TH-30

A STUDY OF BACKGROUND NOISE IN TRANSCERVICAL ELECTROLARYNX

A dissertation
submitted in partial fulfilment of the
requirements for the degree of

Master of Technology

by

Hiren R. Shah

(97314007)

Guide: **Prof. P. C. Pandey**

School of Biomedical Engineering
Indian Institute of Technology, Bombay

January 1999

Indian Institute of Technology, Bombay

M. Tech. dissertation approval

Dissertation entitled **"A study of background noise in transcervical electrolarynx"** by Hiren R. Shah, is approved for the award of the degree of Master of Technology in Biomedical Engineering.

Guide	:	P C Pandey	(Dr. P. C. Pandey)
Internal Examiner	:	Gadre	(Dr. V. M. Gadre)
External Examiner	:	VKMadan	(Dr. V. K. Madan)
Chairman	:	M S Agarwal	(Dr. M. S. Agarwal)

Date: 25-1-99

Abstract

Hiren R. Shah, "A study of background noise in transcervical electrolarynx", M.Tech. dissertation, School of Biomedical Engineering, Indian Institute of Technology, Bombay, January '99, Guide: Dr. P. C. Pandey.

Transcervical electrolarynges make possible adequate communication for people who are not able to use their larynx. The device is held against the neck, and provides pulsatile vibration to the neck tissue, and the vibration gets coupled to the air in the vocal tract. The resulting speech has low intensity, an unnatural quality and is significantly less intelligible than normal speech. Main source of the degradation in quality is the presence of leakage sound. The objective of this project is to study the leakage background noise and methods for its cancellation.

Spectral analysis of alaryngeal speech and leakage sound has been carried out. An attempt was made to develop a leakage sound cancellation system in which the sound is picked up by a microphone and processed for removal of the direct leakage, and the resulting sound is amplified, and output through a speaker. Results of real time implementation of single input adaptive leakage sound canceller using LMS algorithm were not satisfactory due to non-stationarity of the leakage sound. Three methods to estimate leakage sound were implemented through off-line processing: LMS adapting, ensemble averaging, and inverse Fourier transform of ensemble averaged spectrum. All the three methods were compared. None of them could fully remove leakage sound because of variation in shape of pulses coming from vibrator. Reasons for this variation may be change in the application pressure and transducer dynamics.

Acknowledgments

I would like to express my deep gratitude to my guide, Pror. P. C. Pandey, for his invaluable guidance and personal interest in this project. I sincerly thank Dr. G. K. Bachher of Tata Memorial Hospital for discussions on the problem and also for providing artificial larynges, Pror. S. R. Devasahayam for valuable suggestions during the project work, and Prof. V. M. Gadre for discussion on the problem and for providing a TI/TMS320C50 DSP board from the "TI DSP Elite Lab" of EE Dept. for this project. I also thank Harshal Nemade and Mandar Chitnis for many interesting and helpful discussions. I also thank my lab-mates, Ashish, Prashant, Mr. Chaudhari, Mrs. Jangamashetti, Swati, Abhijit, and Munish for their memorable and enjoyable company. I also take this opportunity to thank Mr. A. D. Apte and Mr. L. D. Muntode of the Instrumentation Lab of EE dept. for their kind help during this project work.

Hiren R. Shah

Contents

List of Figures

List of Tables

Chapter 1

INTRODUCTION

1.1 Problem overview

In normal speech production chain, lungs provide the air stream, the vocal folds in the larynx provides the vibration source for sound, and the oral and nasal cavities provide the spectral shaping of the sound spectrum. Movements of articulators in the oral cavity result in different configuration of the cavity. When the vocal folds become dysfunctional due to disease or injury and/or the surgical removal of the larynx (laryngectomy), the integrity of this pathway is significantly disturbed, and alternative methods for providing a non-laryngeal voice source are required.

A person, who has undergone laryngectomy or who has dysfunctional vocal folds, has three alternatives for communication. The first alternative is nonverbal communication such as writing or gestures. These methods are tiring and are often useful for face-to-face situations only. The second alternative is esophageal speech, a technique for producing speech without using larynx by expelling air from the esophagus instead of trachea and then move this air across the upper tissue of esophagus which creates vibrations. This air then moves up into the oral and nasal cavities resulting in speech sounds. Esophageal speech is preferred on the basis of naturalness, convenience, and economy. It is generally not as loud as normal speech, and takes a lot of time to learn.

The third choice is the use of an artificial larynx. It is very useful in the immediate postoperative period for laryngectomees.

Artificial larynx is a device used to simulate an approximation to normal laryngeal tones. The function of artificial larynx is analogous to vocal folds i.e. they provide vibration in vocal tract which are necessary for sound production. One of the widely used artificial larynx is the transcervical electrolarynx. The stages of speech production by this method start with the production of vibrations, which is mechanically coupled to the neck tissues, through which vibrations are transmitted to the vocal tract. From there, normal speech production process takes over. The various resonances of the vocal tract shape the harmonic spectrum of the vocal tract vibrations, which results in speech. The resulting speech has low intensity, an unnatural quality and is significantly less intelligible than normal speech. Despite the fact that they have been available for a long time, the design of transcervical electrolarynx has remained essentially unchanged and many of the problems associated with these devices remain unsolved [1]. The speaker has to exercise pitch and intensity control using knobs/ switches which takes a lot of practice. The unvoiced speech segments get substituted by voiced segments. The harmonic content of the mechanically coupled vibration is very different from that of the glottal pulses, and therefore intensity relationship between formants becomes unnatural. In addition to these problems, major problem is the presence of background interference due to the leakage of acoustic energy from backside of the vibrator.

1.2 Project Objective

Transcervical electrolarynx is one of the widely used artificial larynx, but background interference in this device deteriorates the speech quality. Main source of background interference is transducer of vibrator. Effective shielding of the directly radiated sound from vibrator should reduce the background noise. But, the use of acoustical shielding yielded only a marginal reduction in the noise [1]. The objectives of this project are to study background interference in transcervical electrolarynx and to find out suitable leakage sound canceller methods to reduce it. Study of speech produced by alaryngeal speaker using transcervical type electrolarynx should be helpful to understand the leakage sound. So, spectral analysis of speech produced by alaryngeal speaker using transcervical electrolarynx is carried out.

The system, in which the sound is picked up by a microphone and processed for removal of the direct leakage, the resulting sound is amplified, and output through a speaker, is developed to remove background noise. Radiated sound is processed by the leakage sound canceller, which is trained to remove background leakage sound. Three methods are implemented and tested for estimation of leakage sound. Those methods are single input adaptive leakage sound cancellation method, ensemble-averaging method, and inverse Fourier transform of ensemble averaged spectrum method.

1.3 Outline of the dissertation

Voice production using transcervical type electrolarynx and problems with transcervical electrolarynx are described in the second chapter. Study of speech produced by alaryngeal speaker using transcervical electrolarynx and an outline of the leakage sound cancellation schemes are given in the third chapter. Real time implemen-tation of single input adaptive leakage canceller is described in fourth chapter. In the fifth chapter, off-line implementation of three leakage sound estimation methods and their testing with results are discussed. Summary and suggestions for further work are covered in the last chapter. A review of artificial larynges, an overview of LMS algorithm, a description of spectrographic package and the TI/TMS 320C50 digital signal processor and its starter kit are given in the appendices.

Chapter 2

TRANSCERVICAL ELECTROLARYNX

2.1 Voice production

The larynx is situated in the mid line of the neck between the pharynx above and the trachea below, as shown in Fig 2.1. The cavity of the larynx expands into a wide vestibule. Two folds of mucous membrane project from the sides of the cavity into its interior. The lower pair is involved in the production of the voice, and is therefore named the vocal folds. In order to generate voice, the vocal folds move toward axial line of larynx and close the glottis during exhalation as shown in Fig 2.2. This movement of the vocal folds obstructs the outward flow of air from the lower respiratory tract, so that, exhaled air accumulates in the trachea and the subglottal pressure rises. At some point, the increased subglottal pressure blows the vocal folds apart and a small quantity of air escapes to the pharynx. This release of air slightly decreases the subglottal pressure. At the same time, it causes a negative pressure between the medial edges of the vocal folds. This negative pressure is the Bernoulli effect, which results from the great velocity with which air passes through the glottal constriction. The decrease in subglottal pressure and the negative pressure between the medial edges cause the vocal folds to move towards one another, and the glottis shuts with a snap. As soon as the glottis is closed, the Bernoulli effect ceases. As the subject continues to exhale, the subglottal pressure rises again and blows the vocal folds apart. These vibrations due to quick oscillating movements of the vocal folds divide the air column from the lungs into a series of puffs that are released into the vocal tract. These result in a vocal wave, which generally has the shape of a heavily damped oscillation. This type of excitation of the vocal tract results in "voiced" sounds, and the rate of vibration of the vocal folds is the "pitch". Pitch value is dependent on the length and thickness of the vocal folds. The acoustical filtering properties of the oral and nasal cavities provide the spectral shaping of the sound spectrum. Noise like excitation involves the generation of turbulence by rapid flow of air through the constriction. Noisy or aperiodic type excitation results in "unvoiced" sounds. [2].

2.2. Transcervical electrolarynx

When the vocal folds become dysfunctional, the integrity of speech production chain is significantly disturbed. Several types of artificial larynges have been developed for the purpose of providing excitation source to the vocal tract, in order to restore the speech production. A brief review of these devices is given in Appendix A. Artificial larynges can be classified in two broad areas, as determined by the method of vibration coupling to the vocal tract: internal or external. Both types of artificial larynges are further divided in electronic type or pneumatic type depending upon its energy source. Transcervical electrolarynx provides external source of sound and uses electronic energy source. Transcervical electrolarynx does not have direct contact with any part of vocal tract so it is more hygienic than any other type of artificial larynx. Also, studies of acoustic properties of alaryngeal speech show a significant increase in intelligibility with the use of transcervical devices [3].

Transcervical electrolarynx consists of a pulse generator with pitch and intensity control, and vibrator as shown in Fig. 2.3 (a). It is held by the hand against the throat as shown Fig 2.3 (b). Vibrating diaphragm is located at the end of the device. The stages of speech production by transcervical electrolarynx start with the production of a vibration, which is mechanically coupled to the neck tissues, through which the vibration is transmitted to the vocal tract. From there, normal speech production processes take over. The various resonances of the vocal tract shape the harmonic spectrum of the vocal tract vibrations, which results in speech.

From a physical examination of the available transcervical electrolarynges, we found that transducer for vibrator unit is of mainly two types: electrodynamic and electromagnetic.

A typical electrodynamic type transducer unit consists of a light diaphragm of paper or plastic connected to a coil located in a magnetic field set up by a permanent magnet as shown in Fig 2.4 (a). The signal from pulse generator goes to coil and produces magnetic field that interacts with permanent magnetic field and results in vibrations of the diaphragm. A typical electromagnetic type transducer unit shown in Fig 2.4 (b), consists of coils wound around two permanent magnet pole pieces. A steel plate is located a small distance away from the pole pieces. The permanent magnets in contact with the pole pieces supply the steady magnetic field to the steel plate. The signal from pulse generator to coil produces a flux superimposed upon the steady magnetic field. The resultant force acting upon the steel plate produces vibrations of the diaphragm. The vibrations of the diaphragm are transmitted to the neck tissue by a diaphragm coupled to the plate.

In both types of transducer assemblies, the vibrations in addition to getting coupled to the neck tissue also get coupled to the surrounding air from the backside of the diaphragm. This results in sound leakage. Due care is necessary to seal the cavity for reducing this leakage.

2.3 Problems in a transcervical electrolarynx.

Transcervical electrolarynx does not provide highly intelligible sound over a wide range of acoustic environments. The problems in transcervical electrolarynx can be divided into three types (i) improper spectral characteristics (ii) difficulties in co-ordinating the controls of transcervical electrolarynx with manner and articulation of speech production, and (iii) background noise due to leakage of sound from the vibrator.

2.3.1 Low frequency spectral deficit

Once a sound wave is introduced into the anterior surface of the neck tissue, it must propagate through the neck tissue to the anterior pharyngeal wall. The transducer excites a highly localised area of neck tissue, which is a highly non-uniform mass of muscle and membrane. The properties of sound propagation through such an inhomogeneous medium is such that there is much spreading of the sound wave from the localised source. There is a shift of phase and an amplitude variation of the various harmonics of the impressed wave over the different paths from the transducer to the pharynx because of mass-spring viscous damping effects. Transmission loss for the acoustic energy is inversely proportional to frequency. Thus, low frequency components of excitation signals generated by transcervical electrolarynges do not get transmitted to the vocal tract [4, 5].

Sometimes laryngectomees can not get good resonance for intelligible speech with neck type artificial larynx because the sound waves are absorbed before they can be transmitted to the vocal tract due to a thickening of neck after radiation therapy of cancer treatment.

The low frequency energy deficit in speech produced by this type of device results in decreased voice quality. Qi and Weinberg [4] designed and implemented a second order digital filter to compensate for the low frequency energy deficit in speech produced by this type of device. Results of the speech perception tests carried out by them indicated that low frequency enhancement is important to the perception of naturalness and pleasantness. In their conclusion, they commented

> "The development of new materials and more efficient electromechanical transducers would be expected to foster significant improvement in low frequency enhancement. Although low frequency enhancement of speech signals should be accomplished more easily on an electronic basis, an extra or added electronic

device may be necessary".

In order to achieve maximum quality, complete coupling must take place between the neck tissue and the neck type artificial larynx. This is not always achieved with every attempt at speaking. The lack of proper coupling reduces the level of speech and also creates background noise (which will be discussed later) [6].

2.3.2 Difficulties in co-ordinating the controls

Pitch variation is an important factor for naturalness of speech. Speech produced with electrolarynx is generally monotonous and mechanical, even when pitch control is used. Electrolarynx users are not able to produce systematic pitch variations in speech because the method of varying pitch is not efficient or waveform fluctuation does not take place in speech produced by them. The device has to be held in the hand in order to be coupled with neck tissue. It seems that controlling continuous pitch movements by hand while speaking is too complex. Also, constantly moving the hand to the neck and holding it there becomes tiring and tedious. It is difficult to co-ordinate use of the electrolarynx with articulation and manner of production to achieve high degree of speech communication proficiency.

2.3.3 Background noise

Major problem in this type of device is a steady background noise due to the leakage of acoustic energy [1]. Even if the user has not started to speak, it is observed that this type of device makes considerable amount of sound when it is turned on. The main reason is that one side of transducer of vibrator is coupled with anterior part of neck tissues but other side gets coupled to the surrounding atmosphere. It results in direct leakage sound that produces background interference as shown in Fig 2.5. This problem is known to many of transcervical electrolarynx device manufactures. They tried to reduce this noise by reducing coupling from inside the device to the surrounding atmosphere. Another solution is using a personal amplifier unit along with the transcervical electrolarynx. In this amplifier, a highly directional microphone picks sound from the mouth of alaryngeal speaker and it goes to amplifier unit, which amplifies that sound and outputs through a speaker.

In conclusion, we can say that there is a need to develop transcervical electrolarynx that has low background interference, provides naturality to sound, and make speech of laryngectomees more understandable.

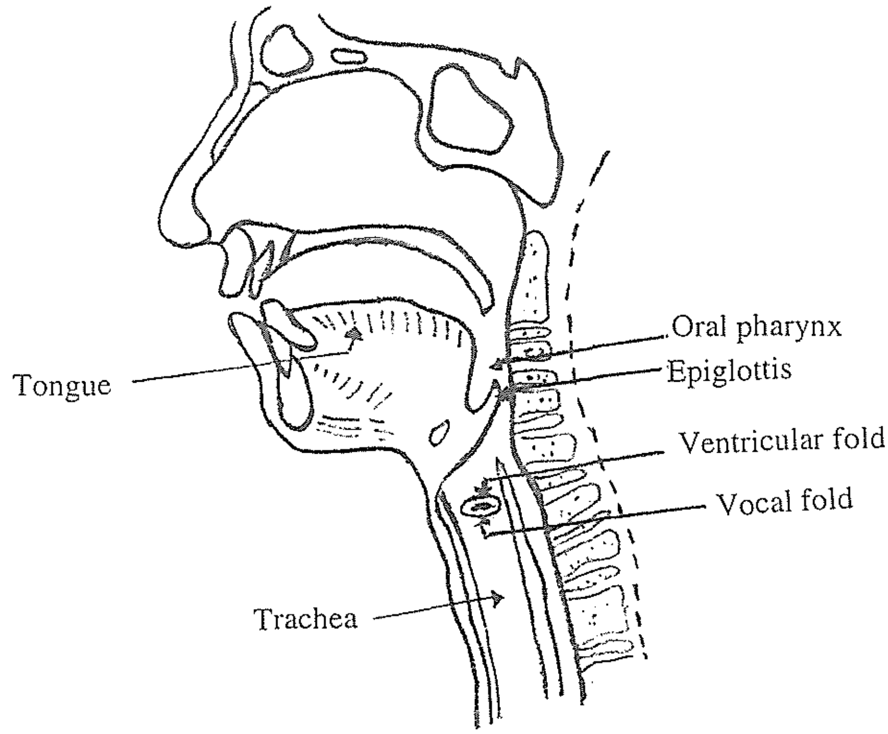


Fig 2.1 Segittal section of the human head and larynx [7]

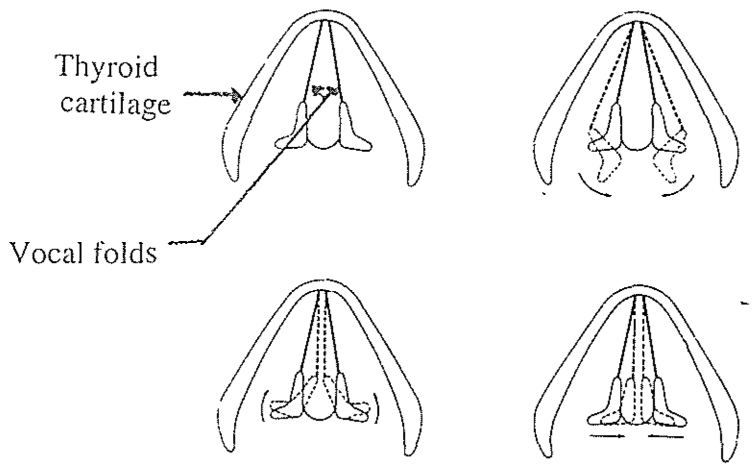


Fig 2.2 (a) Position of vocal cords in quiet respiration
(b) Abduction of vocal cords
(c) & (d) Closure of vocal cords [2]

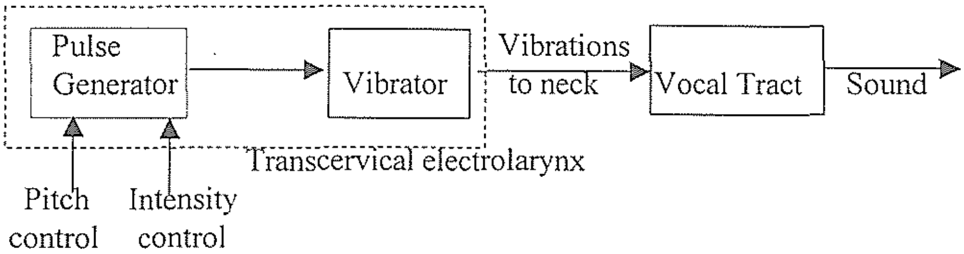


Fig 2.3 (a) Block diagram of sound generation using a transcervical electrolarynx.

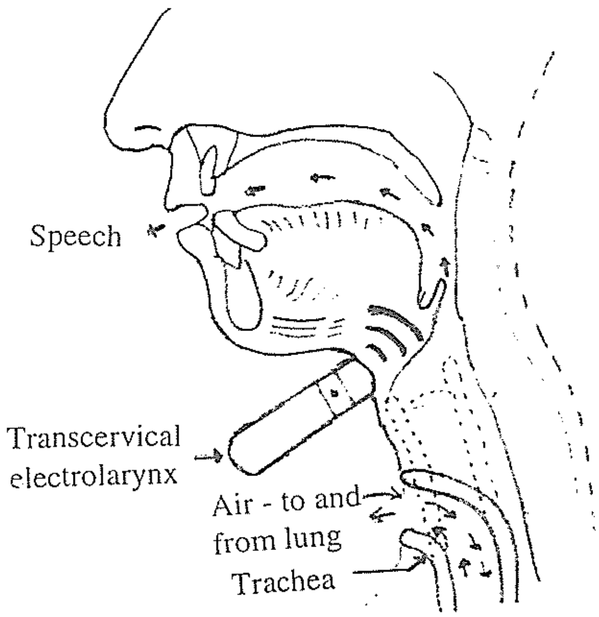
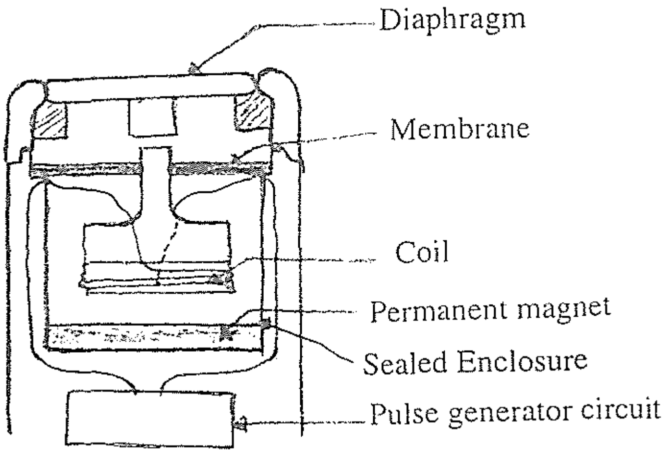
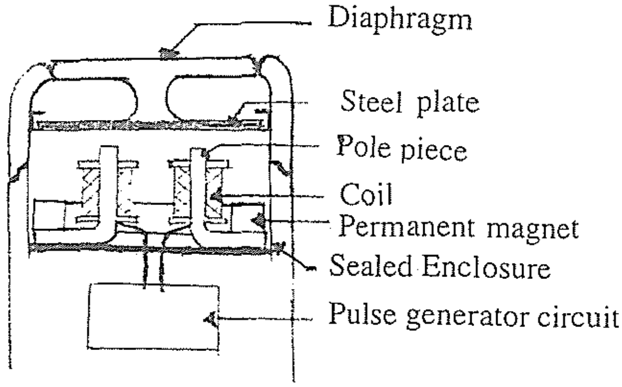


Fig 2.3 (b) A transcervical electrolarynx on throat [7]



(a)



(b)

Fig 2.4 (a) Electrodynamic type transducer (as used in Servox device)
(b) Electromagnetic type transducer (as used in AT&T device)

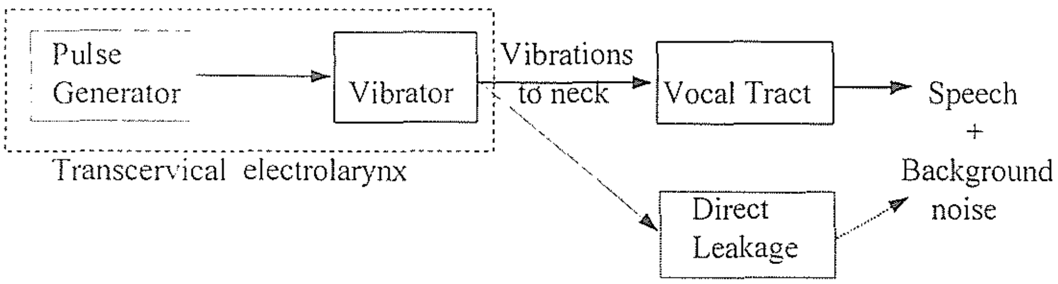


Fig 2.5 Generation of sound with background noise

Chapter 3

ANALYSIS OF ALARYNGEAL SPEECH

3.1 Introduction

The transcervical electrolarynx and its associated problems were reviewed in Chapter 2. Background noise due to leakage of sound from the vibrator significantly degrades quality of alaryngeal speech. In this chapter, background noise has been studied using spectrographic analysis. This is followed by a description of various methods for cancellation of the background noise.

3.2 Spectrographic analysis

The analysis of speech signals involves measurement of temporal features and spectral features. Spectrogram displays the time varying magnitude spectrum with time in horizontal direction, frequency in vertical direction, and intensity by the relatively blackness or brightness [8, 9]. One convenient method among the various methods of spectrogram generation is to compute the short-time Fourier transform from the sampled waveform. The choice of window duration in short time Fourier transform trades off time and frequency resolutions. Wideband spectrograms use a bandwidth of 300 Hz (window duration $\approx$ 3 ms) and smooth out harmonic structure. These are useful in observing pitch period as vertical striations and for seeing formant transitions. Narrow band spectrograms use a window with 45 Hz bandwidth (window duration $\approx$ 20 ms). These enable resolution of individual harmonics but smooth time behavior over a few pitch periods, and are useful for observing the pitch harmonics and formant frequencies during vowel segments. A spectrogram software package, which we used in our analysis, is described in Appendix C. Narrowband spectrogram using this software package for naturally uttered vowel /a/ is shown in Fig 3.1.

Spectrographic analysis is useful to understand the characteristics of alaryngeral speech and background noise. Thus, we have recorded speech of a speaker LP (laryngeal patient) and a speaker HRS (normal) using two different types of Servox make electrolarynx (TE1, TE2). The transcervical electrolarynx TE1, which was used by a speaker LP (laryngeal patient), has three pitches 90 Hz (low pitch 1), 110 Hz (medium pitch 1) and 125 Hz (high pitch 1). The transcervical electrolarynx TE2, which was used by a speaker HRS (normal), has two pitches 80 Hz (low pitch 2) and 90 Hz (high pitch 2). A spectral analysis has been done by taking narrowband spectrogram of sound recorded under two conditions: (a) vibrator on but the person not speaking ("no speech" mode), and (b) person uttering vowel /a/ ("speech /a/" mode). The objective of the analysis is to get an idea about the nature of leakage sound and spectral characteristics of the alrayngeal speech. In the spectrogram package, the magnitude of the various frequency components can be obtained using the cursors. Individual harmonics are seen in the form of horizontal lines in narrowband spectrogram. Formants are seen as bands, each band having a number of pitch harmonics. The center frequency and 6-dB bandwidth of formants are found out putting a cursor at a time instant and measuring the magnitude of the various frequency components. The center frequency and 6-dB bandwidth of formants for both devices are tabulated in Table 3.1 and Table 3.2. The center frequency and 6-dB bandwidth of formants for naturally uttered vowel /a/ is tabulated in Table 3.3.

In "low pitch 1", the sound waveforms along with the narrowband spectrograms for "no speech", and "speech /a/" is given in Fig 3.2. For "no speech", the sound obtained is the direct leakage from the vibrator. The fundamental frequency was found to be 90 Hz. The harmonics are concentrated in the band centered at 0.98, 1.68, 2.11, 3.51, and 4.53 kHz. For sound /a/, the formant frequency are estimated to be 0.70 and

1.09 kHz and the band centered at 1.68 and 4.26 kHz in "speech /a/" spectrogram are mainly contributed by the leakage.

In "medium pitch 1", the sound waveforms along with the narrowband spectrograms for "no speech" and "speech /a/" are given in Fig 3.3. The fundamental frequency was found to be 110 Hz. The harmonics are concentrated in the band centered at 0.98, 2.26, 3.59, and 4.33 kHz. For sound /a/, the formant frequency are estimated to be 0.66 and 1.01 kHz and the band centered at 1.56 and 3.24 kHz in "speech /a/" spectrogram are mainly contributed by the leakage.

In "high pitch 1", the sound waveforms along with the narrowband spectrograms for "no speech" and "speech /a/" are given in Fig 3.4. The fundamental frequency was found to be 125 Hz. The harmonics are concentrated in the band centered at 1.05, 1.68, and 4.29 kHz. For sound /a/, the formant frequency are estimated to be 0.66 and 1.05 kHz and the band centered at 1.71, 3.32 and 4.26 kHz in "speech /a/" spectrogram are mainly contributed by the leakage.

In "low pitch 2", the sound waveforms along with the narrowband spectrograms for "no speech" and "speech /a/" are given in Fig 3.5. The fundamental frequency was found to be 80 Hz. The harmonics are concentrated in the band centered at 0.80, 1.56, 2.89, and 3.63 kHz. For sound /a/, the formant frequency are estimated to be 0.80 and 1.12 kHz and the band centered at 2.54 kHz in "speech /a/" spectrogram is mainly contributed by the leakage.

In "high pitch 2", the sound waveforms along with the narrowband spectrograms for "no speech" and "speech /a/" are given in Fig 3.6. The fundamental frequency was found to be 90 Hz. The harmonics are concentrated in the band centered at 0.78, 1.29, 2.15, and 3.12 kHz. For sound /a/, the formant frequency are estimated to be 1.17 and 2.30 kHz and the band centered at 3.04 and 3.61 kHz in "speech /a/" spectrogram are mainly contributed by the leakage.

The low frequency energy deficit in speech produced by this type of device was seen in narrowband spectrogram. When vibrator is on but person is not speaking, the sound that comes only from vibrator has energy deficit below 500 Hz. This energy deficit also remains in speech produced by speaker using this device.

It is observed that alternate pulses from vibrator have different intensity, and shape in device TE1. Pulses are varying in shape in "low pitch 2" in device TE2.

3.3 Leakage cancellation

Leakage sound generation in transcervical electrolarynx can be modeled as shown in Fig 3.7. The vibration from the device is an excitation $e(t)$ for speech production system using transcervical electrolarynx. The sound from the device is speech signal $s(t)$ mixed with leakage sound $l(t)$.

$$s(t) = h_v(t) * e(t)$$

$$l(t) = h_l(t) * e(t)$$

where $h_l(t)$ is the impulse response of leakage path and $h_v(t)$ is the impulse response of vocal tract.

The normal method of estimating a signal corrupted by additive interference is to pass it through a filter which can suppress the interference while leaving the signal relatively unchanged. The function of the leakage sound canceller is to estimate speech corrupted by leakage sound assuming that leakage sound is additive interference.

A two input leakage sound canceller is shown in Fig 3.8. The microphone 1 is positioned very closed to the lips, and it primarily pick-up the speech signal, with a small amount of leakage sound. The microphone 2 is used to provide a reference signal $r(t)$ for the leakage sound canceller. It is positioned very closed to the place where the device is applied to the neck so that r(t) contains mainly leakage component $l(t)$. Thus, we can write.

$$w(t) = s(t) * h_1(t) + l(t) * h'_1(t)$$

$$r(t) = l(t) * h_2(t) + s(t) * h'_2(t)$$

The reference signal $r(t)$ is given to filter in the leakage canceller as a input. Several approaches for reduction of interference have been reported. But all these methods basically assume that the interference is uncorrelated to the signal, and the adaptive filter uses the minimization of the error. In this particular case, both the signal of interest and the background interference are highly correlated, resulting due to the pulse train from the vibrator. Hence, it will require developing a scheme for adaptive cancellation that in some way makes use of properties of speech signal. Further, microphone with high directivity may be helpful in this regard. If we assume the leakage waveform to be stationary, then we can use a single input adaptive leakage sound canceller, as shown in Fig 3.9. In contrast to two input adaptive leakage sound canceller, it uses only one microphone and works in two modes: (a) training mode and (b) use mode. In training mode, person is silent but vibrator is on and is coupled with speaker's throat. Assuming that sound picked up by the microphone in training mode is only due to leakage, sound picked by the microphone is given to leakage sound canceller. Leakage sound canceller estimates the impulse response of the leakage path using an adaptive algorithm. In use mode, person is speaking and vibrator is also on. The sound picked by microphone in use mode is the sound with the leakage sound. Leakage canceller subtracts estimated leakage sound from the sound picked by the microphone. The output of leakage canceller which is sound with reduced leakage sound is amplified and outputted through speaker. Single input adaptive leakage canceller can be easily incorporated with existing devices without any modification compared to two input leakage canceller.

There are many simple ways of implementing the adaptive leakage canceller. It was decided to implement a leakage canceller based on LMS algorithm on a TI/TMS320C50 based DSP board. The algorithm and its real time implementation is described in Chapter 4. Results of single point adaptive leakage canceller were not found satisfactory. It was difficult to study and analyze the performance of the algorithm in its real time implementation. Hence it was decided to implement and test in off-line mode. Two other methods, for estimating the impulse response of the leakage path: (a) Ensemble averaging and (b) Inverse Fourier transform of ensemble averaged spectrum were also implemented for off-line processing. The implementations and test results are presented in fifth chapter.

Table 3.1 The center frequency and 6-dB bandwidth of formants for transcervical electrolarynx device TE1, as used by speaker LP (laryngeal patient)

Device TE1	When vibrator is on but Person is not speaking		When vibrator is on but person is uttering vowel /a/	
	Center Frequency (kHz)	6-dB bandwidth (kHz)	Center Frequency (kHz)	6-dB bandwidth (kHz)
Low pitch 1 (90 Hz)	0.97	0.66-1.05	0.70	0.62-0.74
	1.67	1.60-1.75	1.09	1.15-1.13
	2.10	2.07-2.18	1.68	1.64-1.72
	3.51	2.93-4.02	2.03	1.95-2.07
	4.53	4.22-4.65	4.26	4.02-4.33
Med pitch 1 (110 Hz)	0.97	0.94-1.01	0.66	0.62-0.70
	2.23	2.07-2.30	1.15	0.94-1.05
	3.59	3.08-3.67	1.56	1.48-1.60
	4.33	4.06-4.49	3.24	3.16-3.28
High pitch 1 (125 Hz)	1.05	0.97-1.09	0.66	0.54-0.70
	1.68	1.64-1.71	1.05	0.98-1.09
	3.63	3.59-3.71	1.71	1.64-1.75
	4.29	4.21-4.33	3.32	3.28-3.39
			4.25	4.18-4.33

Table 3.2 The center frequency and 6-dB bandwidth of formants for transcervical electrolarynx device TE2, used by speaker HRS (normal)

Device TE2	When vibrator is on but Person is not speaking		When vibrator is on but person is uttering vowel /a/	
	Center frequency (kHz)	6-dB bandwidth (kHz)	Center frequency (kHz)	6-dB bandwidth (kHz)
Low pitch 2 (80 Hz)	0.80	0.74-0.97	0.80	0.74-0.94
	1.56	1.48-1.76	1.13	1.01-1.25
	2.89	2.69-3.04	2.54	2.46-2.61
	3.63	3.56-3.75		
High pitch 2 (90 Hz)	0.78	0.74-0.86	1.17	1.05-1.25
	1.29	1.25-1.37	2.30	2.22-2.34
	2.15	1.90-2.38	3.04	2.92-3.12
	3.12	2.92-3.20	3.61	3.55-3.71

Table 3.3 The center frequency and 6-dB bandwidth of formants of vowel /a/ for speaker HRS (normal) without using transcervical electrolarynx

Center frequency (kHz)	6-dB bandwidth (kHz)
0.78	0.74-0.82
1.05	0.89-1.09
2.19	2.14-2.27

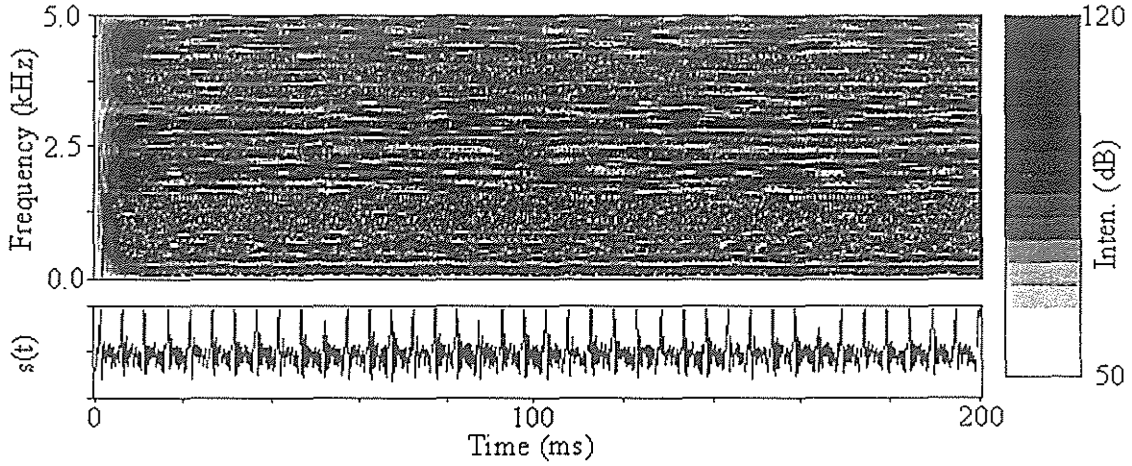


Fig. 3.1 Spectrogram for naturally uttered vowel /a/

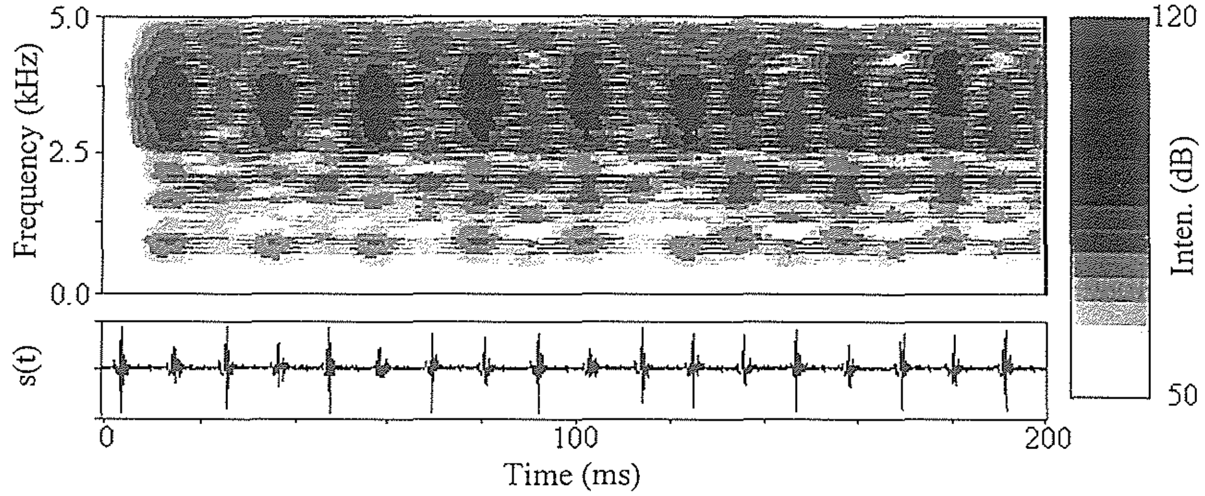


Fig. 3.2 (a) Narrow band spectrogram for a speaker LP (laryngeal patient) using transcervical electrolarynx device TE1 for low pitch 1 in "no speech" mode

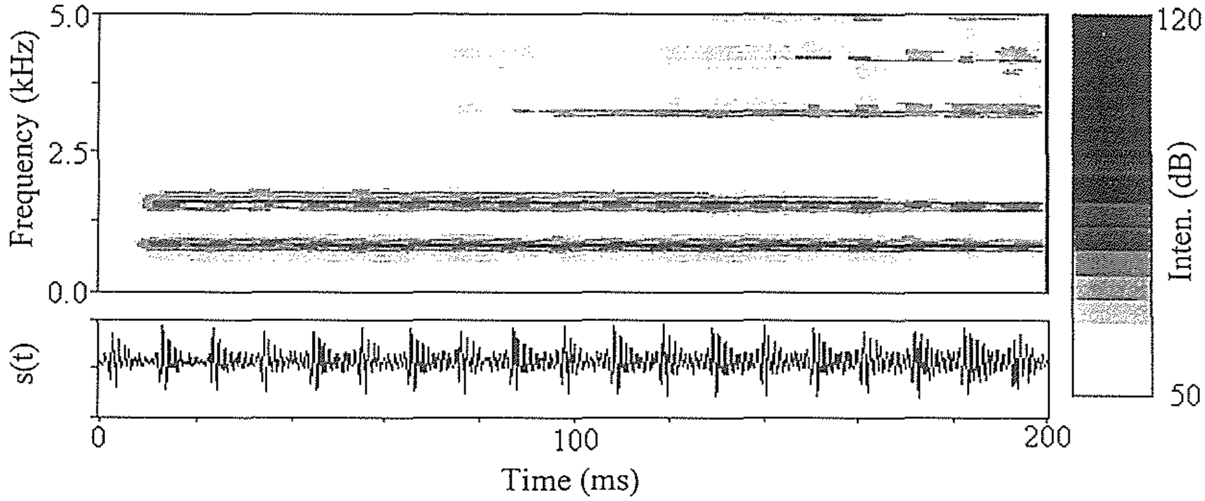


Fig. 3.2 (b) Narrow band spectrogram for a speaker LP (laryngeal patient) using transcervical electrolarynx device TE1 for low pitch 1 in "speech /a/" mode

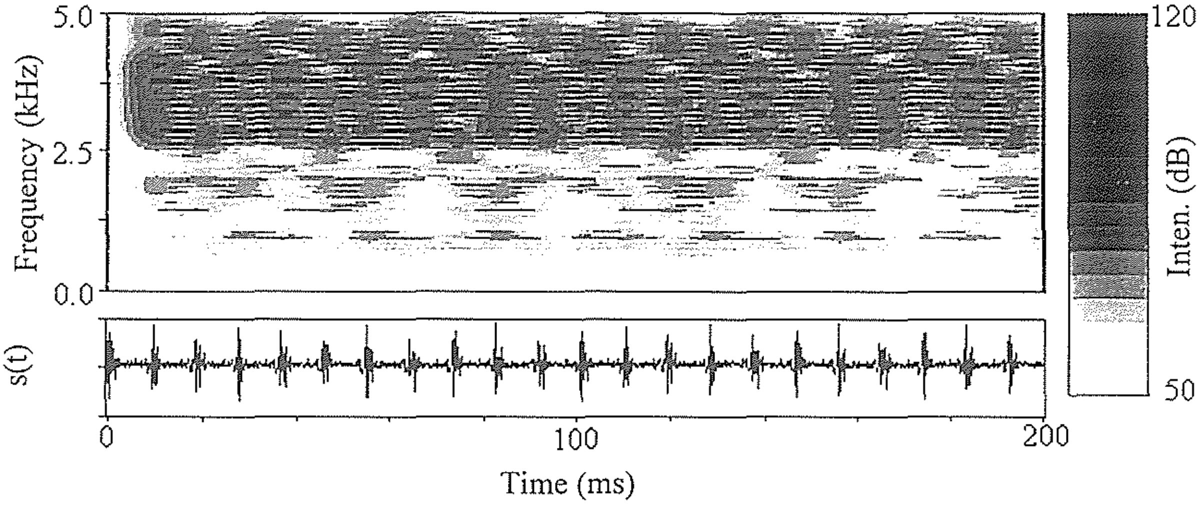


Fig. 3.3 (a) Narrow band spectrogram for a speaker LP (laryngeal patient) using transcervical electrolarynx device TE1 for med pitch 1 in "no speech" mode

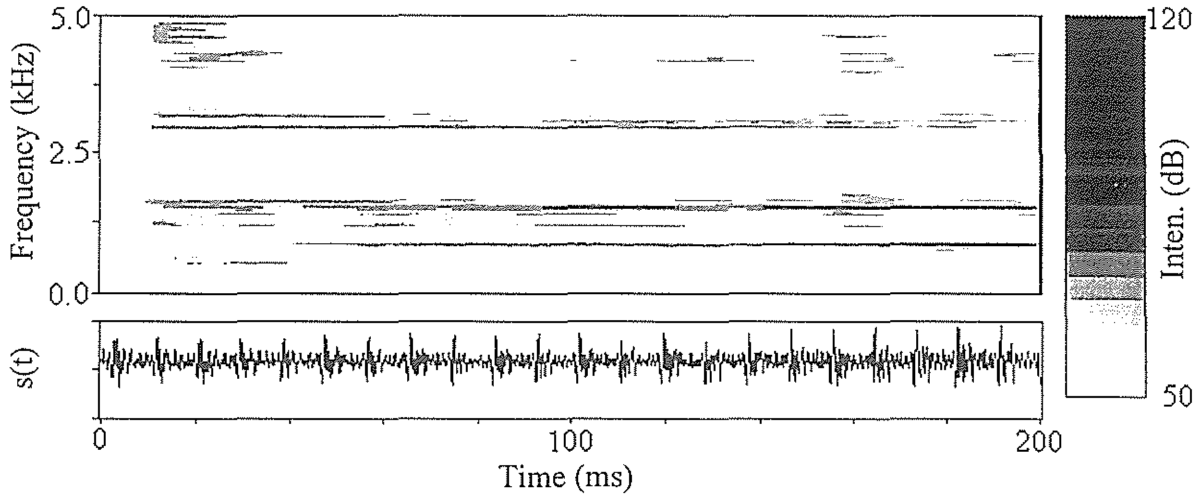


Fig. 3.3 (b) Narrow band spectrogram for a speaker LP (laryngeal patient) using transcervical electrolarynx device TE1 for med pitch 1 in "speech /a/" mode

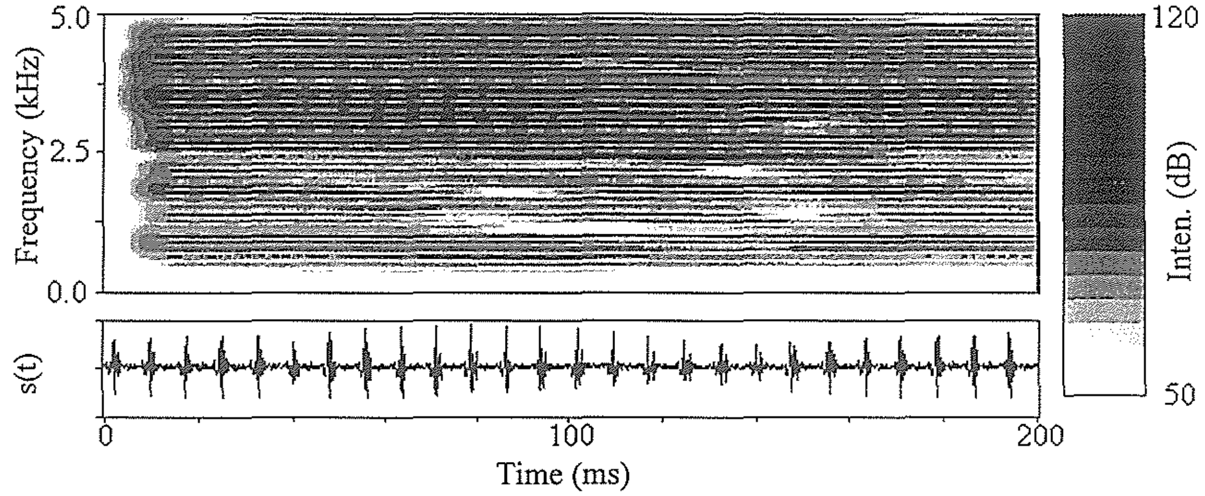


Fig. 3.4 (a) Narrow band spectrogram for a speaker LP (laryngeal patient) using transcervical electrolarynx device TE1 for high pitch 1 in "no speech" mode

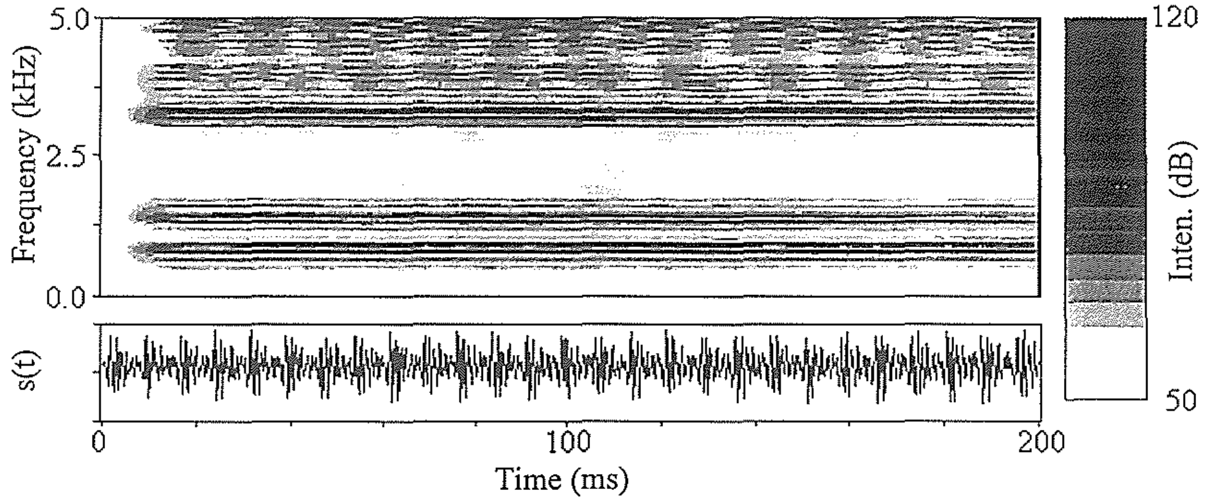


Fig. 3.4 (b) Narrow band spectrogram for a speaker LP (laryngeal patient) using transcervical electrolarynx device TE1 for high pitch 1 in "speech /a/" mode

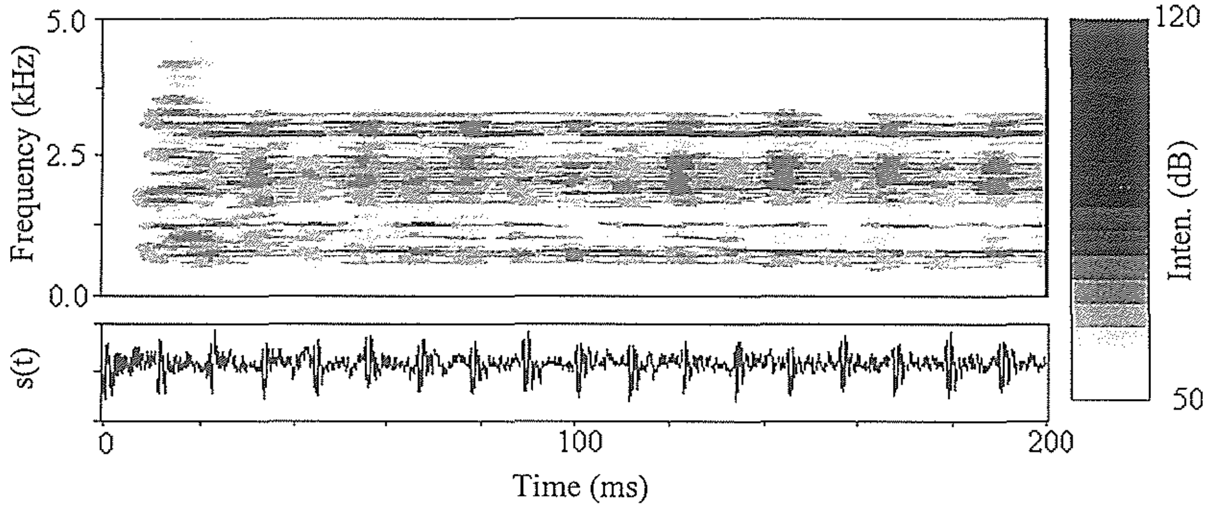


Fig. 3.5 (a) Narrow band spectrogram for a speaker HRS (normal) using transcervical electrolarynx device TE2 for low pitch 2 in"no speech" mode

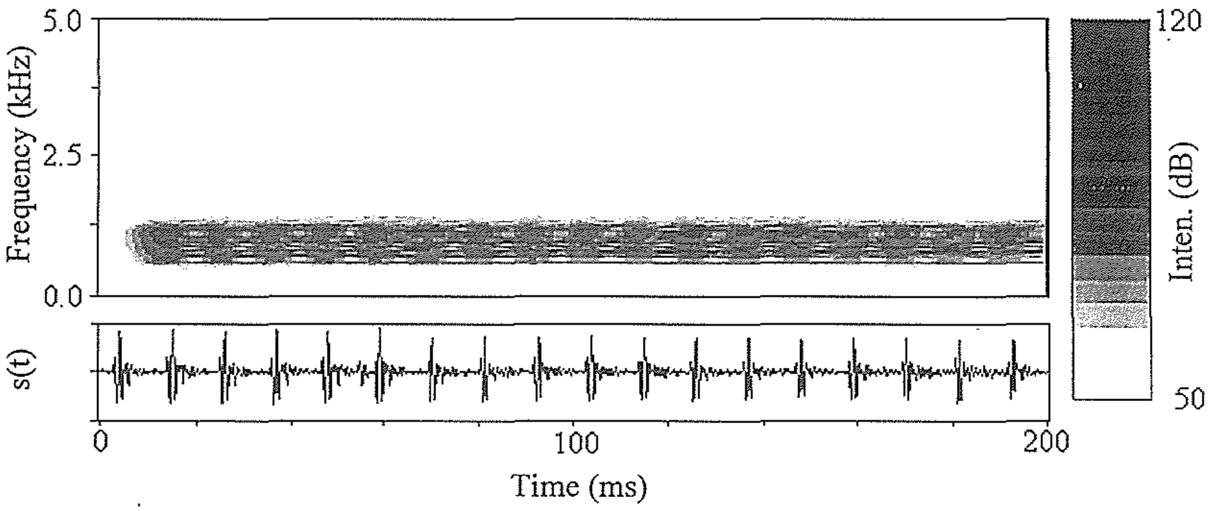


Fig. 3.5 (b) Narrow band spectrogram for a speaker HRS (normal) using transcervical electrolarynx device TE2 for low pitch 2 in"speech /a/" mode

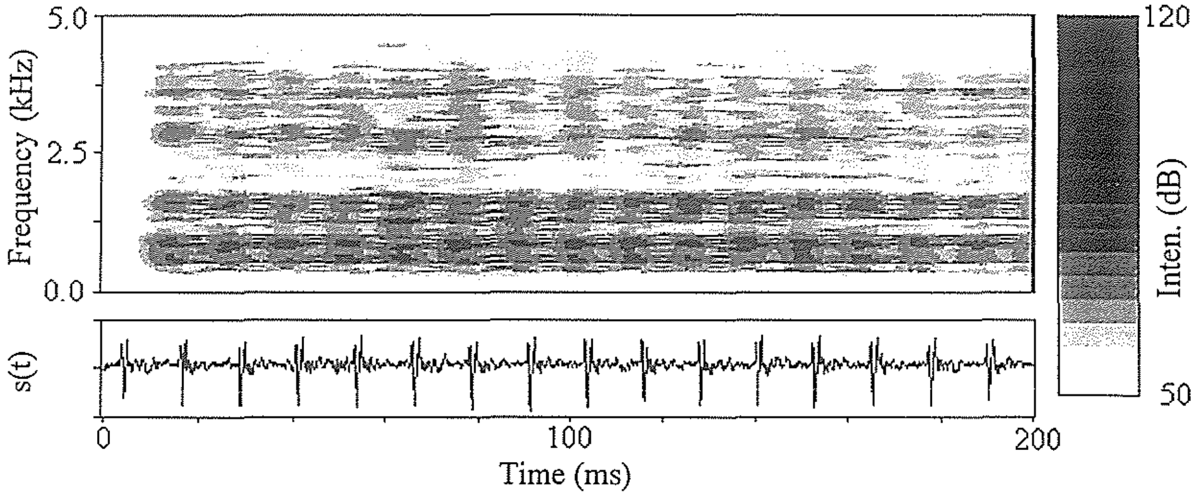


Fig. 3.6 (a) Narrow band spectrogram for a speaker HRS (normal) using transcervical electrolarynx device TE2 for high pitch 2 in "no speech" mode

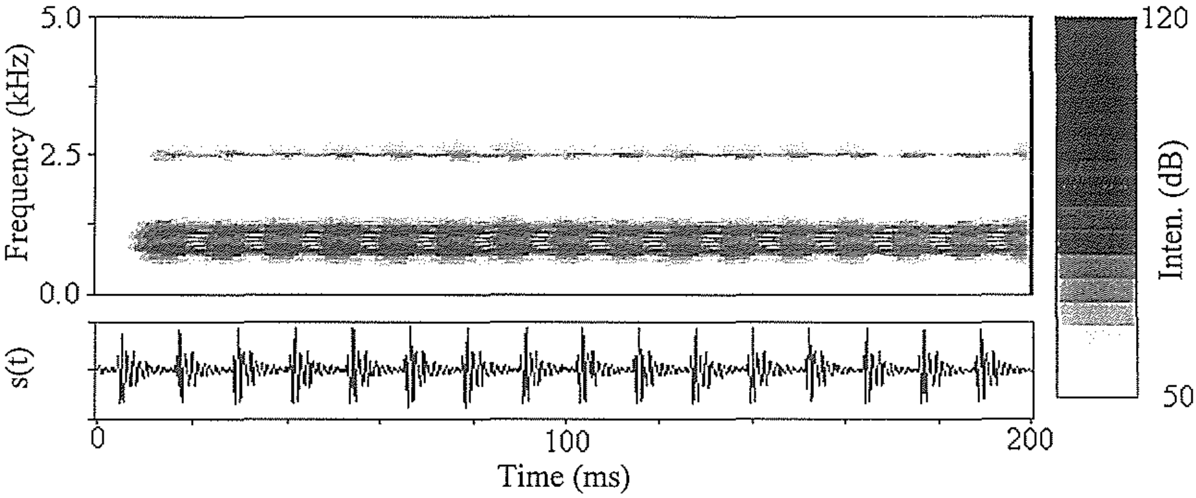


Fig. 3.6 (b) Narrow band spectrogram for a speaker HRS (normal) using transcervical electrolarynx device TE2 for high pitch 2 in "speech /a/" mode

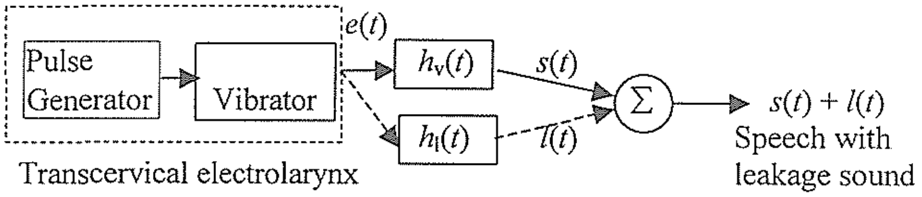


Fig 3.7 Model of leakage sound generation in transcervical electrolarynx

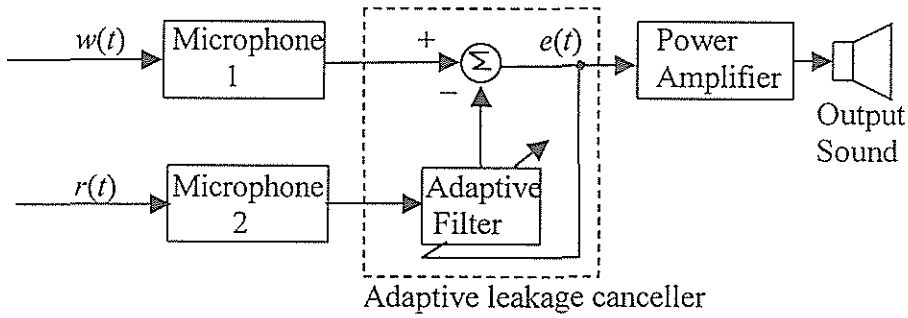


Fig 3.8 Two input adaptive leakage sound canceller

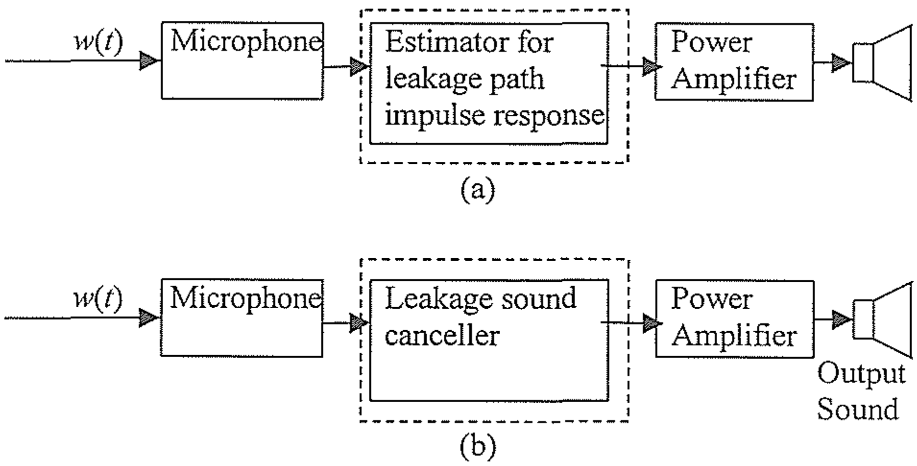


Fig 3.9 Single input adaptive leakage sound canceller
(a) training mode (b) use mode

Chapter 4

REAL TIME IMPLEMENTATION OF SINGLE INPUT LEAKAGE SOUND CANCELLER USING LMS ALGORITHM

4.1 Introduction

In the last section of the previous chapter, the basic scheme of the single input leakage sound canceller using LMS algorithm has been presented. In this chapter, this scheme is outlined, a real time implementation of the scheme using TI TMS320C50 (32-bit fixed-point accumulator based DSP) is described, and test results with this implementation are presented.

4.2 Single input adaptive leakage sound canceller

The scheme of the single input leakage canceller is based on the background noise model shown in Fig 3.7. It is further assumed that the impulse response for the additive noise path does not change with time. Hence, the artificial larynx is first operated, with the speaker not making any attempt at speaking. The sound waveform picked up by the microphone for each excitation pulse of the vibrator will be the impulse response of the leakage path. In the training mode, LMS algorithm can be used to estimate this impulse response over a large number of excitation pulses. The algorithm tries to obtain the estimate so that the error between the input and estimate is minimised in the mean square sense. In the use mode, the impulse response of the leakage path for each excitation pulse is used as an estimate of the leakage sound. This estimate is subtracted from the input signal, in order to cancel the background noise.

In both the training and use modes, estimation of the leakage sound requires the reference position of the excitation pulse of the vibrator. The reference position of excitation pulses can be found by tapping it from the pulse generator circuit of the vibrator but this requires hardware changes in the existing devices. To avoid this, we developed an impulse generator algorithm, which uses the input signal to locate the reference position of the excitation pulses. In the following subsection, we will describe the impulse generator, estimation of leakage sound during training mode and cancellation of leakage sound during use mode.

4.2.1 Impulse generator

The impulse generator gives an impulse that refers to each excitation pulse position in the input signal. This impulse generator algorithm is based on the dynamic threshold method for pitch estimation [10]. A block diagram of the impulse generator using the magnitude thresholding method is shown in Fig 4.2. The threshold value for impulse generation is given by some constant percent the average squared value of the input signal. The average squared value of input signal is found out by passing squared input through a low pass filter. The input is compared with the threshold, and if the input is greater than the threshold an impulse is generated. To avoid more than one impulse at single pitch period, we provide a refractory period after each first impulse at that pitch period. During the refractory period, there is no impulse even if input exceeds the threshold. If the input does not exceed the threshold, no impulse is generated. The threshold value could also be given by some constant percent the average magnitude value of the input signal. The average magnitude value of the input signal is found out by passing full wave rectified input through a low pass filter. Experimentally we found that the use of the average squared value of the input for finding the threshold gives more accurate and stable reference position of the input excitation pulse than the use of the average magnitude of input. Hence, we used the average squared input for finding the threshold. The algorithm can be summarised as under.

1 Obtain the squared magnitude of the input $w(n)$.

$$v(n) = [w(n)\,]^2 \tag{4.1}$$

2 Obtain the average squared value of N samples of the input.

$$p(n) = \frac{1}{N}\sum_{m=0}^{N-1} v(n\text{-}m)$$

which can be rewritten as under.

$$p(n) = p(n\text{-}1) + [v(n) - v(n\text{-}N)\,]\,/\,N \tag{4.2}$$

3 When input exceeds the threshold during non-refractory period, obtain the impulse.

$$\begin{aligned} x(n) &= 1 \quad \text{if } v(n) > \alpha\, p(n) \text{ and } n \text{ is in non-refractory period} \\ &= 0 \qquad \text{otherwise} \end{aligned} \tag{4.3}$$

where α is constant and the value of refractory period is fixed.

4.2.2 Estimation of leakage sound

A block diagram of single input adaptive leakage sound canceller during "training" mode is shown in Fig. 4.2. The input signal is given to the impulse generator. The impulse generator output is given to the adaptive filter as input. The signal $w(n)$ delayed by a fixed delay N_d is given as reference input to the adaptive filter. The adaptive filter uses an FIR filter, where coefficients are adjusted by using LMS algorithm. FIR filter output is subtracted from input, and the error signal is given by the LMS block to calculate the filter coefficient. We have used the LMS algorithm using instantaneous error, and the algorithm is briefly described in Appendix B. The LMS algorithm changes the filter coefficients to minimise mean-square error. When the error is minimised, the impulse response of the filter gives estimate of leakage sound.

There is certain delay in the impulse generator block. The delay "N_d" has been introduced to compensate for it. However, there is no need to adjust delay N_d exactly as required. If the value of N_d is slightly greater than required, FIR filter will take care of additional delay, FIR filter output $y(n)$ is given as

$$y(n) = \sum_{m=0}^{N-1} b_m(n)\, x(n-m) \tag{4.4}$$

The error is given as

$$e(n) = w(n-T_d) - y(n) \tag{4.5}$$

The error $e(n)$ and the output of the impulse generator $x(n)$ are used for updating the coefficients of FIR filter $b_m(n)$, using LMS algorithm, on the basis of previous coefficients as under.

$$b_m(n) = b_m(n-1) + \mu \cdot e(n)\, x(n) \tag{4.6}$$

where μ is convergence parameter [Appendix B]. When the LMS algorithm minimises mean square error $e(n)$, an impulse response of FIR filter gives estimate of leakage sound.

$$y(n) \approx l(n) \tag{4.7}$$

4.2.3 Cancellation of leakage sound

A block diagram for the single point adaptive leakage sound canceller in use mode is shown in Fig 4.3. An impulse generated from the impulse generator according to input signal is given to filter. An impulse response of filter is estimation of the leakage sound as shown in equation 4.7. This estimation is subtracted from the delayed input signal, in order to cancel the leakage sound.

4.3 Implementation

The real time implementation of single point adaptive leakage sound canceller consists of interfacing of the TI/TMS320C50 DSP board to the input analog signal conditioning unit and implementation of the adaptive sound cancellation algorithm on the TI/TMS320C50 DSP board. The DSP board is a standalone application board that allows us to experiment and use the DSP for real time signal processing. The DSP board consists of the TI/TMS320C50 DSP for full speed verification of the source code and analog interface circuit TLC32040 that interfaces to the TMS32C50 serial port. The TMS320C50 processor and the TMS320C50 DSP board are briefly described in Appendix D.

The experimental set-up for the real time implementation of the scheme is shown in Fig 4.4. The input analog signal conditioning unit was developed by P. S. Gavankar as part of his M. tech. project [11]. The input analog signal-conditioning unit consists of preamplifier followed by a low pass filter. The preamplifier amplifier amplifies the speech signal picked up from the microphone to the voltage range used by the ADC. After amplification the signal is fed to the low-pass filter. The filter is used to cut-off any frequency components above 5 kHz, as most of the energy in normal speech is concentrated below 5 kHz. The output from the filter is then fed to the A/D converter of the TMS320C50 DSP board. The output from the TMS320C50 DSP board is fed to oscilloscope and/or headphone.

To implement the single point adaptive leakage sound cancelling method in real time, the TI/TMS320C50 DSP board is used. The software for the DSP board is written in the DSP's assembly language. An assembler is available which can generate machine language code. The machine language code of the program can then be loaded on the DSP board and run. A debugger is also available, which can be used to debug the program. The debugger helps in viewing the register and data memory contents. Single stepping and break point facilities are also available.

The sequence of steps of steps necessary to write the programs are summarised as follows:

(1) Write the assembly language program using any editor.

(2) Assemble the above program using dsk5a.exe (the DSK assembler).

(3) Load the object file (*.dsk) using dsk5d.exe (the DSK debugger interface).

(4) If necessary, make the changes in assembly code and repeat steps (2) and (3).

The single-point adaptive leakage sound canceller is implemented on the TMS320C50 DSP board. The sampling rate is selected as 20,000 samples/sec with RX Counter A and RX Counter B of AIC TLC32040C of the TMS320C50 DSP board. Whenever an interrupt is received, the input data is stored in serial data buffer and one processed sample is outputted to the D/A converter. The input data in serial data buffer is stored in starting register of registers stack in TMS32C50 processor. The Average squared value of the input data is found out by passing squared input through a low pass filter as shown in Equation 4.2.

The threshold value is set equal to constant α time the average squared value of the input. The value of α is established empirically 0.4. The present value of input data is compared with the threshold value. If the input is greater than the threshold during non-refractory period, the impulse is generated and refractory counter starts to decrease. After each first impulse at that pitch period, refractory flag is set to indicate starting of refractory period. When refractory counter becomes zero, refractory flag is reset to indicate starting of non-refractory period. During the refractory period, there is no impulse even if input exceeds the threshold. The value of refractory counter is established empirically. The value of refractory period is established 2.4ms.

The impulse generator output is given to adaptive FIR filter as input. The window length of adaptive FIR filter is greater than the minimum length required representing leakage sound. The window length in our case is 100. The difference between the output of adaptive filter $y(n)$ and the delayed input signal is calculated as shown in equation 4.5. In training mode, the coefficients of FIR filter $b_m(n)$ are updated using LMS algorithm., in which the coefficients of filter are updated on the basis of previous coefficients as shown in equation 4.7. The value of convergence parameter μ is established to 0.64. When FIR filter converges, the updating of coefficients is stopped and the impulse response of the FIR filter is an estimate of the direct sound leakage radiated from the device. The difference between the output of the FIR filter and the delayed input is outputted from DAC of TMS32C50 board. Output sound can be heard using speaker or can be seen on CRO.

The tasks to be carried out are summarised in the following modules.

(1) Acquisition of the data by sampling the incoming speech signal.

(2) Generation of impulse to locate excitation pulse position in input signal using impulse generator algorithm.

(3) Compute the error signal between delayed input signal and output of FIR filter.

(4) Compute coefficients of FIR filter using LMS algorithm in training mode

(5) Cancellation of leakage sound in use mode.

4.4 Testing and results

The real time implementation of single-input adaptive leakage sound canceller was tested for satisfactory performance of its various blocks using different types of signals. Testing of whole operation was done using speech files recorded from two different devices and speech spoken by normal person using transcervical device.

To test the data acquisition software module, different types of signals from function generator were given to the TMS320C50 DSP board and the acquired data was outputted to CRO where it was compared with input signal. This procedure was repeated for different sampling rates that were adjusted by software. It is observed that output waveform was similar to the input waveform.

Impulse generator module was tested using recorded speech waveform and ECG waveform from an arbitrary waveform generator (HP 33120A). The value of constant α and refractory time was established such that the output of impulse generator was an impulse for each QRS waveform of ECG.

Adaptive filter was tested using recorded speech waveform and ECG waveform from function generator. Number of coefficients for FIR filter was greater than minimum number required representing a pulse from corresponding waveform. In case of recorded alaryngeal speech waveform, number of coefficients varies between 85 to 128. When FIR filter converges, the impulse response of adaptive filter is the estimate of the input waveform. This fact was easily tested by seeing output of FIR filter on CRO. The variation in the RMS value of the error as obtained using an oscilloscope with built in rms display (HP 54601A) with time for ECG signal with different magnitude is shown in Fig 4.5. The variation in the rms value of error with time for different value of the convergence parameter is shown in Fig 4.6. Small value of the convergence parameter increases convergence time by algorithm. Large value of the convergence parameter results in significant reduction in the error but time taken to converge is increased.

After testing every module, the whole single point adaptive leakage canceller software was tested using recorded speech and actual speech through microphone. As the noise cancellation does not work satisfactory, there is a need to further investigate. It was decided to study the implementation by off-line processing, and to compare the results as obtaining by two other schemes. These are presented in fifth chapter.

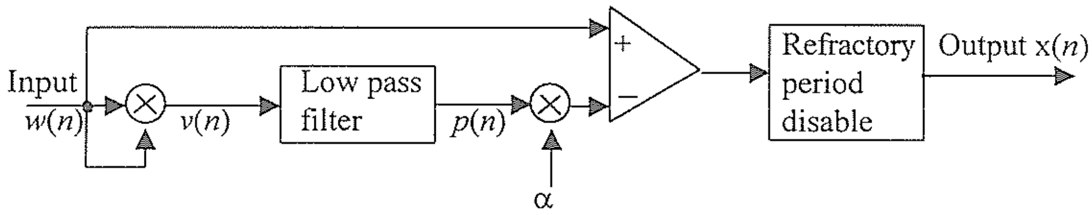


Fig. 4.1 Block diagram for the impulse train generator

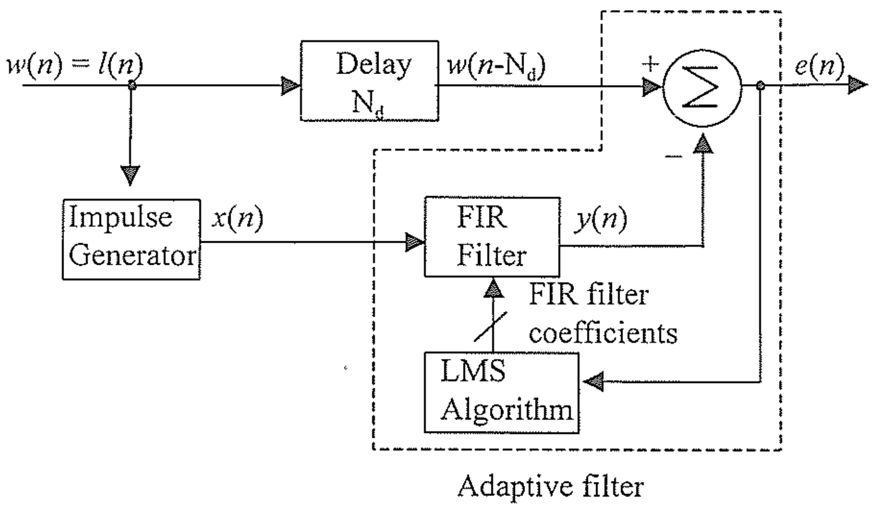


Fig. 4.2 Block diagram for leakage canceller during training mode

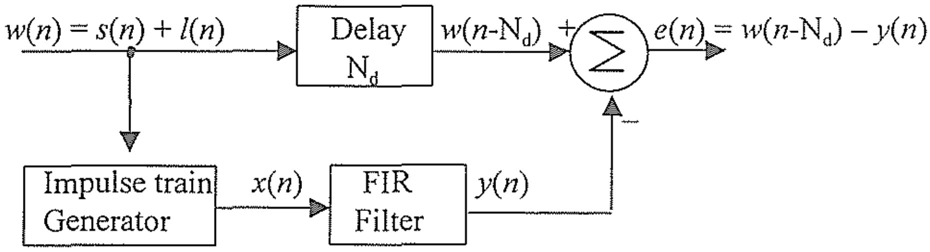


Fig. 4.3 Block diagram for leakage canceller in use mode.

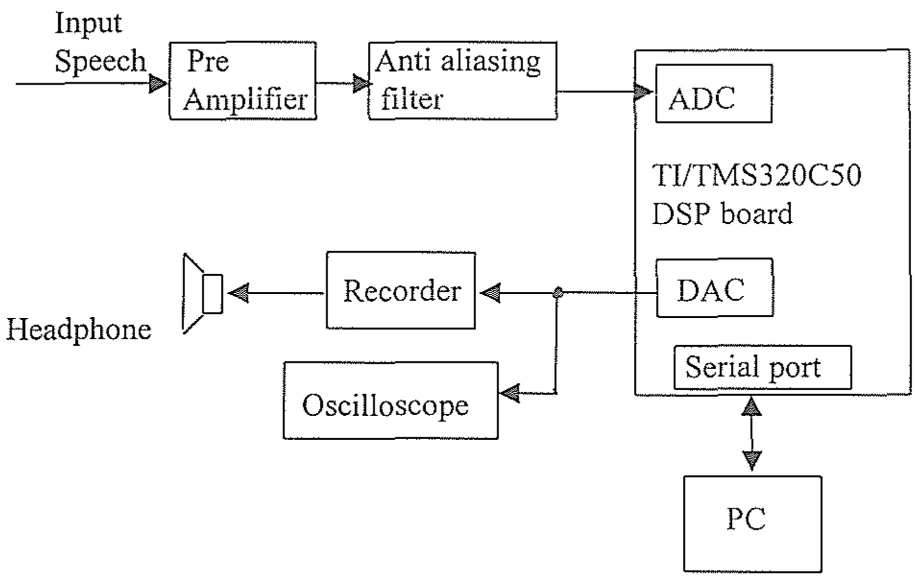


Fig. 4.4 Experimental set up for real time implementation of the leakage canceller

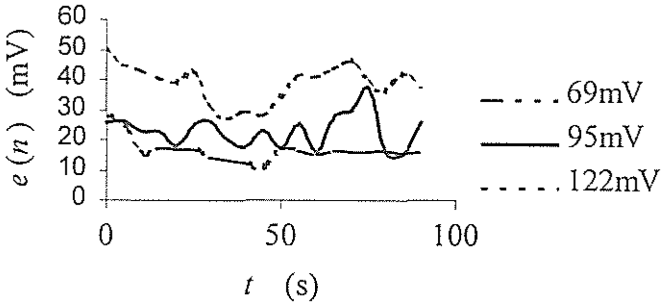


Fig 4.5 Time variation of the rms value of error for ECG signal for different input magnitudes
(Vin rms =69mV, 95 mV,122mV for $\mu = 0.64$)
(training mode 0-50 s then use mode)

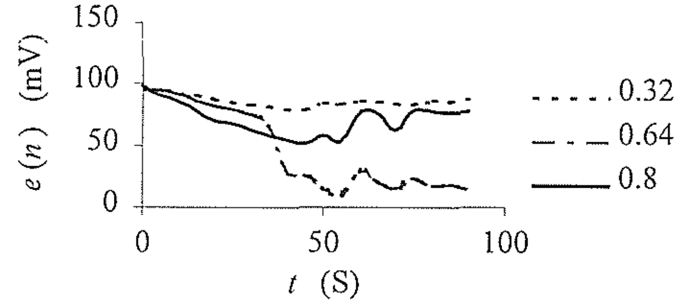


Fig 4.6 Time variation of the rms value of error for ECG signal for different convergence parameter
($\mu = 0.32, 0.64, 0.8$ for Vin rms =122mV)
(training mode 0-50 s then use mode)

Chapter 5

OFF-LINE IMPLEMENTATION OF LEAKAGE SOUND CANCELLER

5.1 Introduction

The results of real time implementation of single input adaptive leakage sound canceller using LMS algorithm were not satisfactory as seen in Chapter 4. This could be due to non-stationarity of the background noise. caused by variation in hand pressure while holding the device, or due to interaction of the vocal tract muscles with the diaphragm of the artificial larynx. There is also a possibility that LMS algorithm does not converge properly to obtain a good estimate of the impulse response of the leakage path. It was difficult to study the dynamics of convergence of LMS algorithm in the real time implementation. Hence, we decided to carry out an off-line implementation of the algorithm.

The impulse response of the leakage path can also be obtained by the following two methods.

(a) Ensemble averaged the leakage sound with respect to excitation pulses of the vibrator.

(b) Inverse Fourier transform of the ensemble averaged Fourier transform of segments of the input signal with respect to excitation pulses of the vibrator.

It was decided to study above three leakage sound estimation methods, using off-line processing. All the three methods require the impulse generator. The impulse generator gives an impulse that refers to each pulse position in the input signal. A block diagram of the impulse generator is shown in Fig 4.2. The input is compared with the threshold. If the input is greater than the threshold during non-refractory period, an impulse is generated. During the refractory period, there is no impulse even if input exceeds the threshold. If the input does not exceed the threshold, an impulse is not generated. The impulse generator algorithm as described in Chapter 4 was implemented using C language. Estimation and cancellation of leakage sound using the three methods are described in the following two sections. The test results are presented in the last section.

5.2 Estimation of leakage sound

After the pulses in the input signal have been properly located, estimation is an important stage in the training mode of the leakage canceller. The following three methods estimate the leakage sound.

(1) LMS adapting

(2) Ensemble averaging

(3) Inverse Fourier transform of ensemble averaged spectrum

In the training mode, the vibrator of the transcervical electrolarynx is on, but the person is silent. Sound from the vibrator, which is picked by a microphone, is sampled and stored. The three methods were implemented for off-line processing on a PC using C programming language.

5.2.1 LMS adapting

A block diagram of LMS adapting method for estimation of the leakage sound is shown earlier in Fig. 4.2 for real time implementation. The impulse generator output is given to the adaptive filter as input. The delayed signal is given as reference input to the adaptive filter. The adaptive filter uses FIR filter, where coefficients are adjusted by using LMS algorithm. FIR filter output is subtracted from input, and the error signal is given by the LMS algorithm block to calculate the filter coefficient. When the error is minimised, the impulse response of the filter gives estimate of leakage sound.

5.2.3 Ensemble averaging

A block diagram of the ensemble averaging method to estimate leakage sound is shown in Fig. 5.1. The input signal is given to the impulse generator. Whenever the impulse generator generates an impulse, the block selector selects a block of samples from the input signal with reference to the impulse position. The block starts a certain number of samples before the impulse. The selected block is passed to the ensemble averager. This process is repeated for the specified number of blocks after which the averager outputs the ensemble averaged waveform for these blocks. This ensemble averaged waveform is an estimate of the impulse response of the leakage path.

Let N_1 be the number of samples before impulse position and N_2 be the number of samples after impulse position to represent the significant part of the waveform corresponding to the excitation pulse from vibrator. For the impulse positioned at sample "n", the block for averaging consist of

$$g_n(m) = w(n+m)\,, \quad -N_1 \le m \le N_2$$

Ensemble average is obtained as

$$y(m) = \frac{1}{M}\sum_{k=0}^{M-1} g_{n_k}(m) \qquad -N_1 \le m \le N_2$$

This $y(m)$ is estimate of the impulse response of the leakage path.

5.2.3 Inverse Fourier transform of ensemble averaged spectrum

A block diagram of the inverse Fourier transform of average spectrum method to estimate leakage sound is shown in Fig. 5.2. The input signal is given to the impulse generator. Whenever the impulse generator generates an impulse, the block selector selects a block of samples from the input signal with reference to the impulse position, as in the previous method. Zero valued extra samples are padded to make block size in an exact power of 2. Spectrum of these blocks is obtained by taking its FFT. The complex spectra for the specified number of blocks are ensemble averaged. Inverse Fourier transform of ensemble averaged spectrum is an estimation for the impulse response of the leakage path

5.3 Cancellation of leakage sound

A block diagram for leakage cancelling in normal mode for real time implementation of the LMS adaptive method has been given earlier in Fig 4.4. An impulse generated from the impulse generator according to input signal is given to FIR filter. An impulse response of FIR filter is estimation of the leakage sound. The output of FIR filter is subtracted from the delayed input signal, in order to remove the leakage sound.

The ensemble averaging method and the inverse Fourier transform of ensemble averaged spectrum method give estimate of the leakage sound in the form of a block of samples in training mode. Cancellation of the leakage sound in normal mode for these two methods is shown in Fig 5.3. The input signal is sound spoken by person using the transcervical electrolarynx. An impulse from the impulse generator gives reference position for the input pulse. Whenever an impulse is generated, the estimated leakage sound is subtracted from input signal with reference to the position of an impulse.

5.4 Testing and results

Three methods for estimation of leakage sound as described in section 5.2 were implemented on PC using C language. The implementation was tested for satisfactory performance of its various blocks and for overall operation, using speech files recorded from two different devices.

Fig 5.4 shows sound waveform $w(n)$ for two different devices as referred in Chapter 3 and their output $x(n)$ from the output of the impulse generator. The position of pulses in the input was located properly by the impulse generator algorithm.

Alternate pulses were similar in TE1 but that was not the case in TE2. So, the estimation of leakage sound was done using (1) all pulses from the vibrator, (2) alternate pulses from the vibrator, and (3) every third pulse from the vibrator. Estimated leakage sound using these three methods in training mode is shown in Fig 5.5.

It is observed that estimated leakage sound is different for each method. To find out which one is better estimation of leakage sound, same signal was given in normal mode. In first case, estimated leakage sound was subtracted from each pulse of the input for the cancellation of leakage sound. In second case, estimated leakage sound from even pulses was subtracted from each even pulse of the input and estimated leakage sound from odd pulses was subtracted from each odd pulse of the input for the cancellation of leakage sound. Same thing was done in the third case. The input signal and the difference between that signal and the estimated leakage sound are shown in Fig 5.6, Fig 5.7 and Fig 5.8 for various conditions.

In single-input adaptive leakage cancelling method, the peak magnitude of the output signal is decreased in condition shown in Fig 5.6 (b). Some pulses of the input signal remained in the output signal in condition shown in Fig 5.6 (c). The output signal is fully deteriorated in conditions shown in Fig 5.6 (d).

In ensemble averaging method, some pulses of the output signal are fully cancelled by estimated leakage sound in condition shown in Fig 5.7 (b). In conditions shown in Fig 5.7 (c) and (d), magnitude of the output signal is decreased but most of the pulses do not get fully cancelled. In inverse FT of ensemble averaged spectrum method, there is no significant reduction in the output signal in all conditions shown in Fig 5.8 (b), (c), and (d).

It is noted that skipping of pulses does not improve results but it deteriorates original signal in some cases. Ensemble averaging method gave better estimation than other two methods but still it could not remove background noise fully. All the three methods could not give satisfactory results because of variation in shape of pulses coming from vibrator. This variation may change from device to device. Reasons for this variation are change in positioning of the device on the throat, change in application pressure, transducer dynamics, and fluctuation in the battery voltage during use of device.

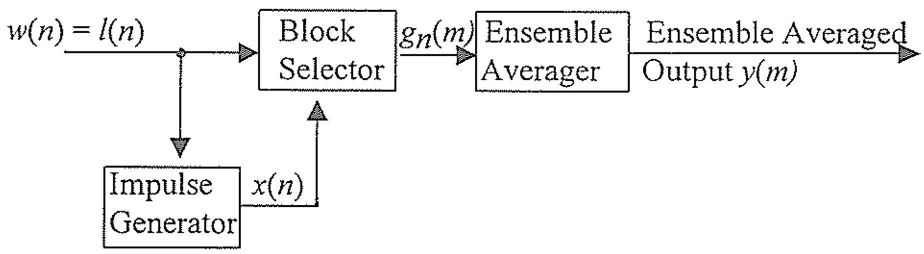


Fig. 5.1 Block diagram for ensemble averaging for estimation of the leakage sound

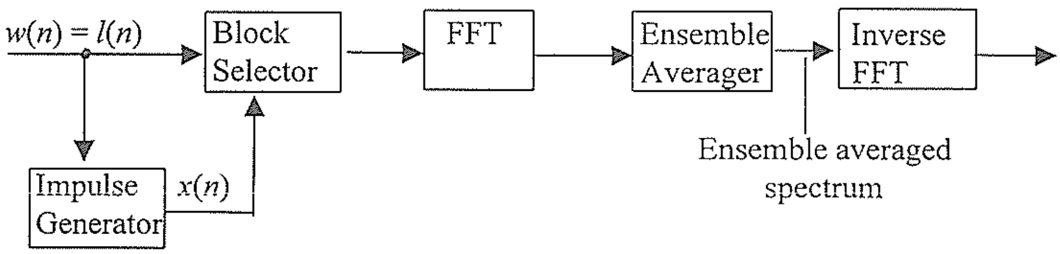


Fig. 5.2 Block diagram for inverse FT of ensemble averaged spectrum for estimation of the leakage sound

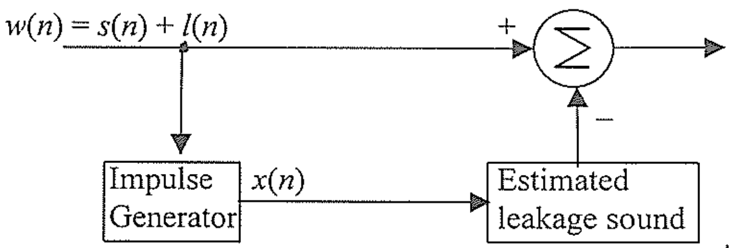
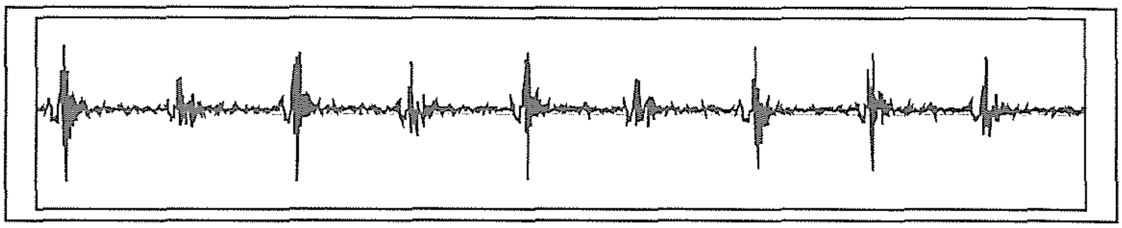
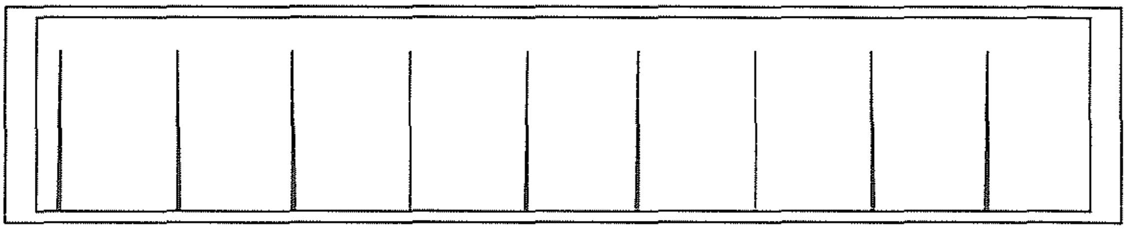


Fig. 5.3 Block diagram for cancellation of leakage sound in normal mode

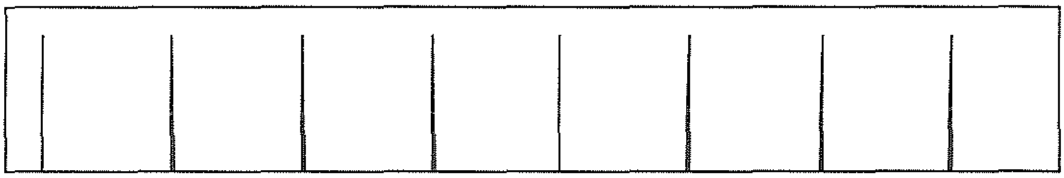


(a)



(b)

(c)



(d)

Fig 5.4 (a) Input $w(n)$ from device 1 (low pitch 1) to the impulse generator
(b) Output $x(n)$ from the impulse generator for $w(n)$ shown in (a)
(c) Input $w(n)$ from device 2 (high pitch 2) to the impulse generator
(d) Output $x(n)$ from the impulse generator for $w(n)$ shown in (c)
(Total time period of all above shown signals is 100ms)

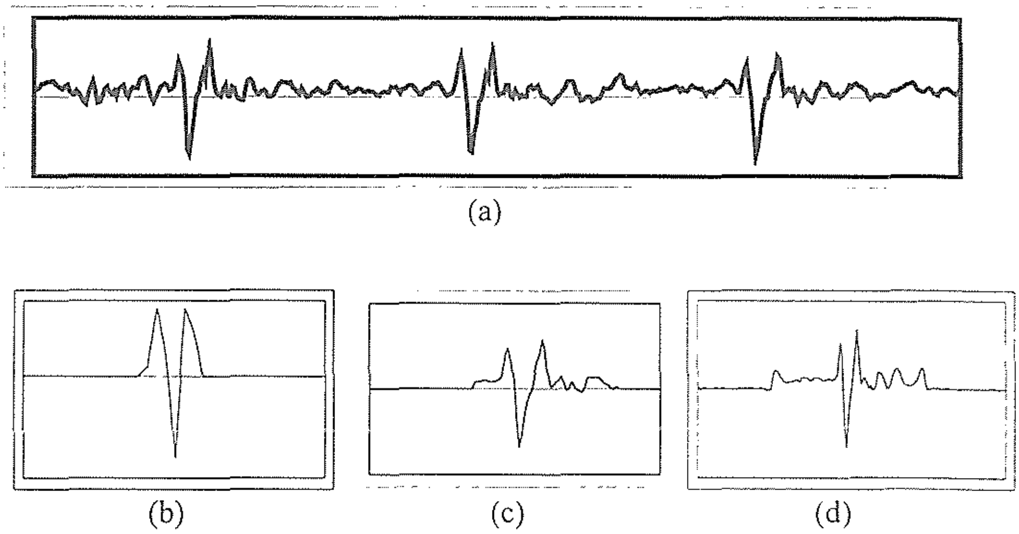
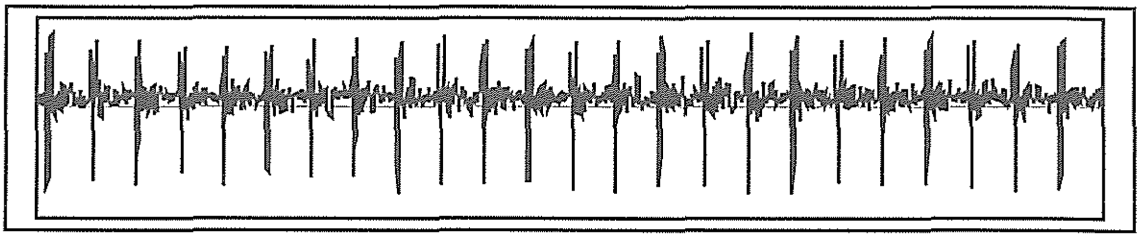
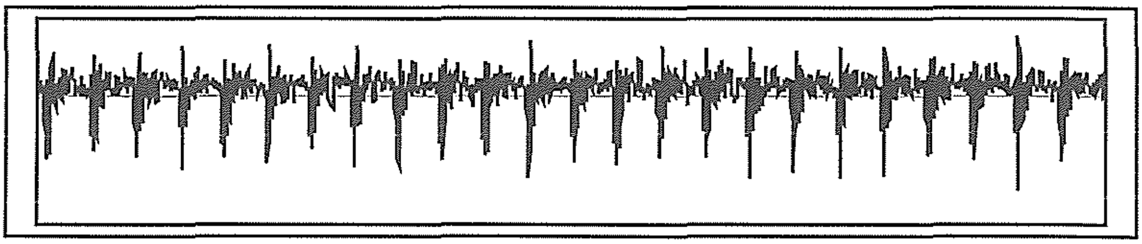


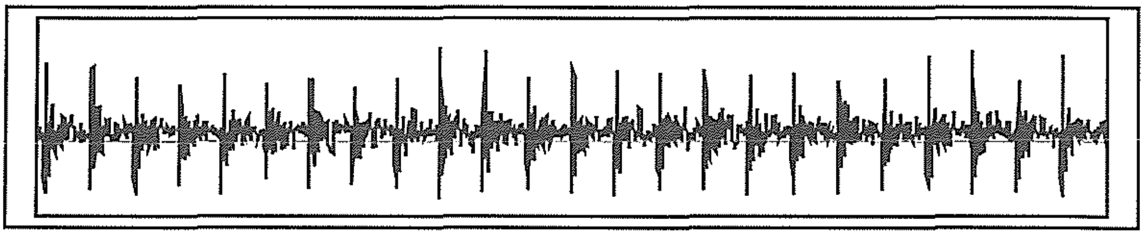
Fig 5.5 (a) Leakage sound from devise2 (high pitch 2) Estimated leakage sound using (b) Single point adaptive leakage sound cancelling (c) Ensemble averaging (d) Inverse FFT of average spectrum (Total time period of signal shown in (a) is 38.4 and signals shown in (b), (c), and (d) is 12.8 ms)



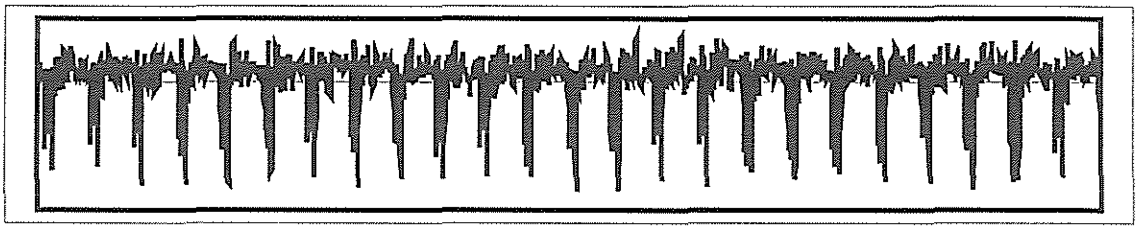
(a)



(b)

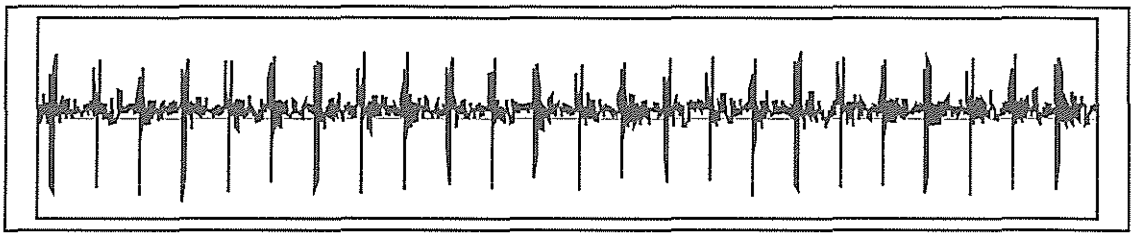


(c)

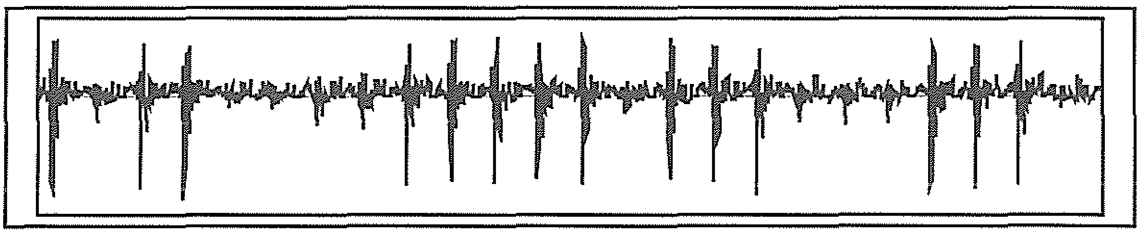


(d)

Fig 5.6 (a) Input $x(n)$ from vibrator of TE2 when person is silent. During estimation, output $e(n)$ from leakage canceller with adaptive leakage cancelling method, using (b) all pulses from the vibrator (c) alternate pulses from the vibrator (d) every third pulse from the vibrator (Total time period of all above shown signals is 300ms)



(a)

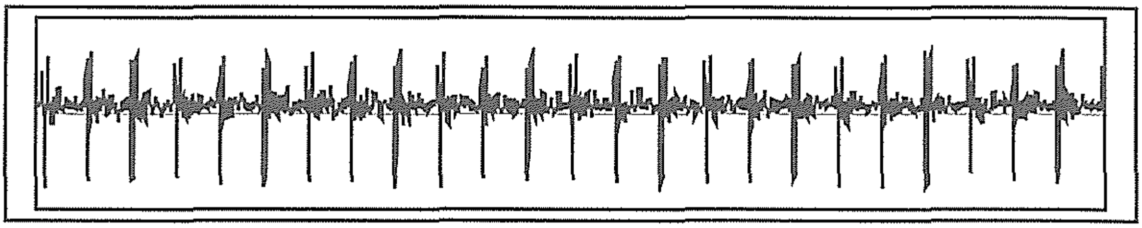


(b)

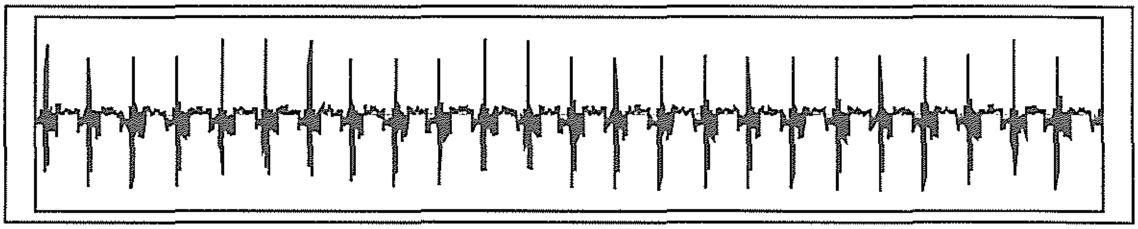
(c)

(d)

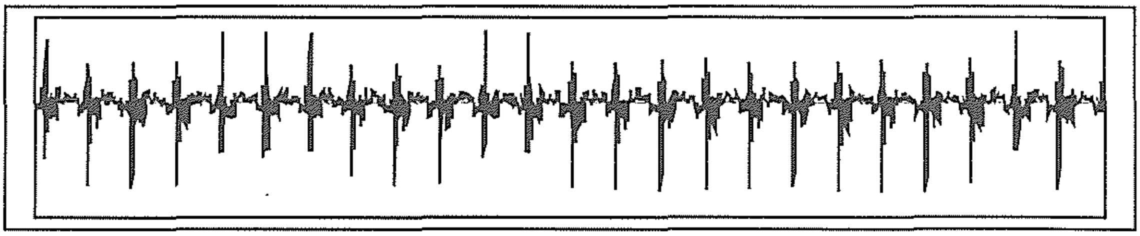
Fig 5.7 (a) Input $x(n)$ from vibrator of TE2 when person is silent During estimation, output $e(n)$ from leakage canceller with ensemble averaging method, using (b) all pulses from the vibrator (c) alternate pulses from the vibrator (d) every third pulse from the vibrator (Total time period of all above shown signals is 300ms)



(a)



(b)



(c)

(d)

Fig 5.8 (a) Input $x(n)$ from vibrator of TE2 when person is silent. During estimation, output $e(n)$ from leakage canceller with inverse FT of average spectrum method, using (b) all pulses from the vibrator (c) alternate pulses from the vibrator (d) every third pulse from the vibrator (Total time period of all above shown signals is 300ms)

Chapter 6

SUMMARY AND SUGGESTIONS

6.1 Summary

Literature survey was carried out to understand the problems in transcervical electrolarynges. The speech obtained using transcervical electrolarynx has low intensity, an unnatural quality and is significantly less intelligible than normal speech. Main source of the degradation in quality of the alaryngeal speech is the presence of leakage sound. The objectives of this project were to study the leakage sound and to investigate methods for cancelling it.

A spectral analysis was done by taking narrowband spectrogram of sound recorded under two conditions: (a) a speaker is not speaking but vibrator is on, and (b) a speaker is uttering vowel /a/ using transcervical device. The resonance bands for the leakage sound are present in the spectrogram for /a/ also, and can be considered as superimposed as the formants for /a/. Also, the low frequency energy deficit in speech produced by this type of device was seen in narrowband spectrogram. The sound output from two devices were analysed. It was observed that alternate pulses from vibrator had different intensity, and shape in device TE1 and pulses were varying in shape in device TE2 for low pitch.

Real time implementation of single input adaptive leakage sound canceller using LMS algorithm was done on TI/TMS32C50 DSK board. There are two modes: training and use. During the training mode, the devices is applied to the throat and kept on, but the speaker keeps his lips closed. Thus the microphone signal corresponds to the leakage sound only. The impulse response of the leakage path is estimated by the LMS algorithm over a large number of excitation pulses from the vibrator. In the use mode, the impulse response of the leakage path for each excitation pulse, which is used as an estimate of the impulse response of the leakage path, is subtracted from the input signal, in order to cancel the background noise. The implementation was not very effective in reducing the background noise. This could be due to non-stationarity of the leakage sound.

It was difficult to study the dynamics of convergence of LMS algorithm in the real time implementation. Hence, an off-line implementation of LMS algorithm along with other two methods namely ensemble averaging, and inverse FT of ensemble averaged spectrum were carried out to estimate the impulse response of the leakage path. None of the three methods could fully cancel background noise because of variation in shape of pulses coming from vibrator. This variation may change from device to device. Reasons for this variation could be change in positioning of the device on the throat, change in the application pressure, and transducer dynamics.

6.2 Suggestions for future work

The spikes in the current drain of the vibration circuit can result in periodic fluctuation in the battery voltage, which will in turn affect the vibrator output. It could be resulting into non-stationary background noise as well as degradation of excitation to the vocal tract. This needs to be studied.

Another approach to estimate and cancel non-stationary background noise is a two input adaptive leakage sound canceller. In a two input leakage sound canceller, one microphone is positioned very close to the lips, and it primarily pick-up the speech signal, with a small amount of leakage sound. The other microphone is used to provide a reference signal. It is positioned very closed to the place where the device is applied to the neck. Several approaches for reduction of interference have been reported. But all these methods basically assume that the interference is uncorrelated to the signal, and the adaptive filter uses the minimisation of the error. In this particular case, both the signal of interest and the background interference are highly correlated, resulting due to the pulse train from the vibrator. Hence, it will require developing a scheme for adaptive cancellation that in some way makes use of properties of speech signal.

The source of background interference is the leakage from the transducer of the vibrator. In the present devices, an electrodynamic or electromagnetic transducer is used. The vibrating diaphragm couples the vibration to the neck tissue. The other surface couples vibration to the air, and this results in background noise. This noise is generally reduced by appropriate shielding. There is a need to investigate the use of piezoelectric or magnetostrictive transducers. In case of these transducers, the vibrations are generated due to deformation in a solid piece. One side of this piece will act as the diaphragm for coupling the vibration to the neck tissue. The other side can rest again a solid object of larger inertia, which is coupled to the body of the artificial larynx, in such a way that vibrations are damped, resulting in very low leakage.

Appendix A

TYPES OF ARTIFICIAL LARYNGES

A review of various types of artificial larynges has been carried from various articles by L.P.Goldstein [6] and Y. Lebrun et al. [7]. Artificial larynges can be classified into two broad areas, as determined by the method of vibration coupling to the vocal tract: internal or external.

A.1 Internal type artificial larynx

Internal larynx can be either pneumatic or electronic depending upon its energy source. Internal pneumatic type artificial larynx is inserted between the trachea and the pharynx. It uses pulmonary air for voicing which leads to a more natural method of speech production but it has leakage problems from the fistulas, which becomes embarrassing and annoying to the patient [6].

Internal electronic artificial larynx can be either implantable type or intraoral type. In the implantable device, excitation sound source is placed at approximately the same location below the pharynx as the natural sound source, the larynx. Since the transducer acts as vibrational source in the vocal tract, distributed excitation effects and harmonic phase-amplitude effects may be avoided [5]. The first intraoral type electrolarynx was developed by Gluck around 1909 [6, 7]. It consisted of an Edison type phonograph cylinder driven by an electromotor. The phonograph output was connected with a receiver metal reed inserted in tube produced sound as air passed through it. However, the air puffs produced by the reed would be blown into the mouth instead of into the pharynx. The quality of sound was poor with the patient only producing single syllables. Czermark's device is considered the prototype of all later external pneumatic devices. In 1892, Hockenegg developed a device that consisted of a bellows, for air supply, connecting with a tube that was inserted through the nose into the pharynx as shown in Fig A.2. Midway into the tube was a reed for sound production. In 1899, George Gottstein designed a device that consisted of a tracheal cannula, rubber tubing, a reed, a valve for air intake, and a mouthpiece as shown in Fig A.3. From aesthetic point of view, his device was pleasing. But the patient could not move his mouth freely due to this feature.

Study on local make pneumatic larynx was also carried out by B.Weinberg et al. [12]. and S.E.Chalstrey et al. [13]. The "Tokyo artificial larynx" was an inexpensive, mechanically simple pneumatic artificial larynx. This device consisted of a stoma cover, a vibratory mechanism and a sound conduction tube. It was used by placing the cover against stoma cover against user's stoma, directing pulmonary air through a vibrator to produce voice, and transmitting the psuedoglottal sound to the vocal tract through a sound conduction tube placed in user's mouth. The "Taiwan tube" was a tracheo-oral shunt consisting of three main parts joined by connecting tubes. Air was exhaled from the trachea through the tube and the resulting sound from the resonator was articulated by the mouth and tongue to produce speech.

The Italian inventor R. Ticehioni concealed an electromagnetic vibrator and oscillator in a tobacco pipe bowl and used the pipe stem as an air duct as shown in Fig A.4. It was provided with frequency control and volume control. A rechargeable battery attached to the bottom of the bowl. Thus, this device was cosmetically pleasing.

Vocal wave may also be produced by holding a vibrator driver against the skin of the neck. The neck tissues transmit the blows received from the vibrator to the air in the vocal tract. An external transcervical electrolarynx was developed by Wright on this principle in 1950 [7]. It consisted of a vibrating electromagnetic diaphragm located at the end of a cylinder which was held by the hand against the throat. A cord connected the cylinder with a portable battery power pack which could be carried in the pocket. The great advances in electronics brought about by the transistor resulted in a transistorized mechanical larynx by Barney, Hawork and Dunn in 1959. It became the Western Electric No.5 type artificial larynx as shown in Fig A.5 which featured a control for pitch variation.

Right now, transcervical electrolarynges are available from several sources: Servox, Nu-voice, Tru-tone, etc. The Servox device (German) is also based on Wright's invention. It is a cylindrical shaped. self contained device as shown in Fig A.6. It has facility for the presetting of volume (intensity) and pitch (fundamental frequency), and there are controls for pitch and volume during use. The Nu-Voice device also has control of pitch and volume along with intra-oral adapter, rechargeable battery, and charger as shown in Fig A.7 (information obtained through Internet). The True-tone device has a single pressure sensitive button for natural voice intonation as shown in Fig A.8. Their sound quality is quite good as made available by the manufacturer over the Internet.

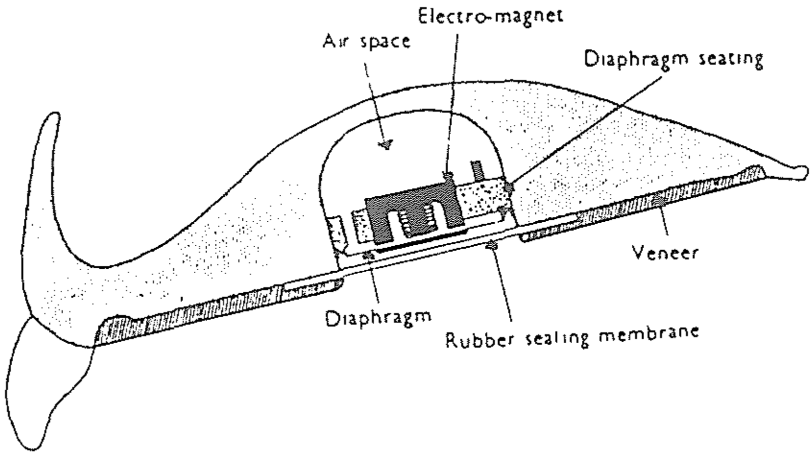


Fig A.1 Tait's oral vibrator [7]

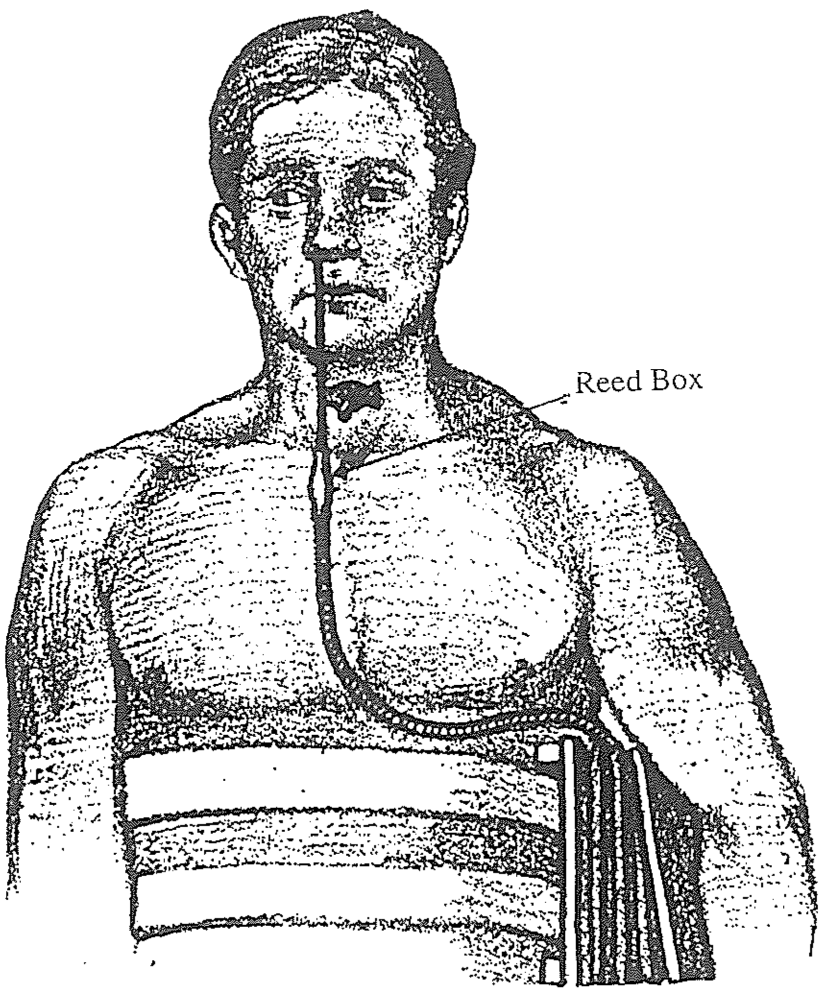


Fig A.2 Hochenegg's external type pneumatic artificial larynx [7]

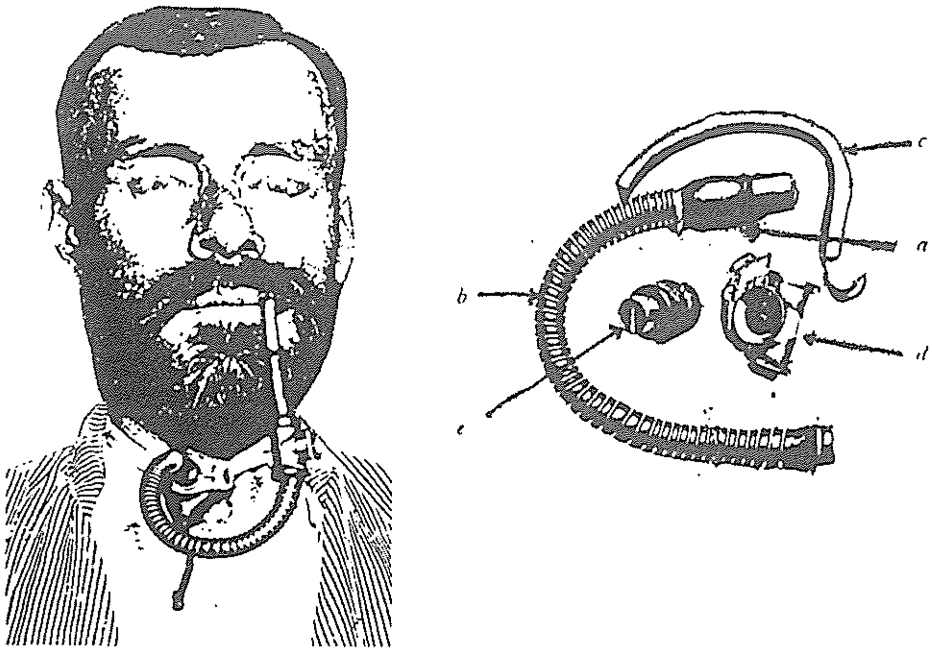


Fig A.3 Goltstein's external type pneumatic artificial larynx
a : connection to the trachea cannulla b : wired rubber tube
c : mouth piece d : inhalatory valve e : reed case [7]

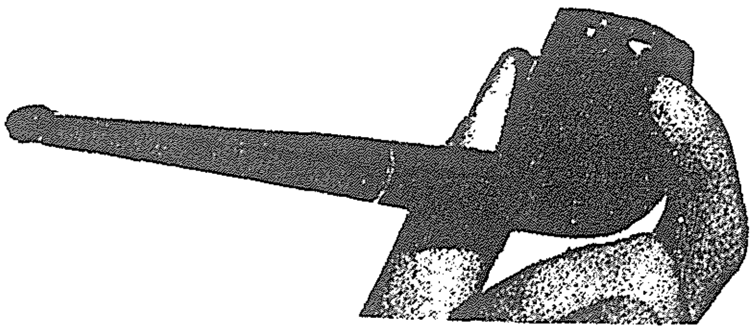


Fig A.4 The Dana pipe [7]

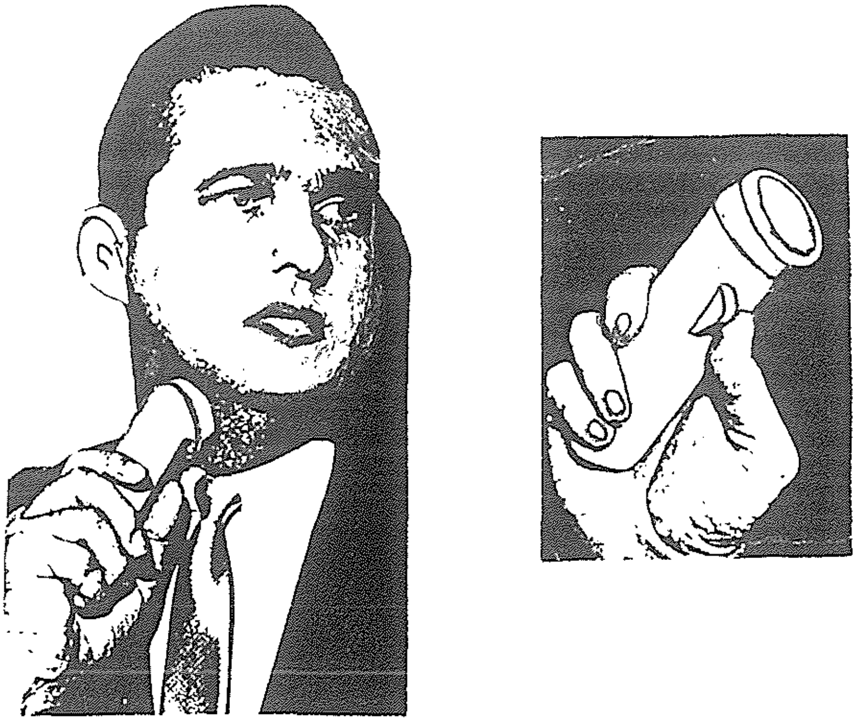


Fig A.5 The Western Electric No-5 transcervical electrolarynx [7]

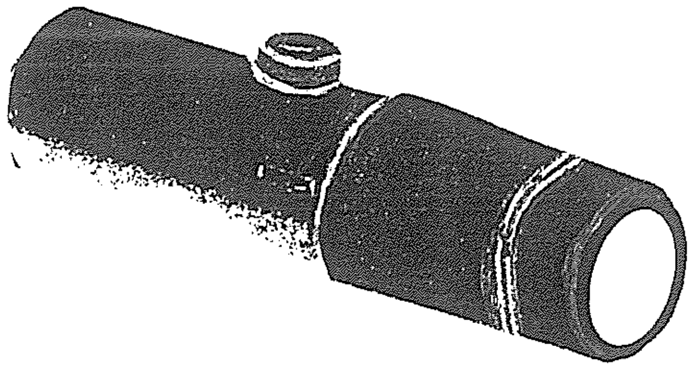


Fig A.6 The Servox transcervical electrolarynx [7]

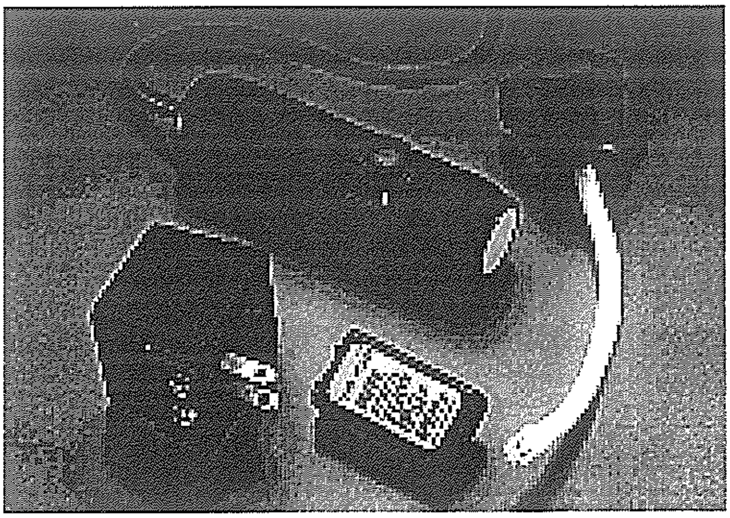


Fig A.7 The Nu-voice transcervical electrolarynx [Internet 1997]

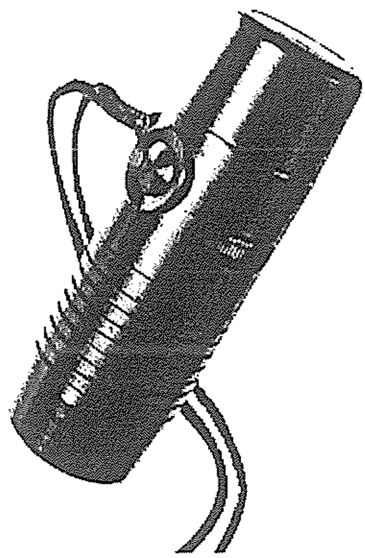


Fig A.8 The True-tone transcervical electrolarynx [Internet 1997]

Appendix B

LEAST MEAN SQUARES ALGORITHM

An adaptive filter structure is shown in the block diagram form in Fig B.1. It consists of two basic parts (1) a transversal filter with adjustable tap weights whose values at time n are $w_0(n)$, $w_1(n)$, $w_2(n)$,...,$w_{M-1}(n)$, and (2) a mechanism for adjusting these tap weights in an adaptive manner. During the filtering process, an additional signal $d(n)$, called the desired response, is supplied along with the usual tap inputs. Estimate of the desired response at the filter output is $y(n)$. By comparing this estimate with the actual value of the desired response $d(n)$, an estimation error denoted by $e(n)$ is given as under.

$$e(n) = d(n) - y(n) \tag{B.1}$$

where

$$y(n) = \sum_{m=0}^{M-1} w_m(n)\, x(n-m) \tag{B.2}$$

We can write the tap weight vector $w_m(n)$ as

$$w^T(n) = [w_0(n),\ w_1(n),\ w_2(n),\ .\ .\ .\ .\ , w_{M-1}(n)] \tag{B.3}$$

and the tap input vector $x(n)$ as

$$x^T(n) = [x(n), x(n-1),\ .\ .\ .\ .\ , x(n-\overline{M}-1)] \tag{B.4}$$

The error $e(n)$ is used to control the adaptive process.

The "least mean squares" (LMS) algorithm is one of the various algorithms for implementing the adaptive process. A significant feature of the LMS algorithm is that it does not require measurements of the pertinent correlation functions, nor does it require matrix inversion [14, 15]. The goal in this method is to adapt toward the minimum mean-square error (MMSE) solution, i.e. adjust the tap weights in order to minimize $E[(e(n))^2]$. When the adaptive filter is realized in the direct form FIR structure, mean squares error performance surface is a quadratic function of the tap weight vector and thus has a single minimum. This is illustrated for the simple one tap weight case (M=0) where $w_0(0)$ denotes the initial condition and w_0* corresponds to the optimal, i.e., the MMSE solution. The goal of the adaptive process is to adjust the tap weights in such a way that they move from the initial condition $w_0(0)$ to the MMSE solution w_0*. In general, for M tap filter, the mean square error surface would be a parabolic function in (M+2) dimensional space.

In non-stationary environments, MMSE solution varies as signal conditions change. So we have to deal with varying performance surface. Thus, the adaptive process must continually adjust the tap weights in order to track the MMSE solution. In adaptive system analysis, however, we typically assume that signal statistics change very slowly, so that signal stationarity may be assumed. The technique utilized by the LMS algorithm to update tap weight vector is based on the method of steepest descent [15]. This can be described in the algorithmic form using vector notation, as follows:

$$w(n) = w(n-1) + (-\nabla(n))\,\mu/2 \qquad \text{(B.5)}$$

where gradient vector $\nabla(n) = \dfrac{\partial E[(e(n))^2]}{\partial w(n)}$

$$= \left[\frac{\partial E[(e(n))^2]}{\partial w_0(n)}, \ldots\ldots, \frac{\partial E[(e(n))^2]}{\partial w_{M-1}(n)}\right]^T \qquad \text{(B.6)}$$

and μ is a parameter that controls the rate of convergence. The tap weight updates are proportional to the negative gradient ($-\nabla(n)$) of the performance surface. Thus, when $\nabla(n)$ is known at each step of the adaptive process, the adjustment always results in a better filter. In addition, once the MMSE solution is found, the gradient reaches zero so the coefficients remain at their optimal values.

Due to the inexact knowledge of the performing surface gradient $\nabla(n)$, it is difficult to implement the above-mentioned algorithm. Various techniques are available for estimating $\nabla(n)$. The approach taken in the LMS algorithm is to use a gradient estimate based on the instantaneous squared error,

$$\nabla'(n) = \frac{\partial\,((e(n))^2)}{\partial\, w(n)} = 2\, e(n)\, \frac{\partial(d(n) - y(n))}{\partial\, w(n)} \qquad \text{(B.7)}$$

Desired response $d(n)$ is independent of filter coefficients $w_m(\mathrm{n})$, and hence

$$\frac{\partial\,(d(n))}{\partial\, w_m(n)} = 0$$

Using equation (B.2) which relates output $y(n)$ to input $x(n)$, we have

$$\frac{\partial\,(y(n))}{\partial\, w_m(n)} = x(n)$$

And therefore we can write equation (B.7) as

$$\nabla'(n) = -2\, e(n)\, x(n) \qquad \text{(B.8)}$$

Using this, we can write equation (B.5) as

$$w(n) = w(n-1) + \mu\, e(n)\, x(n) \qquad \text{(B.9)}$$

Operation of the LMS algorithm requires selection of the adaptation constant μ, the filter length M, and the initial filter tap weight vector $w(0)$. The selection of these parameters is application dependent [14].

The convergence parameter μ plays an important role in determining the performance of an adaptive system. The gradient estimate used in the LMS update is based on the instantaneous error value $(e(n))^2$ rather than the mean value $\mathrm{E}((e(n))^2)$. Although, on the average, this form of update moves the coefficients towards the

MMSE solution, it is apparent that a single update of the $w(n)$ vector could contain a considerable error [15]. Thus, a large μ could result in an adaptive process that never converges to the MMSE solution. Conversely, if μ is too small, the coefficient vector adaptation is very slow. The effects of the inaccurate gradient estimate tend to average out in this case, and although the coefficient vector eventually converges if the signals are stationary, it may not converge in the non-stationary environments. In other words, if μ is too small, the system may not react rapidly enough to cope with the changing signal statistics. To avoid either of these conditions, μ must be neither too large nor too small. It has been shown that the stable range of μ varies according to the input signal power [15], and is given as

$$0 < \mu < \frac{1}{(M+1)\,\sigma^2} \tag{B.10}$$

where M is the number of filter coefficients, and σ^2 is the input signal power.

In practice, μ is generally restricted to a small fraction of this stable range in order to smooth the noisy instantaneous gradient estimate. The tap weight vector, $w(n)$, must be initialized to zero or some other desired initial condition.

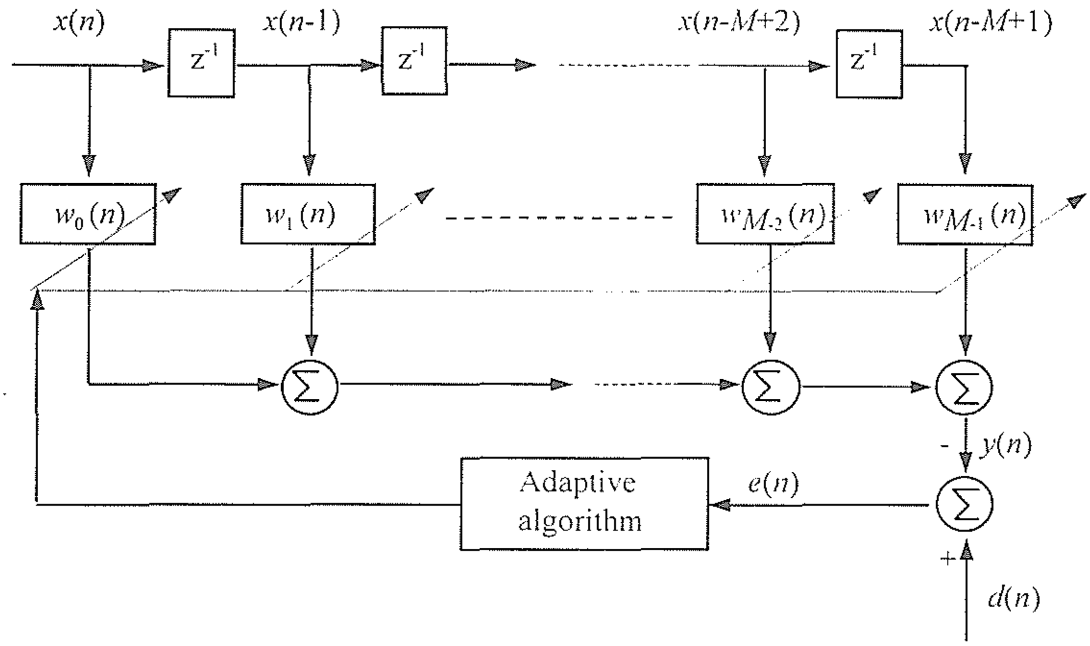


Figure B.1 Structure of adaptive filter
Source Clerkson 1993

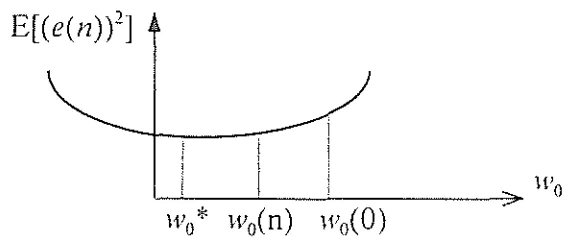


Figure B.2 MSE performance for one-coefficient filter
Source Stearns 1988

Appendix C

SPECTROGRAM

Spectrographic analysis of speech is an important means of studying the characteristics of speech waves. Spectrogram displays the time varying magnitude spectrum with time in horizontal direction, frequency in vertical direction, and intensity by the relatively blackness or brightness. One convenient method among the various methods of spectrogram generation is to compute the short-time Fourier transform from the sampled waveform. A spectrogram software package, which we used in our analysis, has been developed by Baragi [16] and Prasad [17] and has been modified by Chaudhari [18]. This spectrograph generation software package uses a PC with a VGA card, a video display of 640x480 pixel wide, and a DSP card based on TMS320C25 DSP processor. This package has two modules, one module developed in C language will be working on the PC while the other one developed in assembly language of the TMS320C25 DSP processor will be working on the DSP card. Partitioning the tasks appropriately between the PC and the DSP board results in improvement in the speed of analysis and display. The spectrogram using this program is 500 pixels wide and 128 pixel high and is displayed above the display of the selected time segment of the speech signal. The gray scale of the dynamic range is also displayed on the monitor. After displaying the spectrogram, the magnitude of the various frequency components can be obtained using the cursors. This program also provides a facility to store the spectrogram in a file, and also to acquire the speech signal from the microphone and then to generate its spectrogram. Narrowband spectrogram for naturally uttered vowel /a/ using this program is shown in Fig C.1

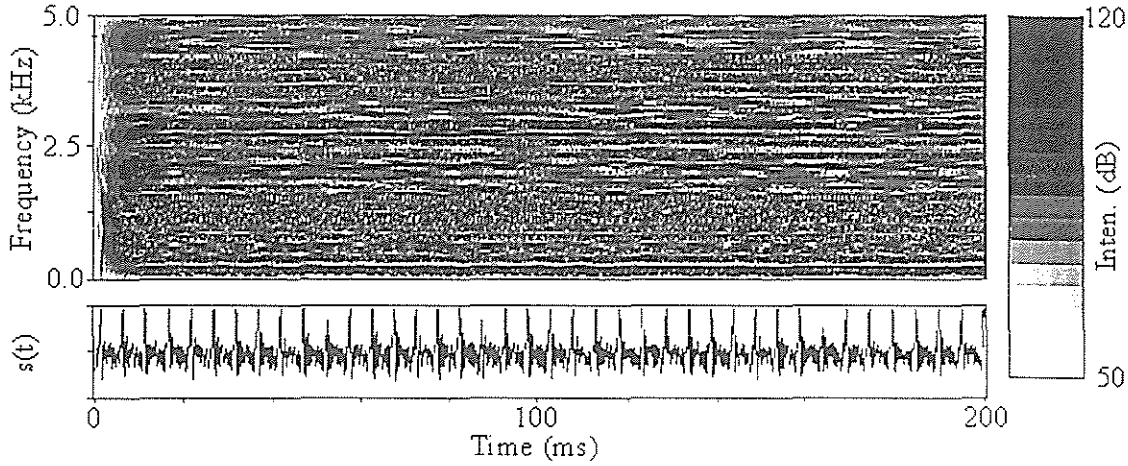


Fig. C.1 Narrowband spectrogram for naturally uttered vowel /a/

In above spectrogram, the fundamental frequency and its harmonics can be clearly seen in the form of horizontal band, as the frequency resolution is good in narrowband spectrogram.

Appendix D

TI/TMS320C50 DSP KIT

The Texas Instruments TMS320 product line contains a family of digital signal processors, designed to support high speed and numeric intensive DSP applications [19]. The key features of the TMS320 family are a highly pipelined architecture, and a comprehensive, efficient, and easily programmable instruction set. It can execute greater than five million instructions per second. It incorporates special instructions e.g. multiply, multiply and accumulate with fast data move etc. It has different addressing modes. It also includes Boolean instructions for bit testing.

Architecturally, the TMS320 utilizes a modified Harvard architecture. In a strict Harvard architecture, the program and data memories lie in two separate spaces, permitting a full overlap of instruction fetch and execution. The TMS320 family's modification allows transfer between program and data space. This eliminates the need for a separate coefficient ROM and also maximizes processing power.

The TMS320C50 is a 32-bit fixed-point accumulator based microprocessor of the TMS320 family. It can execute instructions at a speed of greater than 28 MIPS with a 35-50 ns fixed point instruction execution time. It has a 9K X 16 bit single cycle on chip program/data RAM and its memory operation is RAM based. Its key features are:

- Enhanced and modular architectural design.
- Enhanced instruction set for faster algorithms.
- 32 bit ALU, accumulator and accumulator buffer.
- 16 bit parallel logic unit.

- 32 bit ALU, accumulator and accumulator buffer.
- 16 bit parallel logic unit.
- 16 X 16 bit parallel multiplier with 32 bit product capability.
- Single cycle multiply/accumulate instructions.
- Eight auxiliary registers for indirect addressing.
- Eleven registers for storing CPU controlled registers during Interrupt Service routine.
- Eight level Hardware Stack.
- 0 to 16 bit left and right data barrel shifters and 64-bit incremental data shifter.
- Two indirectly addressed circular buffers.
- Single instruction repeat and block repeat instructions.
- Block memory move instructions.
- Bit reversed and index addressing mode.
- Full duplex synchronous serial port for direct communication between C50 and another serial device.
- 4 deep pipelined operation for delayed branch, call, and return instruction.

The TMS32C50 performs two's complement arithmetic and Boolean operations using 32 bit ALU and accumulator (ACC). The accumulator buffer is used for temporary storage. The parallel logic unit (PLU) executes logic operations on data for bit manipulation ability. The multiplier performs 16 X 16 bit two's complement multiplication with a 32-bit result in a single instruction cycle. The scaling shifter produces a left shift of 0 to 16 bits on input data. Additional abilities include numerical scaling, bit extraction, extended arithmetic and overflow protection options. Eight levels of hardware stack save the contents of program counter during interrupts and subroutine calls. On interrupts, the strategic registers are pushed onto a one deep stack and popped upon interrupts.

The TMS32C50 has a high degree of parallelism, i.e. while data is being operated by ALU, arithmetic operations may also be executed in the arithmetic register unit. Such parallelism results in a powerful set of arithmetic, logic and bit manipulation operations that may all be performed in a single machine cycle.

The TMS320C5x DSP Starter Kit (DSK) is a standalone application board that allows us to experiment and use the DSP for real time signal processing [20]. The DSK consists of the TMS320C50 DSP for full speed verification of the source code and analog interface circuit TLC32040 that interfaces to the TMS32C50 serial port. Bi-directional communication between a PC and the starter kit is through an RS232 serial port by using BIO and XF of the TMSC50 chip. The starter kit needs a 9V ac, 250-mA power supply.

The DSK assembler and debugger are software interfaces that help to develop, test and refine DSK assembly language programs. The assembler translates assembly language source code into executable object code allowing us to work with mnemonics rather than hexadecimal machine instructions and to reference memory locations with symbolic address. The debugger is windows oriented and is capable of loading and executing code with single step breakpoint and runtime halt capabilities. It separates code, data and commands into manageable portions.

The TLC32040 Analog Interface Circuit (AIC) interfaces to the TMS320C50 serial port [20]. It provides a single channel input/output voice quality analog interface. It integrates a bypassable band-pass switched capacitor anti-aliasing input filter, a low-pass switched capacitor output reconstruction filter, 14 bit resolution A/D and D/A converter and four microprocessor compatible serial port modes. It offers numerous combinations of master clock input frequencies and conversion/sampling rates. which can be changed via digital processor control (up to 19,200 samples per second). This is done by changing counts in the TX Counter A, RX Counter A, TX Counter B, RX Counter B as the sampling rate is derived from values contained in these registers. We can also choose between synchronous and asynchronous ADC and DAC conversion rates, with programmable incremental timing adjustments. There is also a serial port interface to serial-parallel shift register for parallel interfaces to DSPs.

References

[1] C. Y. Espy-Wilson, V. R. Chari, "Enhancement of alaryngeal speech by adaptive filtering", *Proceedings ICSLP 96*, pp. 764-771, 1996

[2] R. Warwick, P. L. Williams, *Gray's Anatomy*. London: Longman, pp. 1229-1239, 1978

[3] N. C. Roussel, "Acoustic properties of alaryngeal speech", *Acoustic Society of America*, pp. 1250-1256, 1995 (Abstract)

[4] Qi Yingyong, B. Weinberg, "Low-frequency energy deficit in electrolaryngeal speech", *Journal of Speech and Hearing Research*, pp. 1250-1256, 1991

[5] R. L. Evevett, B. J. Bailey, "Implanted electrolaryngeal prosthesis", *Electrostatic Analysis and Enhancement of Alaryngeal Speech*, pp. 205-235, year not known

[6] L. P. Goldstein, "History and development of laryngeal prosthetic devices", *Electrostatic Analysis and Enhancement of Alaryngeal Speech*, pp. 137-165, year not known

[7] Y. Lebrun, Swets and Zeitlinger, "History and development of laryngeal prosthetic devices", *The Artificial Larynx*, pp. 19-76. 1973

[8] D. O'Shaughnessy, *Speech Communication: Human and Machine*. Reading, Massachusetts: Addision-Wesley, 1987

[9] L. R. Rabiner, R. W. Schafer, *Digital processing of speech signals*. Englewood Cliffs, New Jersey: Prentice-Hall, 1978

[10] S. Gracias, "*A speech training aid for the deaf*", M. Tech. dissertation, Electrical Engineering Department, IIT Bombay, 1991

[11] P. Gavankar, "*A speech training aid for the deaf*", M. Tech. dissertation, Electrical Engineering Department, IIT Bombay, 1995

[12] B. Weinberg, A. Riekena, "Speech produced with the Tokyo artificial larynx", *The journal of speech and hearing disorders*, vol. 38, pp.383-389, 1973

[13] S. E. Chalstry, C. A. Van Haselt, "A pneumatic artificial larynx popularized in Hongkong", *The journal of laryngology and otology* vol. 108, pp. 852-854, 1994

[14] P. M. Clerkson, *Optimal and Adaptive Signal Processing.* Boca Raton, Florida: CRC press, pp. 164-166, 1993

[15] S. D. Stearns, *Signal Processing Algorithms.* Englewood Cliffs, New Jersey: Prentice-Hall, pp. 252-269, 1988

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

[17] V. V. Siva Rama Prasad, "*Speech processing for single channel sensory aid*", M. Tech. dissertation, Electrical Engineering Department, IIT Bombay, 1996

[18] D. S. Chaudhari, P. C. Pandey, "A DSP based spectrographic analyzer", *Proc. National Conference on Electronic Materials, Devices & Systems,* Gulbarga, Karnataka, Jan. 18-20, 1999 (to be published).

[19] *TMS320C5x User's guide*, Texas Instruments U.S.A., 1993

[20] *TMS320C5x DSP Starter Kit User's guide*, Texas Instruments U.S.A., 1994