REDUCTION OF BACKGROUND NOISE IN ARTIFICIAL LARYNX

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

Master of Technology

by

Santosh M. Bhandarkar
(00307053)

Guide: Prof. P.C. Pandey

Department of Electrical Engineering
Indian Institute of Technology, Bombay
January 2002

INDIAN INSTITUTE OF TECHNOLOGY, BOMBAY

M. TECH DISSERTATION APPROVAL

Dissertation entitled : Reduction of background noise in artificial larynx

by **Santosh M. Bhandarkar**

is approved for the award of the degree of MASTER OF TECHNOLOGY.

Guide	: ……	(Prof. P. C. Pandey)
Internal Examiner	: ……	(Prof. Preeti Rao)
External Examiner	: ……	(Dr. V. K. Madan)
Chairman	: ……	(Prof. S. S. Pande)

Date:

Santosh M. Bhandarkar / Prof. P.C. Pandey (Guide), "Reduction of background noise in artificial larynx", *M.Tech dissertation*, Department of Electrical Engineering, Indian Institute of Technology, Bombay, January 2002.

ABSTRACT

The transcervical artificial larynx is of great help to people who cannot use their natural voice production mechanism. The device is held against the neck, and the vibrations generated move up the vocal tract to produce useful speech. The resulting speech has poor quality due to the presence of background noise. The background noise is due to the leakage of the acoustic energy. The objective of the project is to investigate signal-processing techniques for reducing the background noise, thereby improving the quality of the speech output. The spectral analysis of the speech generated with a transcervical electrolarynx was carried out. An earlier proposed method estimates the impulse response of the leakage path during the training phase and then subtracts estimated noise from the noisy speech during use mode. It did not result in any significant noise reduction, possibly due to significant changes in the impulse response of the leakage path.

After formulating a theoretical basis, spectral subtraction method is used for noise reduction. Average magnitude spectrum of noise, obtained with lips closed in training mode, is subtracted from the magnitude spectrum of the noisy speech and the signal is reconstructed using the original phase spectrum. It is observed that effective noise cancellation is obtained, if the noise estimation and subtraction is done using 2-pitch frames. The method improves the intelligibilty of the speech output.

ACKNOWLEDGEMENTS

I am grateful to my guide Prof. P. C. Pandey for his constant encouragement throughout the project tenure. His thorough guidance helped me in studying the subject of artificial larynx, obtain good results, and prepare this report.

I thank Dr. Gurmeet Baccher from the Tata Memorial Hospital, for discussions on the problem, and help in recordings using the artificial larynx. I am also thankful to the inmates of the SPI lab, Mr. Dipak Patel, Mr. Bharat Nihalani, Ms. Alice Cheeran, Mr. Parveen Lehana, Ms. Dakshayani Jangamashetti, Mr. Dinesh Choudhary, and Mr. Vidhyadhar Kamble for their support and valuable suggestions.

January 2002 Santosh M. Bhandarkar

CONTENTS

LIST OF FIGURES

Chapter 1

INTRODUCTION

1.1 Background

In normal speech production mechanism, the lungs provide the air stream, the vocal chords in the larynx provide the vibration source for the sound, and the vocal tract provides the spectral shaping of the resulting speech. In some cases of disease and injury, the larynx is surgically removed by an operation known as laryngectomy, and the patient (often known as a laryngectomee) needs external aids to communicate. In the post-operation period, he can use visual signs, or esophageal speech, or artificial larynx for communication. The esophageal speech method requires the laryngectomee to force the air through the esophagus instead of trachea, and produce sound. This method produces natural sound, but is difficult to learn.

An artificial larynx is a device used to provide excitation to the vocal tract (as a substitute to that provided by a natural larynx). The external electronic larynx or the transcervical electrolarynx is the widely used type of device. It is hand held and pressed against the neck. The vibrations produced get coupled to the neck, move up the vocal tract and produce useful speech, when spectrally shaped by the vocal tract articulators. The device is easy to use and portable. However the speaker needs to control the pitch and volume switches to prevent monotonic speech, and this needs practice. The speech produced is generally deficient in low frequency energy due to lower coupling efficiency through the throat tissue. The unvoiced segments generally get substituted by the voiced segments. In addition to these, the major problem is that the speech output has a background noise, caused by the leakage of the acoustic energy from the vibrator.

1.2 Project objective

The external electrolarynx suffers from the problem of the background noise, which reduces the quality of the output speech considerably. The transducer of the vibrator is the primary source of background noise. Effective acoustic shielding of the vibrator should reduce the noise. However studies have shown that shielding results only in a marginal reduction of the noise [1]. The objective of this project is to

investigate signal processing techniques for reducing the background noise, so as to improve the quality of the output speech.

The implemented algorithms take a single input and have two modes of operation namely the training mode and the use mode. In the training mode, the noise forms the input, while during the use mode, the noisy speech forms the input. The algorithm estimates the noise in the training mode and subtracts it from the noisy speech in the use mode, resulting in better quality sound output. The algorithms implemented are based on (a) estimation of the impulse response of the leakage path and (b) magnitude spectrum of background noise. These are tested for reduction of the background noise and for improving speech intelligibility.

1.3 Outline of the report

Chapter 2 provides a brief description of the different types of artificial larynxes. This chapter also deals with the problems associated with transcervical electrolarynxes and the methods to overcome them. Chapter 3 deals with the characteristics of the alaryngeal speech, and reviews the various noise reduction methods which use signal processing techniques. Chapter 4 describes the implementation of the algorithms described in Chapter 3. The results obtained by implementing various algorithms are presented and compared with each other. Chapter 5 gives a summary of the report, the conclusions that can be drawn from the results and the scope for future work.

Chapter 2

ARTIFICIAL LARYNX

2.1 Voice production

Fig. 2.1 shows the section of the human head and the larynx [2]. The vocal chords are positioned in between the trachea and the pharynx. The region above the larynx, known as the vocal tract, acts as an acoustic filter shaping the spectrum of the radiated sound. It consists of the oral and nasal tracts. The oral tract begins at the vocal chords or the glottis, and ends at the lips. The cross-sectional area of the oral tract is determined by the positions of the tongue, lips, jaw, and velum. The nasal tract begins at the velum and ends at the nostrils. The contraction and expansion of the lungs supply the energy required for the speech production. The speech is the acoustic wave, propagated either through the oral tract or the nasal tract or both, and radiated from the end of the vocal tract. Various speech sounds are produced by varying the vocal tract configuration (by changing the position of the velum and movement of articulation in the oral cavity) and by varying the modes of excitation of the vocal tract.

Fig. 2.2 shows a speech production system [3], comprising of the lungs, the vocal chords, and the vocal tract. The production of the voiced sounds requires the vocal chords to operate in relaxation oscillation mode, thereby producing quasi-periodic pulses of air, which vibrate the vocal tract. The unvoiced sounds are produced by forming a constriction at the vocal chords or at some point in the vocal tract, and forcing air through the constriction at high velocity to cause turbulence. The production of the plosive sounds requires complete closure (generally at the end of the vocal tract), building up the pressure behind the closure and then releasing it suddenly. As the sound generated propagates down the vocal or the nasal tract, the frequency selectivity of the tract shapes the frequency spectrum of the sounds. The resonant frequencies of the sound are known as formant frequencies, and depend upon the shape and size of the vocal tract. Different sounds are produced by varying the shape of the vocal tract.

2.2 Artificial larynx

There are various causes for the malfunctioning of the larynx, such as larynx cancer, yelling, smoking etc. In many cases of throat cancer, the larynx is surgically removed to prevent the cancer from spreading. The operation, known as laryngectomy, results in a hole in the neck called stoma. Such a patient, known as laryngectomee, breathes through the stoma, and hence loses the natural speech production mechanism. In order to communicate, such persons need external aids.

In the post operation period, a laryngectomee can use either written means or signs, or esophageal speech, or an artificial larynx to communicate. Written means or signs are inconvenient as the other person has to understand what the laryngectomee intends to say, by looking at the signs. The esophageal speech method is a good alternative as it produces speech closer to the natural speech. In this method of speech production, the air is forced through the esophagus [4]. The resulting air vibrates the walls of the throat, thereby producing sound. Learning to produce esophageal speech is difficult, and cannot be done easily in the post operation period. Artificial larynx is the commonly used mode of communication in the immediate post operation period. It is a device that generates vibrations, which are substituted for the air puffs generated by the vocal chords. Learning to produce speech using an artificial larynx is often convenient and done easily. The speech produced is more intelligible than that by other methods. However one has to bear with an unnecessary crutch, noisy speech, tube maintenance, battery dependence and so on. Despite these problems, artificial larynx is most widely used.

A number of artificial larynxes have been developed [2][5][6], and these can be broadly classified into: external and internal pneumatic, intra-oral and implantable electronic, and external electronic (or transcervical) types.

2.3 Pneumatic larynx

The pneumatic artificial larynxes make use of the air exhaled out from the lungs to produce the vibrations. Based upon the placement of the artificial larynx, these are sub-classified into two groups as external pneumatic larynxes and internal pneumatic larynxes.

The first prototype of the external pneumatic larynx was developed by Johann Czermark in 1854 [2][5]. This device consists of a tube fitted from stoma to the

mouth. A vibrating reed is fitted into the tube. During exhalation, the air from the lungs moves out through the stoma. This air passing through the tube makes the reed vibrate and produces vibrations. These vibrations are moved to the posterior of mouth. The quality of the sound produced with such an artificial larynx is inferior, as one could produce single syllables only. In 1892, Hockenegg developed a device consisting of bellows for air supply, connecting with a tube that was inserted through the nose into the pharynx [2][5], as shown in Fig. 2.3. A vibrating reed was placed at the center of the tube. The devices like Western Electric No. 1, Osaka artificial larynx, Tokyo artificial larynx, Van Humen artificial larynx, belong to the external pneumatic type. Most of these devices are only of historical importance today [5].

The first internal pneumatic larynx was developed by Dr. Billroth in 1873 [5]. He removed the larynx, but left a passage between the windpipe and the pharynx. Later Joseph Leiter devised a prosthesis that was inserted between the trachea and the pharynx. It consisted of three cannulas namely tracheal, laryngeal, and phonatory cannula, with a metal reed. Fig. 2.4 shows such a device. The upper end of the laryngeal cannula had a lid held open by a spring. This behaved like an epiglottis. To speak, the speaker closed the stoma with his fingers and exhaled through the laryngeal cannula. The exhaled air set the reed into vibrations. Since the pulmonary air was used for the production of the sound, this method of sound production closely resembled the natural method of sound production. In 1874, Gussenbaeur developed an improvement over Leiter's artificial larynx. Several other types of internal artificial larynxes were reported [5]. Among these, the prominent were the ones developed by Wolf (1893), Taub (1972), and Sisson (1975). All these devices had the risk of leakage and aspiration [5].

2.4 Internal electronic larynx

In an internal electronic larynx, a vibration generator or a transducer produces the vibrations. The vibration generator consists of an electronic circuit, which produces periodic pulses of a particular frequency. The vibrations produced move up the vocal tract and produce speech. There are two types of internal electronic larynxes: implantable and intra-oral.

In the implantable device, excitation source is placed at approximately the same location below the pharynx as the natural sound source, the larynx. In 1957,

Pickler invented an electrolarynx comprising of four parts, a battery powered signal generator carried in the patient's pocket, an antenna worn around the patient's neck, a tiny receiver affixed to the patient's denture, and a switch located in the patient's tracheal cannula [2][5]. The drawback of the Pickler's device was a large energy loss between the generator and receiver.

The first intra-oral type artificial larynx was developed by Gluck in 1909 [2][5]. It consisted of an Edison type phonograph cylinder, driven by an electromotor. The output of the phonograph was connected to a receiver, which directly fitted the patient's nose. The cylinder of the phonograph consisted of damped vowel sounds, which energized the receiver membrane. The air vibration in the oral cavity resulted in voice production.

The electrolarynx developed by Tait in 1959, consisted of an artificial palate fitted on to the upper part of the denture [2][5]. The palate consisted of an oscillator and a battery. A rubber membrane covered the palate. This prevented the palate from coming into contact with saliva. The artificial larynx could be turned off by moving the tongue over the posterior part of the palate and onto the central part to turn on. When turned on, the vibrations produced were let into the posterior part of the mouth. However the sound produced with this device was of poor quality [5].

2.5 External electronic larynx

An external electronic larynx or a transcervical larynx consists of an electronic vibration generator, the output of which is tightly coupled to the neck. The vibration generator consists of a pulse generator circuit, which powers a coil wound around a pole piece. A permanent magnet is present near the coil. A steel plate connected to the diaphragm vibrates in accordance with the net magnetic field [7]. Fig. 2.5 shows an electromagnetic transducer. The vibrator is properly housed, with the vibrating steel plate at the top. The steel plate is pressed against the neck and the device is turned on. Among the various external electronic larynxes that were developed, the Western Electric No.5, Aurex, Neovox, and Servox were popular [5][6]. The larynxes, which are popular currently, are mentioned in Appendix A.

The pneumatic artificial larynxes are bulkier in size, but the quality of sound resembles the natural sound, due to the use of pulmonary air for speech production. Electronic larynxes, on the other hand, are smaller in size and convenient to use.

However they produce a strong background noise, which degrades the quality of speech output considerably. The problems associated with external electronic larynxes are described in next section.

2.6 Problems in artificial larynx

The main problems associated with the external type electronic artificial larynx [7] are

1. Difficulty in coordinating controls.
2. Spectral deficit.
3. Background noise generation.

The electronic artificial larynx is a hand held device, which has to be coupled to the neck during operation. The speaker, in addition to turning on the device, has to manipulate the pitch control to prevent a monotonous speech. This requires simultaneous movement of the vocal tract articulators (hand, mouth, lips, and teeth), which are difficult to coordinate and may either lead to a monotonous sound or improper words until the speaker attains the necessary expertise.

The artificial larynx when coupled to the neck causes the vibrations to propagate through the neck tissue on to the vocal tract. The neck tissue is a highly non-uniform mass of muscle and membrane. When the sound propagates through such a medium, there is an amplitude variation and phase shift of various harmonics of the impressed sound wave. This change in amplitude and phase is because of the mass-spring-viscous damping effect [8]. Secondly, since the transmission loss is inversely proportional to frequency, the low frequency components in the signal are attenuated. Sometimes the vibrations may not propagate through the medium at all. Such is the case when the neck muscles have thickened due to the radiation generally given after the laryngectomy operation [8]. So the speech produced by an artificial larynx lacks in low frequency components.

The major problem encountered in an electronic artificial larynx is a steady background noise [1][2]. This background noise is generated due to the leakage of the vibrations produced. The front end of the vibrator membrane/plate is coupled to the neck. The back end of the membrane/plate is coupled to the air in the instrument housing. Leakage of the acoustical energy from the housing to the air outside is responsible for the production of the noise. This noise is present even if the speaker's

lips are closed. The leakage noise gets added to the sound from speaker's lips. The listener is presented with the speech (from the lips) along with the leakage noise. Leakage of vibrations from the front end of the vibrator membrane/plate due to improper coupling of the vibrator to the neck tissue also contributes to the background noise.

The problem of background noise generation is considered and means of reducing it are discussed in the following sections.

2.7 Leakage model

Fig. 2.6 shows the block diagram of sound production mechanism of the artificial larynx. The pulse generator produces the pulses, which excite the vibrator, thereby producing vibrations in its diaphragm. Two controls, namely the pitch and the intensity, are provided. The pitch setting determines the frequency of the pulses and hence the vibrations. By varying the pitch the monotonic speech can be prevented. The intensity control determines the magnitude of the pulses, magnitude of the vibrations and hence the intensity of the speech produced. The vibrations produced are coupled to the neck, from which they move up the vocal tract, and produce speech.

Fig. 2.7 shows the model of the leakage sound generated during the use of the external electronic artificial larynx. The vibrations generated by the diaphragm have two paths. The first path is through the neck on to the vocal tract and the second one is through the surroundings known as the leakage path. The impulse response of the first path $h_v(t)$ is dependent on various factors such as the length and configuration of the vocal tract of the speaker, the place of coupling of the vibrator, the amount of coupling, etc. The excitation pulse $e(t)$ when passed through the vocal tract filter delivers the useful speech $s(t)$. The leakage of the vibrations is perceived as the component $l(t)$, which gets added to the useful speech $s(t)$, thereby deteriorating its intelligibility.

2.8 Noise reduction

The background noise, discussed in the previous section, can be reduced or cancelled by acoustic shielding, vibrator design, or noise cancellation techniques. These methods are described below in the following subsections.

2.8.1 Acoustic shielding

The casing of an artificial larynx is designed to provide a good seal against the sound leakage. Its effectiveness reduces with time and servicing. Further a sound proof shield may be wrapped around the artificial larynx but it fails to provide effective noise isolation [1]. The shielding increases the size of the artificial larynx and causes inconvenience in its holding. It also counterbalances the damping provided by hand holding the device.

2.8.2 Redesigning the vibrator

By proper vibrator design, the noise at source can be reduced. Vibrators based on piezoelectric or magnetostrictive effect can be used. One end of the vibrator can be attached firmly to the instrument case, while the vibrations at the other end are coupled to the neck tissue. Due to the mass of the case and damping due to holding, leakage from the fixed end can be reduced. However such vibrators have poor efficiency. Author could locate only one reference about the use of piezoelectric crystal [9]. It is to be noted that vibrator design cannot help in reducing the leakage from the vibrator to tissue interface.

2.8.3 Noise canceller algorithms

Noise reduction can be done using various signal-processing algorithms. These algorithms estimate the noise present in the signal. In the algorithm's used, there are two modes of operation namely training mode and the use mode. In the training mode, only the background noise is the input. This is done, by turning on the artificial larynx, and the speaker keeping his lips closed. The algorithm operates on this noise segment to obtain a noise estimate. In the use mode, the speaker speaks with the artificial larynx on. The speech corrupted with the background noise is the input. The algorithm subtracts the previously estimated noise from the speech corrupted by noise. This subtraction leads to the cancellation /reduction of the noise from the useful speech. Chapter 3 provides a review of some of these techniques. The following chapter provides implementation of an earlier technique and a new technique, and experimental investigations.

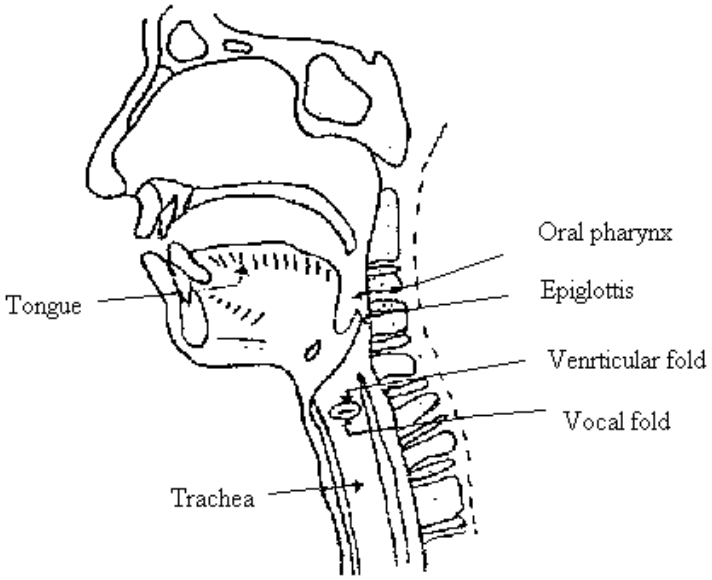


Fig. 2.1 Section of human head and larynx [2]

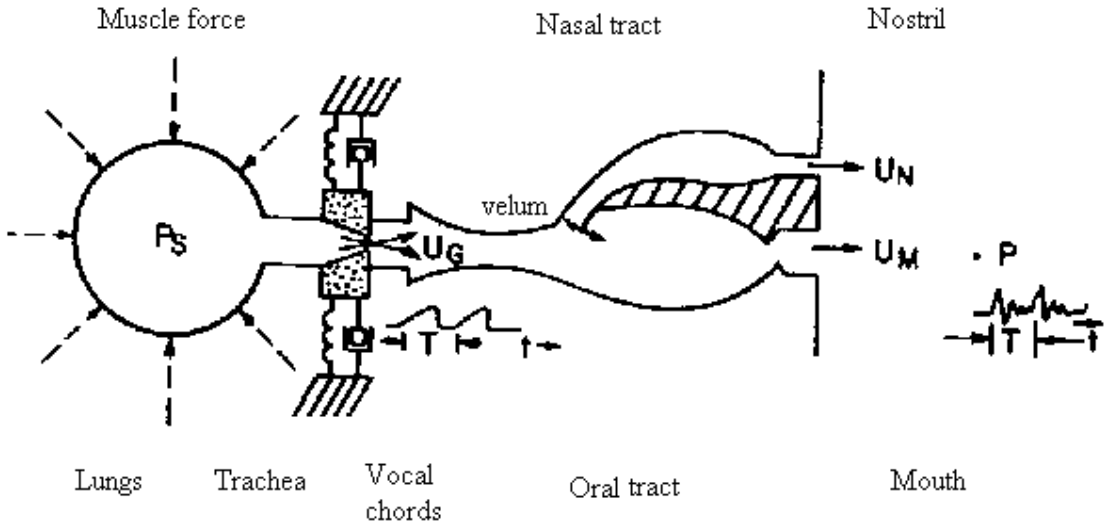


Fig. 2.2 Sound production mechanism [3]

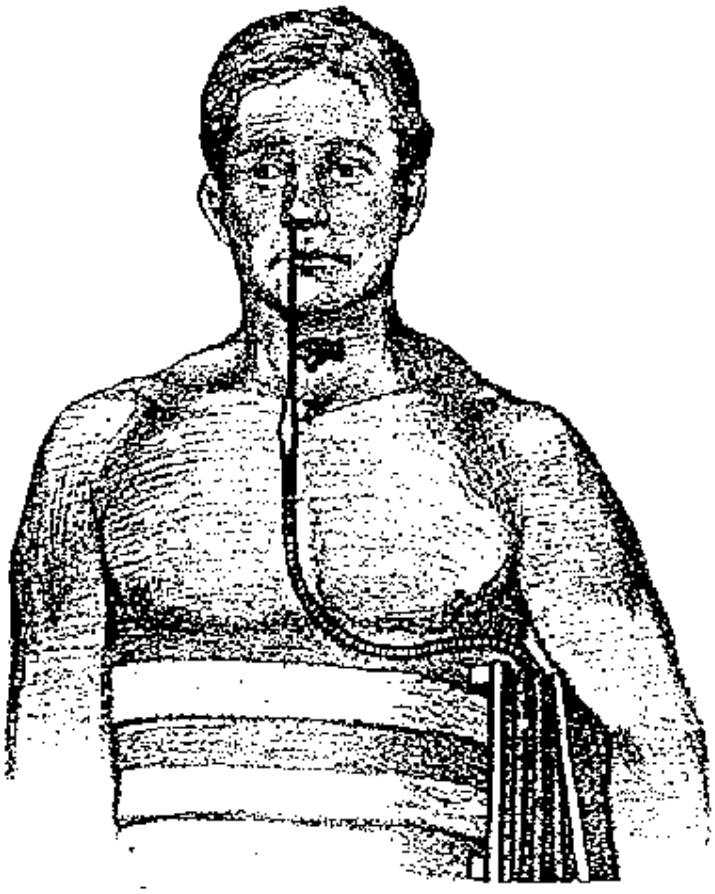


Fig. 2.3 Hockenegg's external pneumatic larynx [5]

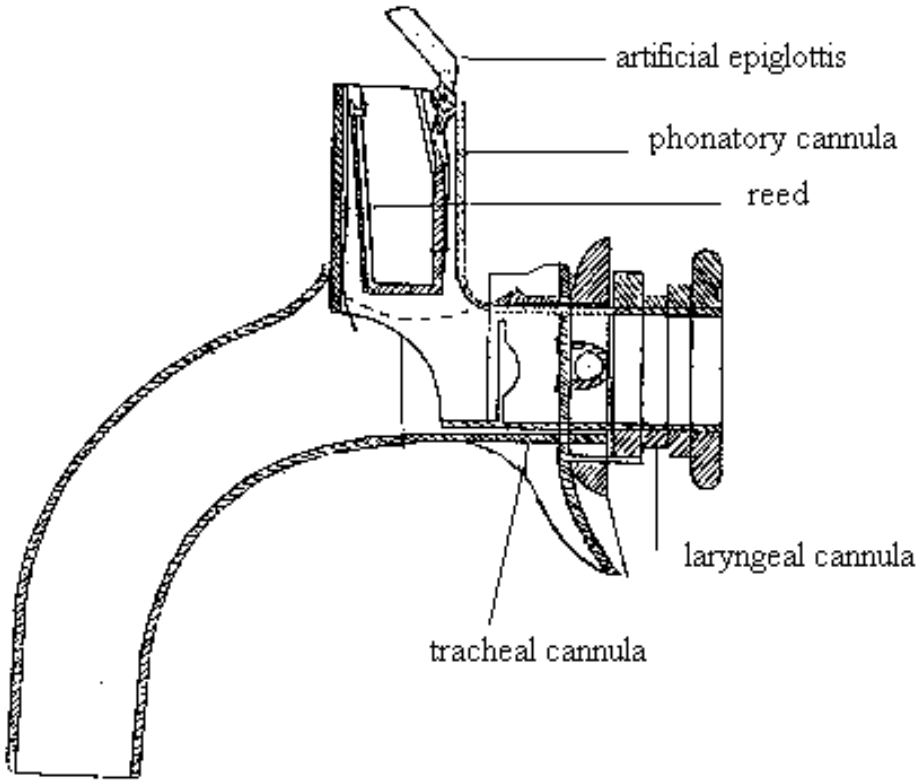


Fig. 2.4 Internal pneumatic larynx [5]

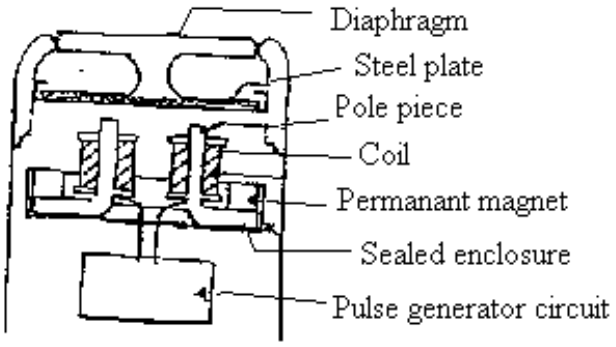


Fig. 2.5 Electromagnetic type transducer [7]

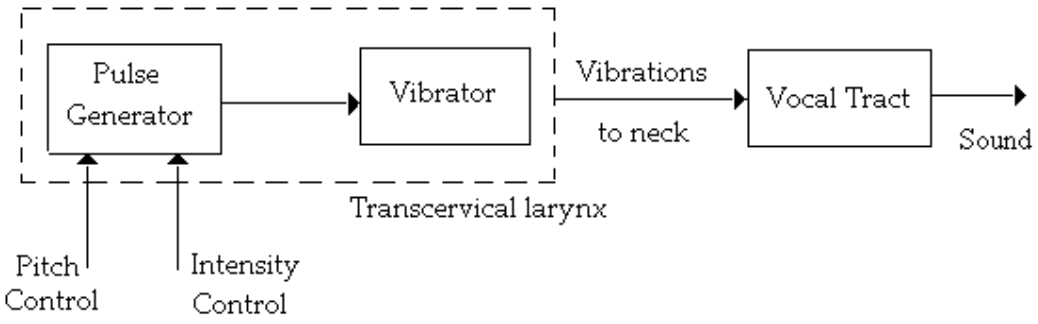


Fig. 2.6 Generation of sound with a transcervical electrolarynx [7]

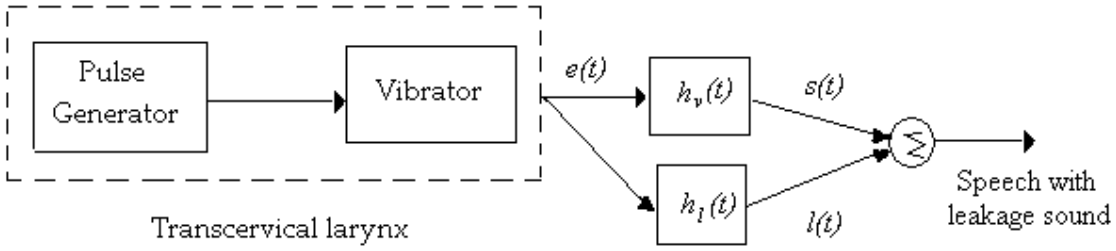


Fig. 2.7 Model of background noise generation due to leakage of vibrations in a transcervical electrolarynx [7]

Chapter 3

REVIEW OF NOISE REDUCTION TECHNIQUES

3.1 Introduction

This chapter starts with a description of acoustic and perceptual characteristics of alaryngeal speech (speech produced with the help of an artificial larynx).The perceptual characteristics determine the confusions inherent in the alaryngeal speech. The reasons for such confusions are discussed. The signal-to-noise ratio (SNR) of the alaryngeal speech also plays an important role in the intelligibility of the speech. The effect of the SNR is discussed.

This chapter also reviews the signal processing techniques for reducing the background noise. The work done earlier at IIT Bombay by Hiren Shah [7] as part of his M.Tech. dissertation is reviewed. His work centered around single input algorithms of noise cancellation. The single input noise canceller algorithms have two modes of operation namely training and the use mode. In the training mode the input to the canceller is the noise, and the algorithm estimates the noise. In the use mode the speech corrupted by the noise forms the input, from which the estimated noise is subtracted to yield the clean speech. The two input LMS algorithm makes use of two inputs, one the noise and the other being noisy speech. The algorithm adjusts the coefficients of a filter based on minimum mean square error criterion. A review of the two input LMS algorithm is done in this chapter.

3.2 Characteristics of alaryngeal speech

Weiss et al. [10] reported a detailed study of perceptual and the acoustical characteristics of the speech produced with the help of external electronic artificial larynx. The study was carried out using Western Electric Model 5. The test material recorded included the vowels /*a*/, /i/, and /u/, the first paragraph of the Rainbow Passage, and two lists of the Modified Rhyme Test (MRT). The experiments conducted were aimed at studying the perceptual confusions in the alaryngeal speech. The recordings were presented to a group of listeners, who were given the response forms. Based on the intelligibility of the word presented, the listeners chose one of the word from the rhyming 50 word set. The results indicated that there were very few vowel errors. However in the case of consonants, the errors depended on the position

of the consonant. More errors were seen in case of the word-initial consonants than with the word-final consonants. The word-initial errors were prevalent in case of the stops and the fricative consonants. The glides and affricates were more intelligible than stops and fricatives. The nasals were equally perceived in both the word-initial and the word-final positions.

The experiments indicate that the results obtained with the nasal consonants, /r/, /l/, /w/, and words with vowels had a high degree of intelligibility. The errors were more in case of voiceless stops, because of voicing confusions. There was ambiguity in perceiving /p/, and it was perceived as /b/ most of the time. Similarly /t/ was confused with /k/ and /d/, and /k/ was perceived as /g/. The voiced stops were correctly perceived most of the times. Among the affricates and fricatives in the word-initial position, /f/ and /s/ were correctly identified most of the times. The consonant /h/ is difficult for laryngectomees to produce. However /h/ produced along with a vowel was easily identified.

The perceptual confusions for stop consonants in word-final position were lesser compared to the word-initial position. This is because of greater accuracy in the voicing feature identification. Hence the perceptual errors arose basically due to the voicing decision. One reason for this is the continuous operation of the artificial larynx, which makes it difficult to distinguish between voiced and unvoiced sounds. Theoretically the on/off characteristics is controllable, but coordinating manual triggering of the device with voicing characteristics of individual phonemes in continuous speech is extremely difficult.

Analyses were conducted to determine the speech-to-noise ratio (SNR) in the alaryngeal speech and its effect on intelligibility. The SNR in this case [10][11] is defined as the ratio of the average level of the vocal peaks in the alaryngeal speech (inclusive of background interference) and the level of radiated sound measured with the speaker's mouth closed. The results indicated that for the same device, the SNR varied over 4-15 dB across the subjects. It is to be noted that leakage from the casing of the instrument should be speaker independent. Large variation in SNR indicates that leakage from the vibrator-tissue interface varies across speakers, and it also significantly contributes to the background interference. In an early report of the development model of a similar device, Barney et al. [11] reported a similarly defined speech-to-noise ratio to be approximately 20-25 dB.

The results of the identification tests revealed that SNR should be greater than 4 dB to reduce the occurrence of the error. The alaryngeal speech with SNR lower than 4 dB has significantly lower intelligibility compared to the speech with higher SNR's. A look at the spectrum of the direct-radiated noise indicated that most of the energy was concentrated in the frequency region 400-800 Hz. A second peak was found between 1-2 kHz, with magnitude down the previous value by 5-10 dB. There were usually 2 or 3 additional peaks between 2 and 4 kHz. The frequency and magnitude of these peaks were speaker dependent. In case of alaryngeal speech with poor SNR's, there is significant auditory masking of the vowel formants, which could lead to vowel identification errors. However the noise spectrum is steady in nature in contrast to the rapidly changing formant frequencies, as the vocal tract configuration is altered. Because of this reason, the listeners were able to track the formant trajectories and perceive speech in the presence of background noise for relatively higher SNR's. Another observation is that for alaryngeal speech, the spectrum begins to decrease rapidly below frequency of 500 Hz, with an average roll-off of 14 dB/decade. The spectrum of the natural speech rolls off earlier at 100-200 Hz. The effect of the roll-off is a reduction of the first formant (F1) intensity for vowels [10].

3.3 Two input LMS algorithm

An adaptive filter for noise removal is based on the premise that the desired signal is corrupted by an uncorrelated noise, and a reference signal is available that is in some way correlated with the noise, but uncorrelated with the desired signal [12][13]. Fig. 3.1 shows the block diagram of such a scheme of adaptive filter. There are two inputs to the filter, one is the noisy speech $x(n) = s(n) + l(n)$, where $s(n)$ is the speech signal and $l(n)$ is the background interference or the noise. The second signal $r(n)$ is correlated with the noise $l(n)$. The error $e(n)$ between $x(n)$ and $r(n)$ is used to modify the coefficients of the filter. The coefficients of the filter, b_m's are adaptively modified based on the minimum mean square error criterion. When the error is minimized, the output of the filter is a good estimate of the noise and is subtracted from the raw speech signal to output noise-free speech. The adaptive nature of the filter allows it to react against any changes in the signal and noise characteristics.

Espy-Wilson et al. [1] applied the two input adaptive filtering of Fig. 3.1 for enhancement of alaryngeal speech. The adaptive filter was based on LMS algorithm.

For the sound produced by an artificial larynx, the signals $x(n)$ and $r(n)$ are from the same source, and are strongly correlated when $x(n)$ corresponds to voiced sounds. This is because the vibrations are derived solely from the artificial larynx. However in case of consonants (unvoiced segments, to be more precise) there will be weaker correlation between $x(n)$ and $r(n)$ on account of the vocal excitation being caused by the turbulation at constrictions. During the vocal sounds, if the adaptation is allowed then the cancellation will result in an output that contains no information at all. However when the signals are not correlated, the filter removes the noise from the speech signal.

In the reported work [1], the decision to turn on and off the adaptation is based on whether the segment is voiced or unvoiced. For this purpose a windowed average energy detector is used. Whenever the energy exceeds a threshold, it is classified as a voiced segment and the adaptation is prevented. The filter coefficients are retained to the last value. Whenever the segment is unvoiced, the filter coefficient's adaptation continues from the last value. The equations of LMS algorithms are as given below.
The FIR filter output is given as

$$y(n) = \sum_{m=0}^{N-1} b_m(n) r(n-m) \tag{3.1}$$

The error is given as

$$e(n) = x(n) - y(n) \tag{3.2}$$

The coefficients of the FIR filter, b_m's are updated on the basis of the previous coefficients as

$$b_m(n) = b_m(n-1) + \mu e(n) r(n-m) \quad m = 0,\ldots \text{N-1} \tag{3.3}$$

where μ is the convergence parameter. When the LMS algorithm minimizes mean square error $e(n)$, the impulse response of the FIR filter gives estimate of the leakage sound $y(n) \approx \text{l}(n)$, and the error $e(n)$ is the noise removed signal output.

The adaptation size plays an important role in determining the behavior of the LMS algorithm. Increasing the magnitude of the adaptation constant increases the size of the iteration step thereby increasing the speed with which the algorithm converges. However it also increases the likelihood of the algorithm responding to spurious events and increases the mean squared error. Increasing the value beyond certain

value results in instability of the algorithm. Hence a proper choice of μ is necessary for proper operation of the algorithm. The bounds on the μ are given by

$0 < \mu < \frac{2}{NE\{r^2(n)\}}$, where $E\{r^2(n)\}$ is the power of the reference input.

The test setup used by Espy-Wilson at el. [1] consisted of two microphones, the first one positioned to the left of the mouth and approximately 6 cm from its center. The second microphone was placed approximately 2 cm right of the artificial larynx. The recordings comprised of the first paragraph of the Rainbow Passage uttered by normal persons using the artificial larynx, with their glottis closed. The results obtained indicate that there is a considerable noise cancellation. During the non-sonorant intervals or the low energy intervals, the noise cancellation is effective, and most of the background noise is cancelled. However during the sonorant intervals there was an improvement in the output quality, though the background noise was not removed fully. The intelligibility of the processed output was better than that of its unprocessed counterpart.

The LMS algorithm described above uses two inputs, and the assumption that the signal $r(n)$ is an estimate of the noise present in the signal $x(n)$. However the input $r(n)$ consists of noise as well as some portion of speech. A model of the speech and background noise picked up as two inputs is shown in Fig. 3.2. The signal $x(t)$ predominantly consists of the speech and a small amount of noise, while the signal $r(t)$ contains a major portion of the noise and a small amount of speech. The presence of the speech in the noise affects the quality of the output processed with the LMS algorithm. The author has not been able to locate other literature (published on the web) on the use of two input LMS algorithm for background noise cancellation in an artificial larynx.

3.4 Single input leakage canceller

In a single input leakage canceller, a microphone placed in front of the lips forms the input. From the noisy speech input, an estimate of the noise (derived from the input itself) is subtracted, to get the noise-free speech output. The work done by Hiren Shah [7] as part of M.Tech. dissertation at IIT Bombay is reviewed. The cancellation scheme implemented by Shah [7] consists of two modes of operation namely training mode and use mode. In the training mode, the input to the filter is the

background noise alone. The speaker keeps his lips closed during this period. The leakage canceller algorithm estimates the impulse response of the leakage path. In the use mode the person speaks, and the input to the canceller will be noisy speech. The leakage canceller subtracts the estimated noise waveform from the noisy speech to deliver the speech. The basic assumption here is that the leakage noise remains stationary. In both the modes mentioned above, estimation of the leakage noise requires the reference position of the excitation pulse of the vibrator. The reference position is determined by using an excitation impulse generator algorithm, which uses the input signal to locate the reference position of the excitation pulses [7].

An excitation impulse generator is used to generate an impulse with reference to the excitation pulse of the vibrator. The impulse generator algorithm is based on the dynamic threshold method [3]. Fig. 3.3 shows the block diagram of the impulse generator. The input signal $x(n)$ is squared and low pass filtered. A fraction c of this average is set as threshold. Whenever the squared value of the input exceeds this threshold, an impulse is generated. Once an impulse is generated, a refractory period is enabled, which prevents further impulses in the vicinity of the earlier impulse. The refractory period may be set to one half the pitch period. During the refractory period, no impulse is generated even if the squared input exceeds the threshold. With such an algorithm, exactly one impulse is generated during each pitch period. Different impulses correspond to the similar point in different pitch periods. Low pass filtering of the squared input is done by taking a N-point moving average as

$$p(n) = \frac{1}{N}\sum_{m=0}^{N-1} v(n-m) \tag{3.4}$$

In a variation of the above algorithm, square root of the average squared value is compared with magnitude of the input sample.

Two algorithms, namely, the ensemble averaging and single input LMS algorithm, were used by Hiren Shah [7] to get an estimate of the impulse response of the leakage path during the training mode. Both these algorithms make use of the impulse generator, and are described in the subsequent sub-sections.

3.4.1 Ensemble averaging algorithm

Fig. 3.4(a) shows the block diagram of the ensemble averaging in training mode [7]. The input signal is given to the impulse generator. The impulse generator generates an impulse whenever the input across it increases beyond a certain preset

threshold. Whenever an impulse is generated, the block selector selects a block of samples from the input signal with reference to the impulse position. The block starts a certain samples before the impulse. The selected block is passed to the ensemble averager. This process is repeated for a number of blocks, during the training mode, after which the ensemble averager outputs the averaged waveform for these blocks. This ensemble-averaged waveform is an estimate of the impulse response of the leakage path. Let N_1 denote the number of samples of the block before the impulse position and N_2 be the number of samples of the block after the impulse. For an impulse stationed at n, the block consists of

$$g_n(m) = x(n+m), \quad -N_1 < m < N_2 \tag{3.5}$$

Ensemble average is obtained by

$$h(m) = \frac{1}{N}\sum_{k=0}^{N-1} g_{n_k}(m), \quad -N_1 < m < N_2 \tag{3.6}$$

Where N is the number of blocks selected during the training period. This $h(m)$ is taken as impulse response of the leakage path.

In the use mode, the noise estimate is subtracted from the input signal, with reference to the impulse as illustrated in Fig. 3.4(b).

3.4.2 Single input LMS algorithm

Fig. 3.5(a) shows the block diagram of the LMS based adaptive filter for estimation of impulse response of leakage path [7]. The input signal is given to the impulse generator. The impulse generator output is given to the adaptive filter as the input. The reference to the adaptive filter is the delayed signal $x(n)$. The adaptive filter is a FIR filter whose coefficients are adjusted by the LMS algorithm, which aims at minimizing the mean square error. The equations used are as described below. The FIR filter output is given as

$$y(n) = \sum_{m=0}^{N-1} b_m(n) w(n-m) \tag{3.7}$$

The error is given as

$$e(n) = x(n - N_d) - y(n) \tag{3.8}$$

The delay N_d is introduced to compensate for the delay in positioning of the recovered impulses with respect to excitation impulses. Its value need not be exact and could be

slightly higher than the impulse delay. Difference is automatically compensated in the LMS algorithm by appropriate shifting of the filter coefficient values.

The coefficients of the FIR filter, b_m's are updated on the basis of the previous coefficients as

$$b_m(n) = b_m(n-1) + \mu e(n) w(n-m) \tag{3.9}$$

Where μ is the convergence parameter. When the LMS algorithm minimizes mean square error $e(n)$, the impulse response of the FIR filter gives estimate of the leakage sound.

$$y(n) \approx l(n) \tag{3.10}$$

This filter output is subtracted from the noise corrupted input signal to get the pure sound signal, as can be seen from Fig. 3.5(b).

3.4.3 Results obtained by Shah [7]

Hiren Shah [7] implemented two algorithms namely ensemble averaging and LMS to estimate the impulse response of the leakage path during the training mode as discussed in sub-sections 3.4.1 and 3.4.2. The LMS algorithm was implemented for use in real time using the TMS320C50 DSP board. The results obtained with real-time implementation were not satisfactory, and hence, offline implementation of the algorithms was carried out, as C programs. The results obtained with offline processing indicated that ensemble averaging algorithm worked better than LMS algorithm. However the quality of output obtained with both the methods was poor, as there was no significant noise reduction. This was because of the variation in the leakage sound coming out from the vibrator. The variation could be a result of change in position of the device on the throat, change in application pressure, transducer dynamics and fluctuation in the battery voltage during use of the device.

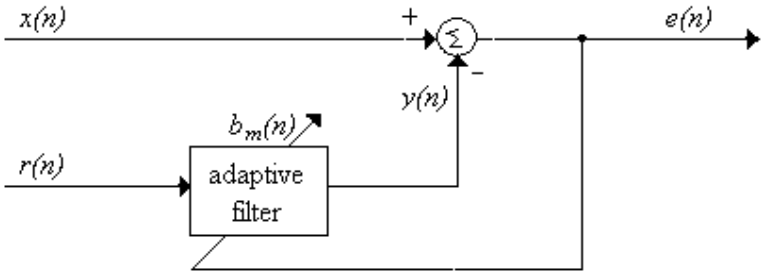


Fig. 3.1 Two input adaptive filter [13]

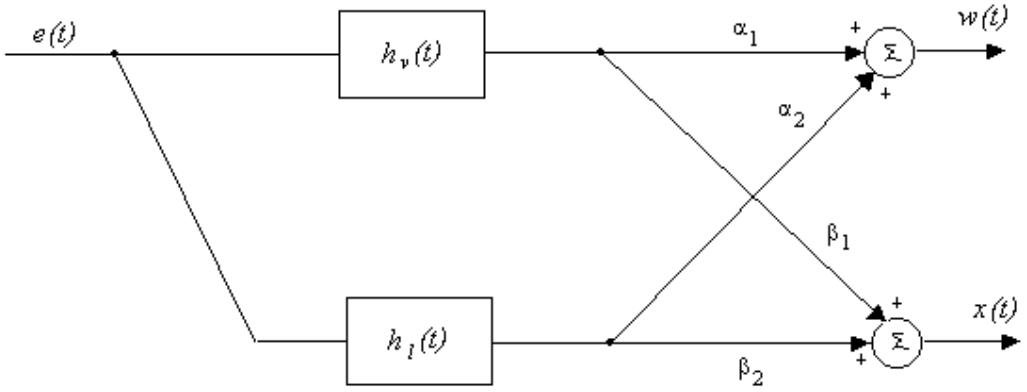


Fig. 3.2 Model of background noise generation using two microphone method

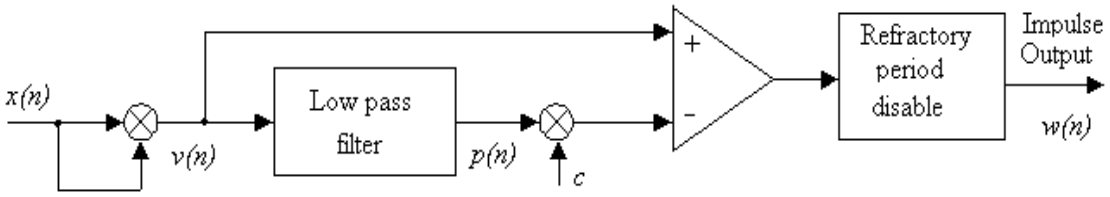


Fig. 3.3 Block diagram of the impulse train generator [7]

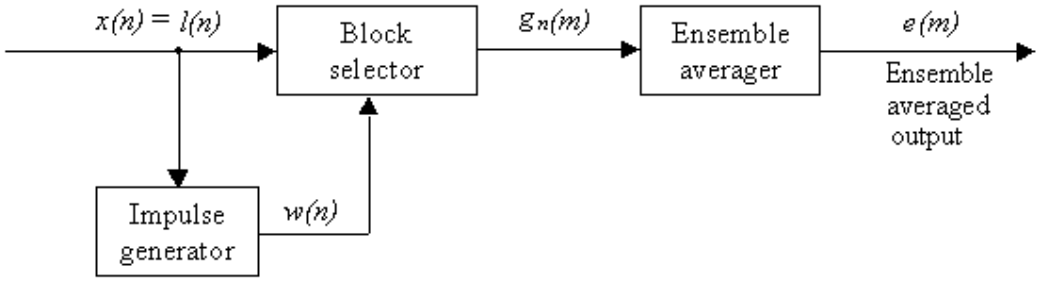


Fig. 3.4(a) Block diagram of "Ensemble Averager" during training mode [7]

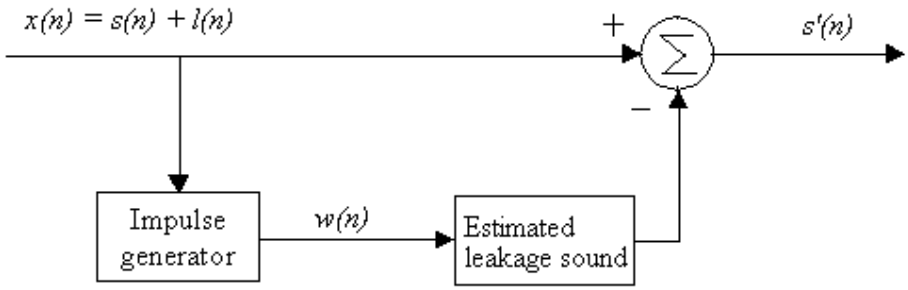


Fig. 3.4(b) Block diagram of "Ensemble Averager" during use mode [7]

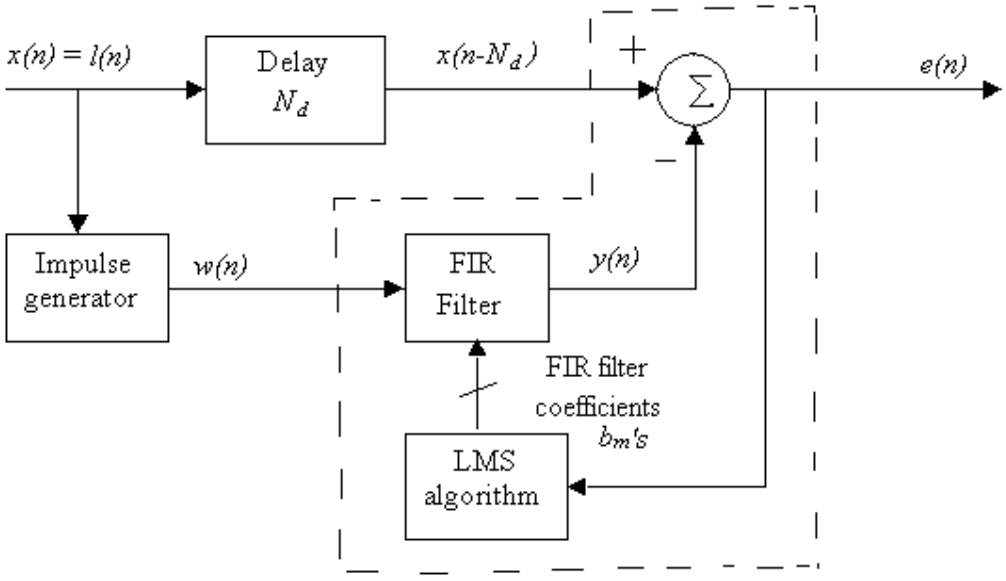


Fig. 3.5(a) Block diagram of leakage canceller during training mode [7]

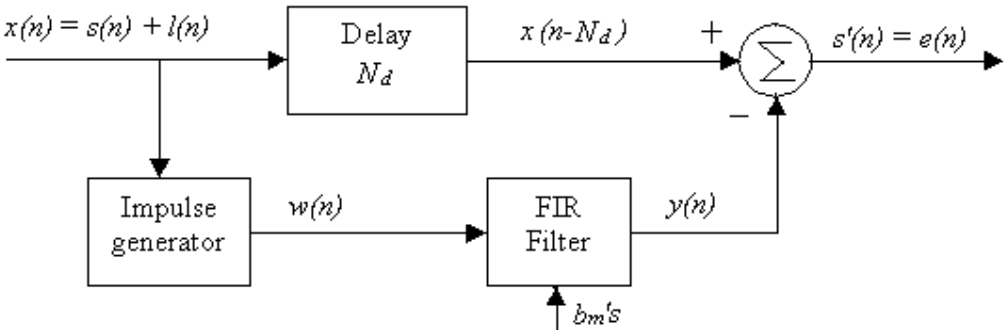


Fig 3.5(b) Block diagram of leakage canceller during use mode [7]

Chapter 4

SPECTRAL SUBTRACTION ALGORITHM

4.1 Introduction

In the previous chapter, two-input noise cancellation based on LMS algorithm and single-input noise cancellation method based on estimation of impulse response of leakage path have been reviewed, and the problems in them outlined. As an alternative to these two, here a single-input noise cancellation based on spectral subtraction is proposed for investigation. Earlier spectral subtraction technique has been successfully employed for reduction of various types of additive noise in speech [14][15]. In this chapter, first the basic method is described. Subsequently its use for cancellation of background noise in artificial larynx is proposed with the help of a model. Next the scheme to be used is described.

4.2 Spectral subtraction for cleaning noisy speech

The basic assumption made in this method is that the clean speech and the noise are uncorrelated, and therefore the magnitude spectrum of the noisy speech signal equals the sum of magnitude spectrum of noise and clean speech. Let $x(n)$ be the windowed noisy speech comprising of the clean speech $s(n)$, and the additive noise $l(n)$. Then the signal $x(n)$ can be expressed as

$$x(n) = s(n) + l(n) \tag{4.1}$$

Taking short-time Fourier transform on either side, we obtain

$$X_n(e^{j\omega}) = S_n(e^{j\omega}) + L_n(e^{j\omega}) \tag{4.2}$$

Assuming $s(n)$ and $l(n)$ to be uncorrelated

$$|X_n(e^{j\omega})|^2 = |S_n(e^{j\omega})|^2 + |L_n(e^{j\omega})|^2 \tag{4.3}$$

Assuming that noise $l(n)$ is stationary, its average magnitude spectrum $|L(e^{j\omega})|$ can be estimated during silence intervals. During speech interval, the estimated magnitude spectrum of noise is subtracted from that of the noisy speech to get

$$|Y_n(e^{j\omega})|^2 = |X_n(e^{j\omega})|^2 - |L(e^{j\omega})|^2 \tag{4.4}$$

Actual noise spectrum during speech can be modeled as the estimated noise spectrum plus error

$$|L_n(e^{j\omega})|^2 = |L(e^{j\omega})|^2 + |\varepsilon_n(e^{j\omega})|^2 \tag{4.5}$$

And therefor the spectrum after subtraction is given as

$$|Y_n(e^{j\omega})|^2 = |S_n(e^{j\omega})|^2 + |\varepsilon_n(e^{j\omega})|^2 \quad (4.6)$$

The basic spectral subtraction algorithm has two modes of operation viz. noise estimation mode and noise subtraction mode. In the first mode, the input is the background noise alone. The squared magnitudes of the FFT of a number of adjacent windowed segments are averaged to get the mean squared noise spectrum. During the speech interval, the noisy speech is windowed by the same window as in earlier mode, and its magnitude and phase spectrum are obtained. The phase spectrum is retained for resynthesis. From the squared magnitude spectrum, the mean squared spectrum of noise, determined during the noise estimation mode is subtracted. The resulting magnitude spectrum from the power spectrum is then combined with the earlier phase spectrum, and its inverse FFT is taken as the clean speech signal $y(n)$ during the window duration. The equations used are

$$|Y_n(k)|^2 = |X_n(k)|^2 - |L(k)|^2 \quad (4.7)$$

$$y_n(m) = IFFT[|Y_n(k)|e^{j\angle X_n(k)}] \quad (4.8)$$

In real practice, assumption regarding speech and noise being uncorrelated may be valid over long duration, but not necessarily over the short window segments. One of the consequences of this is that in the short time spectra, some of the frequency components of $|Y_n(k)|^2$ can go negative. These values are set to zero in the basic algorithm, and this is known as "half wave rectification" [14][15]. However with half wave rectification, there appears a new noise called "musical noise". In addition to this musical noise, a considerable amount of broadband noise is also present in the processed output. To eliminate these two noises from the output, modifications have been carried out.

4.3 Modified spectral subtraction method

The magnitude spectrum contains peaks and valleys. During the noise subtraction mode, if the averaged noise spectrum is subtracted from the noisy speech spectrum, the valleys in the spectrum will get enhanced in the negative direction, which will be set to zero. This will result in narrow random spikes of value between zero and maximum value during non-speech period, known as residual noise. When converted back to the time domain, the residual noise will sound as sum of tone

generators with random frequencies turned on and off. During speech period, this noise residual will be perceived at frequencies, which are not masked by the speech.

In order to reduce the effect of musical noise, Berouti et al. [15] modified the method to reduce spectral excursions.

$$|Y_n(k)|^2 = |X_n(k)|^2 - \alpha |L(k)|^2$$

$$|Y'_n(k)|^2 = |Y_n(k)|^2 \quad \text{if } |Y_n(k)|^2 > \beta |L(k)|^2$$

$$= \beta |L(k)|^2 \quad \text{otherwise} \tag{4.9}$$

where α is the subtraction factor and β is the spectral floor factor.

The block diagram of the modified spectral subtraction algorithm is as shown in Fig. 4.1. With $\alpha > 1$, the noise will be over subtracted from the noisy speech spectrum. This will not only reduce the noise floor, but will also eliminate the peaks of wideband noise, thereby reducing it considerably. However over subtraction may lead to the enhancement of the valleys in the vicinity of the peaks, thereby increasing the noise excursion. This is taken care by the spectral floor factor β. The spectral components of $|Y'_n(k)|^2$ are prevented from going below $\beta |L(k)|^2$. For $\beta > 0$, the spectral excursions are not as large as with the case $\beta = 0$, since the valleys between the peaks are not very deep. This reduces the musical noise to a large extent.

The proper choice of the parameters α and β gives an output free from broadband as well as the musical noise. Another modification by Berouti et al. [15] to the spectral subtraction algorithm is the addition of exponent factor γ in place of 2 for subtraction.

$$|Y_n(k)|^\gamma = |X_n(k)|^\gamma - \alpha |L(k)|^\gamma$$

$$|Y'_n(k)|^\gamma = |Y_n(k)|^\gamma \quad \text{if } |Y_n(k)|^\gamma > \beta |L(k)|^\gamma$$

$$= \beta |L(k)|^\gamma \quad \text{otherwise} \tag{4.10}$$

With $\gamma < 1$, the subtraction of the noise spectrum affects the noisy speech spectrum drastically than with the case when $\gamma = 1$. For $\gamma < 1$, the processed output has a low level, and hence there is a need for normalization of the output level to make it independent of γ. The normalization factor G is given as

$$G = \{(|X_n(k)|^2 - |L(k)|^2)/|Y'_n(k)|^2\}/\gamma \tag{4.11}$$

if the current value of *G* is less than its previous value. If the current value is greater than the previous value, the previous value of *G* is retained. It is to be noted that

implementation of filtering via FFT requires techniques like overlap-add or overlap-save methods [16] for obtaining linear convolution from circular convolution. However spectral subtraction method is a subtraction process, and therefore window segments are processed independently. It is to be further noted that phase spectrum of the noisy speech is coupled with the cleaned magnitude spectrum. Hence a certain degree of distortion is to be accepted. Quality can be improved by obtaining phase spectrum corresponding to the cleaned magnitude spectrum.

4.4 Spectral subtraction for enhancement of alaryngeal speech

In the previous section, spectral subtraction method for enhancement of noisy speech is described. Here a similar method is proposed for enhancement of alaryngeal speech, but the speech and background interference are not uncorrelated. Let $x(n)$ be the noisy speech, $h_v(n)$ be the impulse response of the vocal tract, $h_l(n)$ be the impulse response of the leakage path, and $e(n)$ be the excitation signal. The noisy speech signal is given as

$$x(n) = s(n) + l(n) \tag{4.12}$$

where $s(n)$ is the speech signal and $l(n)$ is the background interference or the leakage noise. If $h_v(n)$ and $h_l(n)$ are the impulse response of the vocal tract path and the leakage path respectively, then

$$s(n) = e(n) * h_v(n) \tag{4.13}$$

$$l(n) = e(n) * h_l(n) \tag{4.14}$$

Taking short-time Fourier transform on either side of 4.12, we get

$$X_n(e^{j\omega}) = E_n(e^{j\omega})[H_{v_n}(e^{j\omega}) + H_{l_n}(e^{j\omega})]$$

Considering the impulse response of the vocal tract and leakage path to be uncorrelated, we get

$$|X_n(e^{j\omega})|^2 = |E_n(e^{j\omega})|^2 [|H_{v_n}(e^{j\omega})|^2 + |H_{l_n}(e^{j\omega})|^2] \tag{4.15}$$

If the short-time spectra are evaluated using pitch synchronous window, $|E_n(e^{j\omega})|^2$ can be considered as constant $|E(e^{j\omega})|^2$. During non-speech interval, $e(n) * h_v(n)$ will be negligible and the noise spectrum is given as

$$|X_n(e^{j\omega})|^2 = |L_n(e^{j\omega})|^2 = |E_n(e^{j\omega})|^2 |H_{l_n}(e^{j\omega})|^2 \tag{4.16}$$

By averaging $|L_n(e^{j\omega})|^2$ during the non-speech duration, we can obtain the mean squared spectrum of the noise $|L(e^{j\omega})|^2$. This estimation of the noise spectra is used for spectral subtraction during the noisy speech segments.

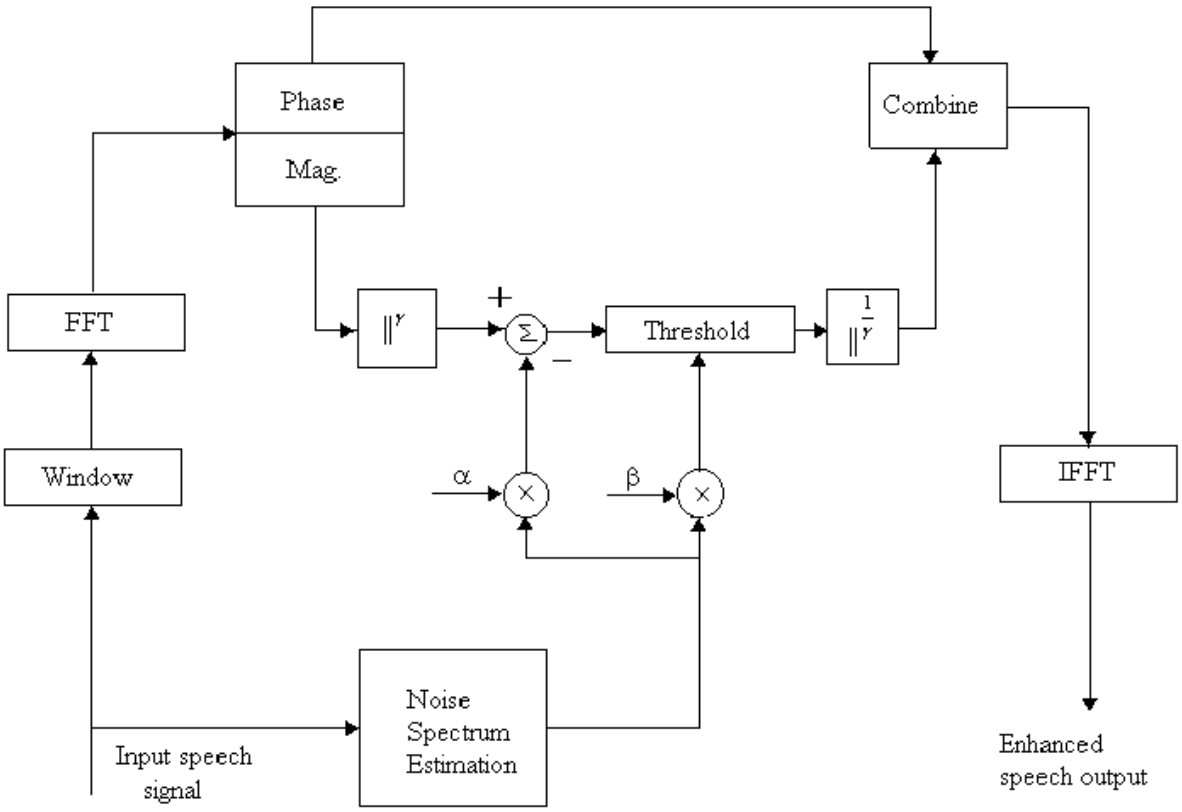


Fig. 4.1 Block diagram of modified spectral subtraction algorithm, adapted from [15]

Chapter 5

IMPLEMENTATION AND EVALUATION OF NOISE REDUCTION TECHNIQUES

5.1 Introduction

Two methods for single input noise cancellation are implemented for experimental evaluation: the method used by Shah [7] based on estimation of impulse response of the leakage path (section 3.4) and the method based on spectral subtraction (Chapter 4). The implementation was carried out for off-line processing of recorded, using MATLAB and as C programs. The signal acquisition was done using a microphone connected to the "Mic." terminal of the sound card of the PC. A sampling frequency of 11.025 kHz was used. The recordings were of 5 s each. The processed sound files were then played back for performance evaluation.

First section discusses the analysis of the alaryngeal speech signal acquired using the transcervical electrolarynx. Subsequent sections present implementation and evaluation of noise reduction techniques.

5.2 Analysis of alaryngeal speech

Spectrographic analysis is a convenient way of analyzing speech signals. The spectrogram displays the time varying magnitude and spectrum of the speech signal, with time in horizontal direction and frequency in vertical direction [3]. The spectrogram is generated by calculating the short time Fourier transform of the sampled speech signal. A spectrogram software package developed by Ratanpal [17] is used to generate the spectrograms. This package allows the user to record the speech signals. It also provides the facility to display and store the spectrogram. Provision for computing wide band as well as narrow band spectrograms is provided. The wide band spectrograms use a smaller window length (bandwidth of 300 Hz) and smoothen out the harmonic structure. Narrow band spectrograms on the other hand, use a 45 Hz bandwidth, thereby increasing the resolution for individual harmonics, but the temporal movement of formants gets smeared [3].

For the purpose of analysis, a number of vowel segments were recorded. For each vowel, two sets of recordings were done. The first recording was done near the mouth of the speaker, where the speech signal was predominant. A second recording

was done near the artificial larynx, where the background noise was severe. The wideband spectrograms of both the speech signal and the noise signal are taken and spectrum analysis was carried out with a bandwidth of 300 Hz. The results are as mentioned below.

Fig. 5.1(a) shows the wideband spectrogram for a 150 ms segment of vowel /*a*/, uttered by a normal person. The microphone is placed near the speaker's lips, and the signal acquisition is done with sampling frequency of 11.025 kHz, using the sound card of the PC. The spectrograms were obtained using package SPEC2000 developed in the laboratory [17]. The fundamental frequency and the formant frequencies are measured from the spectrogram, as well using software program "praat". The fundamental frequency is found to be 99 Hz. The three formant frequencies are visible as distinct bands. The first three formant frequencies were found to be 700, 1480, and 2850 Hz respectively. The formant frequency bandwidths for these three frequencies were found to be 500, 470, and 310 Hz respectively.

Fig. 5.1(b) shows the wideband spectrogram of vowel /*a*/, generated by using an artificial larynx model NP-1 from M/s N P Voice Ltd., Thane. The fundamental frequency is found to be 92 Hz. The first three formant frequencies were found to be 660, 1170 and 2900 Hz respectively. The formant frequency bandwidths for the three frequencies were found to be 510, 290, and 350 Hz respectively. As can be seen from Fig. 5.1(b), the spectrogram lacks low frequency components.

Fig. 5.1(c) shows the wideband spectrogram of the background noise generated by the artificial larynx. This was recorded with the speaker having his lips closed so that no sound from vocal tract is added to the leakage noise. The fundamental frequency is found to be 92 Hz . The first three formant frequencies were found to be 310, 940, and 1680 Hz respectively. The formant frequency bandwidths for the three frequencies were found to be 500, 400, and 400 Hz respectively. Fig. 5.1(c) reveals the range of frequencies present in the noise. It can be seen that background noise has a spectrum covering the entire speech range. Hence superimposition of this noise significantly reduces the intelligibility of alaryngeal speech.

Fig. 5.2 shows narrowband spectrograms, obtained using analysis bandwidth of 45 Hz, for the sounds same as in Fig. 5.1, showing the spectral distribution of energy.

Fig. 5.3 shows the wideband spectrograms for the vowel sequence /*a*/, /*i*/, /*u*/, uttered by a normal person and using NP-1 electrolarynx respectively. The duration of the segment used was 1600 msec. The wideband (300 Hz) spectrogram of Fig. 5.3 exhibits the vertical striations corresponding to glottal pulses. The silence intervals are filled by background noise, which has spectral characteristics similar to that of vowels, with stationary formant structure. It can be seen that in case of vowels, the formant structure of background noise gets superimposed on those of the vowels, masking the distinction between them.

Weiss et al. [10] and Barney et al. [11] have reported speech-to-noise ratio for transcervical larynges. This is measured as the ratio of the short-term energy of noisy speech to the short-term energy of noise (non-speech segment), expressed in dB. Weiss et al. [10] have reported a SNR in the range of 4-15 dB across the subjects. Barney et al. [11] have reported SNR in the range from 20-25 dB for an experimental transistorized artificial larynx. In a recording with the NP-1 electrolarynx, SNR of 8.5 dB was observed. Obviously, this device has a much higher level of background noise.

5.3 Single input leakage canceller

As discussed in section 3.4, Shah [7] developed and tested a method for single input noise canceller. In this method, estimation of the impulse response of the leakage path is carried out in "training" mode, during the non-speech segment at the beginning. Subsequently in the "use" mode estimated noise waveform is subtracted from noisy speech, with the objective of cleaning it of background interference or leakage noise. Operation in both the modes requires location of excitation impulses. In an actual hardware, the excitation impulses can be obtained from the electronic circuit driving the vibrator. However, for the off-line implementation, the impulses are obtained using dynamic thresholding method (section 3.3). Further he used two methods for estimation of impulse response: (a) the ensemble averaging and (b) LMS algorithm for adaptive filtering (also described in section 3.3).

Hiren Shah [7] had carried out the implementation using a TMS320C50 based DSP board for real-time processing. The implementation did not result in any significant noise reduction, or intelligibility improvement. Debugging the program for

obtaining intermediate results turned out to be very difficult. Hence the entire method, for both the approaches, has been reimplemented, using some minor modifications.

The algorithms were implemented as C program. Signal acquisition and playback was done by using PC sound card. A sampling rate of 11.025 kHz was selected for off-line processing of acquired signals. Speech segments of 5 s were acquired, the first 2 s of which corresponded to the training mode, and the next 3 to the use mode. During the training mode the speaker kept his lips closed, while the artificial larynx was turned on. The recording contained the leakage noise. During the use mode, the speaker spoke and the recording now contained both the leakage noise as well as useful speech. The position of the microphone remained unchanged during the entire 5-second duration. The results obtained from the implementation of the above mentioned algorithms are presented in the following three sub-sections.

5.3.1 Impulse generator algorithm

Fig. 5.4(a) shows the speech signal corrupted with the background noise. The recordings are made for the vowel /*a*/. Figures 5.4(b)-(d) show the impulses generated by the impulse generator algorithm as discussed in section 3.4. Fig. 5.4(b) shows the impulses generated, with a fraction of the square root of the average squared value, set as threshold. Figures 5.4(c) and 5.4(d) show the impulses generated with the threshold being a fraction c of the average squared value. Fig. 5.4(c) corresponds to c of 6, while Fig. 5.4(d) corresponds to c of 9. As can be seen, the impulses are located uniquely in each pitch period. Comparison of the Figures 5.4(b)-(d) indicate that the threshold with a fraction of average squared value (c = 9) results in more appropriate placement of impulses. In the analysis of longer sections, it was observed that there was a jitter in the impulse locations obtained by using square root based threshold. Hence it was decided to use average square magnitude based threshold and c = 9.

5.3.2 Ensemble averaging algorithm

Fig. 5.5 shows the results of the ensemble averaging algorithm implementation. Fig. 5.5(a) shows the unprocessed speech and Fig. 5.5(b) the extracted impulses. Fig. 5.5(c) corresponds to the higher value of the threshold (c = 9), Fig. 5.5(d) corresponds to a lower threshold (c = 6). With a lower threshold, impulses are not properly located (may get placed earlier, and erroneous impulse may be generated), and can lead to increase of noise in the use mode. Similarly, a very

high value of threshold can skip a considerable portion of the input, thereby deteriorating the output.

The effect of different length of blocks, selected during the training period, in the ensemble average implementation, can be seen from Figures 5.5(e) and 5.5(f). N_1 points to the length of block selected to the left of impulse, while N_2 corresponds to the length of block to the right side of the impulse. The block length corresponds to N_1+N_2. Fig. 5.5(e) shows the output of ensemble averaging method, with a larger block of samples (130), while Fig. 5.5(f) shows the output with a smaller block (80). A larger length of the block gives better results as compared to the one with smaller length.

5.3.3 Single input LMS algorithm

Example results of LMS algorithm implementation are Fig. 5.6. The impulse generator algorithm is implemented as mentioned above. The effect of the convergence parameter and the filter length were studied. The effect of different filter lengths can be seen from Figures 5.6(c) and 5.6(d). Fig. 5.6(c) corresponds to a higher order filter (60), while Fig. 5.6(d) corresponds to a lower order filter (30). A higher order filter gave better results than a lower order filter. Fig. 5.6(e) shows the output generated with a higher μ (0.01), while Fig. 5.6(f) shows the results with a lower convergence parameter μ (0.001). If μ is too low, the LMS algorithm takes a long time to converge, while with very high values of μ, the LMS algorithm will either settle at spurious values or become unstable.

5.3.4 Discussion on results

Both the methods discussed above did not give good results because of the variation in the leakage sound from the vibrator. Reasons for the variation in the leakage sound 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 the device. Due to relatively slow but random fluctuations in the impulse response of the leakage path, the noise cancellation at times tends to increase the noise. Further a slight shift in the impulse position increased the noise. This can be seen from Fig. 5.7. Fig. 5.7(a) corresponds to the speech 'plug and play' recorded with a NP-1 electrolarynx. Fig. 5.7(b) displays the output waveform processed with the

ensemble averaging method, while Fig. 5.7(c) shows the output of single input LMS algorithm. Playback of processed output indicated that the method did not result in noise reduction, or quality improvement.

5.4 Spectral subtraction algorithm

The spectral subtraction algorithm method for reduction of background interference in alaryngeal speech has been presented (section 4.3). The recordings were done with normal speakers. The speakers used NP-1 electrolarynx, and controlled their breath to avoid natural speech during recording. The phrase 'plug and play' was recorded. The recordings were done with the microphone positioned at the center between the mouth and the artificial larynx position. The total duration of the recording was 5 s, at a sampling rate of 11.025 kHz. Of the 5 s, the first 2 s corresponded to the non-speech interval. During this interval, the subject was advised to keep his mouth closed to prevent any speech from the mouth, and the recorded speech contained only noise. During the other 3 s, the subject pronounced 'plug and play'.

Fig. 5.8 shows the results of the spectral subtraction method on the speech /*a*/ generated with NP-1 electrolarynx. The unprocessed speech /*a*/ waveform is shown in Fig. 5.8(a). The processed output waveform with different parameters set can be seen from the subsequent Figures 5.8(b)-5.8(d). In order to study the effect of various parameters, the processed waveforms corresponding to the entire speech duration were considered, and are described below.

For studying the effect of the window size on the output speech, various window sizes ranging from 60 to 1500 samples, were experimented with, as shown in Fig. 5.9. The vibrator of the NP-1 electrolarynx has a fixed pitch of 90.3 Hz, and this corresponds to pitch period of 122 samples. With the window size less than the pitch period, the output was highly degraded. This can be seen from Fig 5.9(b), which corresponds to approximately half the pitch period (window size = 60). As the window size was increased beyond the pitch period, the output quality got better. This can be seen from Fig. 5.9(c), which corresponds, to window size of twice the pitch period (window size = 244). However with the window size much larger than twice the pitch period, there was slight degradation in the output on account of warbling. This can be seen from Fig. 5.9(d), which corresponds, to a length of around ten times

the pitch period (window size = 1220). The best results were obtained at window size of 244, approximately double the size of samples corresponding to the pitch period. It was observed that the output quality was superior when the window size was a perfect multiple of pitch period. Therefore window length of 244 was fixed in order to see the effect of other parameters.

The effects of different values for subtraction factor α on the output were investigated. Fig. 5.10 shows the effect of varying α while other parameters are fixed. For $\alpha < 1$, there was considerable amount of noise present in the output. This can be seen from Fig. 5.10(b), which corresponds, to $\alpha = 0.5$. For $\alpha > 1$ the results contained a small amount of broadband noise. Figures 5.10(c) and 5.10(d) indicate the effect of increasing α, with α set at 2 and 6 respectively. As can be seen from the figure, with increasing α, the broadband noise in the output decreased considerably. For very large α, some portion of the useful speech information is lost. The range of α from 2 to 5 was considered appropriate.

The effect of the spectral floor factor β values on the output was tested. The value of β was varied between 0 and 1. Fig. 5.11 shows the effect of varying β, with α set at 2. Setting $\beta = 0$ corresponded to the general spectral subtraction algorithm discussed in section 4.2. Fig. 5.11(b) corresponds to the case $\beta = 0$. Increasing the value of β reduces the musical noise. Fig. 5.11(c) corresponds to β value of 0.01. Increasing the β value beyond 0.1 introduces broadband noise in the output. This can be seen from Fig 5.11(d) which corresponds to the case of $\beta = 1$. The best results were obtained with $\beta = 0.001$.

Finally the effects of different values of exponent factor γ on the output were investigated. Fig. 5.12 shows the effect of increasing γ. For $\gamma < 1$, good results were obtained. However there was a need for normalization since the output amplitude of the output was low. This can be seen from Fig. 5.12(b) which corresponds to $\gamma = 0.5$. With $\gamma = 1$, the results obtained were better, as can be seen from Fig. 5.12(c). Increasing γ beyond 1 gave inferior output quality. This can be seen from Fig. 5.12(d), which corresponds to $\gamma = 2$. The value of $\gamma = 1$ seemed to be an appropriate value.

The wideband spectrograms of the input speech and the processed output are given in Figures 5.13(a) and 5.13(b) respectively. As can be seen the frequency components of the background noise are eliminated in the processed output. The

results obtained indicate that with proper choice of parameters, the background noise can be reduced to a large degree. Large values of α, result in effective noise subtraction. However, informal listening tests indicate that noise elimination is accompanied by distinctly perceptible distortion. These tests indicate that for a good balance between perceptual quality and signal-to-noise ratio, the parameters to be used are: $\alpha = 2$, $\beta = 0.001$, and $\gamma = 1$. It is to be noted that the parameters obtained above gave good results for the recordings made by the author using NP-1 electrolarynx. In general, different settings of the parameters may be needed for different devices and users.

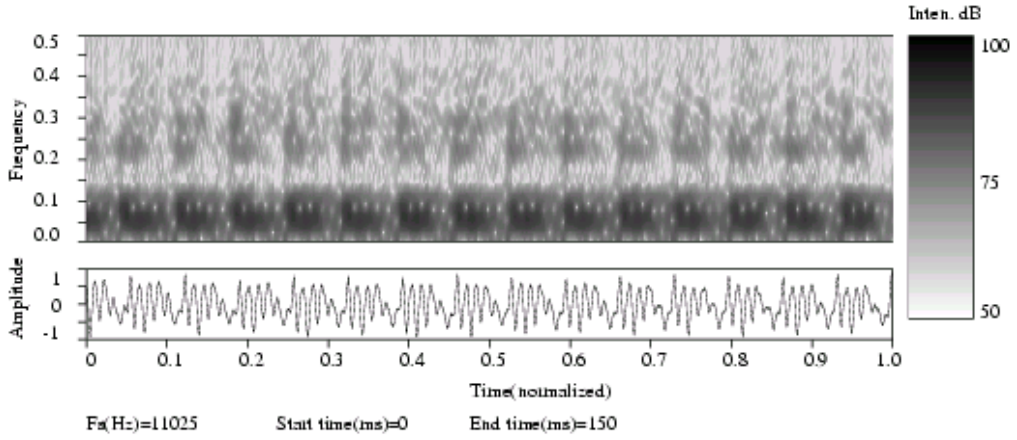


Fig. 5.1 (a) Wideband spectrogram of vowel /*a*/ uttered by a normal person

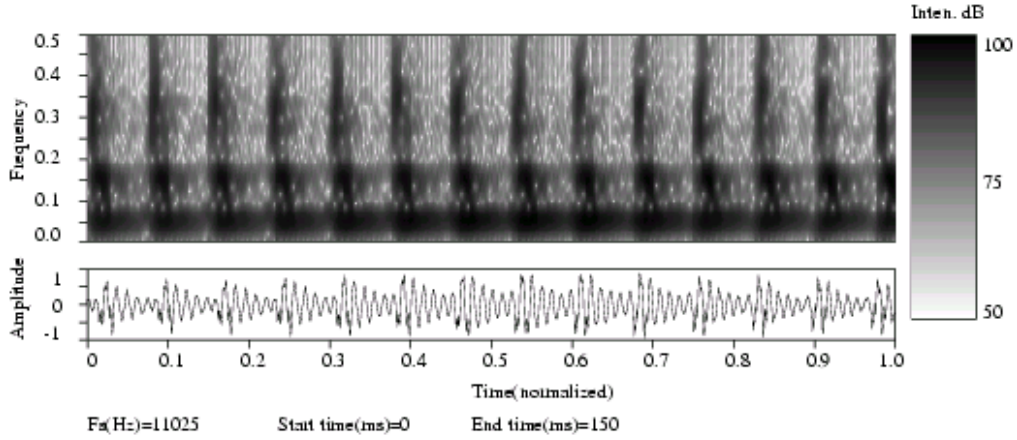


Fig. 5.1 (b) Wideband spectrogram of vowel /*a*/ uttered by a normal person using NP-1 electrolarynx

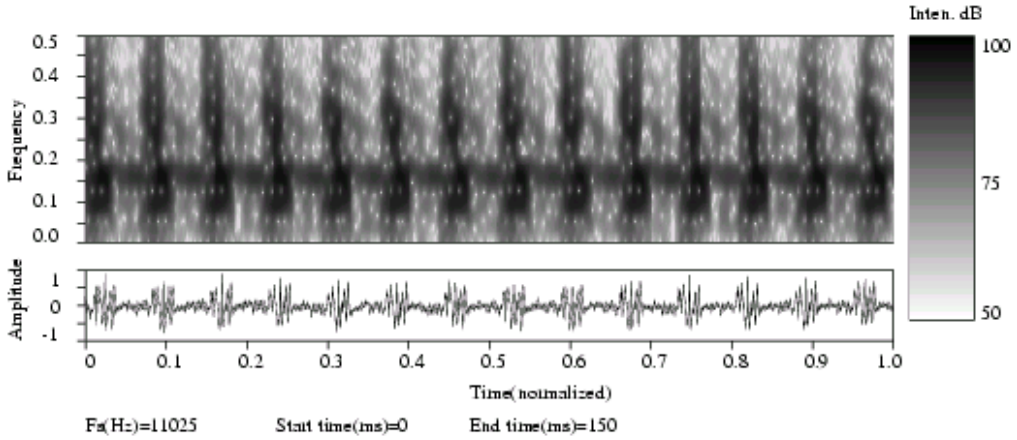


Fig. 5.1 (c) Wideband spectrogram of background noise generated during use of NP-1 electrolarynx, with speaker's lips closed

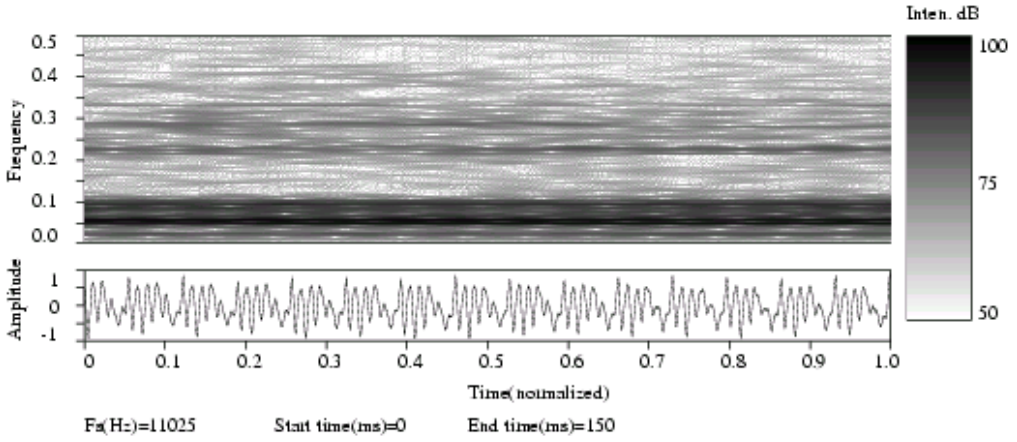


Fig. 5.2 (a) Narrowband spectrogram of vowel /*a*/ uttered by a normal person

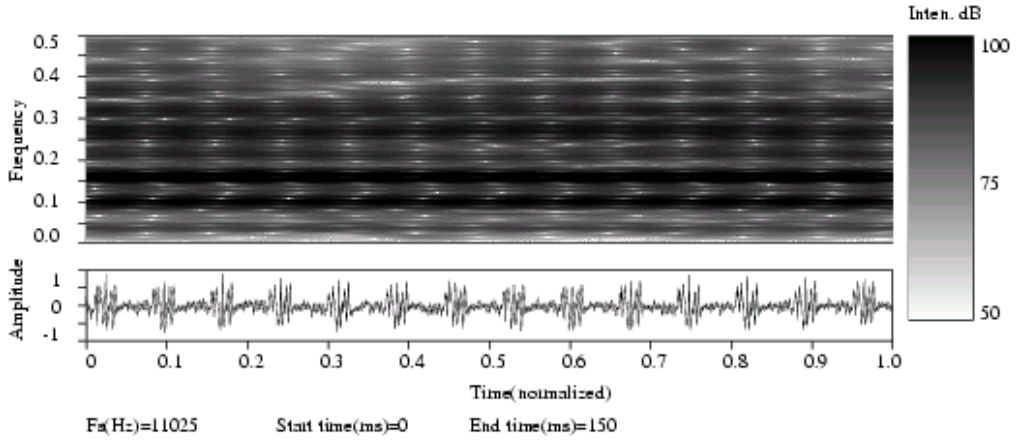


Fig. 5.2 (b) Narrowband spectrogram of vowel /*a*/ uttered by a normal person using NP-1 electrolarynx

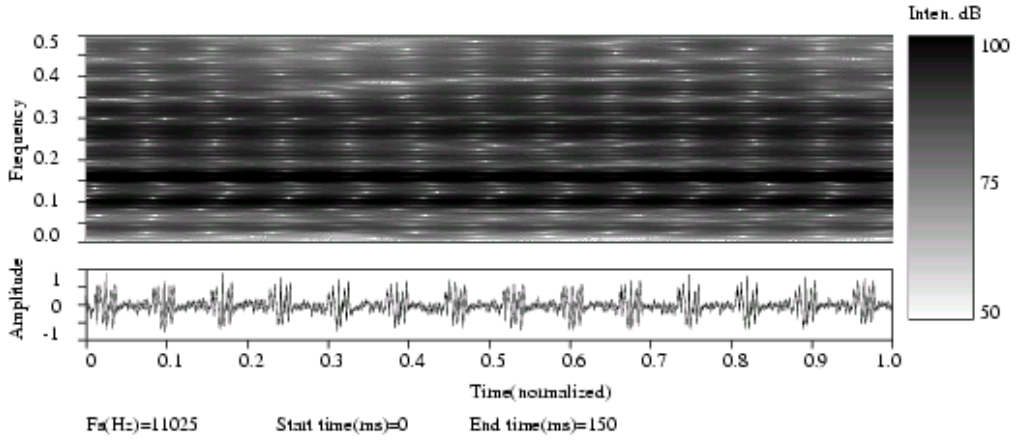


Fig. 5.2 (c) Narrowband spectrogram of background noise generated during use of NP-1 electrolarynx, with speaker's lips closed

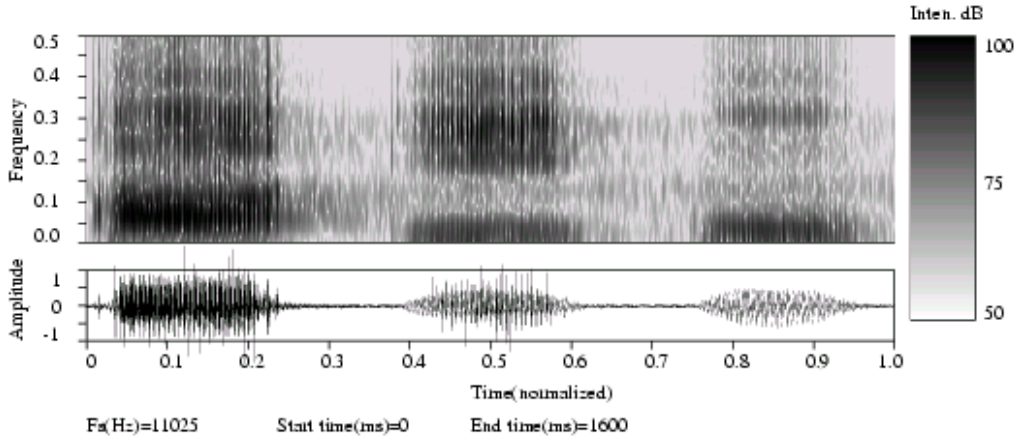


Fig. 5.3 (a) Wideband spectrogram of '/*a*/ /*i*/, and /*u*/' uttered by a normal person

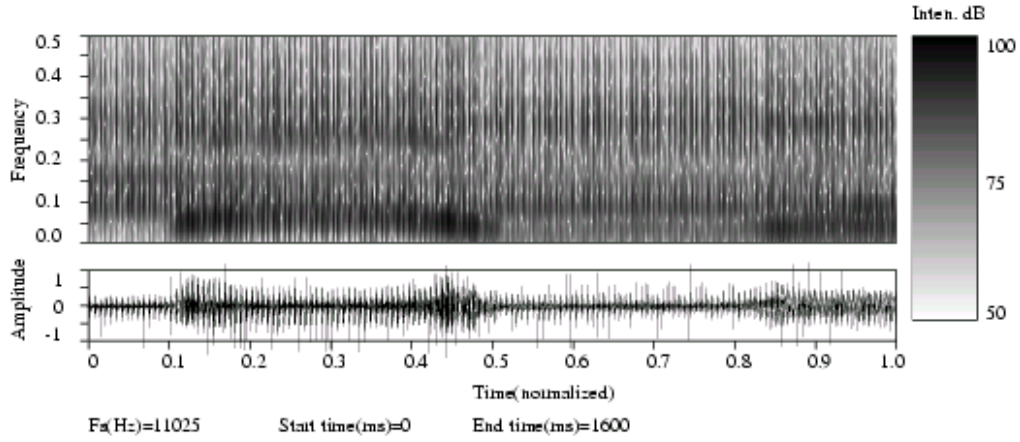
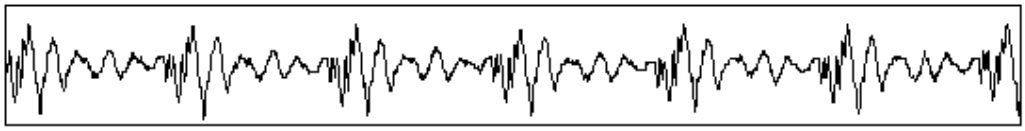
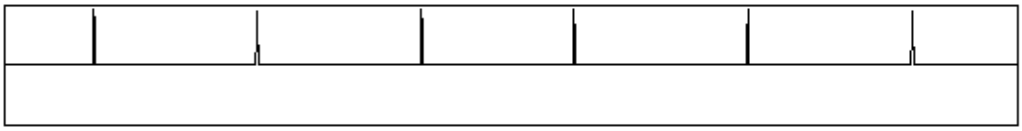


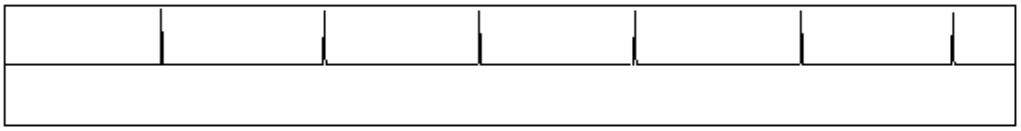
Fig. 5.3 (b) Wideband sprogram of '/*a*/ /*i*/, and /*u*/' uttered by a normal person using NP-1 electrolarynx



(a) Unprocessed speech (/*a*/) generated with NP-1 electrolarynx



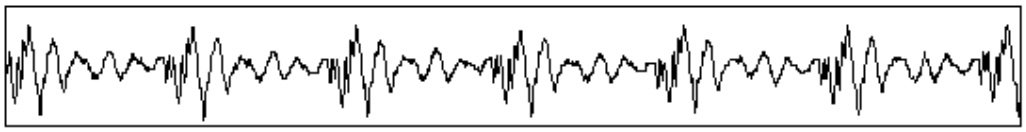
(b) Extracted excitation impulses, square root average threshold, $c = 2.5$



(c) Extracted excitation impulses, average square threshold, $c = 6$

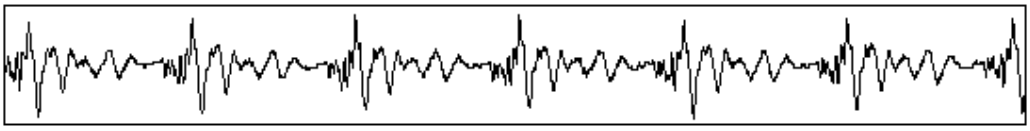
(d) Extracted excitation impulses, average square threshold, $c = 9$

Fig. 5.4 Extracted excitation impulses for alaryngeal speech

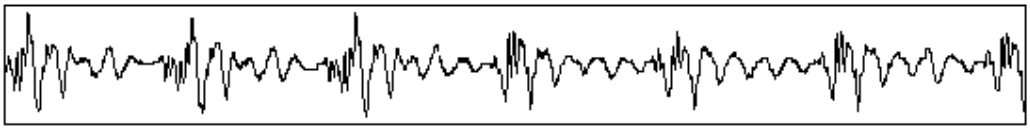


(a) Unprocessed speech (/*a*/) generated with NP-1 electrolarynx

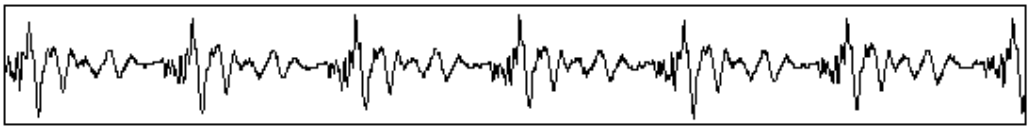
(b) Extracted excitation impulse with $c = 9$



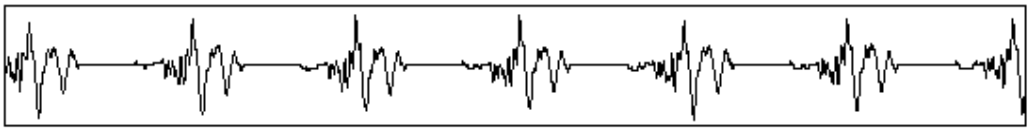
(c) Processed waveform, ensemble method, $c = 9$, $N_1 = 50$, $N_2 = 80$



(d) Processed waveform, ensemble method, $c = 6$, $N_1 = 50$, $N_2 = 80$

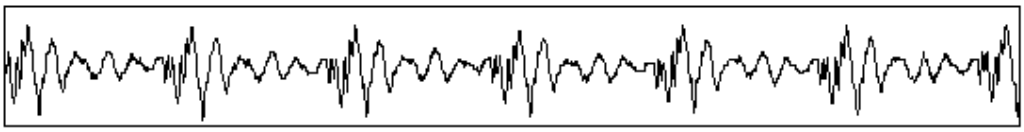


(e) Processed waveform, ensemble method, $c = 9$, $N_1 = 30$, $N_2 = 110$



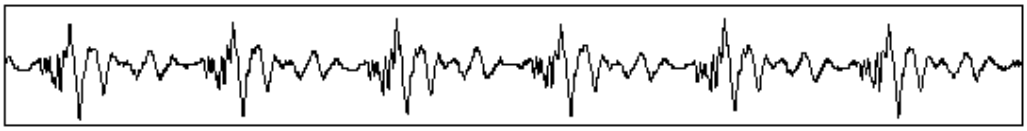
(f) Processed waveform, ensemble method, $c = 9$, $N_1 = 30$, $N_2 = 50$

Fig. 5.5 Waveform of speech output (/*a*/) processed with ensemble averaging method

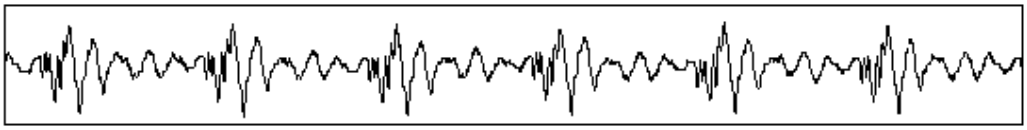


(a) Unprocessed speech (/*a*/) generated using NP-1 electrolarynx

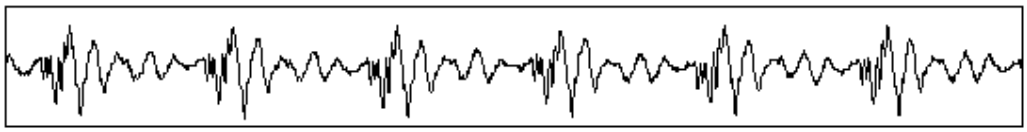
(b) Extracted excitation impulse with $c = 9$



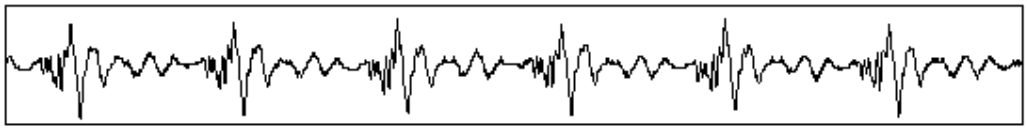
(c) Processed waveform, LMS method, $\mu = 0.01$, N = 60



(d) Processed waveform, LMS method, $\mu = 0.01$, N = 30

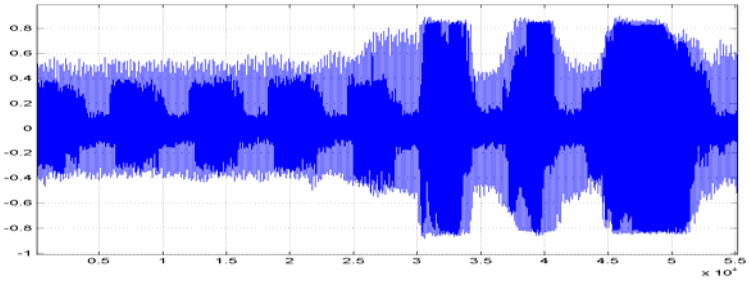


(e) Processed waveform, LMS method, $\mu = 0.01$, N=110

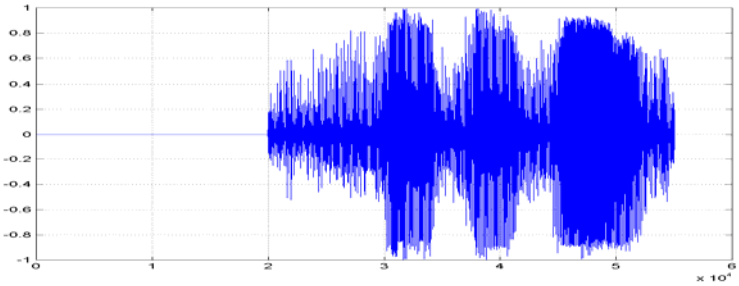


(f) Processed waveform, LMS method, $\mu = 0.001$, N = 110

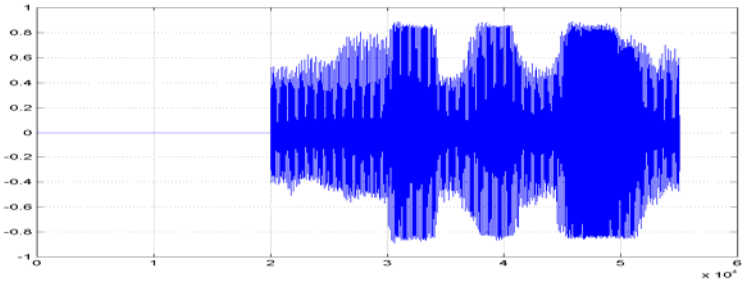
Fig. 5.6 Waveform of speech output (/*a*/) processed using LMS adapting method



(a) Unprocessed speech "plug and play" generated using NP-1 electrolarynx

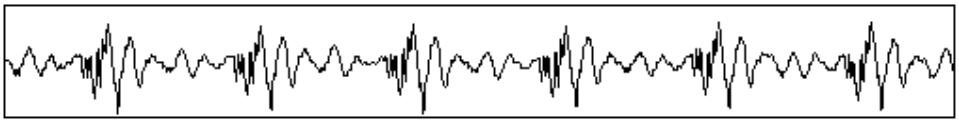


(b) Processed speech using ensemble average method

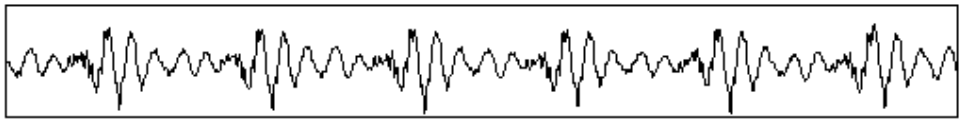


(c) Processed speech using LMS algorithm

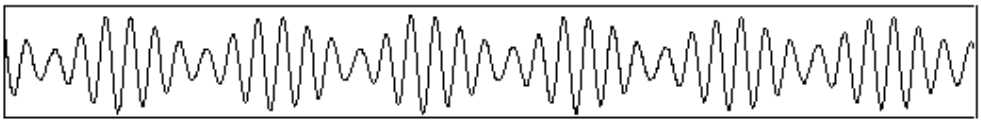
Fig. 5.7 Speech ‘plug and play’ processed using noise canceller algorithms: ensemble average & LMS.



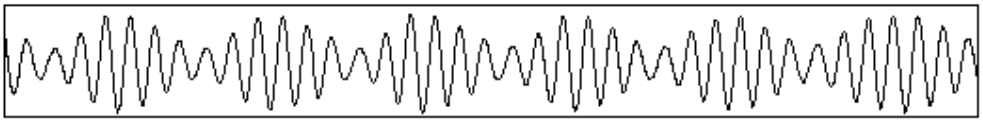
(a) Unprocessed speech (/*a*/) generated with NP-1



(b) Processed speech (/*a*/) using spectral subtraction method with N = 244, $\alpha = 1$, $\beta = 0$, $\gamma = 2$

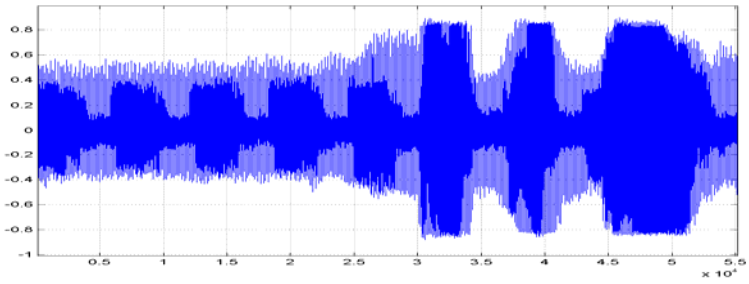


(c) Processed speech (/*a*/) using spectral subtraction method with N = 244, $\alpha = 2$, $\beta = 0$, $\gamma = 1$

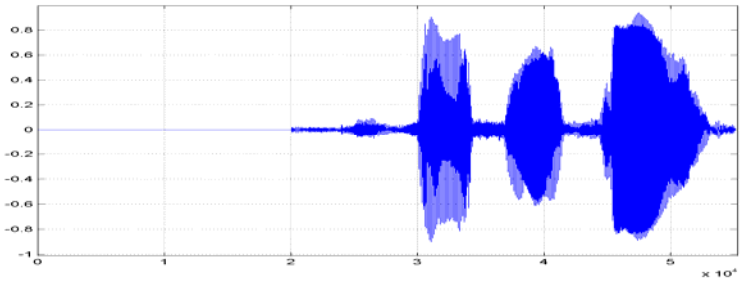


(d) Processed speech (/*a*/) using spectral subtraction method with N = 244, $\alpha = 2$, $\beta = 0.001$, $\gamma = 1$

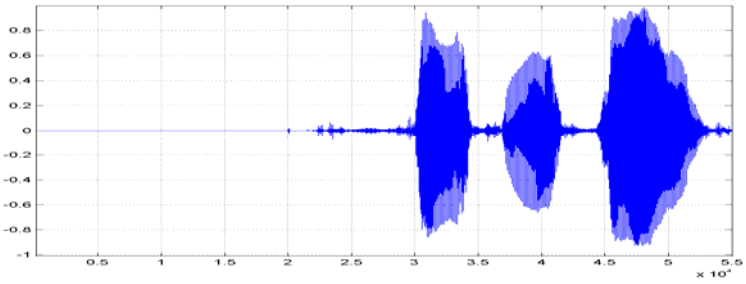
Fig. 5.8 Speech /*a*/ processed using spectral subtraction method



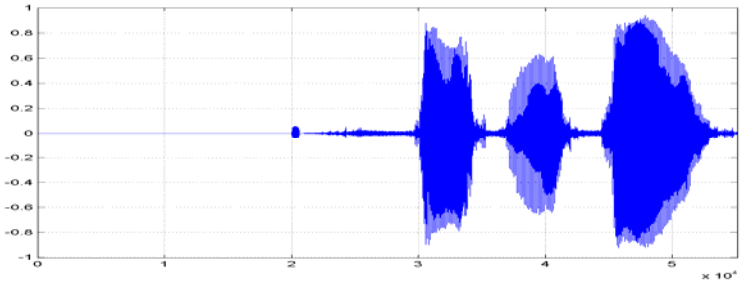
(a) Unprocessed speech "plug and play" generated using NP-1 electrolarynx



(b) Processed waveform, spectral subtraction method, N = 60, $\alpha = 2$, $\beta = 0$, $\gamma = 1$

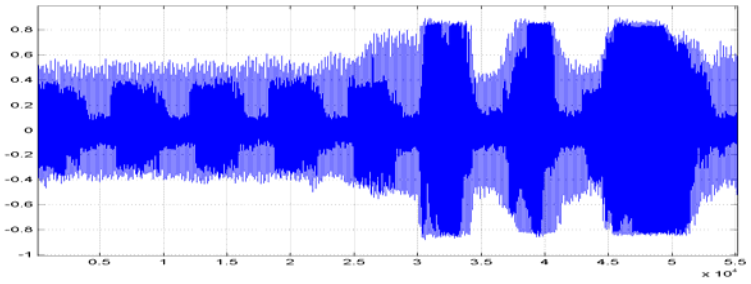


(c) Processed waveform, spectral subtraction method, N = 244, $\alpha = 2$, $\beta = 0$, $\gamma = 1$

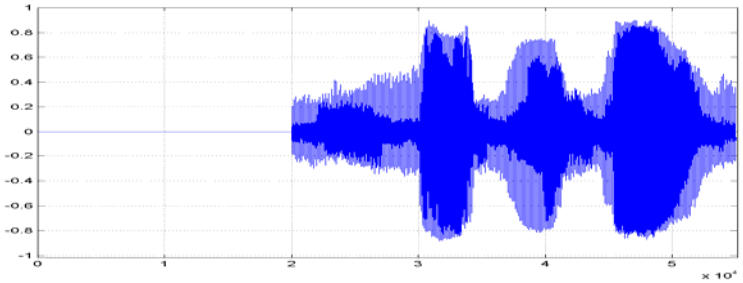


(d) Processed waveform, spectral subtraction method, N = 1220, $\alpha = 2$, $\beta = 0$, $\gamma = 1$

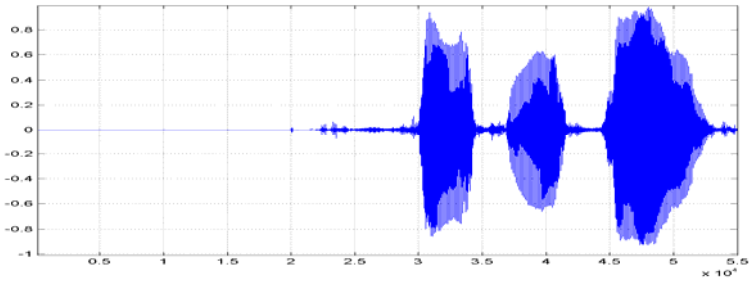
Fig. 5.9 Effect of window length on noise cancellation by spectral subtraction method



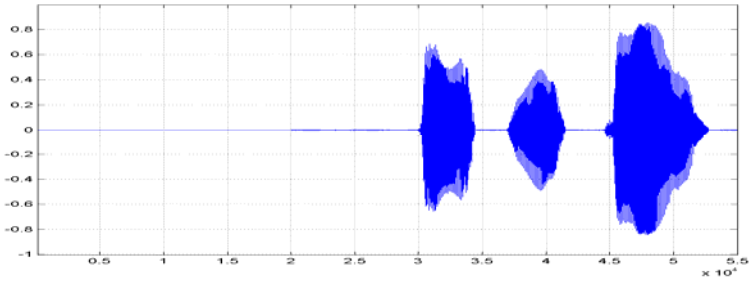
(a) Unprocessed speech "plug and play" generated using NP-1 electrolarynx



(b) Processed waveform, spectral subtraction method, N = 244, $\alpha = 0.5$, $\beta = 0$, $\gamma = 1$

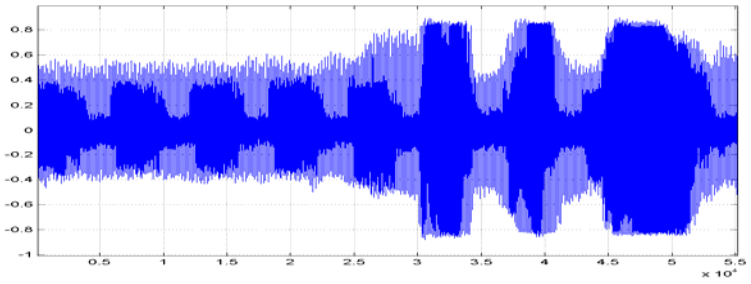


(c) Processed waveform, spectral subtraction method, N = 244, $\alpha = 2$, $\beta = 0$, $\gamma = 1$

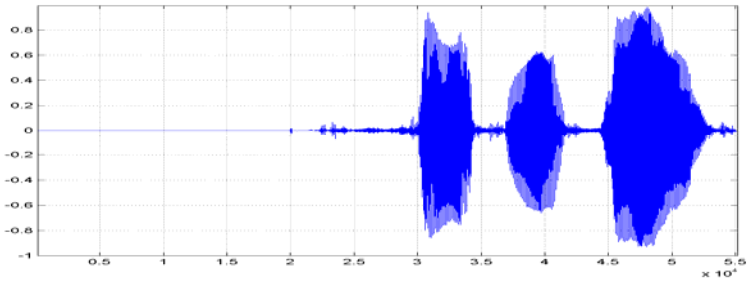


(d) Processed waveform, spectral subtraction method, N = 244, $\alpha = 6$, $\beta = 0$, $\gamma = 1$

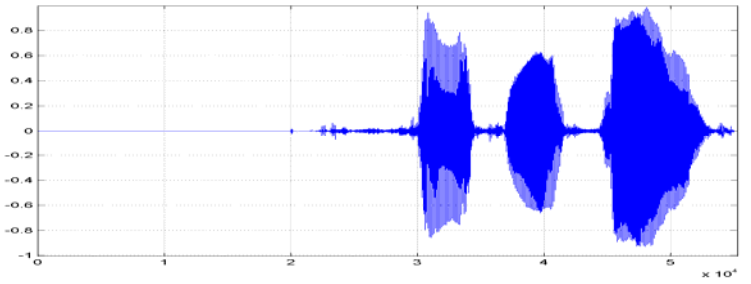
Fig. 5.10 Effect of α on noise cancellation by spectral subtraction method



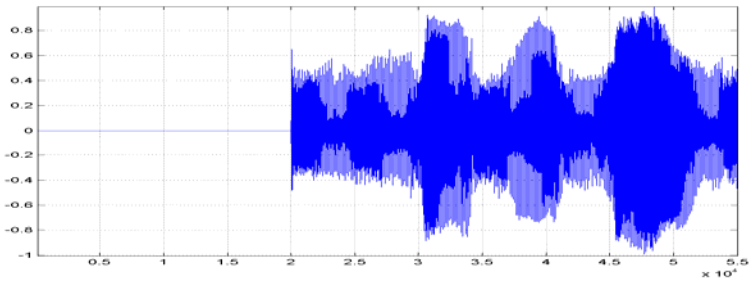
(a) Unprocessed speech "plug and play" generated using NP-1 electrolarynx



(b) Processed waveform, spectral subtraction method, N = 244, $\alpha = 2$, $\beta = 0$, $\gamma = 1$

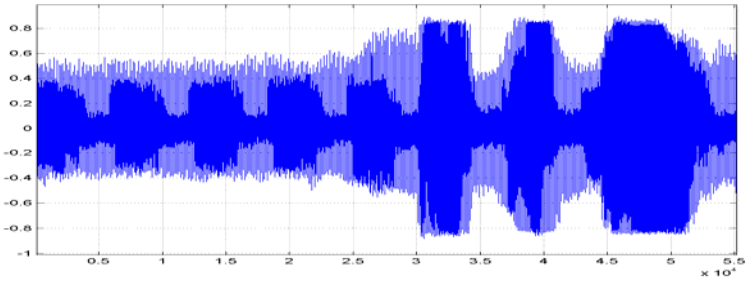


(c) Processed waveform, spectral subtraction method, N = 244, $\alpha = 2$, $\beta = 0.001$, $\gamma = 1$

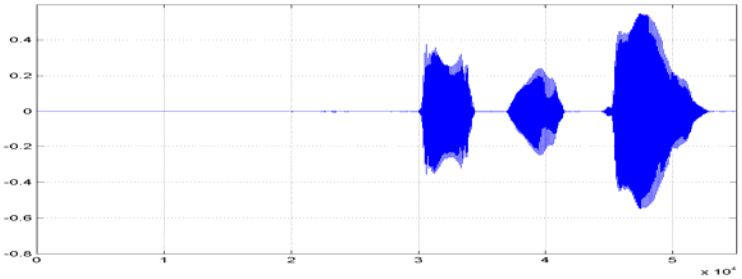


(d) Processed waveform, spectral subtraction method, N = 244, $\alpha = 2$, $\beta = 1$, $\gamma = 1$

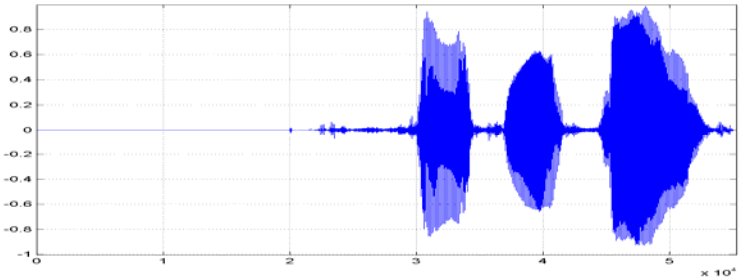
Fig. 5.11 Effect of β on noise cancellation by spectral subtraction method



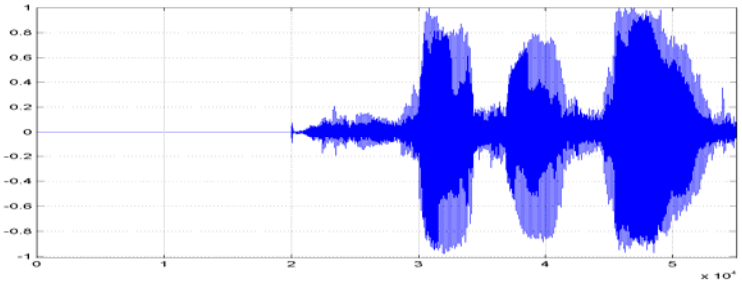
(a) Unprocessed speech "plug and play" generated using NP-1 electrolarynx



(b) Processed waveform, spectral subtraction method, N = 244, $\alpha = 2$, $\beta = 0.001$, $\gamma = 0.5$

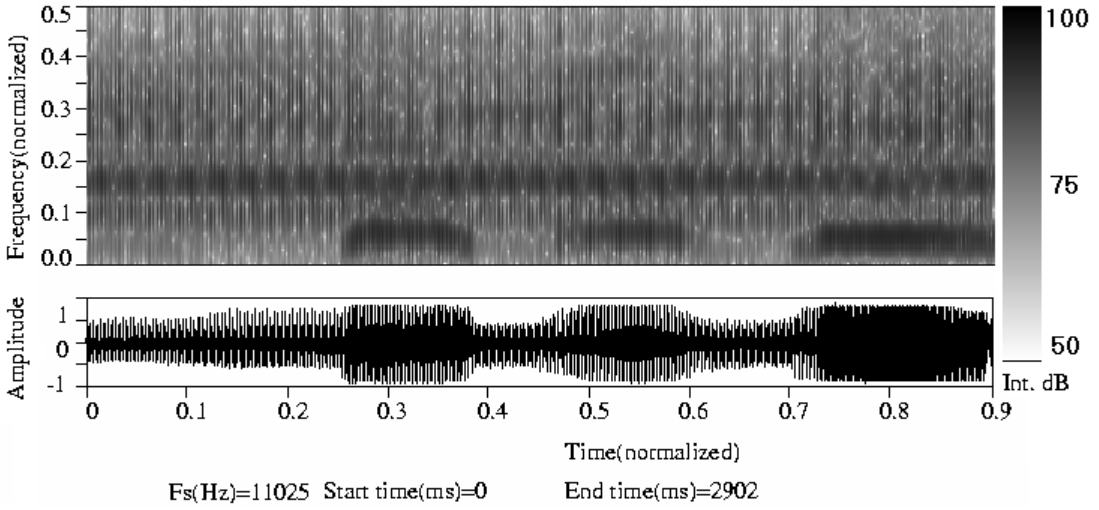


(c) Processed waveform, spectral subtraction method, N = 244, $\alpha = 2$, $\beta = 0.001$, $\gamma = 1$

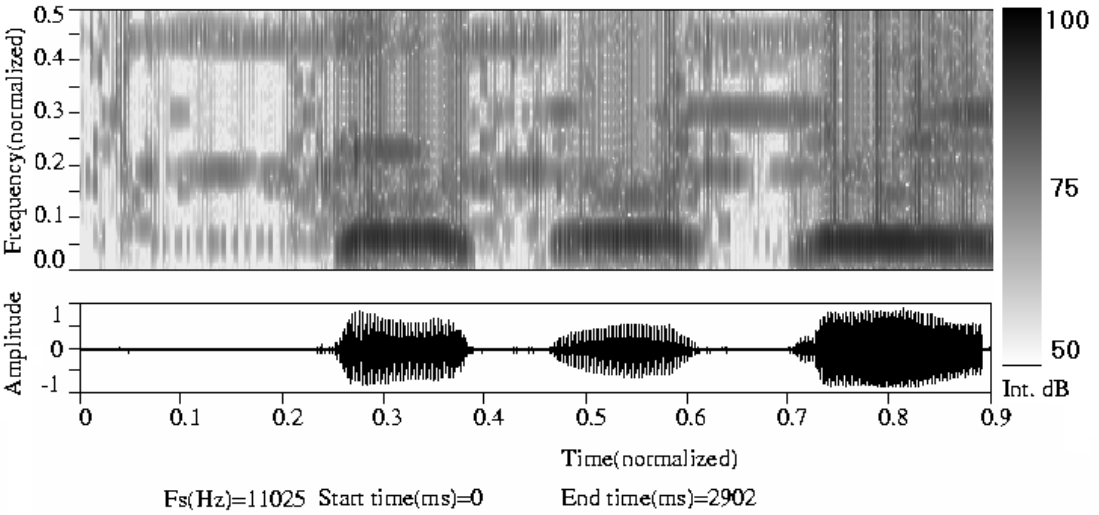


(d) Processed waveform, spectral subtraction method, N = 244, $\alpha = 2$, $\beta = 0.001$, $\gamma = 2$

Fig. 5.11 Effect of γ on noise cancellation by spectral subtraction method



(a)



(b)

Fig. 5.13 (a) Wideband spectrogram of the speech 'plug and play' uttered with the help of transcervical electrolarynx.
(b) Wideband spectrogram of the above input speech file processed using spectral subtraction algorithm with window length of 244, $\alpha = 2$, $\beta = 0.001$, and $\gamma = 1$.

Chapter 6
SUMMARY AND CONCLUSIONS

6.1 Summary

Different types of artificial larynxes were studied. Among the various artificial larynxes available, the transcervical (or external electronic) larynxes are widely used. The problems associated with these larynxes were studied. Of these, the problem of background noise generation is severe and affects the intelligibility considerably. The project objective was to investigate signal processing techniques for reducing the background noise, thereby improving the quality of speech output.

Single input leakage cancellation technique based on waveform subtraction used by Shah [7] was implemented for evaluation. The technique continuously estimates the location of excitation impulses. During non-speech training segment, the impulse response of the leakage path is estimated with the help of excitation impulses and (i) ensemble averaging and (ii) LMS adapting methods. Subsequently (during speech) background noise waveform is estimated as the convolution of the estimated excitation impulses with the earlier estimated impulse response of the leakage path. The estimated background noise is subtracted from the noisy speech in order to obtain cleaned speech signal. The results indicated that there was no consistency in noise reduction, and the technique did not show an improvement in the speech quality.

Single input spectral subtraction algorithm reported earlier for enhancement of speech signal corrupted by uncorrelated noise [14][15] was studied, and adapted for cancellation of background interference in alaryngeal speech. Results from processing of speech recorded using a transcervical larynx indicated that the technique could be used for effective noise reduction. The effect of variation in the values of the various parameters for spectral subtraction was investigated.

6.2 Conclusions

Implementation of single input leakage canceller algorithms based on estimation of the impulse response of the leakage path during non-speech segment did not result in noise reduction. This was primarily due to variations in the impulse response of the leakage path. The variations are most likely caused by change in application pressure while holding the device against the throat. As the method is

dependent on time domain subtraction of estimated noise, even slight jitters may contribute towards further degradation rather than enhancement of speech.

In the spectral subtraction method, only the magnitude spectrum of the noise is estimated. Hence temporal jitters do not affect the processed output. Further by using the modified subtraction method, there is a good tolerance for errors in the estimation of spectral values in the noise spectrum.

It has been found that the best results are obtained when the processing window length equals two pitch periods. Normally in an artificial larynx, the pitch period is fixed, hence the signal processor can use a fixed pitch value. In case of variable pitch control, the signal processing block can be given trigger pulses from the pulse generator, and detection of glottal pulses by signal processing may not be required.

In the modified spectral subtraction method, there are three processing parameters: subtraction factor α, spectral floor factor β, and exponent factor γ. On the basis of theoretical formulation of the spectral subtraction method for application to alaryngeal speech, we should select $\alpha = 1$ (complete subtraction of magnitude speech of noise), $\beta = 0$ (zero noise floor), and $\gamma = 2$ (subtraction of power spectrum). Experimental investigation indicated that for effective noise suppression we have to use over-subtraction with $2 < \alpha < 5$. For a reasonable balance between musical and broadband noises, $0.001 < \beta < 0.01$. Best perceptual quality was obtained with $\gamma = 1$. It is to be noted that the parameters obtained above gave good results for the recordings made by the author using NP-1 electrolarynx. In general, different settings of the parameters may be needed for different devices and users.

6.3 Suggestions for future work

The results obtained with the implementation of spectral subtraction method indicate that, when the window size was a multiple of the pitch period, the quality of the output was better. However in the implementation, the starting position of the window was not aligned with respect to the position of the impulses in the input. The effect of positioning the window, in synchronization with the impulse, needs to be investigated.

In the method investigated, the phase spectrum of the noisy speech was retained and coupled with the "cleaned" spectrum for obtaining processed speech for

each window segment. It is expected that quality can be better is phase spectrum also is noise-free. Towards this, it can be noted that vocal tract as well as leakage path are minimum phase systems, because of their passive nature. For a minimum phase system, the phase response can be restored from its magnitude response. The resynthesis of the phase from the magnitude can be done using the cepstral method [3]. The resulting speech could be of better quality, and needs to be investigated. An attempt for implementing the method resulted in distortion of speech. It is suggested that this implementation should be carried out along with aligning of processing window with excitation impulses.

The method proposed and developed here needs to be implemented for real-time processing. It can be incorporated as part of the communication devices to be used by patients using artificial larynxes. As the next step, the signal processing technique for speech enhancement can be implemented as a real-time processor as shown in Fig. 6.1. Microphone picks up the alaryngeal speech signal, which is input to the ADC of the signal processor. Processed and amplified speech output is directed towards the listeners. Signal processor may use pitch value or trigger pulses from the pulse generation.

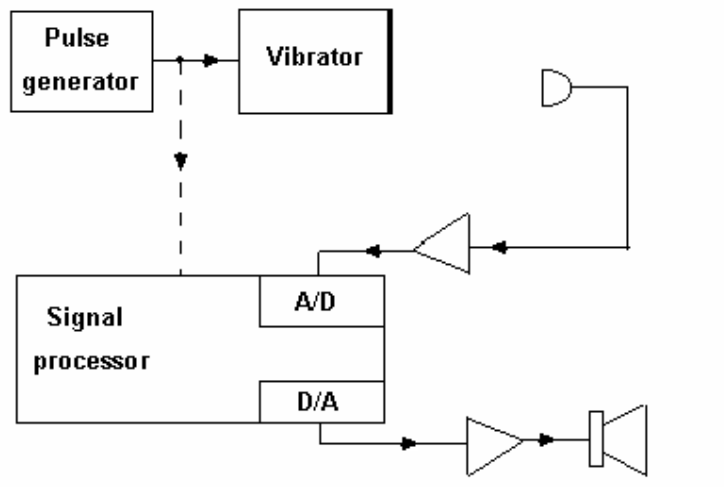


Fig. 6.1 Real-time implementation scheme of noise cancellation with spectral subtraction algorithm

REFERENCES

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

[2] Y. Lebrun, “History and development of laryngeal prosthetic devices,” *The Artificial Larynx*, Amsterdam: Swets and Zeitlinger, pp. 19-76, 1973.

[3] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*, Englewood Cliffs, New Jersey: Prentice Hall, 1978.

[4] "Esophageal voice", http://www.origin8.nl/medical/esophgea.com, Jan 2002.

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

[6] " Speech aids", http://jkemp.larynxlink.com/speechaids.htm, Jan 2002.

[7] H. R. Shah, “A study of background noise in transcervical electrolarynx, ” *M. Tech. Dissertation*, Guide: Dr. P. C. Pandey, School of Biomedical Engineering, IIT Bombay, Jan. 1999.

[8] Qi Yingyong and B. Weinberg, “Low-frequency energy deficit in electro laryngeal speech,” *Journal of Speech and Hearing Research*, vol. 34, pp.1250-1256, Dec. 1991.

[9] "Artificial larynx with PZT ceramics",http://www.nagoya_u.ac.jp/activity/1999-e/VOICE_99E.html, Jan 2002.

[10] M. Weiss, G. Yeni-Komshian, and J. Heinz, "Acoustical and perceptual characteristics of speech produced with an electronic artificial larynx," J. *Acoust. Soc. Am.*, Vol. 65, No. 5, pp. 1298-1308, May 1979.

[11] H. L. Barney, F. E. Haworth, and H. K. Dunn, "An experimental transistorized artificial larynx," *Bell Systems Technical Journal*, vol. 38, No. 6, pp 1337-1356, Nov. 1959.

[12] S. Haykin, *Adaptive Filter Theory*, Englewood Cliffs, New Jersey: Prentice Hall, 1991.

[13] B. Widrow, J.R.Glover, J.M.Mccool, and J. Kaunitz, “Adaptive noise canceling: principles and applications,” *Proc. IEEE*, vol. 63, pp. 1692-1716, Dec. 1975.

[14] S. F. Boll, “Suppression of acoustic noise in speech using spectral subtraction,” *IEEE Trans. Acoust., Speech and Signal Processing*, vol. ASSP-27, pp. 113-120, Apr. 1979.

[15] M. Berouti, R. Schwartz, and J. Makhoul, “Enhancement of speech corrupted by acoustic noise,” *Proc. IEEE. Conf. on Acoust., Speech, Signal Processing*, pp. 208-211, Apr. 1979.

[16] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing*, New Delhi: Prentice Hall of India, 1997

[17] M. S. Ratanpal, “Speech processing for binaural dichotic presentation,” *M. Tech. Dissertation*, Guide: Dr. P. C. Pandey, Department of Electrical Engineering, IIT Bombay, Jan 2000.

[18] Griffin Labs, 27636, Ynez road, #L7199, Temecula, CA 925591, "Hear how Trutone compares with its competitors", http://www.griffinlab.com/compare/compare.htm, Jan 2002.

[19] "Communication products", http://www.kapitex.com/products/communication/products-communication3.htm, Jan 2002.

[20] "Personal communication", http://www.assis-tech.com/products/products-list/communications/personal_communication.htm, Jan 2002.

Appendix A

Comparison of various artificial larynxes based on their specifications [18].

	Trutone	Nuvois	Servox	Optivox	Solatone
Weight	4.5 oz	4.5 oz	6.5 oz	7.5 oz	4.5 oz
Size	1.34 x 4.12" Long	1.38 x 4.4" Long	1.38 x 4.69" Long	1.5 x 5.12" Long	1.34 x 4.12" Long
Battery	9 V Transistor	9 V Transistor	7.2 V Custom	9 V Transistor	9 V Transistor
Speaking Tones Available	1 - 100	1	2	1	1
No. of buttons	1	1	2	1	1
Tone change while speaking without moving thumb?	Yes	No	No	No	No

Trutone artificial larynx [6]

Servox artificial larynx [6]

Optivox artificial larynx [19]

Solatone artificial larynx [6]

Nuvois artificial larynx [20]