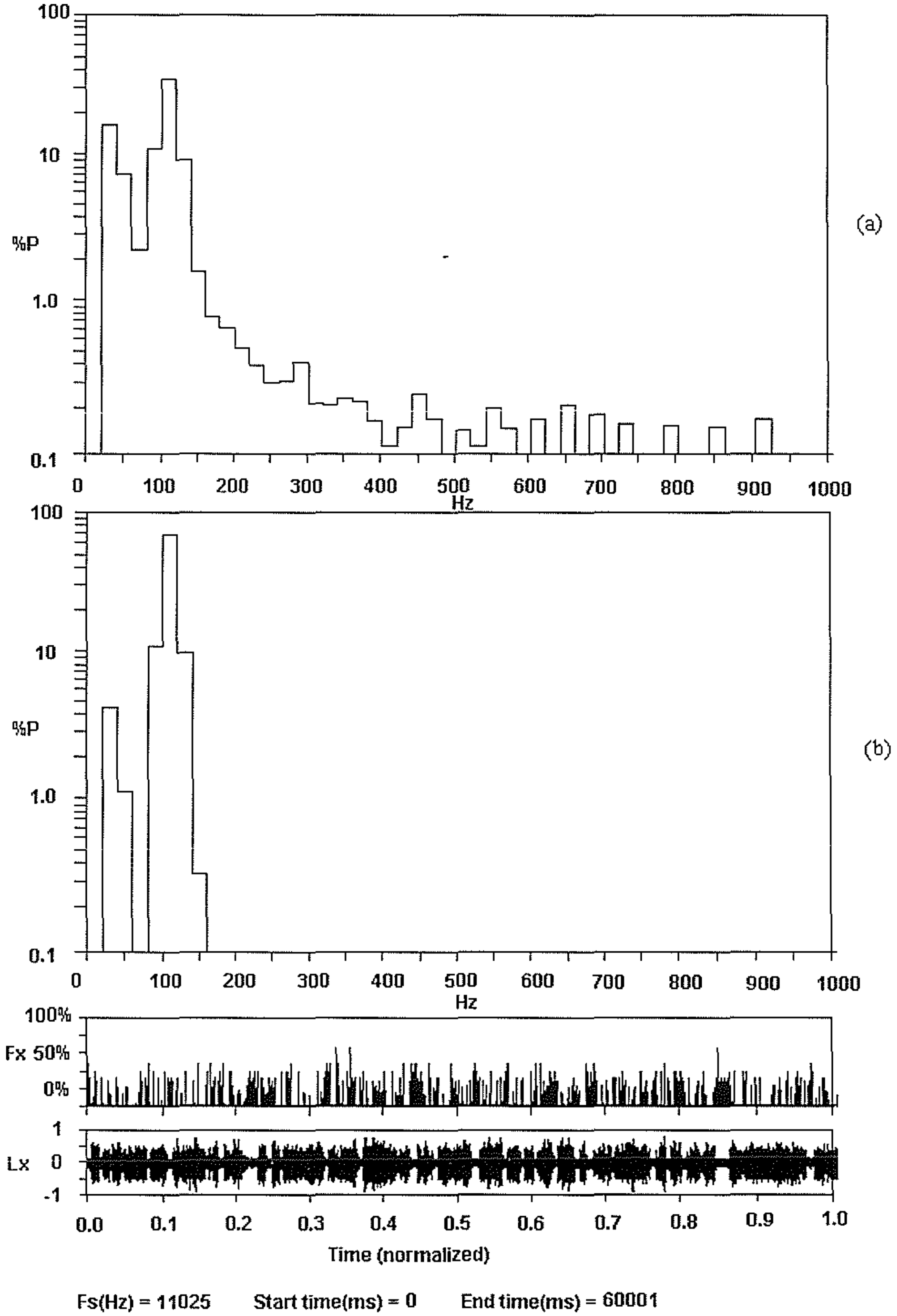


Fig. 5.8 Plot of L_x waveform, pitch F_x, and (a) single period and (b) triple period, F_x histogram (related to time) of male speaker AVP reading continuous text for 1 minute

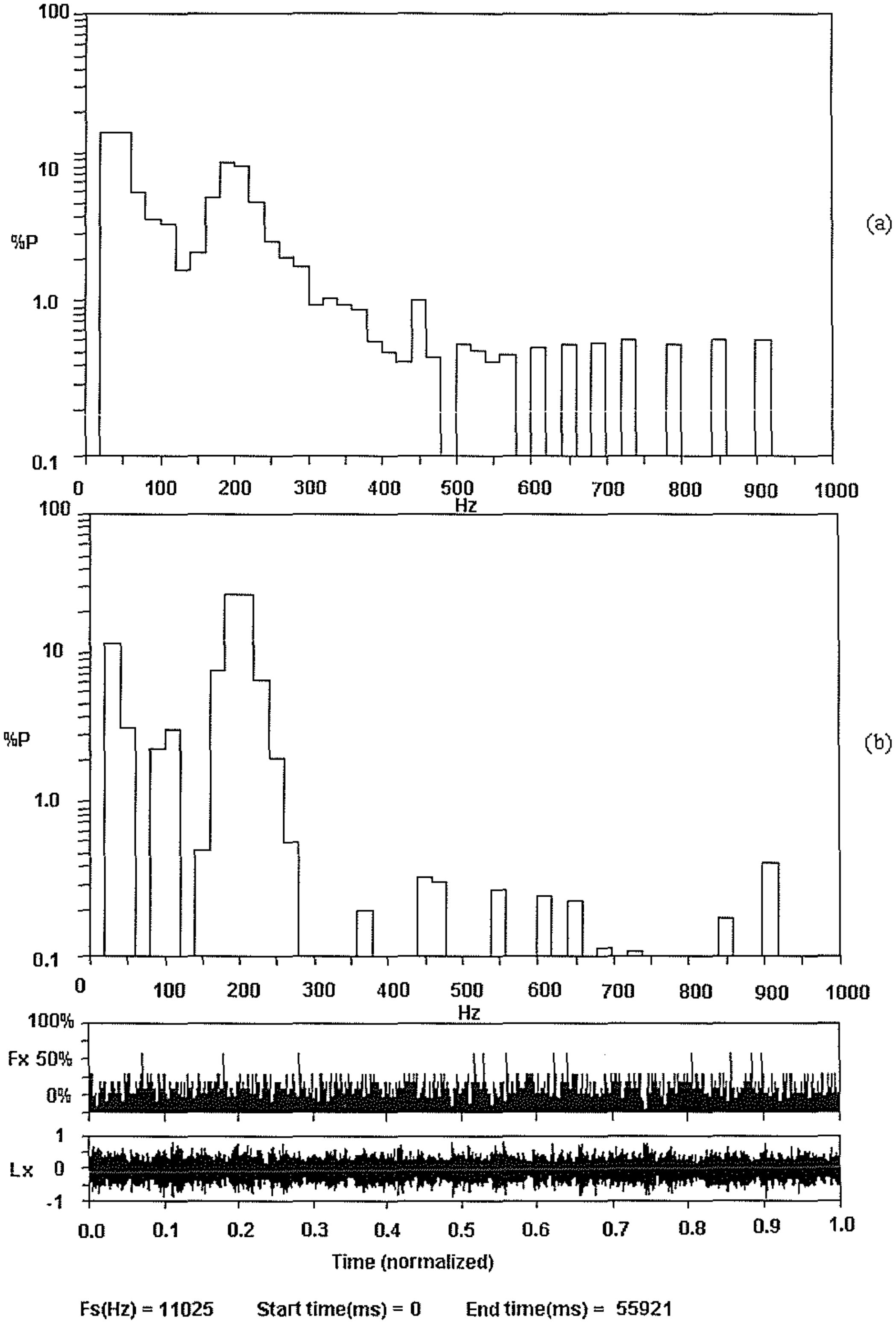


Fig. 5.9 Plot of L_x waveform, pitch F_x, and (a) single period and (b) triple period, F_x histogram (related to time) of female speaker DSJ reading continuous text for 56 sec.

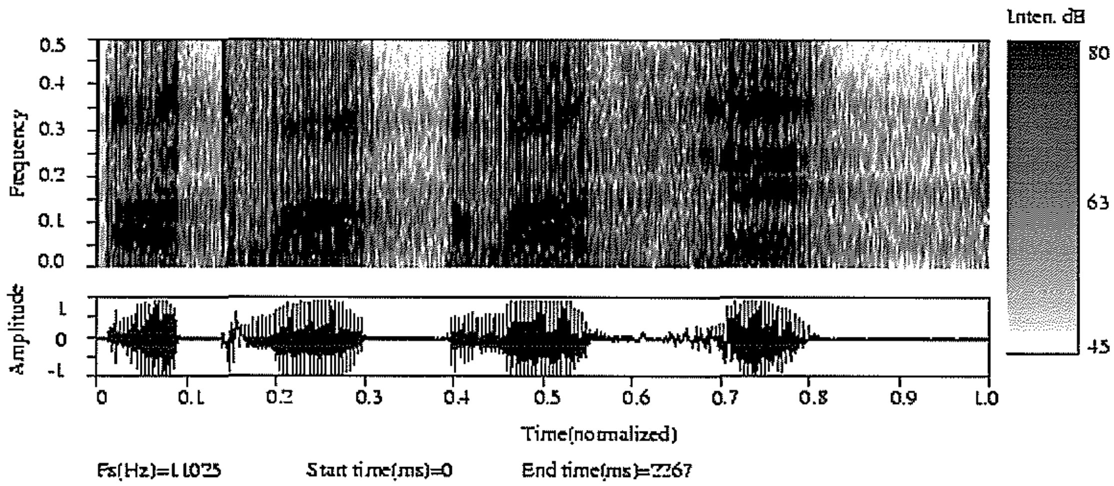


Fig. 5.10a Wideband spectrogram and the speech waveform $s(t)$ for the utterance " *atma amar hai...*"

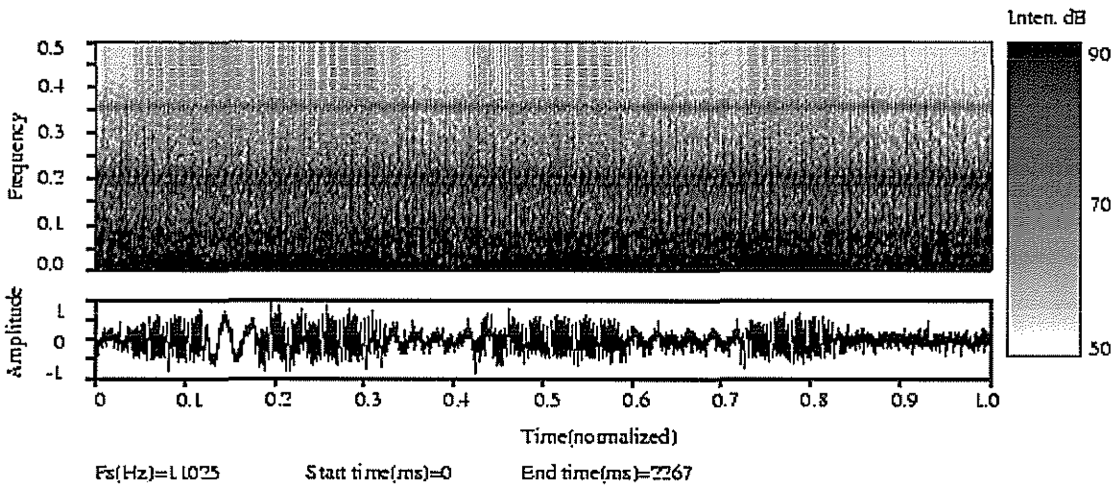


Fig. 5.10b Wideband spectrogram and the L_x waveform for the utterance "*...atma amar hai...*"

Chapter 6
SUMMARY AND CONCLUSIONS

6.1 Work done

The aim of this project was to

(1) develop an electroglottograph, (EGG) or glottal impedance sensor with low level of excitation current, and low power requirement.

(2) study the various electrode configurations and driving/sensing circuit for reducing the distortion of the L_x waveform, and decreasing the noise pick-up.

(3) develop glottal impedance simulator for testing the sensitivity and response characteristics of the impedance sensor.

(4) acquisition, analysis, and display of the L_x waveform using a PC based sound/multimedia card.

The sensitivity of the impedance sensor has been increased, and the injected current is in the range of 0.1 mA to 1 mA. The various electrode and driving/sensing configurations have been tested. Maximum sensitivity is obtained with driven guard rings. The noise pick-up by the circuit has been reduced. The PCB for the glottal impedance sensor and the glottal impedance simulator has been designed and packaged. We have thus two instruments : (i) glottal impedance sensor and (ii) glottal impedance simulator.

The glottal impedance sensor has been interfaced to the PC based sound/multimedia card for acquisition of L_x waveform, and display L_x waveform along with pitch F_x, and the F_x histogram, and spectrographic analysis. For this a program in "C" has been written. This software acquires the signal from the glottal impedance sensor and plots the L_x waveform, pitch F_x, and the F_x histogram.

The instruments (impedance sensor, impedance simulator both) along with the analysis and display set-up have been extensively tested. The glottal impedance sensor has been tested using the simulator. Recording of the waveform and analysis of these waveform have been done for male and female speakers.

6.2 Suggestions for future work

It may be possible to further increase the sensitivity of the instrument, by (i) using a balanced current source that would reduce common mode pick-up, (ii) using an active drive to an indifferent electrode, (iii) increasing the amplifier gain by using low noise circuit components and improved PCB layout.

The bandwidth of the filters used in the amplifier for L_x waveform can be increased, with a sharper roll-off by using a higher order filter design.

There is a scope for improvement in the signal processing for obtaining F_x waveform.

Finally, the instrument has to be used with a number of subjects, with normal larynx, and those having laryngeal disorder of different types.

APPENDIX A

SPECIFICATIONS OF THE INSTRUMENTS

A1. Specifications of glottal impedance sensor

The battery operated glottal impedance sensor has following specifications.

Electrodes	: Neck electrodes
Electrode area	: Circular gold plated glass epoxy PCB, 30 mm in diameter
Electrode distance	: Adjustable
Neck band	: Velcro strap
Electrode voltage	: below 1 V
Electrode current	: 1 mA
Carrier frequency	: 400 kHz
Bandwidth for impedance detection	: 50 Hz to 2.4 kHz
Impedance variation sensitivity	: 1 V/Ω over basal impedance of 591.10 Ω
Power source	: Two Nickel-Hydride batteries. 9 V/ 120 mAh
Current drain	: 44 mA
Use with one charge	: 40 minutes of continuous use
Dimensions	: 200(L)×145(W)×37.5(H) mm

A2 Specifications of glottal impedance simulator

On resistance	: 589.83 Ω
Off resistance	: 590.73 Ω
Resistance change	: 0.9 Ω
Frequency	: 100 –650 Hz
Power source	: One Nickel-Hydride batteries. 9 V/ 120 mAh
Current drain	: 18 mA
Dimensions	:148(L)×100(W)×40(H) mm

APPENDIX B

SOUND CARD SPECIFICATIONS

The project has utilised a sound card manufactured by Creative Technology Ltd for recording and playback of speech / glottographic signal. Fig. B.1 shows a schematic of the sound card connections. The specifications of the sound card are given below:

1] *System requirements*	: a) Intel Pentium or AMD –K5 90 MHz computer with VGA or SVGA card b) 4 MB RAM c) 7.5 MB of free hard disk space d) Windows 95 or Windows 3.1 with MS-DOS 5.0 and a Plug and Play (PnP) configuration manager.
2] *Analog input port characteristics* (established by measurements)	Analog input ia a sum obtained through an internal mixer from two analog inputs: : a) Microphone (Mic) powered by 2.5 V supply from sound card b) Line-in with voltage range of 0-9V and input impedance of 53 kΩ at 1 kHz. Bandwidth = 10 Hz to $f_s/2$ Hz
3] *Output*	: a) Speaker Out/ Line Out connects non-powered speaker by default with output impedance of 0.767 Ω at 1 kHz with 4 watts per channel for 4 Ω. stereo output. Also connects powered speaker and an external amplifier when built-in amplifier is disabled by changing jumper settings on the sound card.

4] *Digitization* : a) For analog input :

Sampling rate = 11025, 22050, or 44100 Sa/sec

No. of quantization bits = 8 or 16.

b) For analog output :

Sampling rate = settable to a value in the range of 5000 Hz to 44100 Hz.

No. of quantization bits = at 8 or 16

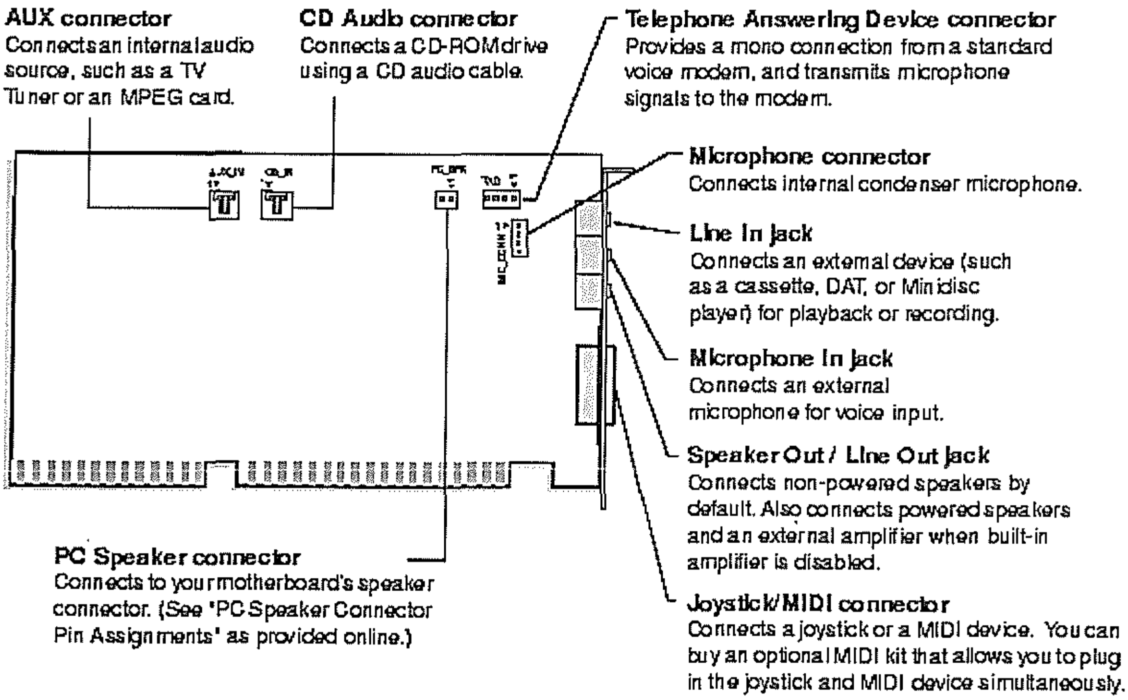
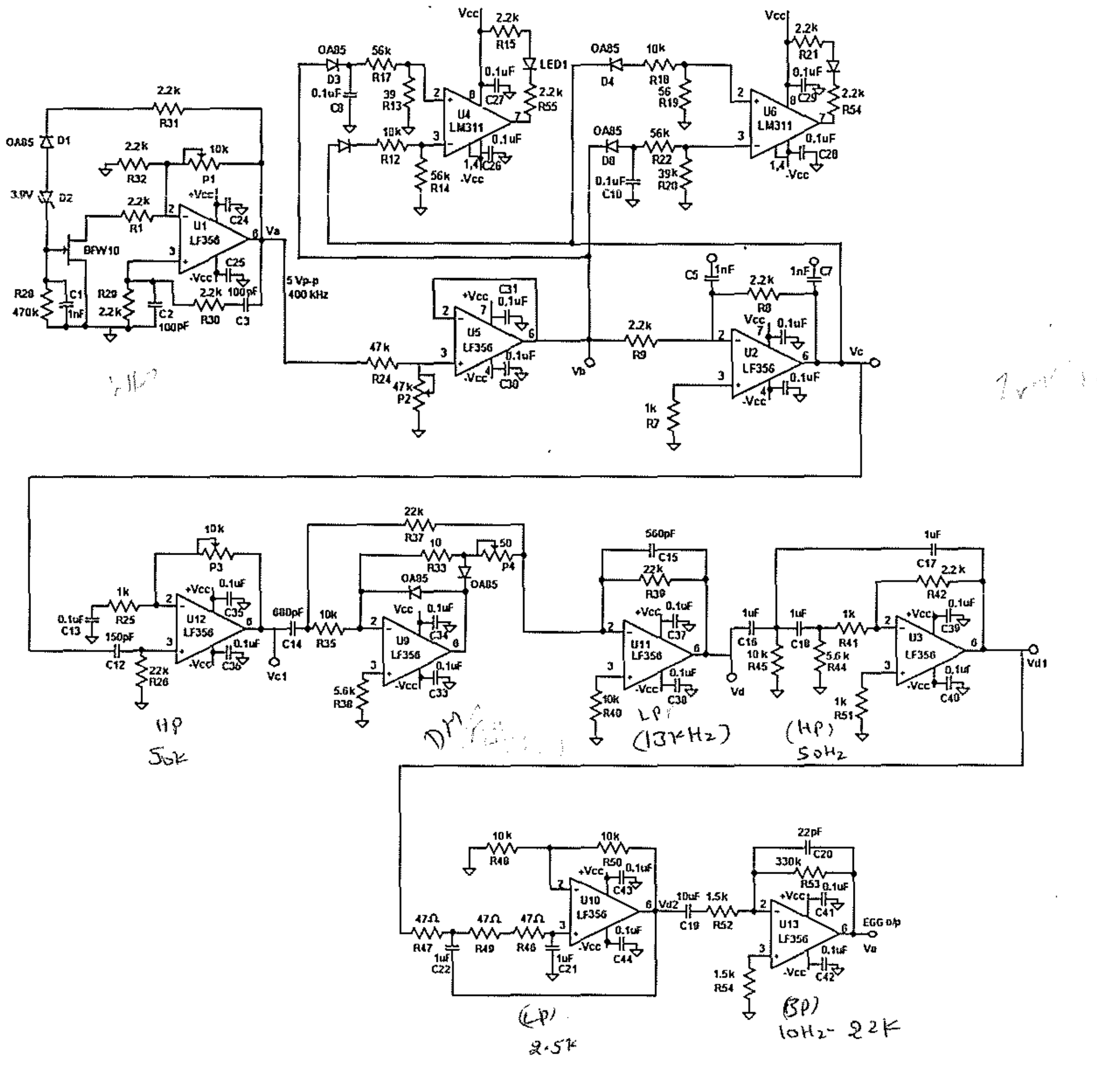


Fig. B.1 Jacks and connections of the sound card. Source : [27]

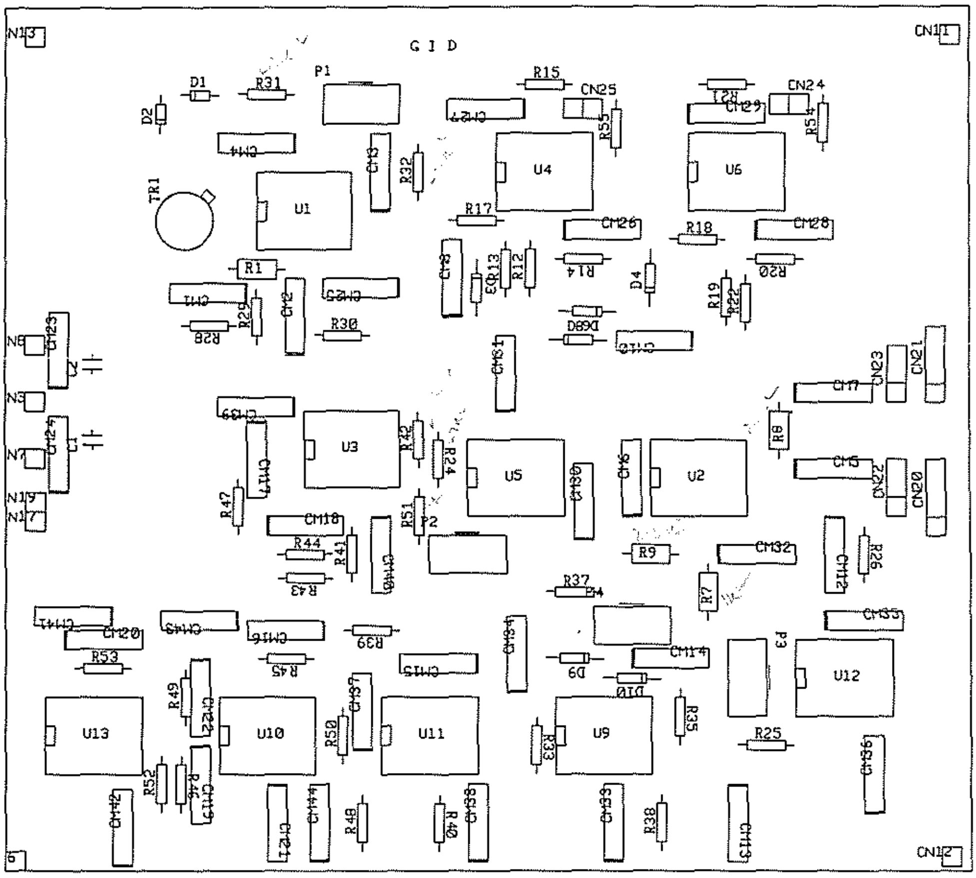
APPENDIX C

CIRCUIT DIAGRAMS AND PCB LAYOUTS

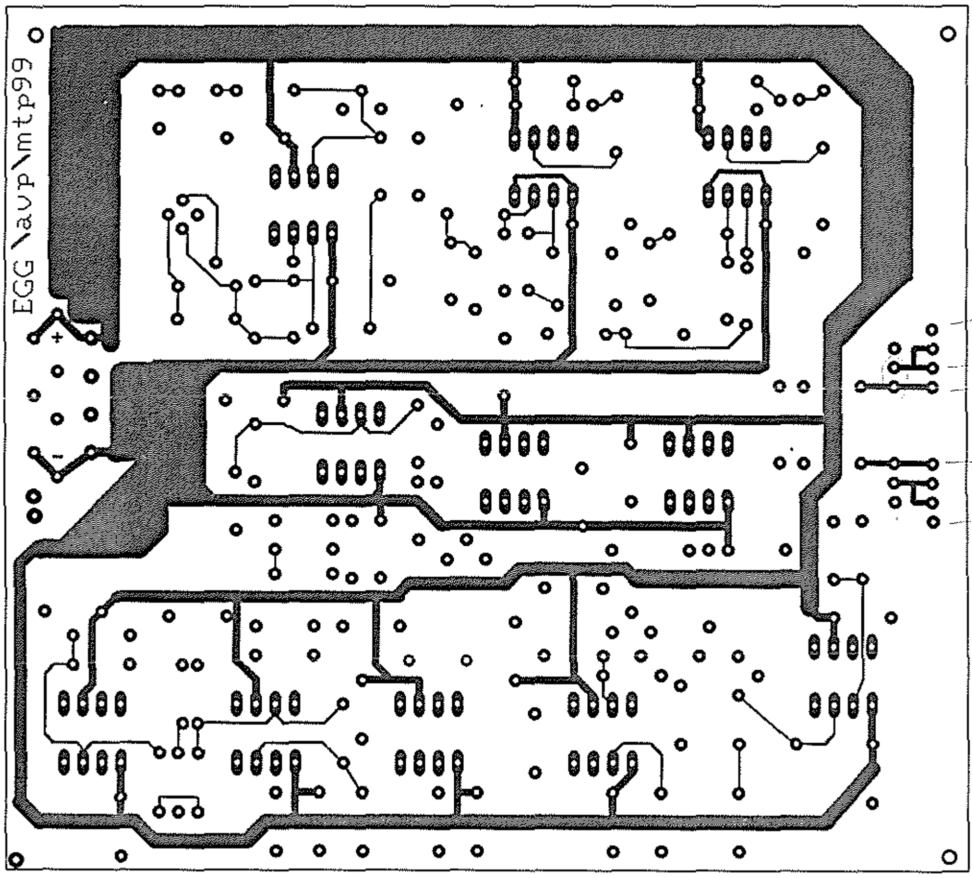
C1 Circuit diagram of glottal impedance sensor



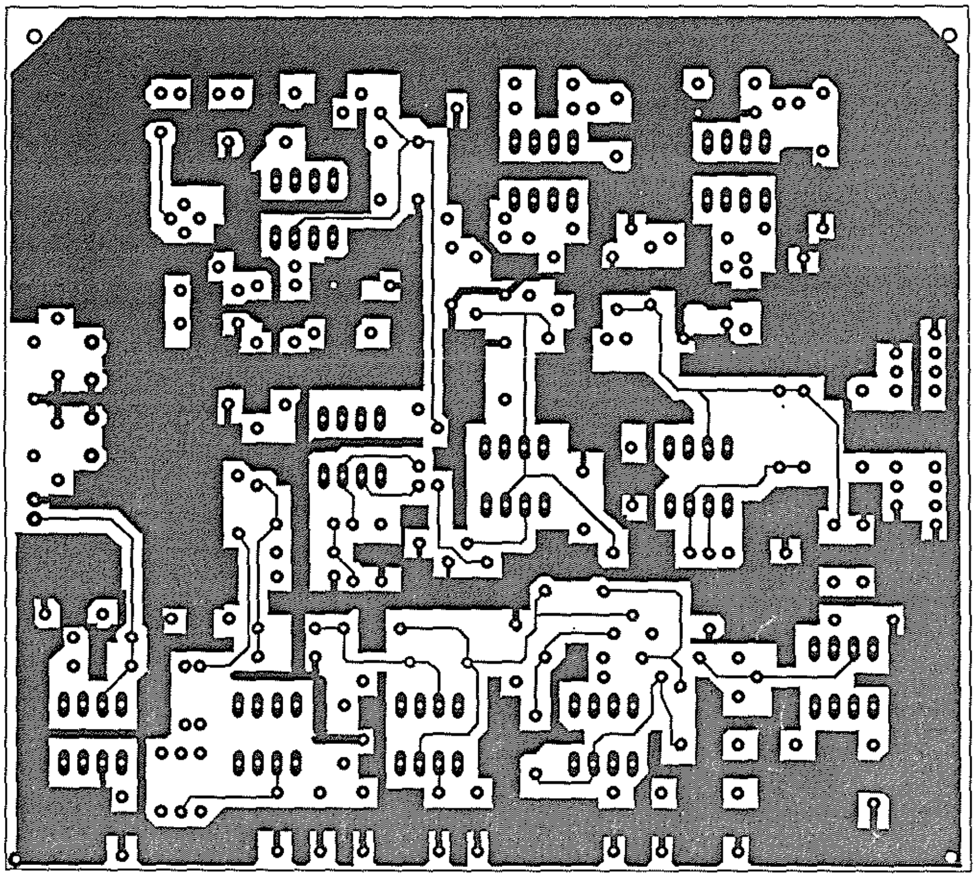
C2 Component placement PCB layout of glottal impedance sensor



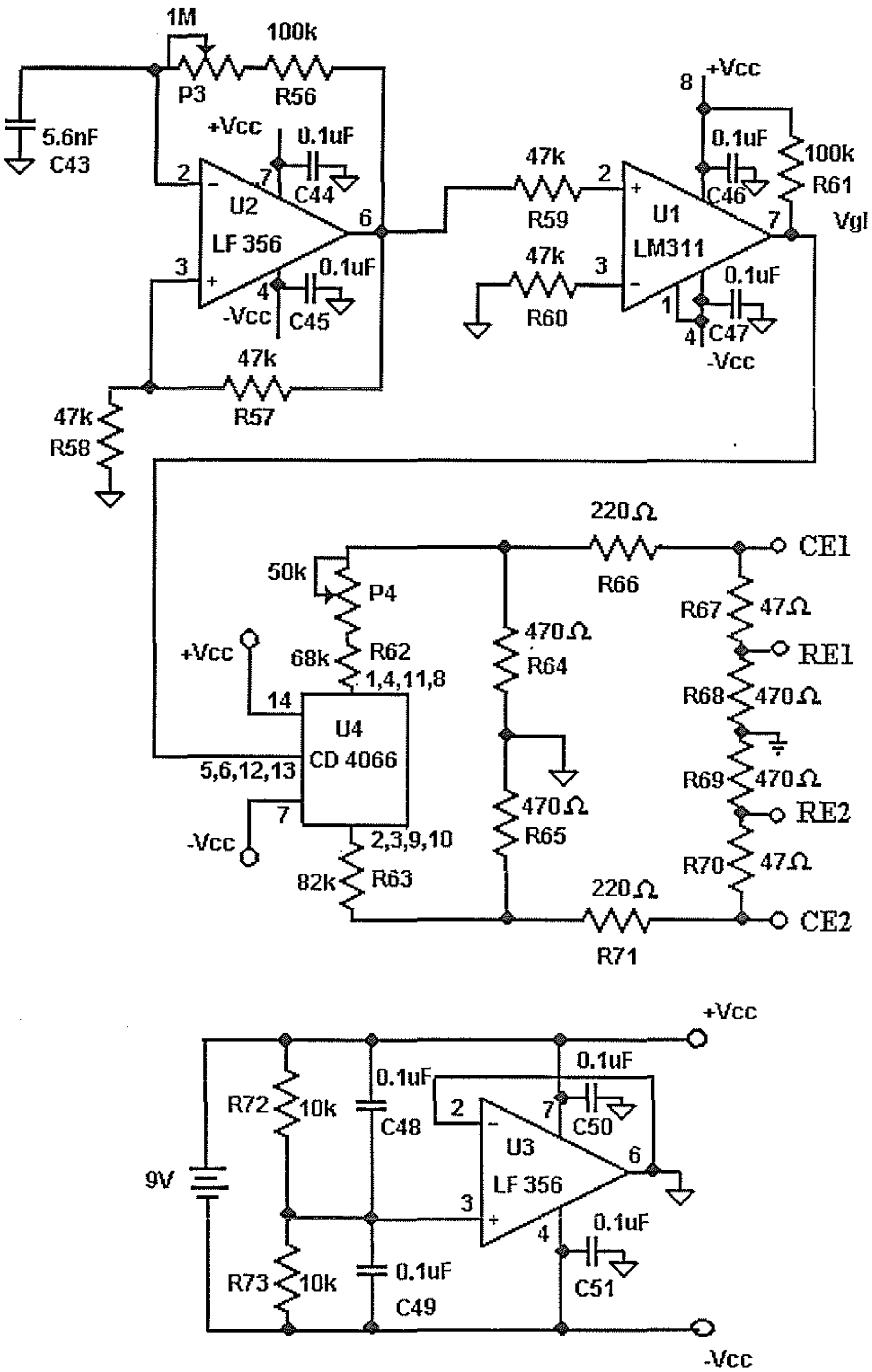
C3 Component side PCB layout of glottal impedance sensor



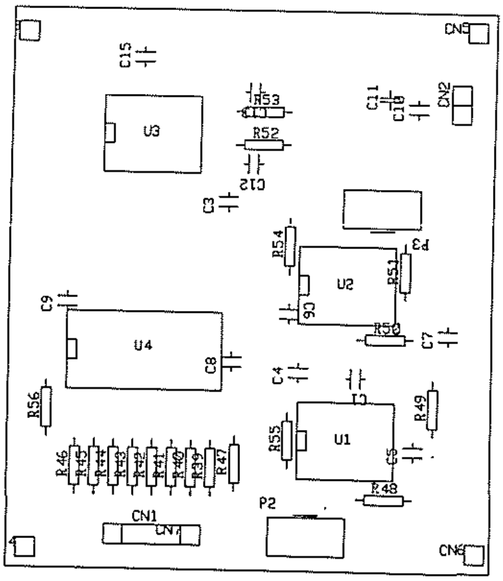
C4 Solder side PCB layout of glottal impedance sensor



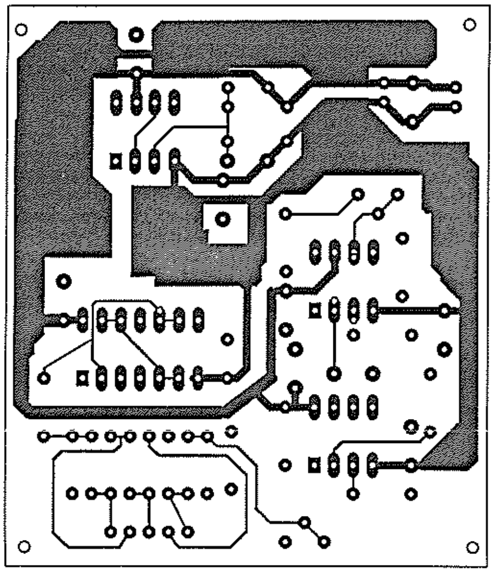
C5 Circuit diagram of glottal impedance simulator



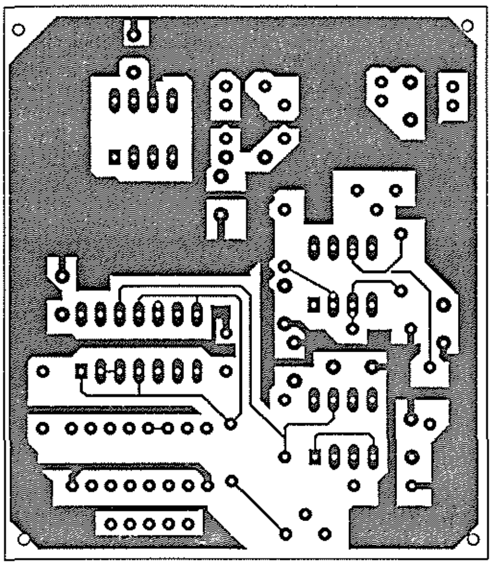
C6 Component placement PCB layout of glottal impedance simulator



C7 Component side PCB layout of glottal impedance simulator



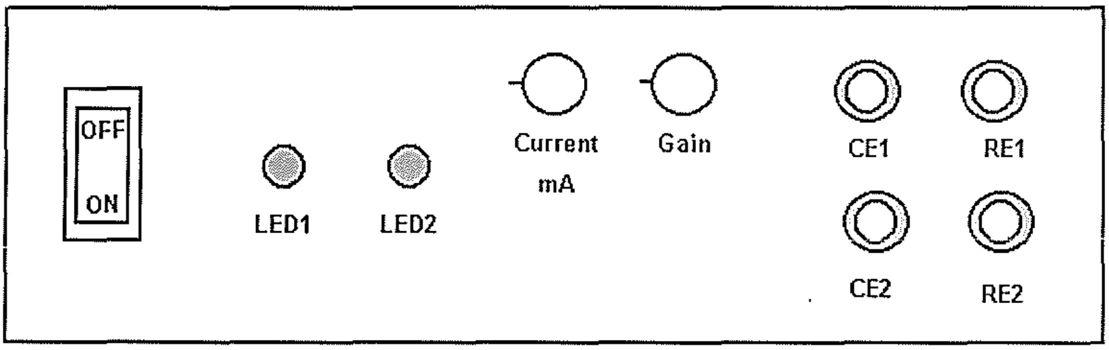
C8 Solder side PCB layout of glottal impedance simulator



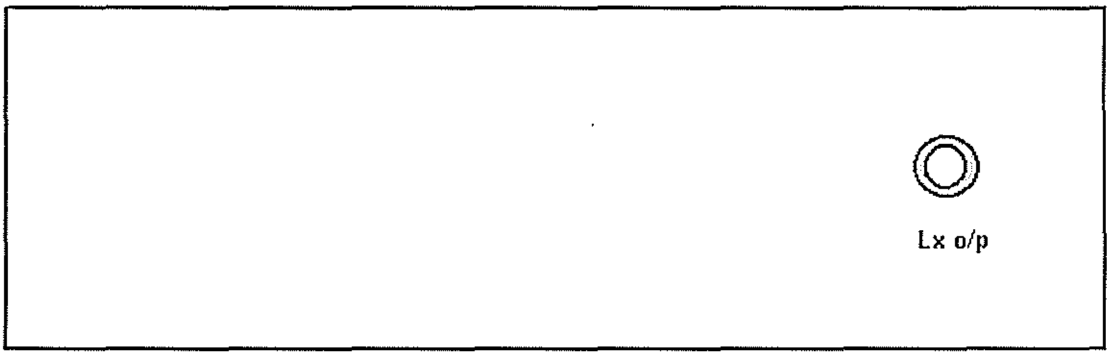
APPENDIX D

PANEL DESIGN AND CABINET DIMENSIONS

D1. Panel for glottal impedance sensor and glottal impedance simulator

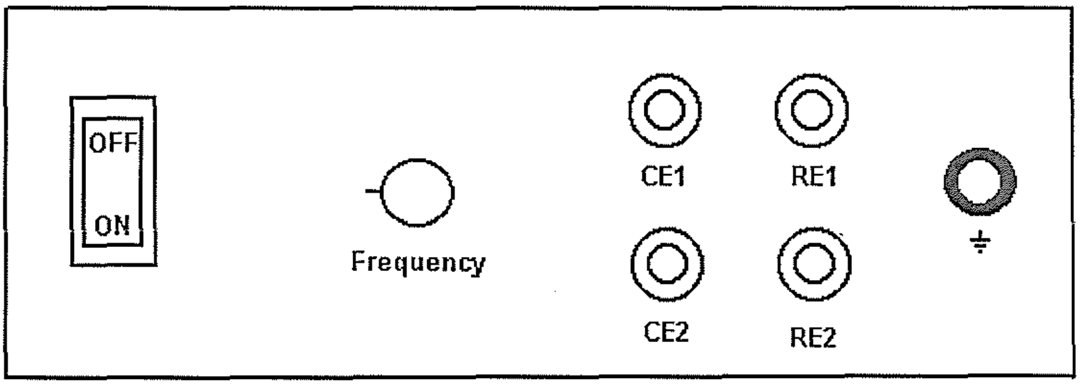


FRONT PANEL



BACK PANEL

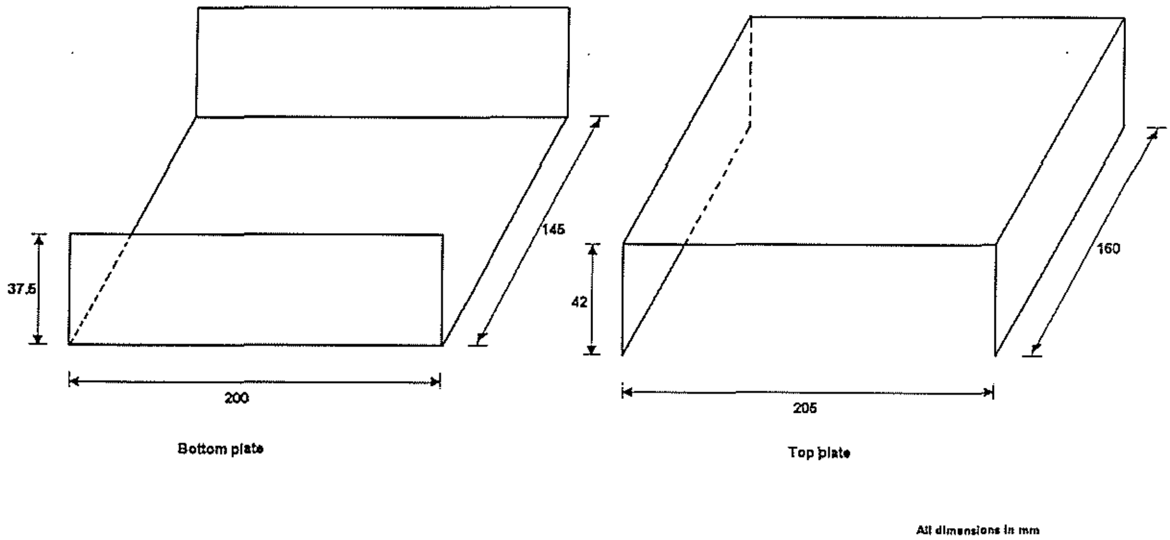
Front and back panel of glottal impedance sensor



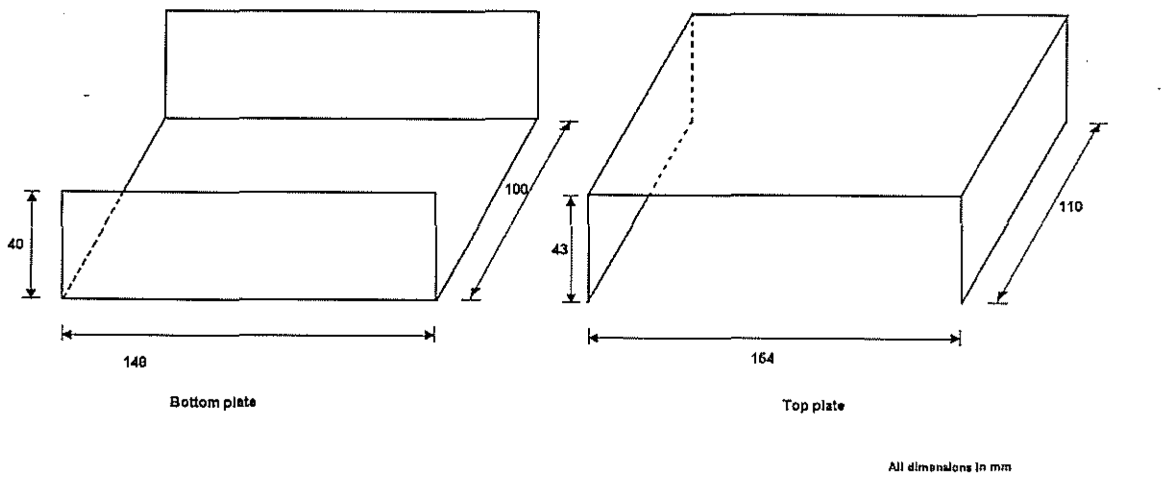
FRONT PANEL

Front panel of glottal impedance simulator

D2. Dimensions of cabinet for glottal impedance sensor and glottal impedance simulator



Cabinet for glottal impedance sensor



Cabinet for glottal impedance simulator

APPENDIX E

COST ESTIMATE

E1. Component list for the glottal impedance sensor

Reference Designator	Part Number/ Value	Part Description	Approximate Price per part (Rs.)
C1, C5, C7	1 nF	Capacitor (ceramic)	1.00
C2, C3	100 pF	Capacitor (ceramic)	1.00
C12	150 pF	Capacitor (ceramic)	1.00
C15	560 pF	Capacitor (ceramic)	1.00
C14	680 pF	Capacitor (ceramic)	1.00
C4, C13, C25, C26, C27, C28, C29, C30, C31,C32, C33, C34, C35, C36, C37, C38, C39,	0.1 uF	Capacitor (ceramic)	1.00
C40, C41, C42	1 uF	Capacitor (electrolytic)	3.00
C16, C17, C18, C21, C22, C19	10 uF	Capacitor (electrolytic)	3.00
R46, R47, R49	47 Ω	Resistor (MFR)	0.75
R12, R18, R25, R41	1 k	Resistor (MFR)	0.75
R52	1.5 k	Resistor (MFR)	0.75
R13, R20	3.9 k	Resistor (MFR)	0.75
R1, R8, R9,R15, R25, R29, R30, R54, R55, R31, R32	2.2 k	Resistor (MFR)	0.75
R19, R17, R22	5.6 k	Resistor (MFR)	0.75

R33, R35, R48, R50 P3 R26, R37,R39 R24 P2, P4 R42 R53	10 k 10 k 22 k 47 k 50 k 220 k 330 k	Resistor (MFR) Pot Resistor (MFR) Resistor (MFR) Pot Resistor (MFR) Resistor (MFR)	0.75 15.00 0.75 0.75 15.00 0.75 0.75
D1, D3, D4, D6, D8 D2	OA85 3.9V	Diode Zener Diode	5.00 2.00
U1, U2, U3, U5, U9, U10, U11, U12, U13 U4, U6	LF356 LM311	Op-Amp Comparator	15.00 15.00
Q1	BFW11	FET	10.00
B1	9V/120mAh	Nickel-Hydride battery	300.00
		PCB	650.00
		Connectors	20.00
		Cabinet	150.00
		Total=	1431.50

E2. Component list for the glottal impedance simulator

Reference Designator	Part Number/ Value	Part Description	Approximate Price per Part (Rs.)
C44, C45, C43, C46, C47, C48, C49, C50, C51	0.1 uF	Capacitor (ceramic)	1.00
R67, R70	47 Ω	Resistor (MFR)	0.75
R66, R71	220 Ω	Resistor (MFR)	0.75
R64, R65, R68, R69	470 Ω	Resistor (MFR)	0.75
R73, R72	10 k	Resistor (MFR)	0.75
R59, R60, R58	47 k	Resistor (MFR)	0.75
P4	50 k	Pot	15.00
R62	68 k	Resistor (MFR)	0.75
R63	82 k	Resistor (MFR)	0.75
R61, R56	100 k	Resistor (MFR)	0.75
P3	100 k	Pot	15.00
U1, U2, U3	LF 356	Op-amp	15.00
U4	CD 4066	CMOS BiQuad Switch	40.00
B2		Ni-Cd Battery	300.00
		PCB	150.00
		Connectors	40.00
		Cabinet	150.00
		Total =	776.00

REFERENCES

1. A. M. Smith and D. G. Childers, "Laryngeal evaluation using features from speech and the electroglottograph", *IEEE Transactions on Biomed. Engg.*, vol. BME-30, pp. 755-759, Nov. 1983.
2. A. C. Guyton, *Textbook of Medical Physiology*, Philadelphia, USA: W. B. Saunders, pp. 478, 1986.
3. E. J. Kniel and M. H. L. Hecker, "Description of the speech of patients with cancer of the vocal folds, Part I: measure of fundamental frequency", *J. Acoust. Soc. Am.*, vol. 49, pp. 1275-1282, 1971.
4. E. J. Kniel and M. H. L. Hecker, "Description of the speech of patients with cancer of the vocal folds, Part II: judgements of age and voice quality", *J. Acoust. Soc. Am.*, vol. 49, No. 4, pp. 1283-1287, 1971.
5. T. Haji, S. Horiguchi, T. Baer, and W. J. Gould, "Frequency and amplitude perturbation analysis of electroglottograph during sustained phonation", *J. Acoust. Soc. Am.*, vol. 80, pp. 58-62, Jul. 1986.
6. D. G. Childers, D. M. Hicks, D. M. Moore, and Y. A. Alaska, "A model for vocal fold vibratory motion, contact area and the electrogram", *J. Acoust. Soc. Am.*, vol. 80, pp. 1309-1320, Nov. 1986.
7. A. J. Fourcin, "Laryngographic assessment of phonatory function", *ASHA Reports*, vol. 11, 1981.
8. S. M. Rosen, A. J. Fourcin, and B. C. J. Moore, "Voice pitch as an aid to lip-reading", *Nature*, vol. 291, pp. 150-152, May 1981.
9. R. Bhagwat, "A Glottal pitch extractor", *B. Tech. project report*, Guide: P. C. Pandey, EE Dept., IIT Bombay, 1990.
10. A. Sriram, "A Glottal pitch extractor", *B. Tech. project report*, Guide: P. C. Pandey, EE Dept., IIT Bombay, 1991.
11. M. Chitnis, "A Glottal pitch extractor", *M. Tech. dissertation*, Guide: P. C. Pandey, EE Dept., IIT Bombay, 1998.

12. M. Chitnis, P. C. Pandey, "Extraction of voice pitch by measuring impedance variation across the thyroid cartilage", *Proceedings of National Conference on Biomedical Engineering*, MIT, Manipal, pp. II 10- II 11, April 1998.
13. D. O'Shaughnessy, *Speech Communication, Human and Machine*, New York: Addison-Wesley, 1980.
14. I. R. Titze, "Physiological and acoustical differences between male and female voices", *J. Acoust. Soc. Am.*, vol. 85, pp. 1609-1707, April 1989.
15. Kay Elemetrics Corp., *The Laryngograph*, product brochure, NJ, USA.
16. J. N. Larar and D. G. Childers, "Electroglottography for laryngeal function assessment and speech analysis", *IEEE Transactions on Biomed. Engg.*, vol. BME-31, pp. 807-817, Dec. 1984.
17. H. Kunov, S. M. Abel, P. C. Pandey, and E. B. Friedman, "Analysis of industrial noise, with special reference to correlation between different indices and transmission of noise through the middle ear", Project Report to the Ontorio Ministry of Labor, Institute of Biomed. Engg., Univ. of Toronto, vol. 1, Sep.1983.
18. J. S. Bendat, *Engineering Applications of Correlation and Spectral Analysis*, New York: John Wiley & sons, 1980.
19. T. Baer, A. Lofquist, and N. S. McGarr, "Laryngeal vibrations: A comparison between high-speed filming and glottographic techniques", *J. Acoust. Soc. Am.*, vol. 73, pp. 1304-1308, April 1983.
20. COSEL, *Power Supplies Catlog*, Toyama: Cosel Ver2.1 S, pp. 410-411, 1999.
21. K. R. Botkar, *Integrated Circuits*, Delhi: Khanna, pp.424- 430, 1992.
22. A. S. Sedra, K. C. Smith, *Microelectronics Circuits*, New York: CBS College, pp. 627-629, 1982.
23. L. A. Geddas and L. E. Baker, *Principles of Applied Biomedical Instrumentation*, New York: Wiley, 1975.
24. M. S. Ratanpal, "Speech processing for binaural dichotic presentation", *M. Tech. dissertation*, Guide: P. C. Pandey, EE Dept., IIT Bombay, Jan. 2000.
25. P. Perry, *Multimedia Developer's Guide*, Indiana: Sams, pp.743-796, 1994.
26. *Sound Blaster 16 Manual*, Creative Technologies Ltd., 1997.

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

28. T. G. Thomas, P. C. Pandey, & S. D. Agashe, "A PC-based multiresolution spectrograph", *J. IETE*, 40, pp. 105-108, 1994.

29. Baragi, B. N. Ashok, "A speech training aid for the deaf", *M. Tech. dissertation*, Guide: P. C. Pandey, EE Dept., IIT Bombay, 1996.

30. D. S. Chaudhari, "Dichotic presentation for improving perception by persons with bilateral sensorinueral hearing impairment", *Ph. D. thesis*, Guide: P. C. Pandey, School of Biomedical Engg. , IIT Bombay, 2000.

31. J. G. Proakis, & D. G. Manolakis, "*Digital Signal Processing Principles, algorithms, and applications*", New Delhi: Prentice Hall of India, 1997.

32. Adobe Systems Inc., *Postscript Language Reference Manual*, Reading: Addison-Wesley, 1998.

33. F-J Electronics A/S, *Electroglottograph type EG830,* product brochure, Denmark.

34. National Semiconductor Corp., *Linear Data Book*, California, vol. 2, 1988.