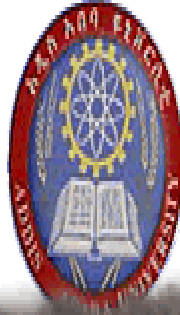

Addis Ababa
University

(Since 1950)



**Convex Combination of Adaptive Linear Finite Impulse Response and
Nonlinear Volterra Filters for Acoustic Echo Cancellation**

**By
Tagel Labena**

**Advisor:-
Dr. Eneyew Adugna**

A thesis submitted to the school of graduate studies of AAiT in partial fulfillment
of the requirements for the degree of Masters of Science in communication
engineering.

Department of Electrical and Computer Engineering

Addis Ababa Institute of Technology

Addis Ababa, Ethiopia

October, 2011

Department of Electrical and Computer Engineering
Addis Ababa Institute of Technology
Addis Ababa, Ethiopia

**Convex Combination of Adaptive Linear Finite Impulse Response
and Nonlinear Volterra Filter for Acoustic Echo Cancellation**

By
Tagel Labena

Approval by Board of Examiners

Chairman, Dept. Graduate Committee

Signature

Dr. Eneyew Adugna
Advisor

Signature

Internal Examiner

Signature

External Examiner

Signature

Declaration

I, the undersigned, declare that this thesis is my original work, has not been presented for a degree in this or any other university, and all sources of materials used for the thesis have been fully acknowledged.

Tagel Labena
Name

Signature

Place: Addis Ababa

Date of Submission: _____

This thesis has been submitted for examination with my approval as a university advisor.

Dr. Eneyew Adugna
Advisors Name

Signature

Dedicated to

my beloved mom

who couldn't live long enough to see how far I have come.

I'll always remember you -rest in peace.

Table of Contents

Acknowledgments.....	i
Abstract.....	ii
List of Abbreviations.....	iii
List of Figures.....	v
List of Tables.....	viii
1. Introduction.....	1
1.1. Need for Acoustic Echo Cancellation.....	1
1.2. Thesis Organization.....	2
2. Background Information.....	3
2.1. Acoustic Echo.....	3
2.2. Hybrid (Electric) Echo.....	4
3. State of the Art in Acoustic Echo Cancellation.....	6
3.1. Acoustic Echo Suppression (AES).....	6
3.2. Acoustic Echo Cancellation (AEC).....	7
3.3. Research Question	9
3.4. Objective of the Thesis.....	10
3.4.1. General Objective.....	10
3.4.2. Specific Objectives.....	10
4. Finite Impulse Response (FIR) Linear Adaptive Filters.....	11
4.1. Introduction.....	11
4.2. Linear Adaptive Filters.....	11

4.2.1. The normalized least-mean-square (NLMS) algorithm.....	13
4.3. Performance Measures.....	14
4.3.1. Echo Return Loss Enhancement (ERLE).....	15
4.3.2. Mean Square Error (MSE).....	15
4.3.3. Algorithm Misalignment Error (AME).....	15
4.4. Simulation Results.....	16
4.4.1. Experimental Setup.....	16
4.4.2. Results.....	17
4.4.3. Discussion.....	23
5. Adaptive Nonlinear Volterra Filters for Nonlinear AEC	24
5.1. Introduction.....	24
5.2. Sources of Nonlinearities in Acoustic Echo.....	24
5.3. The Volterra Filter.....	25
5.4. Adaptive Second Order Volterra Filter.....	26
5.5. Implementation of Second Order Volterra Filter.....	27
5.6. Simulation Setup.....	30
5.7. Results.....	30
5.8. Discussion.....	35
6. Convex Combination of Adaptive Linear and Nonlinear Volterra Filters for AEC.....	36
6.1. Introduction.....	36

6.2. Convex Combination of Linear FIR and Second Order Volterra Filters.....37

6.3. Simulation Setup.....41

6.4. Performance Measure.....41

6.5. Simulation Results.....42

6.6. Discussion.....47

7. Double-Talk Detection (DTD).....49

7.1. Introduction.....49

7.2. Working Principle of DTDs.....49

7.3. DTD Based on Mutual Information (MI).....51

 7.3.1. MI and Its Calculation.....51

7.4. Simulation Setup.....54

7.5. Results.....55

7.6. Discussion.....60

8. Conclusions and Recommendations for Further Research.....61

8.1. Conclusions.....61

8.2. Recommendations for Further Research.....61

References.....62

Appendix A.....66

Acknowledgments

I would like to thank my advisor Dr. Eneyew Adugna not only for his invaluable advice and guidance during the course of my research studies but also his treatment and understanding of the situation I have been during crucial times of my study. I am also grateful for all my family members for their continual encouragement and support. I would also like to recognize the management of Ethio Telecom for the sponsorship not to mention their understanding and support to finish my study. I would like to thank my fellow communication engineering graduate students for their close friendship and support when I was in no man's land. Finally, I am sincerely indebted to my wife, Sosena Seyoum, for her continuous love and support during my studies. Her belief in me gave me the strength needed to keep me going and for this I'm deeply grateful.

Abstract

We all know how much joy and freedom hands-free communication system has brought to our daily life. But associated with this is the problem of acoustic echo which degrades the quality of communication. A number of researches have been conducted in order to eliminate acoustic echo. Most of the acoustic echo cancellers are made on the assumptions of the linearity of the echo. However, this assumption is no longer valid when the nonlinearity along the echo path is taken into account. This led to the invention of Nonlinear Acoustic Echo Cancellation (NLAEC) techniques such as Volterra filters. However, in some situations, NLAEC is inferior to a linear Acoustic Echo Cancellation (AEC), especially when the channel generates a negligible (or no) nonlinear echo. In general, the ratio of the linear to nonlinear echo signal power is unknown a priori, and will vary over time, thus making it difficult to know if NLAEC would improve or degrade the cancellation. Therefore, in this thesis a convex combination of the linear and the nonlinear Volterra filters is studied. Simulation has been carried out and the results indicate the combination scheme follows the best contributing filter regardless of the nature of the echo. Another important problem of AEC is the problem of double talk (DT) which occurs when both parties talk at the same time. If DT is not controlled it results in diverging of the adaptive filter coefficients and as a result audible echoes pass without being cancelled. In order to alleviate this problem most DT controllers have been studied based on correlation of the available signals. Since correlation is valid only for linear relationship; this assumption doesn't hold when the nonlinearity along the echo path is considered. Therefore, in this thesis mutual information (MI) based DT detection has been studied together with AEC. Simulation is carried out by introducing a DT region and the results obtained indicate performance improvement of 97.06% in terms of algorithm misalignment error (AME).

Key words: Acoustic echo, Convex Combination, Volterra, adaptive filters

List of Abbreviations

AEC	Acoustic Echo Cancellation/Canceler
AES	Acoustic Echo Suppression/Suppressor
AFC	Adaptive Filter Coefficients
AME	Algorithm Misalignment Error
AR	Autoregressive
ARMA	Autoregressive Moving Average
DT	Double Talk
DTD	Double Talk Detection/Detector
ERLE	Echo Return Loss Enhancement
EMSE	Excess Mean Square Error
FIR	Finite Impulse Response
IIR	Infinite Impulse Response
LMS	Least Mean Square
LNLR	Linear to Nonlinear Ratio
LRT	Log likelihood Ratio Test
MA	Moving Average
MSE	Mean Square Error
MI	Mutual Information
NLAEC	Nonlinear Acoustic Echo Cancellation

NLMS	Normalized Least Mean Square
RIR	Room Impulse Response
RLS	Recursive Least Square
SM	System Mismatch
ST	Single Talk

List of Figures

<u>Figure</u>	<u>Page</u>
Figure 1.1: Evolutions of Telephone Handsets.....	1
Figure 2.1: Sources of acoustic echo in a room.....	4
Figure 2.2: Source of hybrid echo in telecommunication system.....	4
Figure 2.3: A 2-Wire to 4-Wire Conversion Hybrid Circuit.....	5
Figure 3.1: Acoustic Echo Suppressor Block Diagram.....	7
Figure 3.2: Adaptive Acoustic Echo Canceller Block Diagram.....	8
Figure 4.1: Structure of FIR Linear Filter.....	12
Figure 4.2: Basic setup of Linear AEC Using Adaptive FIR Linear Filter.....	14
Figure 4.3: Plot of Room Impulse Response (RIR) and Linear FIR Adaptive Filter Coefficients (AFC) for number of iterations =2	17
Figure 4.4: Plot of Room Impulse Response (RIR) and Linear FIR Adaptive Filter Coefficients (AFC) for number of iterations =50	17
Figure 4.5: Plot of Mean Square Error [Learning curve] for different number of iterations.....	19
Figure 4.6: Plot of Echo Return Loss Enhancement for different number of iterations.....	20
Figure 4.7: Plot of Room Impulse Response (RIR) and Linear FIR Adaptive Filter Coefficients (AFC) for number of iterations =50	21
Figure 4.8: Plot of Mean Square Error [Learning curve] for number of iterations =50.....	22
Figure 4.9: Plot of Echo Return Loss Enhancement for number of iterations =50.....	19
Figure 5.1: Sources of nonlinearity in the acoustic echo path.....	24
Figure 5.2: Block diagram of adaptive second order Volterra filter.....	26
Figure 5.3: Architecture of Second order Volterra Filter.....	27
Figure 5.4: Plot of Room Impulse Response (RIR) and Linear Volterra kernel response.....	31

Figure 5.5: Plot of the nonlinear component of the acoustic echo to be identified by Volterra quadratic kernel.....	32
Figure 5.6: Plot of the Volterra quadratic kernel response after 50 iterations.....	33
Figure 5.7: Plot of the MSE [Learning Curve] of Adaptive Second Order Volterra Filter.....	34
Figure 5.8: Plot of Echo Return Loss Enhancement for different number of iterations.....	34
Figure 6.1: Block diagram of convex combination of linear FIR filter and second order Volterra filters.....	37
Figure 6.2: Block diagram of Adaptive convex combination of linear FIR filter and second order Volterra filter for AEC.....	38
Figure 6.3: Evolution of the combination adaptation parameter.....	42
Figure 6.4: Plot of ERLE of the Linear FIR, Second Order Volterra Filter and their Convex Combination.....	43
Figure 6.5: Plot of EMSE of the Linear FIR, Second Order Volterra Filter and their Convex Combination.....	44
Figure 6.6: Evolution of the combination adaptation parameter.....	45
Figure 6.7: Plot of ERLE of the Linear FIR, Second Order Volterra Filter and their Convex Combination.....	46
Figure 6.8: Plot of EMSE of the Linear FIR, Second Order Volterra Filter and their Convex Combination.....	47
Figure 7.1: AEC with Double Talk Detector (DTD).	50
Figure 7.2: Double Talk Region Detection using MI for linear acoustic echo.....	55
Figure 7.3: Algorithm misalignment error without DTD for linear AEC.....	56
Figure 7.4: Algorithm misalignment error with MI based DTD for linear AEC.....	56
Figure 7.5: Double Talk Region Detection using MI for Nonlinear acoustic echo.....	57
Figure 7.6: Algorithm misalignment error with MI based DTD for linear AEC.....	58
Figure 7.7: Algorithm misalignment error without DTD for linear AEC.....	58
Figure 7.8: Algorithm misalignment error of quadratic kernel for nonlinear AEC with MI based DTD.....	59

Figure 7.9: Algorithm misalignment error of quadratic kernel for nonlinear AEC without DTD.....59

List of Tables

<u>Table</u>	<u>Page</u>
Table 1: NLMS algorithm.....	13
Table 2: MSE and ERLE as a function of number of iterations.....	20
Table 3: NLMS algorithm for second order Volterra filter.....	29

Chapter 1

Introduction

1.1. Need for Acoustic Echo Cancellation

When telephone was first invented people were forced to use both hands to make a telephone conversation [1]. It was soon found that using both hands was inconvenient and other telephone handsets were created which freed one hand. Nowadays, it's common to see most people use both hands to perform other activities while talking on the phone using loudspeakers and high gain microphones.

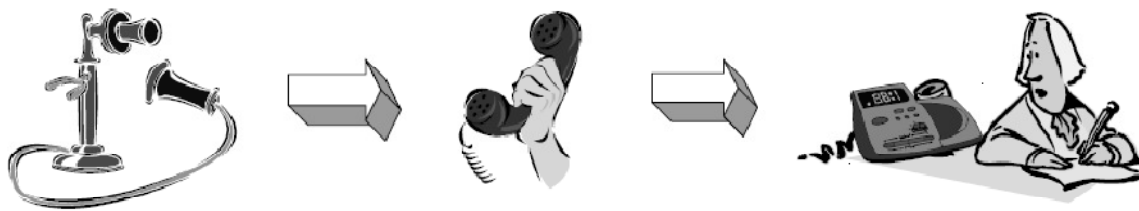


Figure 1.1: Evolutions of Telephone Handsets

The main problem of hands-free devices is that acoustic echo severely degrades the quality of communication. Normally, people hear their own voices through air and bone conduction. This feedback (or echo) is used by the speakers to adjust their volume. During telephone conversations, people find it reassuring to hear an additional echo of their own voices through the earpiece. This is a psychoacoustic phenomenon: most people would find the absence of an echo disturbing, believing that if you cannot hear yourself, then the other person cannot hear you either. However, an echo that is too loud or arrives too late can be annoying to the speaker. The perceptual effects of an echo depend on the time delay between the incident and reflected waves, the strength of the reflected waves, and the number of paths through which the waves are reflected. If the echo is more than a few milliseconds long then it becomes noticeable, and can be annoying and disruptive. In general, echoes with appreciable amplitude and a delay of more than 1 *ms* are noticeable [2]. Provided the round-trip delay is on the order

of a few milliseconds, echo gives a telephone call a sense of “liveliness”. However, echoes become increasingly annoying and objectionable with the increasing round-trip delay and amplitude in particular for delays of more than 20 *ms*. Hence echo cancellation is an important aspect of the design of modern telecommunication systems such as conventional wire line telephones, hands-free phones, cellular mobile phones, or teleconference systems.

1.2. Thesis Organization

The rest of the thesis is organized as follows: - In Chapter 2 the background information regarding Acoustic Echo Cancellation (AEC) is given. In Chapter 3 the state of the art in AEC is thoroughly investigated to find out what has already been done to combat echo problem and the research motivation and objective of the thesis are defined. In Chapter 4 linear FIR filtering and linear AEC are discussed and simulation results are also indicated for linear acoustic echo. In Chapter 5 nonlinear filtering, in particular Volterra filtering, is discussed and how it is used to cancel nonlinear echo is also indicated together with simulation results. In Chapter 6 convex combination of adaptive linear FIR and Volterra filter for AEC is described. Simulation results are also presented and discussed. In Chapter 7 DTD (Double Talk Detector) based on Mutual Information (MI) is discussed and also how it can be applied to detect the presence of near-end speech in AEC scenario is shown. In Chapter 8 conclusions of what has been observed in this thesis is given and also recommendations for future research is indicated. Finally, References and Appendix are also included at the end of the document.

Chapter 2

Background Information

Echo is the repetition of a waveform due to reflection from points where the characteristics of the medium through which the wave propagates changes. Echo is usefully employed in sonar and radar for detection and exploration purposes. In telecommunication, echo can degrade the quality of service, and echo cancellation is an important part of communication systems. The development of echo reduction began in the late 1950s, and continues today as new integrated landline and wireless cellular networks put additional requirement on the performance of echo cancellers. There are two types of echo in communication systems: *acoustic echo* and telephone line *hybrid echo (or electric echo)* [2].

2.1. Acoustic Echo

Acoustic echo results from a feedback path set up between the speaker and the microphone in a mobile phone, hands-free phone, teleconference or hearing aid system. Acoustic echo may be reflected from a multitude of different surfaces, such as walls, ceilings and floors, and travels through different paths as indicated in Figure 2.1. If the time delay is not too long then the acoustic echo may be perceived as a soft reverberation, and may add to the artistic quality of the sound. Concert halls and church halls with desirable reverberation characteristics can enhance the quality of a musical performance. In its worst case, acoustic feedback can result in howling if a significant proportion of the sound energy transmitted by the loudspeaker is received back at the microphone and circulated in the feedback loop.

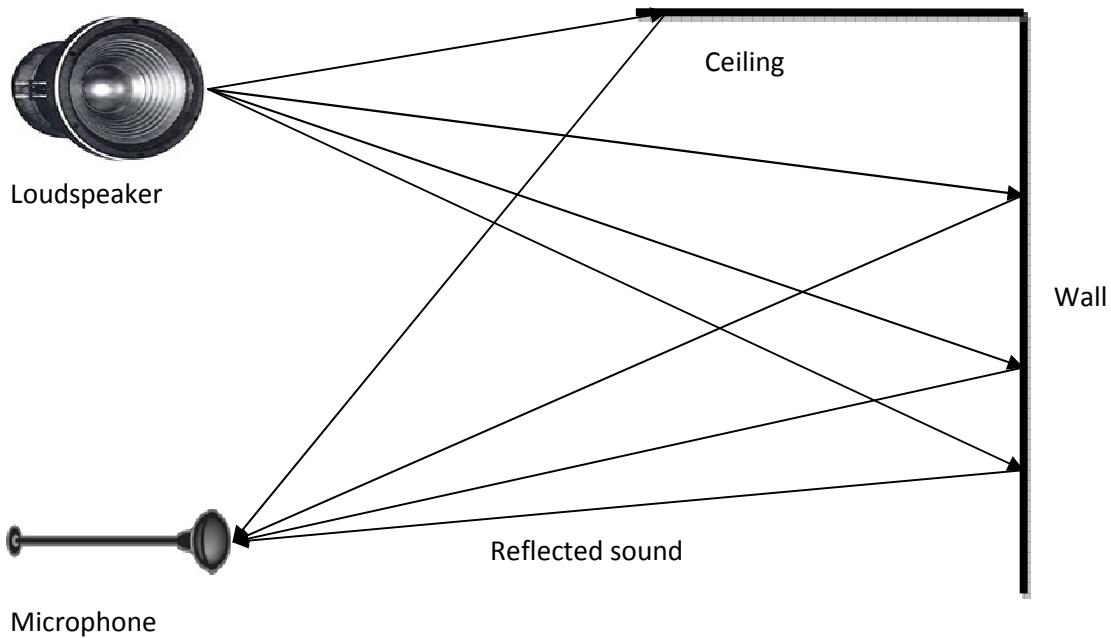


Figure 2.1: Sources of acoustic echo in a room

2.2. Hybrid (Electric) Echo

Telephone line echoes result from an impedance mismatch at telephone exchange hybrids where the subscriber's 2-wire line is connected to a 4-wire line as in Figure 2.2 below. In the early days of expansion of telephone networks, the cost of running a 4-wire line from the local exchange to subscribers' premises was considered uneconomical.

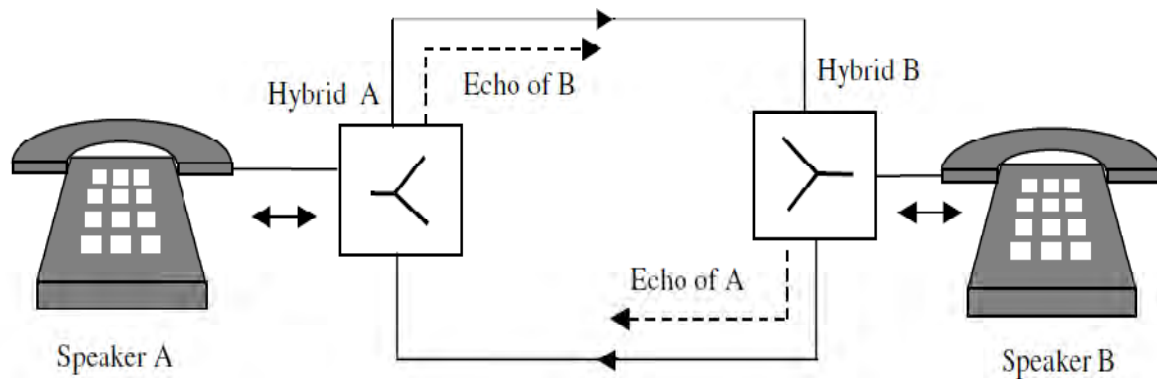


Figure 2.2: Source of hybrid echo in telecommunication system

Hence, at the exchange the 4-wire trunk lines are converted to 2-wire subscribers local lines using a 2/4-wire hybrid bridge circuit as in Figure 2.3 below. At the receiver due to any imbalance between the 4/2-wire bridge circuits, some of the signal energy of the 4-wire circuit bounces back towards the transmitter, constituting an echo signal.

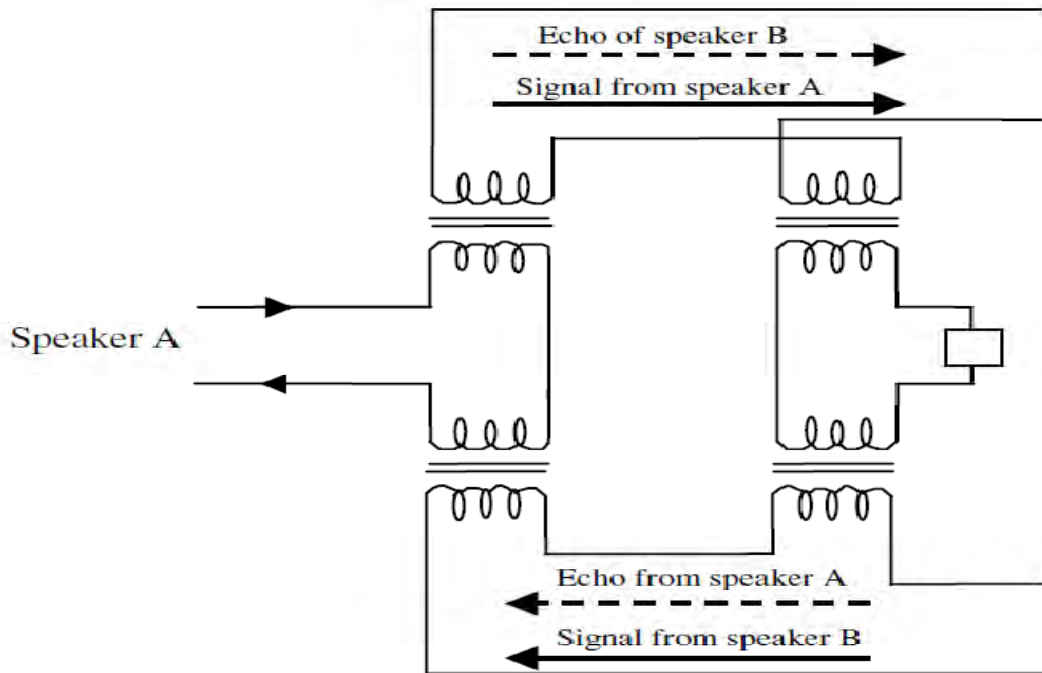


Figure 2.3: A 2-Wire to 4-Wire Conversion Hybrid Circuit

Compared to hybrid echo, acoustic echo has a higher power level and a longer echo path that is also time-varying. Therefore, removing acoustic echo is more challenging than removing hybrid echo. In fact, hybrid echo may be regarded as a special case of acoustic echo where the echo path is short and time-invariant. This thesis focuses only on the control of acoustic echoes.

Chapter 3

State of the Art in Acoustic Echo Cancellation

Ever since people noticed the presence of appreciable amount of acoustic echo posed annoying experience during communications, experts in the area of communication have been trying to combat the problem as early as 1950s [2]. In this chapter we will see what has been done to alleviate the threat posed by echo in telecommunications.

3.1. Acoustic Echo Suppression (AES)

The development of echo reduction began in the late 1950s with the advent of echo suppression systems. AES is an intuitive and straightforward approach to suppress acoustic echo. It uses various devices to control the instantaneous levels of receive and transmit channels. One of the most commonly used AES technique is voice controlled switching. It changes the insertion loss (i.e., negative gains) between both channels according to the direction of main activity, as illustrated in Figure 3.1. In order to suppress acoustic echo, a large loss is placed in the return path when only the far-end speaker is talking. When the second speaker at the near-end begins talking simultaneously with the first speaker (this situation is called *double-talk*), the inserted loss must be reduced so that the speech of the second speaker is not prevented from reaching the first speaker. This adjustment causes the acoustic echo to be attenuated to some extent [3]. In the use of voice controlled switches with high attenuation, a “comfort noise” is often added to simulate a background noise so the users know that the communication is still on-going.

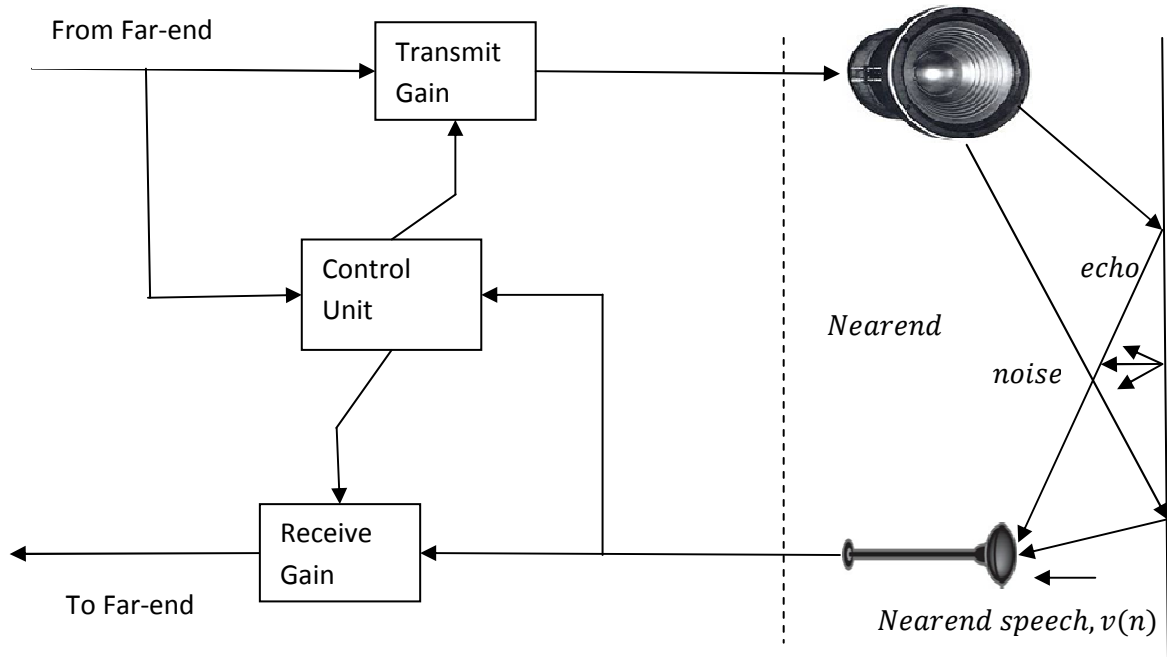


Figure 3.1: Acoustic Echo Suppressor Block Diagram

Echo of speech often has a smaller amplitude level than the speech signal, but otherwise it has mainly the same spectral characteristics and statistics as those of the speech [2]. Therefore, the only basis for discrimination of speech from echo is the signal level. As a result, the speech/echo classifier (control unit) may wrongly classify and let through high-level echoes as speech, or attenuate low-level speech as echo. Although AES is simple in structure and not computationally demanding, it has various drawbacks such as speech clipping, frequent loss of syllables as adjustable gains are turned on, user fatigue resulting from the need to synchronize the conversation. These drawbacks led people to look for other methods and a while later they discovered acoustic echo cancellation methods which is described as follows.

3.2. Acoustic Echo Cancellation (AEC)

AEC is a superior method in which an adaptive filter is used to produce a replica of the acoustic echo signal $y(n)$. This replica is then subtracted from the microphone signal, $d(n)$ prior to its transmission over the communication network so that the acoustic echo is removed. Figure 3.2 below shows the principle of AEC with an adaptive filter.

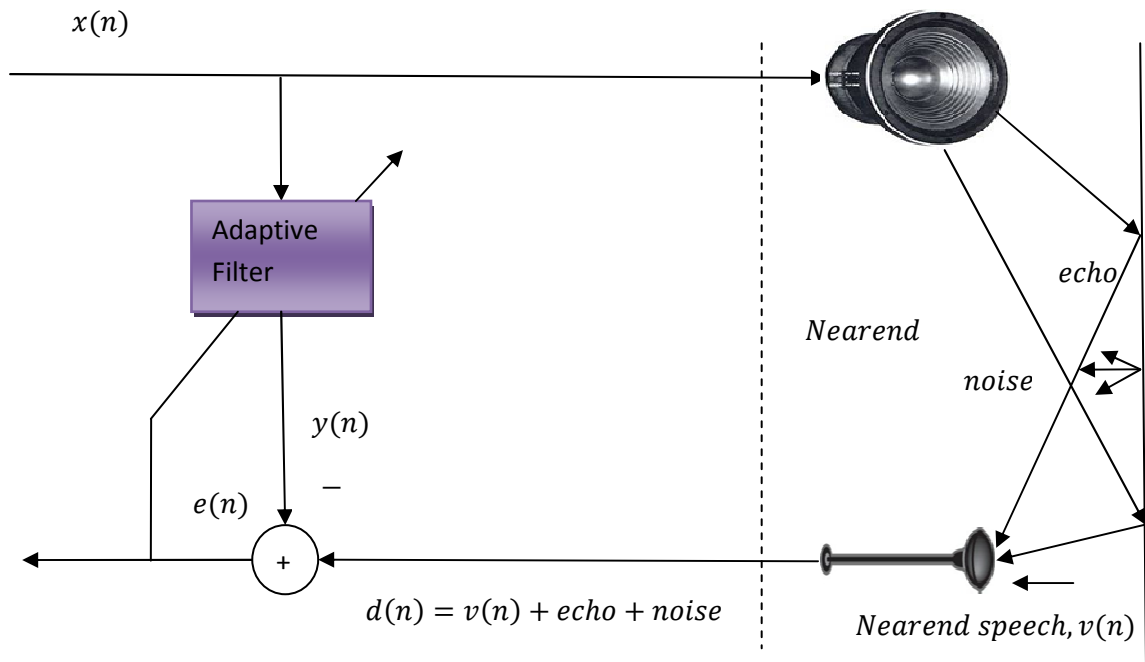


Figure 3.2: Adaptive Acoustic Echo Canceller Block Diagram

Acoustic echo cancellers can be deployed centrally at telecom operators' station for economical reasons or at every hands-free device in order to prevent the acoustic echo produced by that device from transmitting to the other end. AEC is essentially a system identification problem where the Autoregressive Moving Average (ARMA) model can be employed to identify the unknown linear system. In practice, a Finite Impulse Response (FIR) adaptive filter (based on the Moving Average (MA) model) is usually employed in AEC to synthesize the estimated acoustic echo. Infinite Impulse Response (IIR) adaptive filters (based on either the Autoregressive (AR) or the ARMA model) are not as commonly used since they have convergence and stability problems and they do not show obvious advantages over FIR filters [12]. As a result most researchers suggested linear FIR adaptive filtering using common adaptation algorithms such as LMS or RLS as it is necessary to track the time varying nature of the echoes [4, 5]. It wasn't enough only to implement adaptive acoustic echo canceller as there were other main challenges of AEC caused by the following [11]:

- i. The acoustic echo path is long. For instance, it can be several hundred milliseconds for a typical office room.

- ii. Speech is a highly correlated and non-stationary signal.

To deal with the first challenge, an FIR filter with many tap length adaptive filter is used to compensate the long echo path. This results in high computational cost even for the simplest AEC algorithms, such as the normalized least-mean-square (NLMS) algorithm [13]. In order to reduce computational complexity, methods such as subband adaptive filtering [14, 15] and partially updating the adaptive filter coefficients [16, 17] have been developed. To deal with the second challenge, adaptive filters with faster convergence such as the affine projection algorithm [18] and its fast version, fast affine projection algorithms [19] have been developed. Most works on AEC assume that the echo path can be modeled as a slowly time-varying linear system. However, this assumption is no longer valid when nonlinear components along the echo path are taken into account. Major nonlinearities of the echo path may be caused by low-quality loudspeakers and overdriven amplifiers. As a result some researchers developed a nonlinear adaptive filtering approach such as nonlinear Volterra adaptive filter to alleviate the problem [6, 7]. However, in some situations, a nonlinear AEC is inferior to a linear AEC, especially when the channel generates a negligible (or no) nonlinear echo.

The other most important challenge for AEC is double talking situation. If double talk occurs and there is no control mechanism it will make the adaptive filter diverge and audible echoes may pass without being cancelled. A number of researches based on correlation of the available signals [29, 30, 31, and 32] have been carried out in order to tackle this problem. However, since correlation captures only the linear relationship between any two random variables it is no longer valid for nonlinear AEC.

3.3. Research Question

We have seen from the foregoing discussion that a number of researches have been conducted in order to eliminate acoustic echo. Most of the acoustic echo cancellers are made on the assumptions of the linearity of the echo. However, this assumption is no longer valid when the nonlinearity along the echo path is taken into account. This led to the invention of Nonlinear Acoustic Echo Cancellation (NLAEC) techniques such as Volterra filters. However, in some situations, NLAEC is inferior to a linear Acoustic Echo Cancellation (AEC), especially when the channel generates a negligible (or no) nonlinear echo. In general, the ratio of the linear to

nonlinear echo signal power is unknown a priori, and will vary over time, thus making it difficult to know if NLAEC would improve or degrade the cancellation.

Therefore, it is necessary to find a method which can integrate both linear and nonlinear adaptive echo cancellation schemes and favor the appropriate one depending on the nature of the echo path i.e., linear or nonlinear.

We have also seen that there is a need to implement a good DTD which is robust for both linear and nonlinear acoustic echo.

3.4. Objective of the Thesis

3.4.1. General Objective

In the problem statement we have seen that there is a need to find a scheme which can integrate both linear and nonlinear filters and favor one of the two depending on the echo path. Therefore, convex combination of adaptive filters which offers a better tradeoff when compromising nature of filters is involved is studied in this thesis. Together with the convex combination scheme, the problem with DT situation is also solved by implementing Mutual Information based DTD.

3.4.2. Specific Objectives

Therefore, the specific objectives of this thesis are:-

- To design adaptive linear FIR filter for linear AEC
- To design adaptive nonlinear Volterra filter for nonlinear AEC
- To design adaptive convex combination of linear FIR and nonlinear Volterra filters
- To design MI based DTD to combat the DT situation

Chapter 4

Finite Impulse Response (FIR) Linear Adaptive Filters

4.1 Introduction

As explained before, acoustic echo originates from the coupling between the loudspeaker and the microphone of the hands-free device. Random sound fields composed of direct and reflection waves provide a fundamental model for a finite impulse response (FIR) system [41]. The impulse response of this linear system is the superposition of several delayed and attenuated pulses. In practice, due to continual changes in the acoustic environment (e.g., persons/objects moving), the impulse response is a complicated function of time and the corresponding linear system is time-variant. In this section we will see how adaptive linear FIR filters are used to eliminate the linear echo.

Among digital signal processing operations, linear filtering is probably the most common and important one. It is called adaptive if its parameters, the coefficients, are varied according to a specified criterion as new information becomes available. The updating has to follow the evolution of the system environment as fast and accurately as possible, and, in general, it is associated with real-time operation. It's is this behavior of FIR adaptive filters that makes them most appealing to AEC.

4.2. Linear Adaptive Filters

A linear adaptive transversal filter is shown in Figure 4.1, where z^{-1} represents a unit delay. It is based on an FIR filter structure which is inherently stable, as opposed to IIR filters, and therefore is commonly used in practice. In order to discuss linear adaptive filtering algorithms suitable for AEC, it is necessary to first define the far-end speech signal vector \mathbf{x}_n and the adaptive filter coefficient vector \mathbf{w}_n :

$$\mathbf{x}_n = [x(n), x(n - 1), \dots, x(n - N + 1)]^T \dots\dots\dots 4.1$$

$$\mathbf{w}_n = [w_0(n), w_1(n), \dots, w_{N-1}(n)]^T \dots\dots\dots 4.2$$

where N is the length of the adaptive filter, n is the index of the discrete time and T denotes the transposition operator.

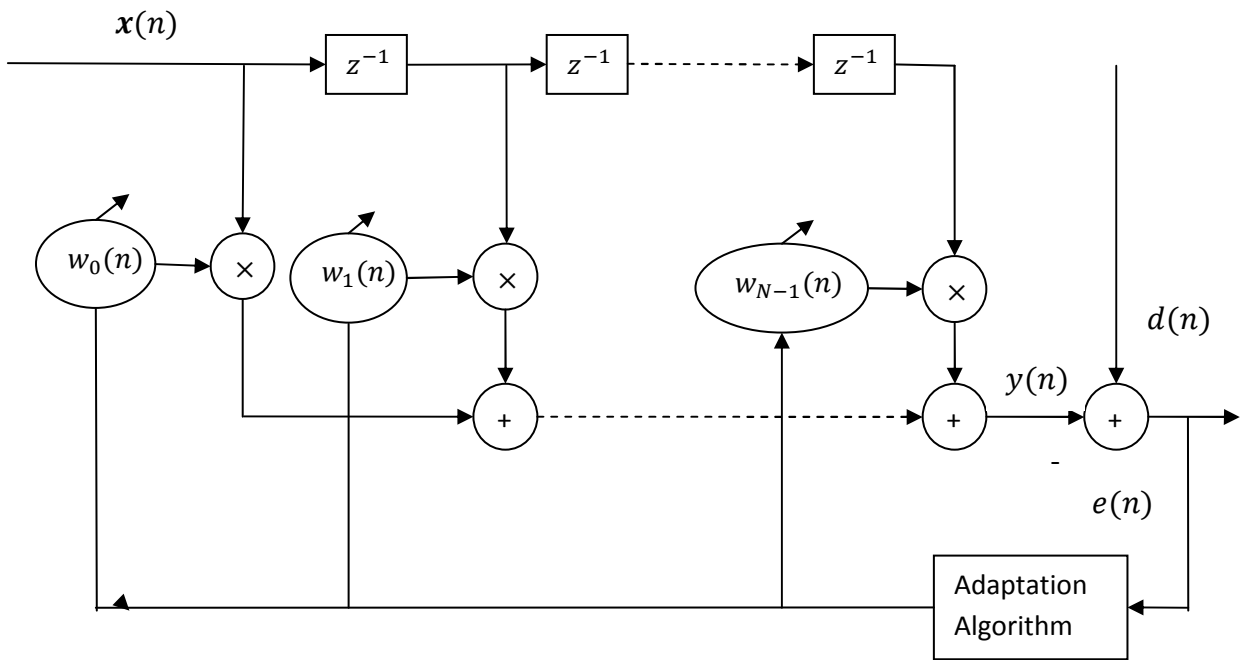


Figure 4.1: Structure of FIR Linear Filter

Where:-

- $x(n)$ represents the input signal sequence to the filter and in our case far-end speech
- $d(n)$ is the desired signal or the reference signal to the adaptive filter
- $y(n)$ is the filter output to match the desired response
- $e(n)$ is the error signal or the difference between the desired signal and filter output

In Figure 4.1 the FIR linear adaptive filter works in such a way as to minimize the mean square error rather than the error itself, using stochastic gradient decent algorithms. In the following section one such algorithm- the NLMS algorithm is discussed.

4.2.1. The normalized least-mean-square (NLMS) algorithm

NLMS is one of the most popular algorithms for AEC due to its simplicity of implementation, low computational complexity, and robust behavior. The computational complexity of NLMS is $O(2N)$ operations per sample (OPS), where one operation is defined as one real multiplication plus one real addition [43].

The normalized least-mean-square (NLMS) algorithm [43] which has the following form:

NLMS Algorithm
<p>Step 1: Initialization</p> $\mathbf{w} = 0$ <p>Step 2: Recursion</p> <p><i>for</i> $n = 1$ to N <i>do</i></p> $e(n) = d(n) - \mathbf{w}_n^T \mathbf{x}_n$ $\mathbf{w}_{n+1} = \mathbf{w}_n + \frac{2\mu \mathbf{x}_n e(n)}{(\mathbf{x}_n^T * \mathbf{x}_n + \delta)}$ <p><i>end</i></p>

Table 1: NLMS algorithm

The step-size μ controls the convergence behavior of the algorithm: the larger the value of μ , the faster the algorithm converges, but this would also cause a greater misadjustment (i.e., larger residual error signal $e(n)$) in steady-state. For the algorithm to be stable, μ must be chosen from $0 < \mu < 2$. The small positive constant δ is introduced in order to prevent the denominator from being too small when the power of the input signal $x(n)$ is very low.

4.3. Performance Measures

In order to define performance measure for the AEC, it's necessary first to understand what is required of the AEC.

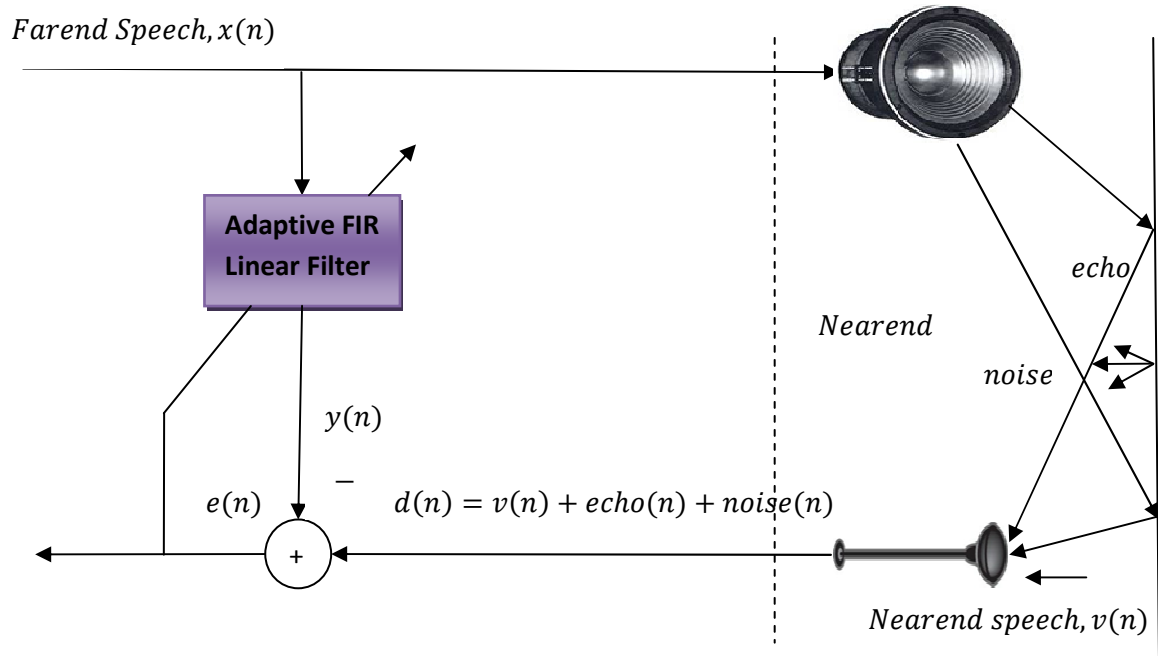


Figure 4.2: Basic setup of Linear AEC Using Adaptive FIR Linear Filter

In Figure 4.2 above we have:-

- Far-end Speech, $x(n)$
- Near-end Speech, $v(n)$
- Background noise, $noise(n)$
- Desired Signal, $d(n) = v(n) + echo(n) + noise(n)$
- echo, the room impulse response
- Adaptive FIR Filter Response, $y(n)$
- Residual Error, $e(n) = d(n) - y(n)$

Now having observed the available signals, it's desirable that a good AEC has to produce $y(n)$ very close to $d(n)$ or minimize $e(n)$ such that the echo is cancelled before it reaches the far-end speaker. According to the above idea the following performance measures are defined:

4.3.1. Echo Return Loss Enhancement (ERLE)

In order to evaluate the performance of an echo canceling system, the ratio of the expected value of the microphone output squared, $E\{d^2(n)\}$, divided by the expected value of the error signal squared, $E\{e^2(n)\}$, is monitored. This quantity, in dB is called the Echo Return Loss Enhancement or ERLE [44]:

$$ERLE(n) \stackrel{\text{def}}{=} 10 \log_{10} \frac{E\{d^2(n)\}}{E\{e^2(n)\}} \dots \dots \dots 4.3$$

Where $d(n)$ and $e(n)$ are as defined above.

This quantity indicates the amount of echo being cancelled after every sample and the greater its value the better our system performs. According to the ITU-T recommendation G167 for acoustic echo controllers [23], ERLE from 30dB to 40dB at steady state and noise free measurement scenario is good.

4.3.2. Mean Square Error (MSE)

The MSE is a measure of how the algorithm converges to the true value in a mean square sense and is given as follows:

$$MSE(n) = E\{e(n)^2\} \dots \dots \dots 4.4$$

This measurement helps us to see if our system model is indeed minimizing the error and it is sometimes called the learning curve of the algorithm.

4.3.3. Algorithm Misalignment Error (AME)

The coefficient based measure, referred to as algorithm misalignment error (AME) or system mismatch (SM) [44], is defined by the ratio of the Euclidean norm of the coefficient error vector and the Euclidean norm of the coefficient vector \mathbf{h} of the system to be identified. It measures the adaptive filter's coefficients ability to converge to the true value at steady state.

- No background noise is assumed as there is no standard set yet for achievable ERLE according to [23].

4.4.2. Results

In this section we will discuss simulation results for the foregoing simulation setup environment.

I. 'Melkamledet' music input as far-end speech, $x(n)$

i. Algorithm Misalignment Error (AME)

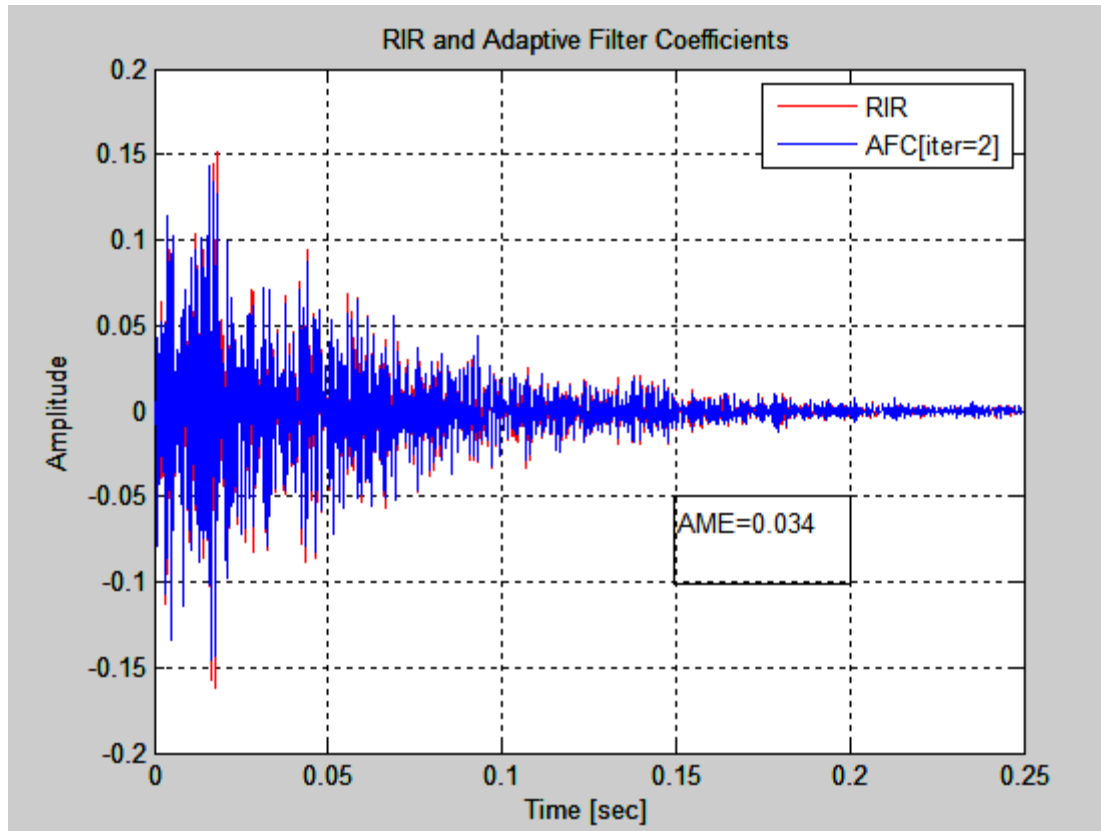


Figure 4.3: Plot of Room Impulse Response (RIR) and Linear FIR Adaptive Filter Coefficients (AFC) for number of iterations =2.

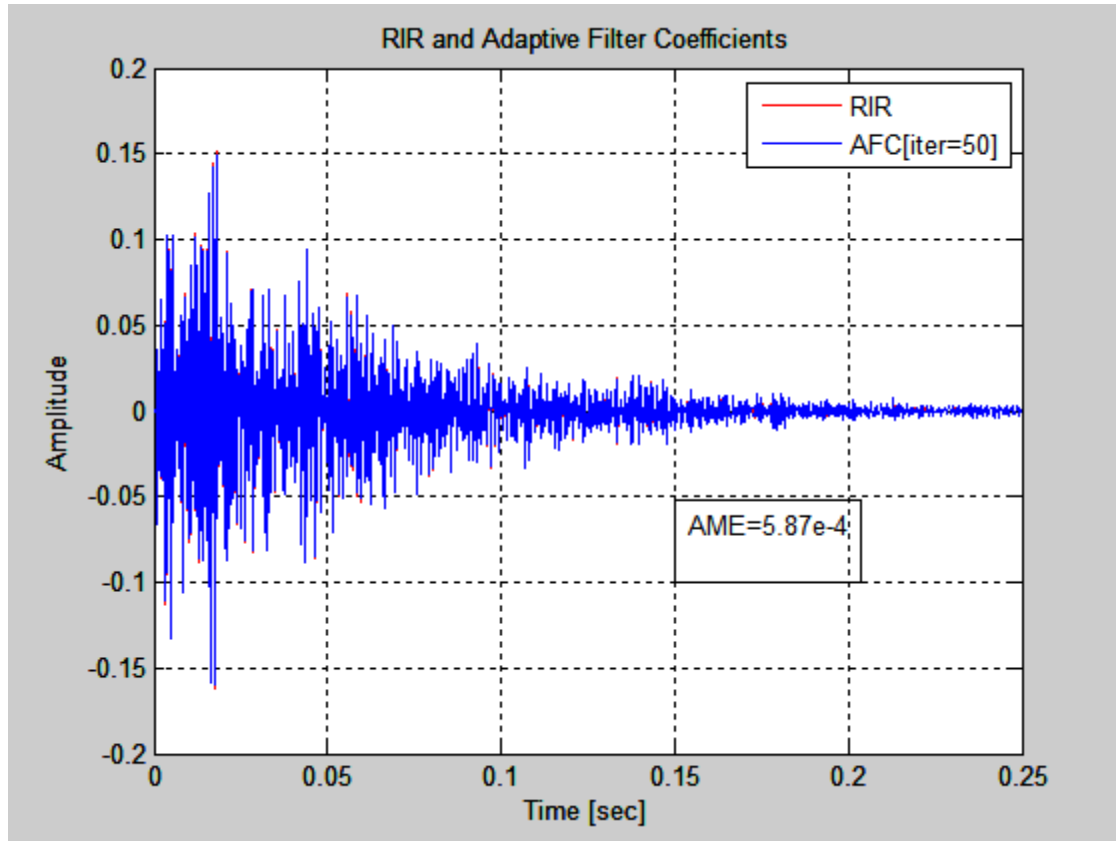


Figure 4.4: Plot of Room Impulse Response (RIR) and Linear FIR Adaptive Filter Coefficients (AFC) for number of iterations =50.

In Figures 4.3 and 4.4 above we see that the AME which is given by equation (4.5) decreases from 0.034 to $5.87e^{-4}$. It shows a great improvement in the adaptive FIR algorithm's capability to match the true room impulse response as the number of iterations is increased. However, we cannot increase the number of iterations indefinitely since it could cost us in terms of memory and computational time (two OPS is required for every iteration for NLMS algorithm). In the following results we will see how much iteration would be enough in terms of the improvement achieved by increasing the number of iterations and the associated computational cost.

ii. Mean Square Error (MSE)

In this experiment six different iterations (iterations =2, 10, 30, 50, 60 and 100) are considered and the results are plotted for MSE (Figure 4.5) and ERLE (Figure 4.6) as a function of number of iterations.

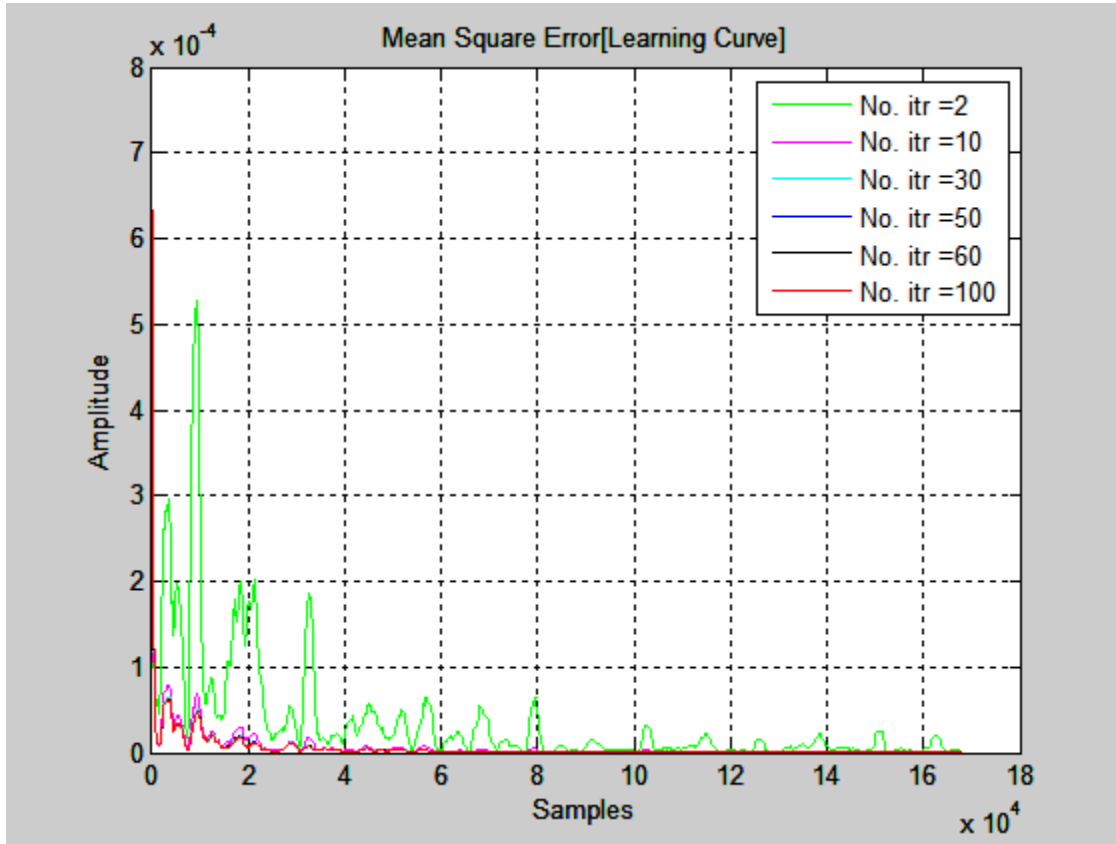


Figure 4.5: Plot of Mean Square Error [Learning curve] for different number of iterations.

In Figure 4.5 above we see that, as expected, we have for all iterations the mean square error decreasing as more and more samples are introduced into the filter. But, as can be seen from Table 2 below for number of iterations 50 and above we do not gain much improvement while the computational cost keeps on increasing as the number of iterations is increased. For NLMS algorithm we have seen that there are 20PS. In Figure 4.6 we see that as the number of iterations is increased from 2 to 100. Table 2 below indicates the achievable improvement in terms of MSE and ERLE as the number of iterations is increased.

iii. Echo Return Loss Enhancement

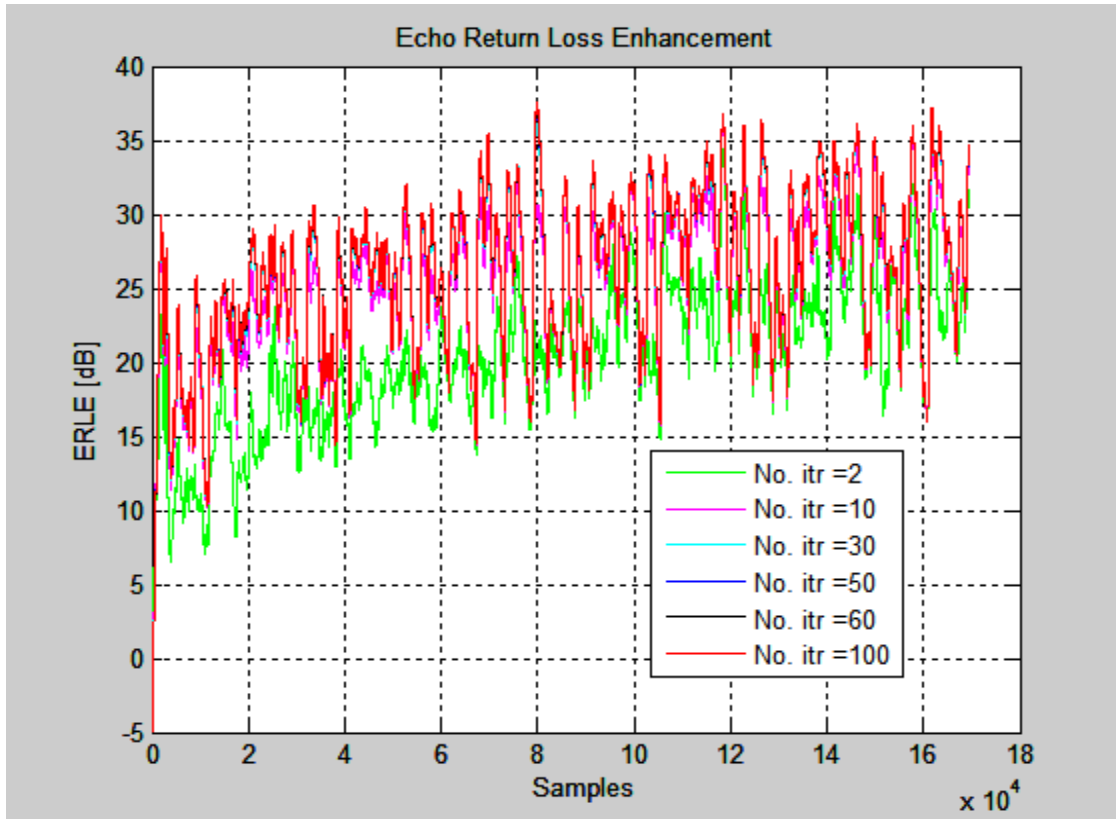


Figure 4.6: Plot of Echo Return Loss Enhancement for different number of iterations

No.	No. Iterations	Average MSE	Difference	Average ERLE	Difference
1	2	3.56e-005 (-44.49dB)	-7.62dB	20.34dB	4.81dB
2	10	6.15e-006 (-52.11dB)	-1.29dB	25.15dB	0.96dB
3	30	4.56e-006 (-53.40dB)	-0.14dB	26.11dB	0.10dB
4	50	4.43e-006 (-53.54dB)	-0.02dB	26.21dB	0.02dB
5	60	4.40e-006 (-53.56dB)	-0.04dB	26.23dB	0.03dB
6	100	4.36e-006 (-53.60dB)		26.26dB	

Table 2: MSE and ERLE as a function of number of iterations

In Table 2 above we see that when number of iterations is increased from 2 to 50, we have MSE reduction of -7.62dB to -0.14dB and ERLE improvement of 4.81dB to 0.10dB; but when the iteration is increased from 50 to 100, MSE reduction of only -0.02dB to -0.04dB and ERLE improvement of only 0.02dB to 0.03dB. From the above values what we see is that the number of iterations equal to 50 is worth the improvement in MSE reduction and ERLE compared to the computational cost it brings, whereas iterations above 50 do not provide enough improvements compared to the computational cost they bring. We know from previous discussions that NLMS has computational cost of 20PS. Therefore in this thesis number of iterations considered is 50 for all simulations.

II. 'farspeech' speech signal input as far-end speech, $x(n)$

i. Algorithm Misalignment Error (AME)

Here, as discussed above, we will only plot the simulation results for the number of iterations equal to 50.

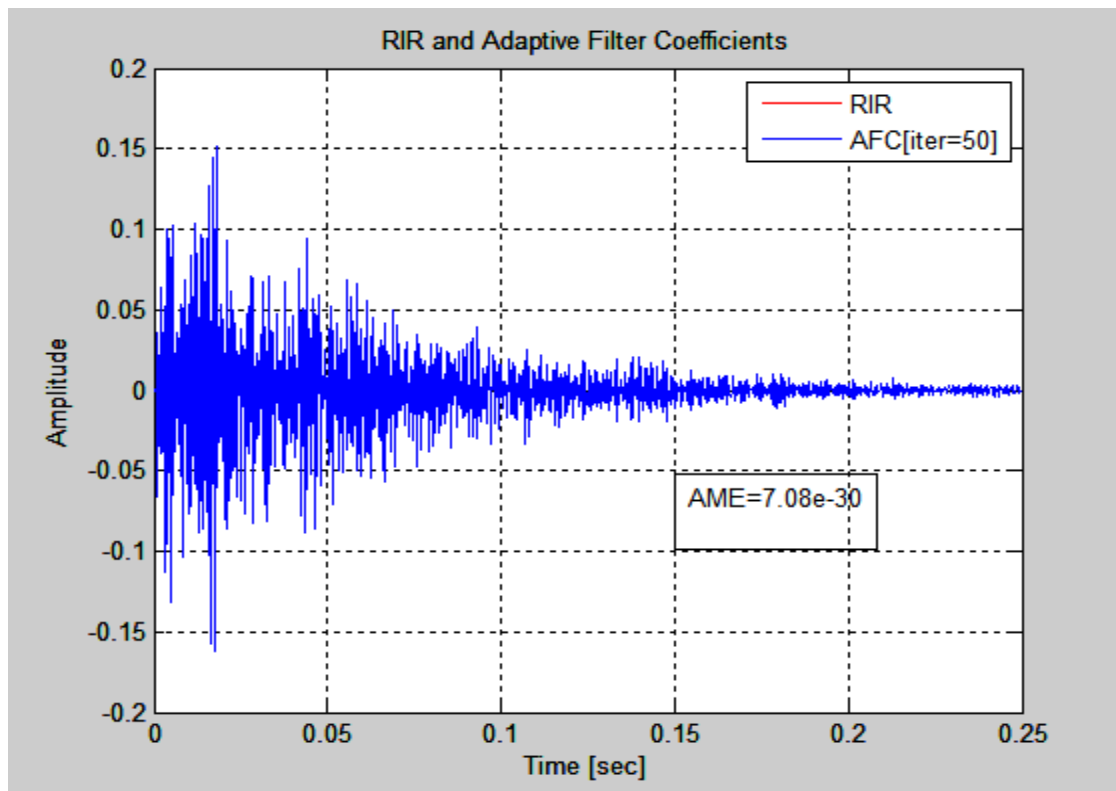


Figure 4.7: Plot of Room Impulse Response (RIR) and Linear FIR Adaptive Filter Coefficients (AFC) for number of iterations =50

As it has been already mentioned in the Experimental setup (Section 4.4.1) the results from Figure 4.7 to Figure 4.9 are plotted here for sake of proving our model can be implemented to different kinds of input speeches. In Figure 4.7 we see that the matching between the true RIR and our AFC is almost perfect (i.e., AME is only $7.08e^{-30}$) that the RIR (red) and AFC (blue) are barely separable.

ii. **Mean Square Error (MSE)**

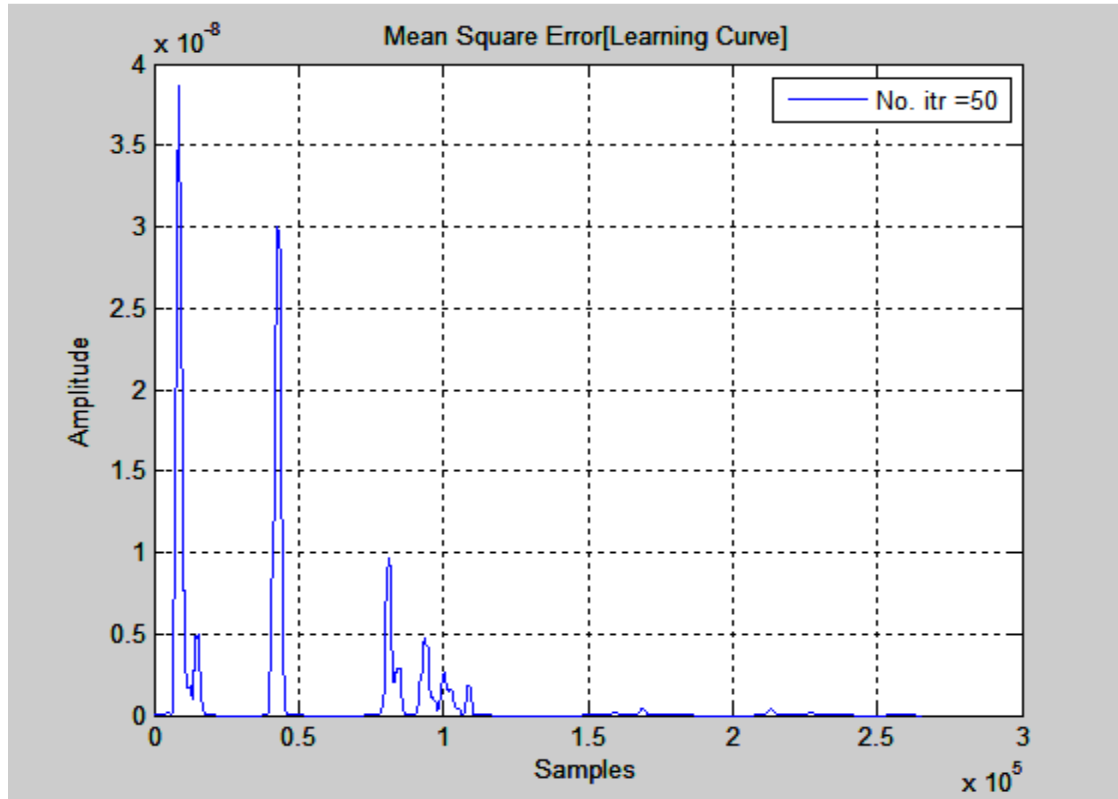


Figure 4.8: Plot of Mean Square Error [Learning curve] for number of iterations =50.

Here, in Figure 4.8, we see that the MSE decreases as more and more samples are introduced into the system. This is similar to the result we obtained for the first kind of input speech (Figure 4.5) which indicates our system is not sensitive to the nature of the input.

In Figure 4.9 also we see the same thing in terms of ERLE as in Figure 4.6 except we have breaks here which are speech silent periods.

iii. Echo Return Loss Enhancement

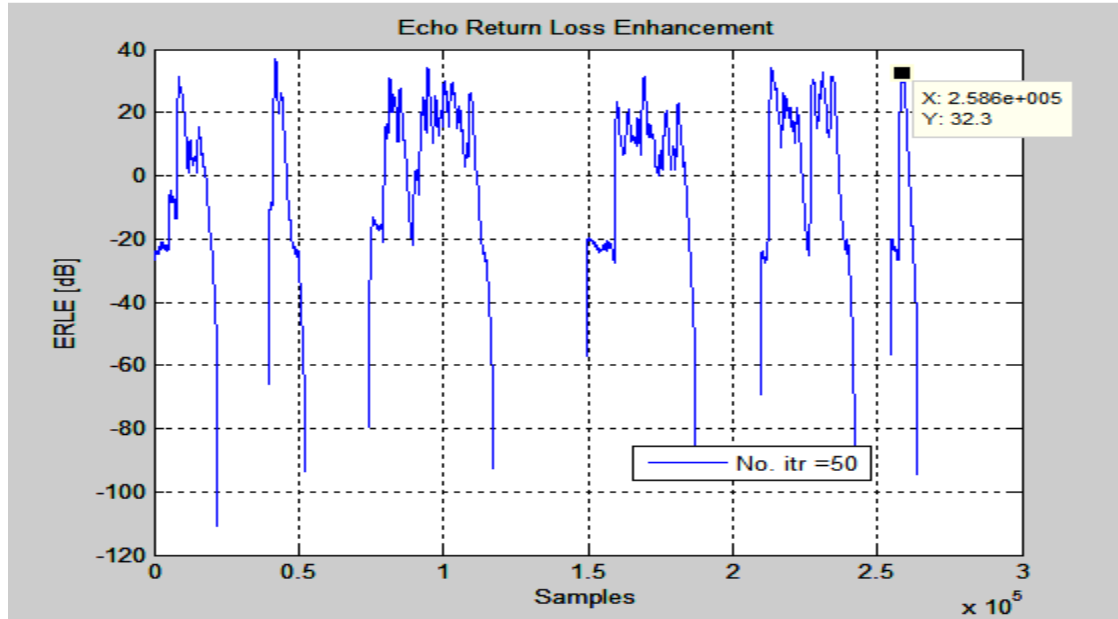


Figure 4.9: Plot of Echo Return Loss Enhancement for number of iterations =50.

4.4.3. Discussion

In Figure 4.3 and Figure 4.4 we see that when the number of iterations increases from 2 to 50 the AME decreases from 0.034 to $5.87e^{-4}$ which confirms that generally, by increasing the number of iterations we can improve the performance of our adaptive FIR linear filter. In Figure 4.5 we see that the MSE also decreases when the iteration is increased and finally in Figure 4.6 we see that the ERLE increases as the number of iterations is increased. The same thing holds for the second type of input that is the real speech signal which indicates our echo canceller is robust for different kind of input signals. In conclusion, even if we can achieve a better performance by increasing the number of iterations, after certain number of iterations the achievable improvement is not worth the additional computational cost incurred by increasing the number of iterations. In this simulation 50 is found to be a good value for number of iterations as discussed above. Most importantly, in both of the above cases (in Figure 4.6 and Figure 4.9) that is for different types of filter input, our AEC performs well to the standard which as discussed earlier states the achievable ERLE has to be between 30dB to 40dB according to [23]. In the next Chapter we will discuss the NLAEC using nonlinear Volterra filter.

Chapter 5

Adaptive Nonlinear Volterra Filters for Nonlinear AEC

5.1. Introduction

While linear filters have been used widely in engineering, many problems are inherently nonlinear and may better be addressed by nonlinear solutions. Acoustic Echo cancellation is also one of those problems when the nonlinearity is considered in the echo path. In this section we will first discuss the possible nonlinearities in the echo path and then the adaptive Volterra filter for cancelling nonlinear acoustic echo.

5.2. Sources of Nonlinearities in Acoustic Echo

For the design of nonlinear echo cancellers, it is essential to have sufficient knowledge about the properties of the underlying physical echo path. It turns out that the loudspeaker and its amplifier constitute the main sources for nonlinear distortion [22].

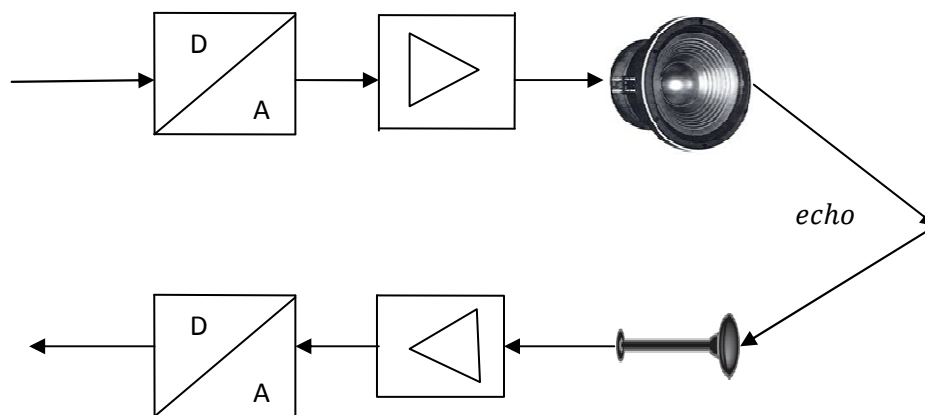


Figure 5.1: Sources of nonlinearity in the acoustic echo path

series one can reduce the computational complexity by sacrificing the accuracy of the series expansion. With reduced order, the Volterra series representation is quite complex even when the orders of the series and the filter are moderate. Therefore, in this thesis only second order Volterra filter is considered.

5.4. Adaptive Second Order Volterra Filter

As it has already been discussed what made Volterra filters most appealing in nonlinear filtering is the ease to extend the adaptive linear filtering theory in a straight forward manner to the Volterra model. Now, before we see how the NLMS algorithm is extended to Volterra filter let's first see how the adaptation is achieved in second order Volterra filter.

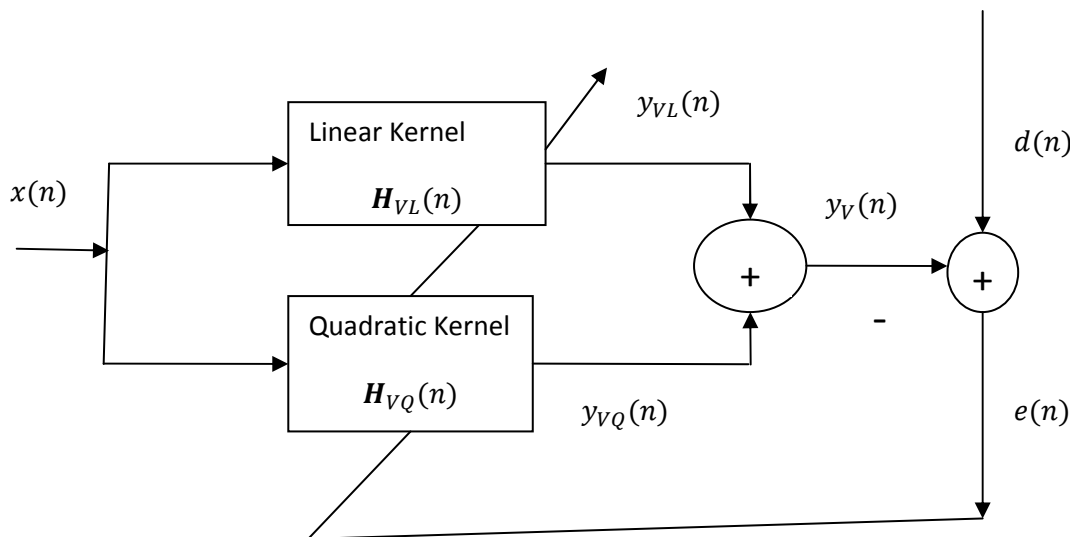


Figure 5.2: Block diagram of adaptive second order Volterra filter

where \mathbf{H}_{VL} is linear Volterra kernel

\mathbf{H}_{VQ} is quadratic kernel

y_{VL} and y_{VQ} are the linear and quadratic kernel outputs respectively

In Figure 5.2 above, the second order Volterra filter has both the linear kernel, \mathbf{H}_{VL} and nonlinear kernel, \mathbf{H}_{VQ} which constitute the second order Volterra filter. The desired signal $d(n)$, the acoustic echo which has both linear and nonlinear components in our case,

constitutes the reference signal for the Volterra filter and the Volterra filter replicates the echo (that is in the form of $y_v(n)$) by the help of the error signal, $e(n)$ which is the difference between the reference signal, $d(n)$ and the Volterra filter output, $y_v(n)$. This way the adaptation continues until the error reaches to a required level.

5.5. Implementation of Second Order Volterra Filter

In this Section, the Volterra NLMS algorithm is presented for a second-order series and Nth-order filter. This choice reduces the computational complexity to an acceptable level for some applications such as AEC and also simplifies the derivations.

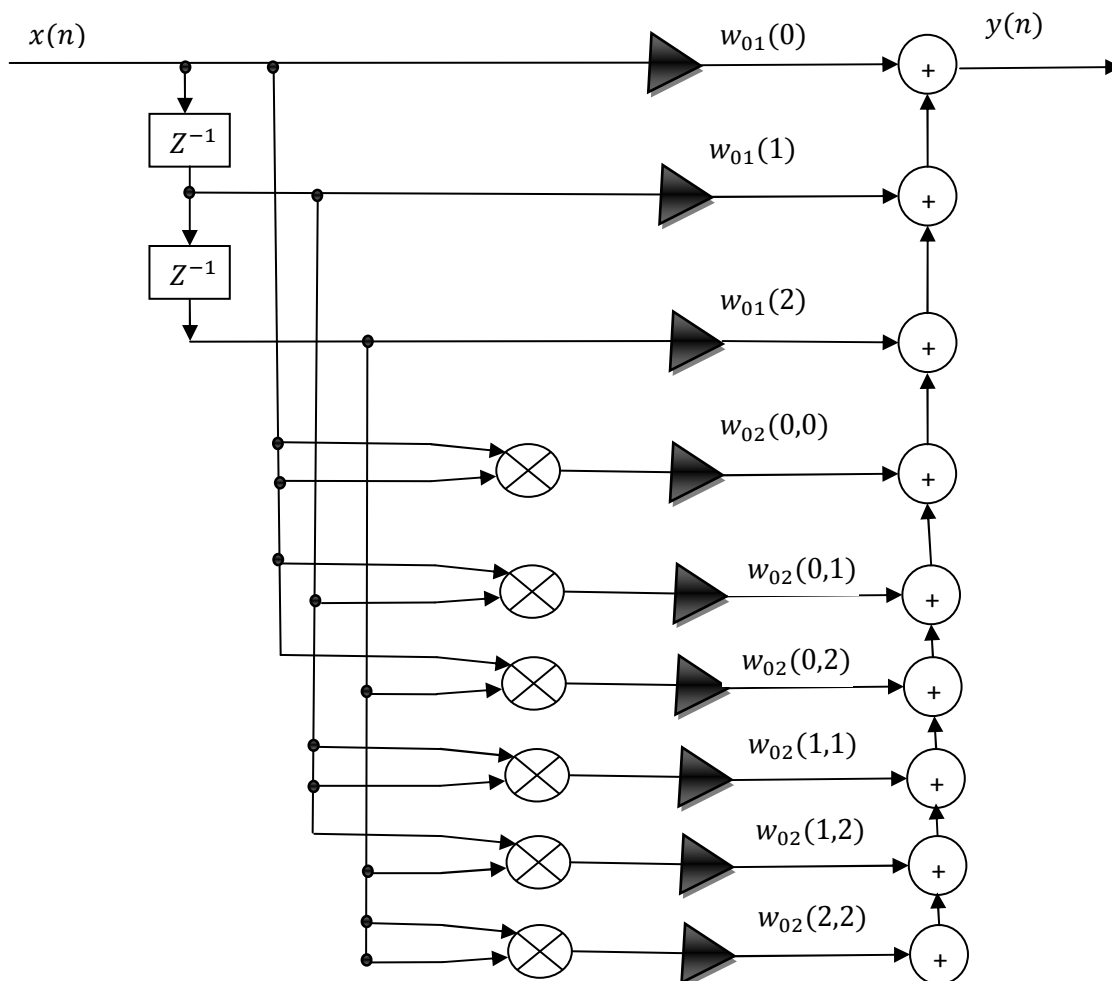


Figure 5.3: Architecture of Second order Volterra Filter

Now, from Figure 5.3 above the adaptive filter that estimates the signal $y(k)$ using a truncated Volterra series expansion of second order, can be described by

$$y(k) = \sum_{l_1=0}^N w_{01}(l_1)x(k - l_1) + \sum_{l_1=0}^N \sum_{l_2=0}^N w_{02}(l_1, l_2)x(k - l_1)x(k - l_2) \dots \dots \dots 5.2$$

where $w_{0i}(l_1, l_2)$, for $i = 0, 1, \dots, N$, are the coefficients of the nonlinear filter model based on the second-order Volterra series expansion, and $y(k)$ represents the adaptive filter output signal. Now, if $X(k)$ is interpreted in the following way

$$X(k) = \begin{bmatrix} x(k) \\ x(k - 1) \\ \vdots \\ \vdots \\ x(k - N) \\ x^2(k) \\ x(k)x(k - 1) \\ \vdots \\ \vdots \\ x(k)x(k - N) \\ \vdots \\ \vdots \\ x(k - N)x(k - N + 1) \\ x^2(k - N) \end{bmatrix} \dots \dots \dots 5.3$$

And the filter coefficients are interpreted in the following way

$$\mathbf{w}(k) = \begin{bmatrix} w_0(k) \\ w_1(k) \\ \vdots \\ \vdots \\ w_N(k) \\ w_{0,0}(k) \\ w_{0,1}(k) \\ \vdots \\ \vdots \\ w_{0,N}(k) \\ \vdots \\ \vdots \\ w_{N,N-1}(k) \\ w_{N,N}(k) \end{bmatrix} \dots\dots\dots 5.4$$

We have the filter output, $y(k)$ as follows

$$y(k) = \mathbf{w}^T(k)\mathbf{X}(k) \dots\dots\dots 5.5$$

The NLMS algorithm can also be implemented as follows

Second Order Volterra NLMS Algorithm	
Initialization	$\mathbf{w}(0) = [0 \ 0 \ \dots \ 0]^T$
Do for $k \geq 0$	$e(k) = d(k) - \mathbf{X}^T(k)\mathbf{w}(k)$
$\mathbf{w}(k + 1) = \mathbf{w}(k) + 2$	$\begin{bmatrix} \mu_L & \dots & 0 & 0 & \dots & 0 \\ 0 & \ddots & 0 & 0 & \ddots & 0 \\ 0 & \dots & \mu_L & 0 & \dots & 0 \\ 0 & \dots & 0 & \mu_{VL} & \dots & 0 \\ \vdots & \ddots & 0 & 0 & \ddots & 0 \\ 0 & \dots & 0 & 0 & \dots & \mu_{VL} \end{bmatrix} \frac{e(k)\mathbf{X}(k)}{\ \mathbf{X}^T(k)\mathbf{X}(k)\ }$

Table 3: NLMS algorithm for second order Volterra filter

In the above algorithm we see that the linear kernel and the quadratic kernel use different adaptation step sizes. This is important because the same step size for the two kernels may not guarantee convergence for both of them at the same time.

5.6. Simulation Setup

To verify the performance of the second order Volterra filter for nonlinear acoustic echo cancellation simulation is carried out using NLMS algorithm with the following settings:

- Here only one kind of far-end speech input is considered because we have seen the algorithm can be implemented successfully for any type of input signal.
 - ✓ ‘Melkamledet’ music is used as far-end speech input, $x(n)$
- The room impulse response is generated using the common image method [20] with delay of 250ms as before.
- The quadratic kernel used is a 4X4 matrix as indicated in [21].
- The filter length for linear kernel is set to 2000 as before.
- The step size (μ) for the linear kernel is chosen as 0.45 and for the quadratic kernel is 0.01 to obtain a better tradeoff between tracking speed and convergence.
- Number of iterations considered is 50 which is found to be a good tradeoff between computational cost and accuracy.
- Here also only single talk situation is considered that is near-end speech is assumed to be zero (the near-end speaker is silent), $v = 0$.

5.7. Results

In this section we will discuss simulation results for the foregoing simulation setup environment as follows

I. Algorithm Misalignment Error (AME)

Here, unlike the linear adaptive FIR filters, the second order Volterra filter has two kernels: - the linear and nonlinear (the quadratic kernel), therefore we will see the AME for both kernels.

A. Volterra linear kernel

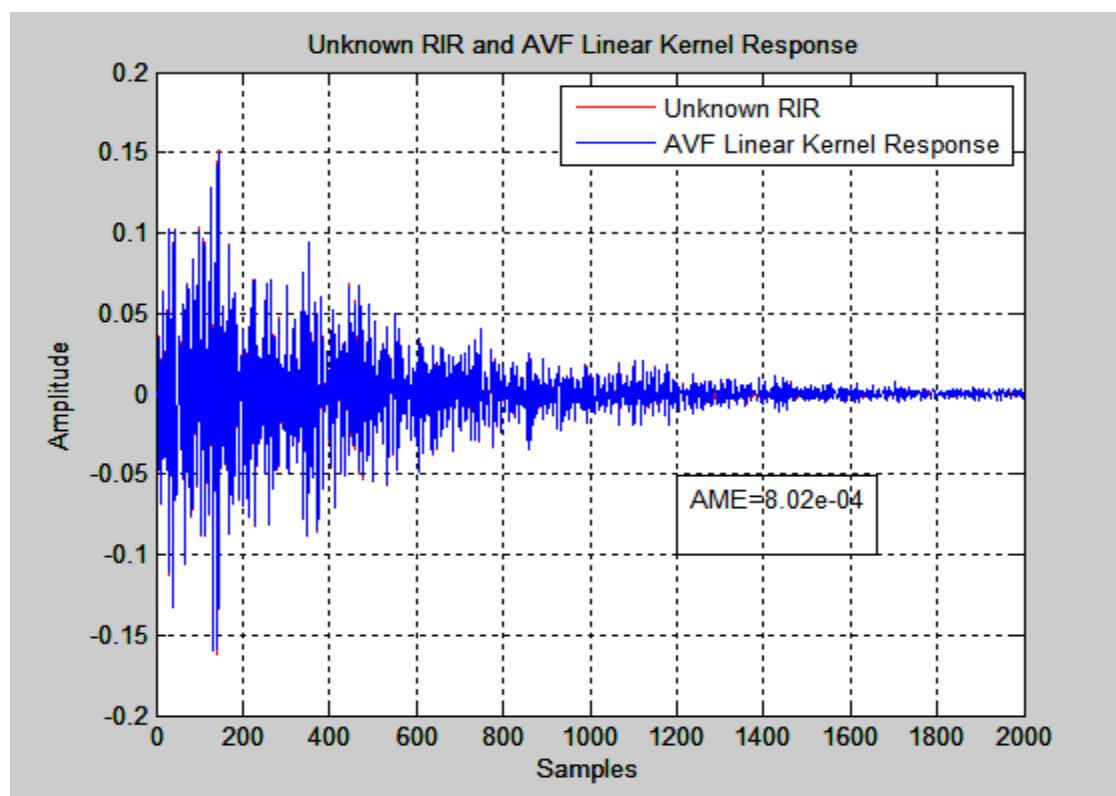


Figure 5.4: Plot of Room Impulse Response (RIR) and Linear Volterra kernel response

In Figure 5.4 above we see that the Room Impulse Response (RIR), in red color and the response of the linear kernel of the adaptive second order Volterra filter in blue color match with the AME of only $8.02e-04$ which together with the quadratic kernel results in a good (to the standard) ERLE as indicated in Figure 5.8.

B. Volterra quadratic kernel

As explained in the simulation setup, Figure 5.5 below indicates the measured loudspeaker nonlinear response which is a 4×4 matrix.

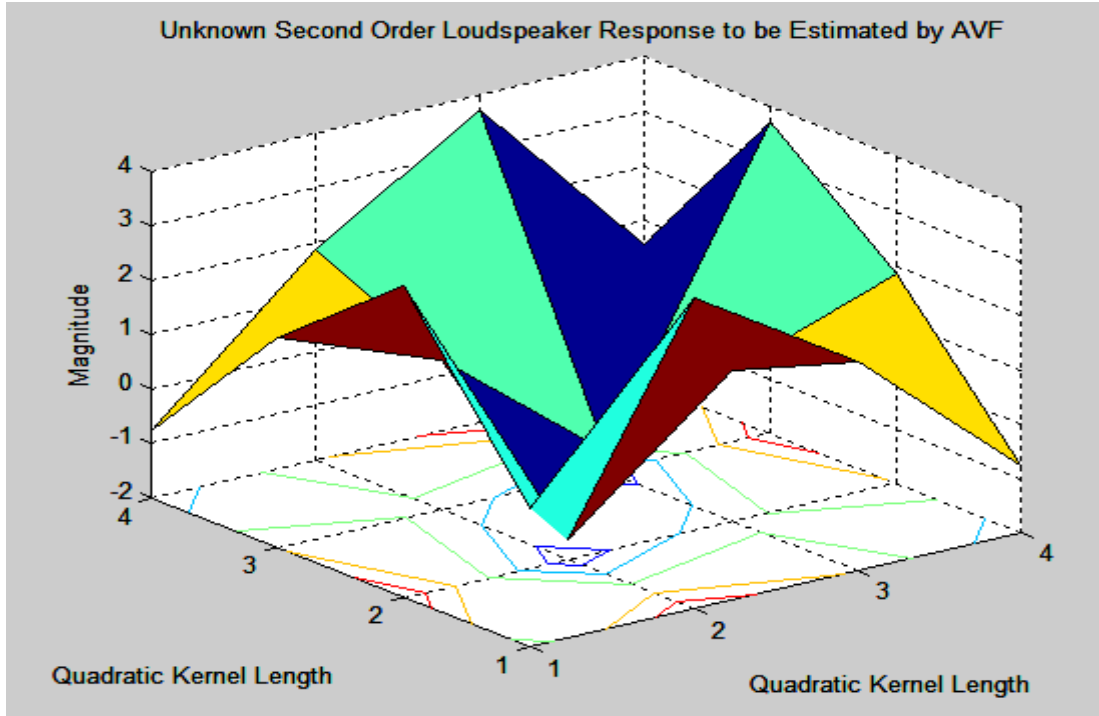


Figure 5.5: Plot of the nonlinear component of the acoustic echo to be identified by Volterra quadratic kernel.

We know that in the Volterra second order filter the second order term is responsible for eliminating the nonlinear terms in our echo. Figure 5.6 below indicates the response of the quadratic kernel of the nonlinear second order Volterra filter which cancels out the nonlinear measured loudspeaker response of Figure 5.5. As indicated in the Figure 5.6 the AME is only $3.49e-05$ which is a good result as it leads to a good (to the standard) ERLE which is indicated in Figure 5.8.

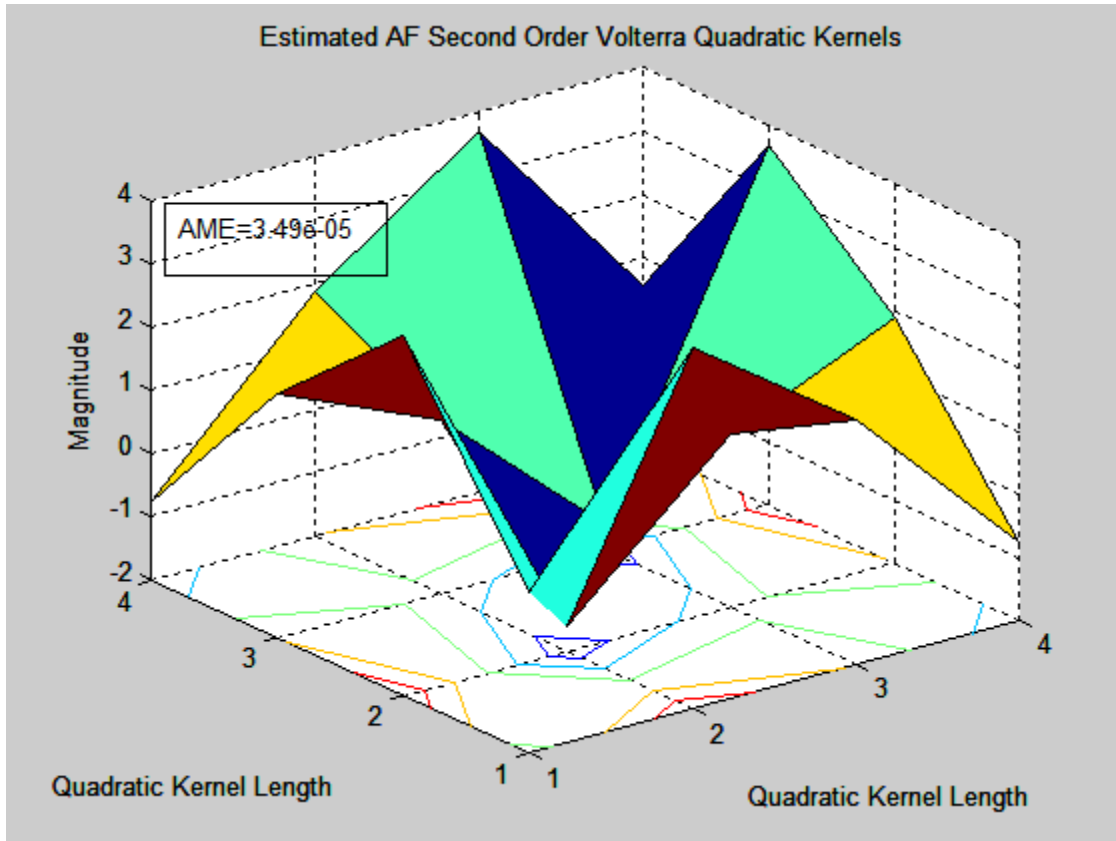


Figure 5.6: Plot of the Volterra quadratic kernel response after 50 iterations.

II. Mean Square Error (MSE)

In Figure 5.7 below the MSE or the learning curve for the algorithm indicates that the MSE decreases as more and more samples are being introduced. Initially, we have large error (echo) but it decreases as more samples are introduced or as the speech progresses in our case.

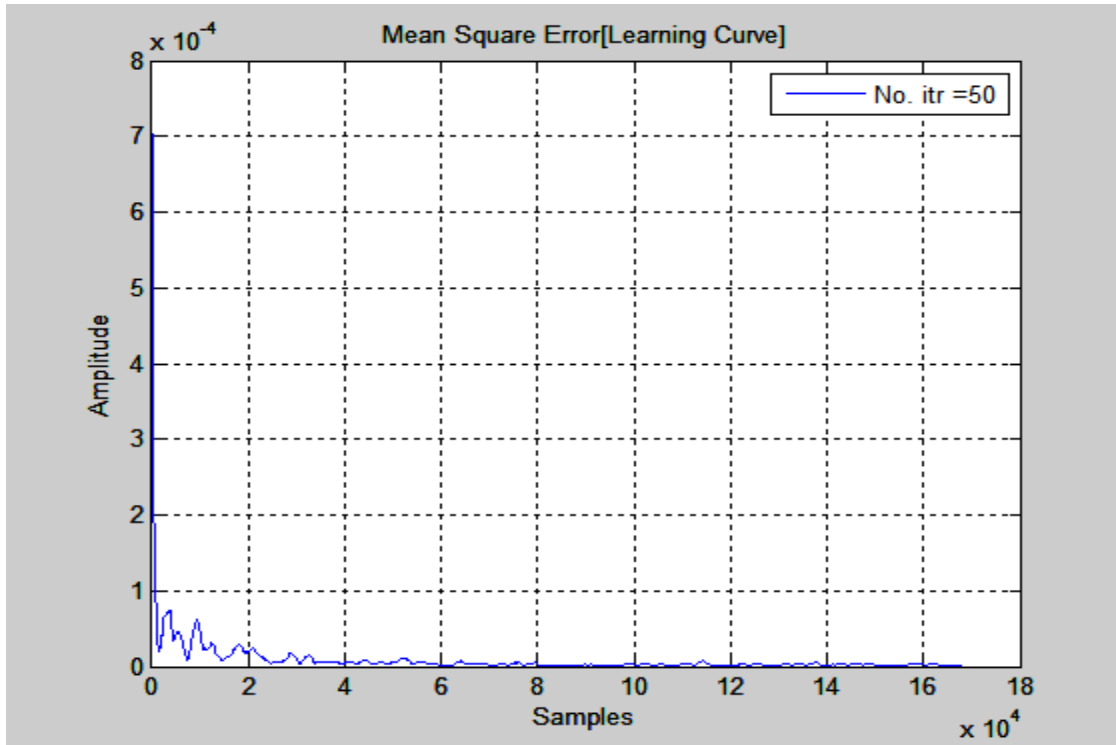


Figure 5.7: Plot of the MSE [Learning Curve] of Adaptive Second Order Volterra Filter

III. Echo Return Loss Enhancement(ERLE)

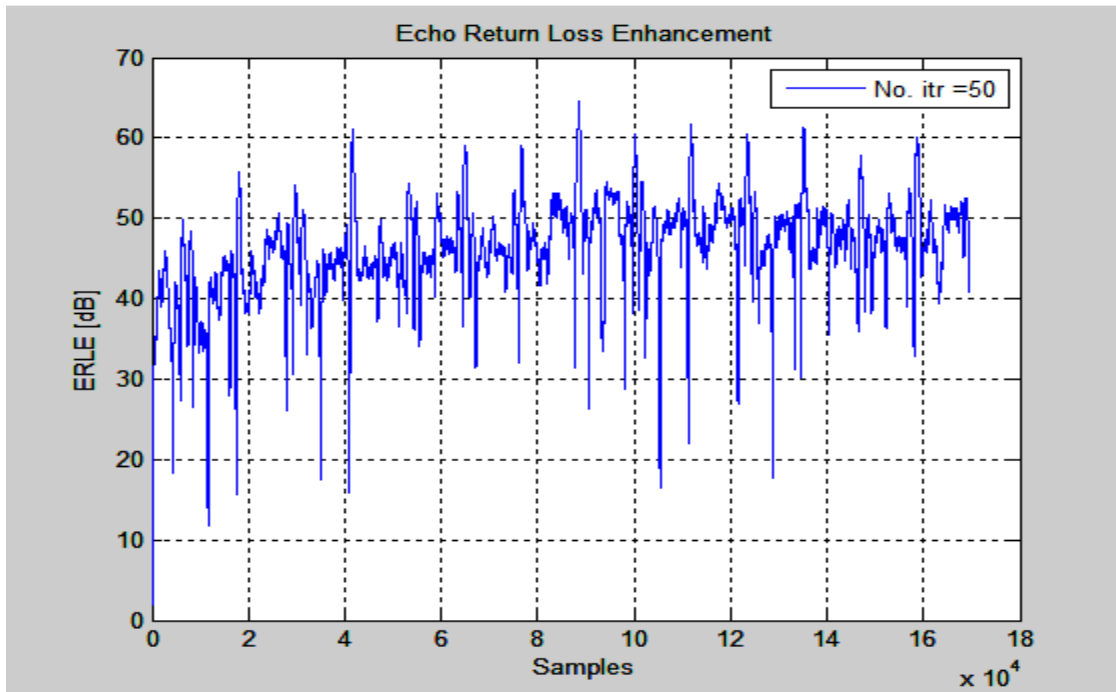


Figure 5.8: Plot of Echo Return Loss Enhancement for different number of iterations.

5.8. Discussion

- Here the same kind of conclusions is drawn as that of the linear FIR adaptive filter, that is, we managed to cancel the echo with AME of only $8.02e^{-4}$ for the linear kernel. Additionally, we have the added quadratic kernel which is used to cancel the nonlinear echo and in Figure 5.6 we see that we have only AME of $3.49e^{-5}$. These results are good results as they lead to ERLE enhancement which is well above the standard by 10dB on average.
- The other most important point is that in Figures 5.8 we see that the steady state value for the ERLE is between 40dB to 50dB which is well above the standard requirement (between 30dB and 40dB).

Chapter 6

Convex Combination of Adaptive Linear and Nonlinear Volterra Filters for AEC

6.1. Introduction

In recent years, nonlinear acoustic echo cancellation (NLAEC) schemes have become increasingly important due to the popularity of hands-free devices and mobile phones that use low-cost amplifiers and loudspeakers introducing significant nonlinearities into the acoustic echo path. Adaptive Volterra filters are widely used for NLAEC because of their generic structure, which can be considered as a straightforward generalization of linear adaptive filters [21].

Although Volterra filters decrease the residual nonlinear echo, they may not always be superior to a plain linear filter: For instance, if the echo cancellation scenario presents a low level of nonlinear echo, non-negligible gradient noise produced by the adaptation of second and higher order kernels degrades the performance of the NLAEC, so that the use of a simple linear adaptive filter would be more efficient. Note that, in general, the power of nonlinear echo is unknown, and will be time-varying for nonstationary signals like speech. Thus, the selection of the most effective adaptive filter, linear or Volterra is not a trivial problem since it requires a priori knowledge about the echo channel and the signal statistics.

A combination of filters constitutes an interesting way to mitigate different kinds of compromises involving adaptive filters [9, 24]. In this approach two or more adaptive filters adaptively combine their outputs obtaining a combined scheme that performs at least as well as the best contributing filter. Due to their simplicity, these schemes have been used in several areas of adaptive signal processing in communications and control applications, including blind equalization [25].

6.2. Convex Combination of Linear FIR and Second Order Volterra Filters

So far we have seen the individual filters i.e. the linear FIR and nonlinear Volterra filters and also what we can expect from the convex combination of the two filters. In this section we will see how the convexity is achieved and is applied to AEC. Figure 6.1 shows the block diagram of how the two filters are combined in convex fashion.

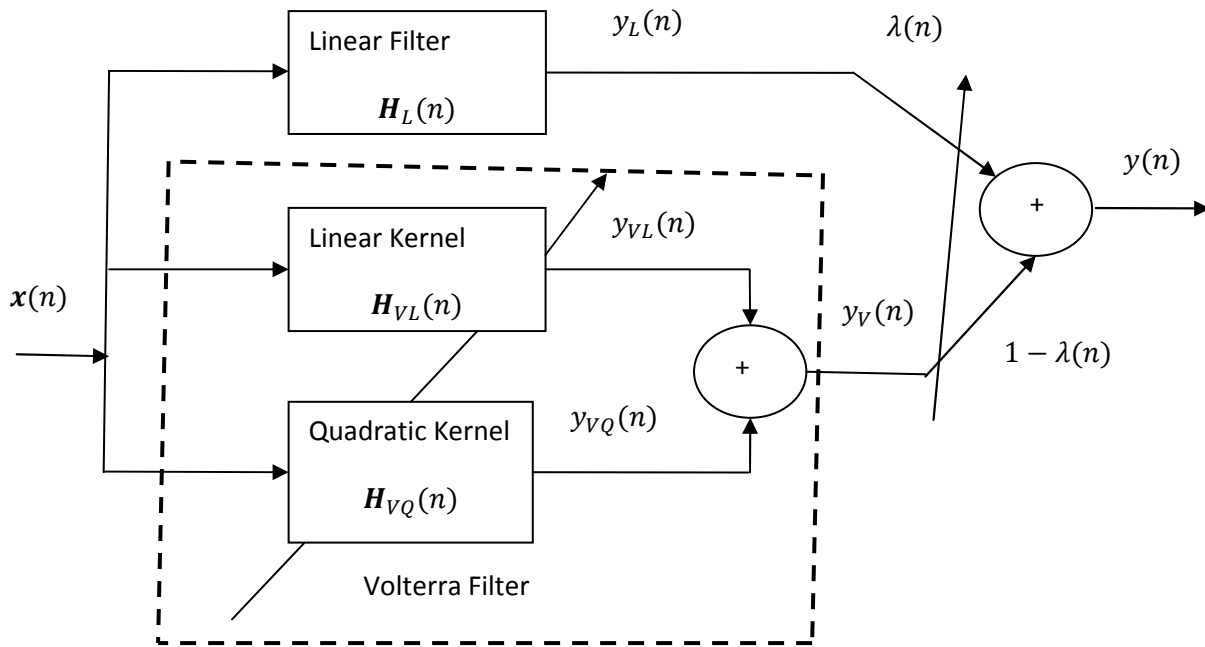


Figure 6.1: Block diagram of convex combination of linear FIR filter and second order Volterra filter.

This scheme consists of a straightforward convex combination of an adaptive linear filter, $H_L(n)$ and an adaptive second order Volterra filter including linear and quadratic kernels, $H_{VL}(n)$ and $H_{VQ}(n)$, respectively. The outputs of the linear and the Volterra filters, can be expressed as

$$y_L(n) = H_L^T(n)x(n) \dots \dots \dots 6.1$$

$$\begin{aligned}
 y_V(n) &= y_{VL}(n) + y_{VQ}(n) \\
 &= \mathbf{H}_{VL}^T(n)\mathbf{x}(n) + \mathbf{x}^T(n)\mathbf{H}_{VQ}(n)\mathbf{x}(n) \dots \dots \dots 6.2
 \end{aligned}$$

where $\mathbf{x}(n)$ denotes the vector of the input signal samples, and $y_{VL}(n)$ and $y_{VQ}(n)$ represent the outputs of the linear and quadratic kernels of the second order Volterra filter, respectively.

The output of the combined filter reads [10]:

$$y(n) = \lambda(n)y_L(n) + [1 - \lambda(n)]y_V(n) \dots \dots \dots 6.3$$

Where $\lambda(n)$ is an adaptive weighting parameter that controls the combination and $[0 < \lambda(n) < 1]$ for the convexity to hold.

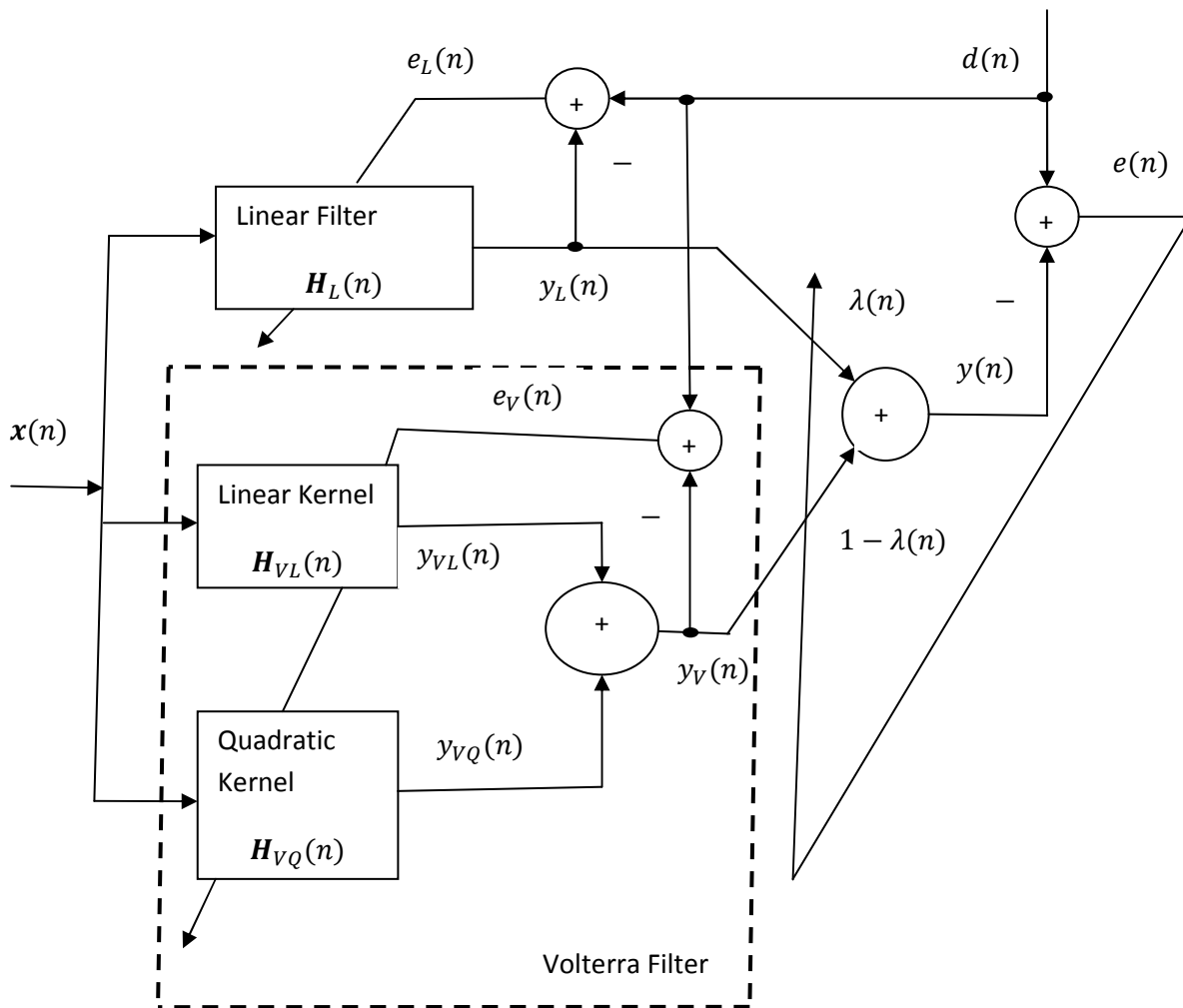


Figure 6.2: Block diagram of Adaptive convex combination of linear FIR filter and second order Volterra filter for AEC.

According to [10], for a good performance of the combination scheme, the contributing filters should update their coefficients following their own rules, in order to minimize the power of their own error signals. When using standard gradient descent rules, this results in

$$\mathbf{H}_L(n + 1) = \mathbf{H}_L(n) + \mu_L e_L(n) \mathbf{x}(n) \dots \dots \dots 6.4$$

$$\mathbf{H}_{VL}(n + 1) = \mathbf{H}_{VL}(n) + \mu_{VL} e_V(n) \mathbf{x}(n) \dots \dots \dots 6.5$$

$$\mathbf{H}_{VQ}(n + 1) = \mathbf{H}_{VQ}(n) + \mu_{VQ} e_V(n) \mathbf{x}(n) \mathbf{x}^T(n) \dots \dots \dots 6.6$$

Where μ_L , μ_{VL} and μ_{VQ} are step sizes for the linear, Volterra linear kernel and Volterra quadratic kernel, and

$$e_L(n) = d(n) - y_L(n) \dots \dots \dots 6.7$$

$$e_V(n) = d(n) - y_V(n) \dots \dots \dots 6.8$$

are the errors produced by the linear and the Volterra filters, respectively, and $d(n)$ is the reference signal to be approximated by the adaptive filters (see Figure 6.2).

The mixing parameter $\lambda(n)$ can also be updated using a gradient descent method with the aim of minimizing the square of the error produced by the combined filter, $e(n)$

$$e(n) = d(n) - y(n) \dots \dots \dots 6.9$$

However, instead of directly adapting $\lambda(n)$, we will rely on the adaptation of another parameter $a(n)$, which defines $\lambda(n)$ via a sigmoid activation function,

$$\lambda(n) = sgm[a(n)] = [1 + e^{-a(n)}]^{-1} \dots \dots \dots 6.10$$

This selection of sigmoid function is justified as it's necessary to keep $\lambda(n)$ between 0 and 1.

Recently, a new update rule for $a(n)$ has been presented in [26]. By normalizing the adaptation of $a(n)$, this rule allows an easier selection of the step size, μ_a and provides improved performance in scenarios with time-varying signal-to-noise ratio (SNR). This normalized rule can be expressed as:

$$a(n + 1) = a(n) + \left(\frac{\mu_a}{p(n)}\right) \lambda(n)[1 - \lambda(n)]e(n)[e_V(n) - e_L(n)] \dots \dots \dots 6.11$$

Where

$$p(n) = \beta p(n - 1) + (1 - \beta)[e_V(n) - e_L(n)]^2 \dots \dots \dots 6.12$$

is rough (low-pass filtered) estimate of the power of $[e_V(n) - e_L(n)]$.

Selection of the forgetting factor β is rather easy [26]. For instance, using $\beta = 0.9$ gives a good enough approximation, and typically ensures that $p(n)$ is adapted faster than any component filter.

The functionality of the presented scheme can be described as follows. When the LNL (Linear to Nonlinear Ratio) is low (i.e., there is a significant level of nonlinear echo), the Volterra filter represents an effective model of the channel, and minimization of the overall error yields $\lambda(n) \rightarrow 0$, so that $y(n) \approx y_V(n)$ and the combination is equivalent to Volterra Filter. The opposite occurs for high LNL, with $\lambda(n) \rightarrow 1$ and $y(n) \approx y_L(n)$, so that the combination is equivalent to a linear filter, avoiding the gradient noise caused by the adaptation of the Volterra quadratic kernel.

From equation (6.3) we know that when the linear echo dominates the nonlinear echo, the adaptation parameter, $\lambda(n)$ should evolve towards one. In Figure 6.3 above what we see is that $\lambda(n)$ evolves towards one as the linearity is dominant which indicates that the linear filter in this case is the best contributing filter and it's the one which cancels the echo.

B. Echo Return Loss Enhancement

Figure 6.4 below indicates the ERLE for the two filters and their convex combination:-

- ERLE for linear FIR adaptive filter [blue curve]
- ERLE for Nonlinear Volterra filter [green curve]
- ERLE for the convex combination of the two [red curve]

Here, since the linearity dominates what we see is that the combination [red curve] follows the best contributing filter that is the linear FIR adaptive filter [blue curve], therefore our model works well when linearity dominates.

In addition to that the achieved ERLE is also between 30dB to 40dB at steady state which is what the standard seeks.

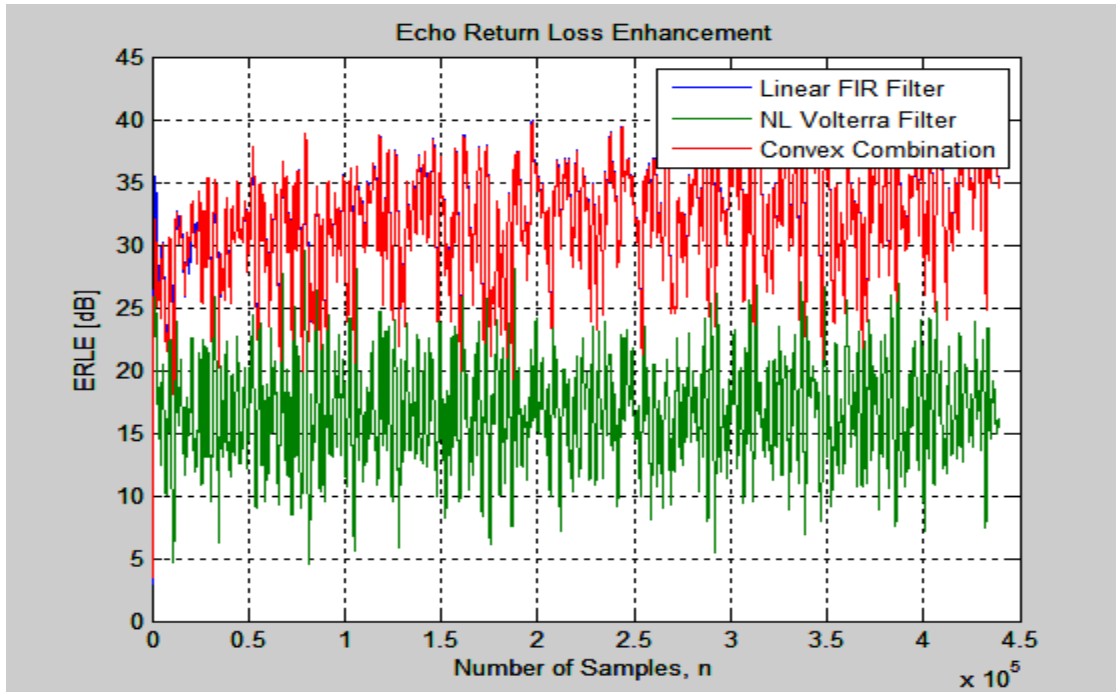


Figure 6.4: Plot of ERLE of the Linear FIR, Second Order Volterra Filter and their Convex Combination

C. Excess Mean Square Error (EMSE)

In Figure 6.5 below, plot for EMSE, the blue, green and red curves indicate the linear FIR adaptive filter, the nonlinear second order Volterra filter and their convex combination respectively. As we know from equation 6.14, EMSE indicates the amount of echo in dB that remains without being cancelled and in Figure 6.5 above we see that the combination scheme follows that of the linear and also they have lower residual echoes compared to second order Volterra filter.

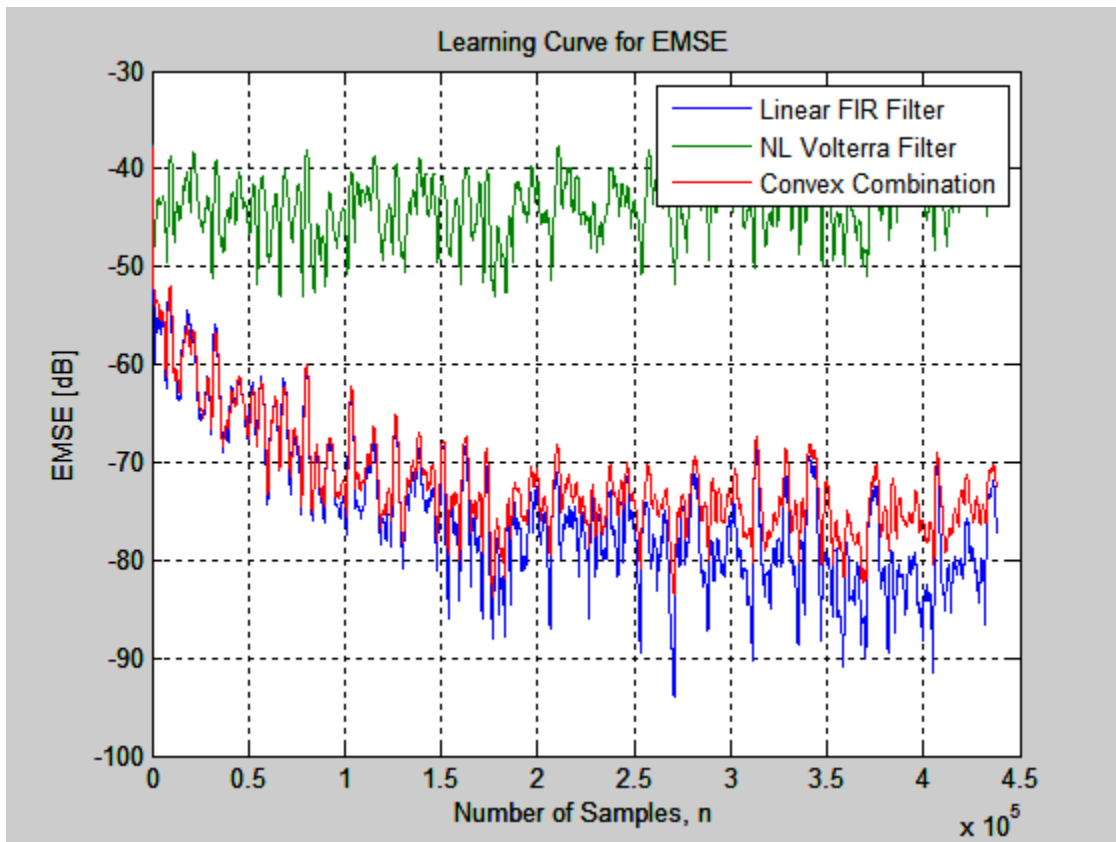


Figure 6.5: Plot of EMSE of the Linear FIR, Second Order Volterra Filter and their Convex Combination

Case II: - When nonlinearity dominates [$LNL R = -23dB$]

A. Adaptation Parameter $\lambda(n)$

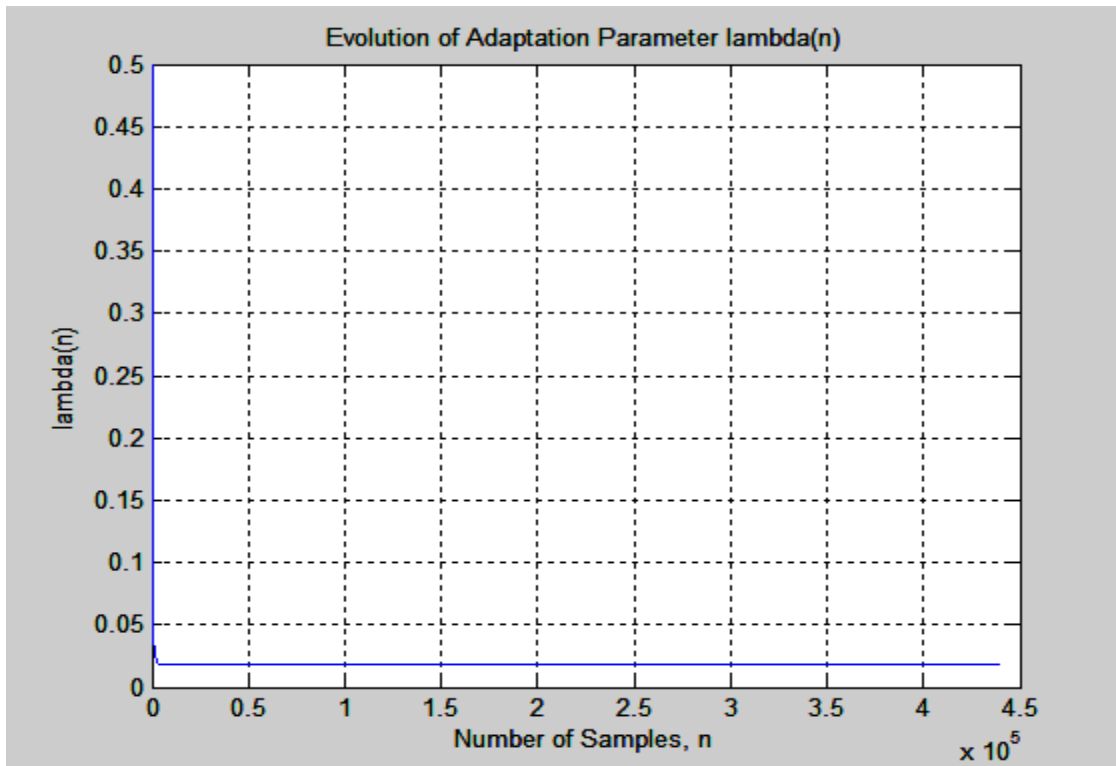


Figure 6.6: Evolution of the combination adaptation parameter [$\lambda(n)$]

From equation (6.3) we know that when the nonlinear echo dominates the linear echo ($LNL R = -23dB$) in this case, the adaptation parameter, $\lambda(n)$ should evolve towards zero. In Figure 6.6 above what we see is that $\lambda(n)$ evolves towards zero and hence the combination scheme is represented by the nonlinear Volterra filter as the nonlinearity is dominant.

B. Echo Return Loss Enhancement

In Figure 6.7 below we see that the red curve that is the combination scheme follows that of the green curve which is the nonlinear Volterra second order filter as nonlinearity is dominant and therefore, our model follows the best contributing filter.

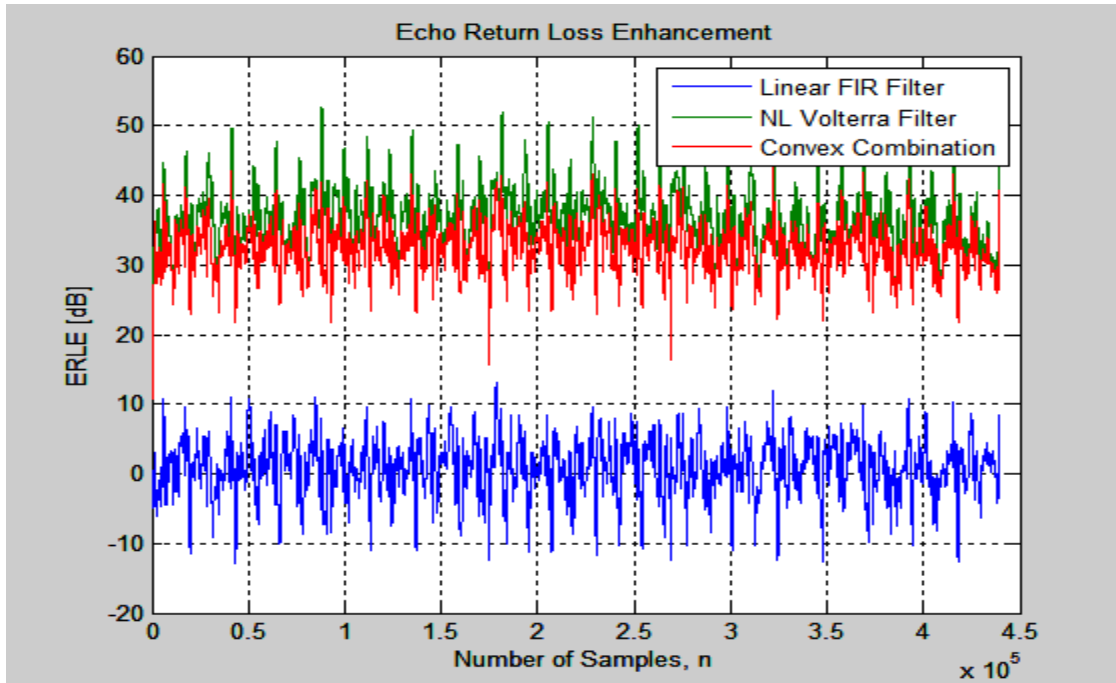


Figure 6.7: Plot of ERLE of the Linear FIR, Second Order Volterra Filter and their Convex Combination

C. Excess Mean Square Error (EMSE)

In Figure 6.8 below, plot for EMSE, the blue, green and red curves indicate the linear FIR adaptive filter, the nonlinear second order Volterra filter and their convex combination respectively. As we know from equation 6.14, EMSE indicates the amount of echo in dB that remains without being cancelled and in Figure 6.8 below we see that the combination scheme follows that of the nonlinear Volterra second order filter and also they have lower residual echoes compared to the linear filter as nonlinearity is dominant.

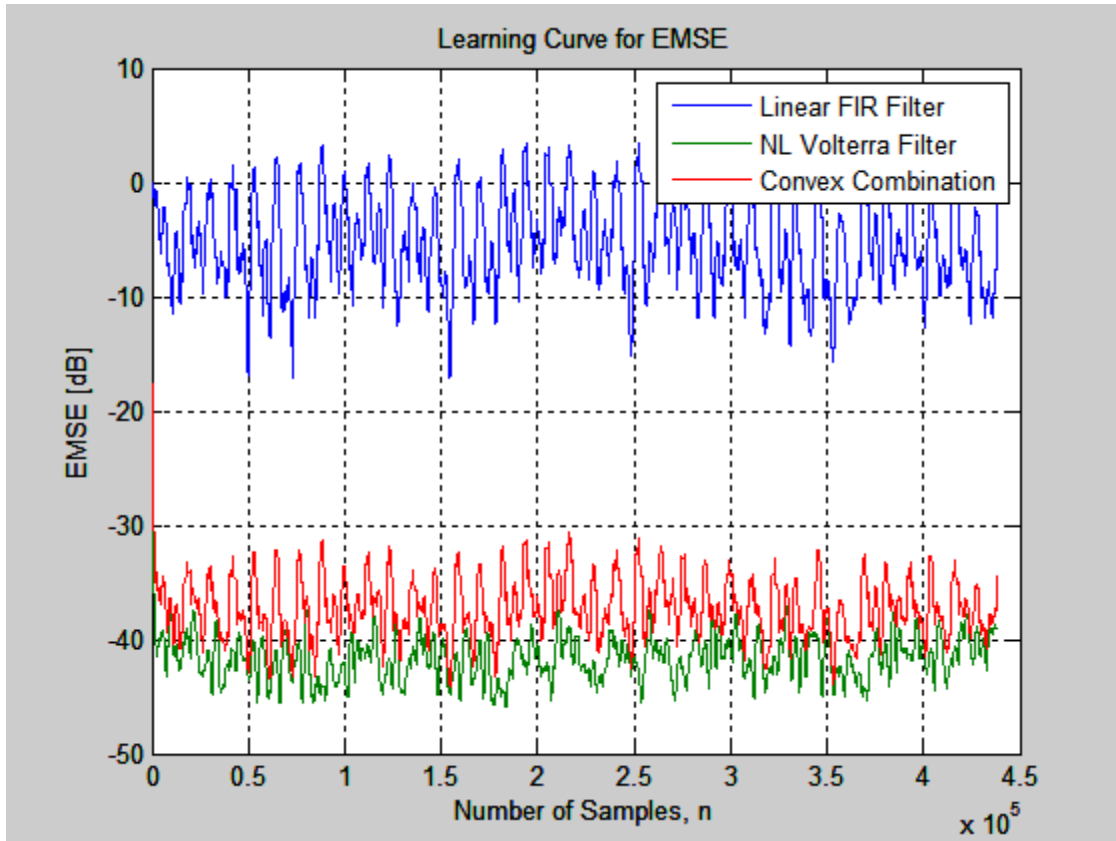


Figure 6.8: Plot of EMSE of the Linear FIR, Second Order Volterra Filter and their Convex Combination.

6.6. Discussion

- In Figure 6.3 the adaptation parameter, $\lambda(n)$ evolves towards one from its initial value that is 0.5 (average of 0 and 1) which implies that the linear filter is the dominant one in this scenario and it's in agreement with the convexity principle as can be inferred from equation (6.3).
- In Figure 6.4 and 6.5 we see that the red curve and the blue curve which indicate the Convex Combination and Linear FIR filter, respectively, overlap which means that the combination model follows the best contributing filter: - in this case the Linear FIR as linearity is dominant.

- In Figure 6.6 the adaptation parameter, $\lambda(n)$ evolves towards zero from its initial value of 0.5 (average of 0 and 1) which implies in this case the dominant filter is the nonlinear second order Volterra filter and the overall filter follows the Volterra model.
- In Figure 6.7 and 6.8 we see that the red curve follows the green curve which indicates the Convex Combination and Nonlinear Volterra filter, respectively. It means that the convexity model acts in the same manner as the best contributing filter: - in this case the nonlinear Volterra as nonlinearity is dominant.
- The other most important point is that in Figures 6.4 and 6.7 we see that the steady state value for the ERLE is as per the standard requirement which is between 30dB and 40dB.

Chapter 7

Double Talk Detection (DTD)

7.1. Introduction

So far we have considered only single talk (ST) situation which occurs when only the far-end or near-end speaker talks. This is not always true in real life, in fact according to [27] double talk (DT) -when both parties talk at the same time, occurs 20% of the conversation period. If double talk situation occurs and the adaptive filter is allowed to update its coefficients without any control, the filter coefficients will diverge instead of converging to optimum value. This is because the near-end speech acts as a corrupting noise to the filter reference which is normally the echo of far-end speaker. Therefore design of a good DTD is essential when designing an Acoustic Echo Canceller.

7.2. Working Principle of DTDs

Most echo cancellers attempt to detect double talk occurrences and then react by freezing the adaptation of the adaptive filter. A DTD employs available signals or estimates to make the decision of whether or not near-end speech $v(n)$ is present. The DTD decision is then utilized to design the control logic for the AEC filter. In general, double-talk detection is handled in the following way [28]:

1. A detection statistic ξ is formed using available signals, e.g., x , y , e , etc., and the estimated filter coefficients.
2. The detection statistic ξ is compared to a preset threshold T and the double-talk is declared if $\xi > T$.
3. Once double-talk is declared, the filter adaptation is disabled.
4. If $\xi \leq T$, the comparison of ξ to T and filter adaptation continue.

Figure 7.1 below shows the implementation of DTD in AEC where the DTD produces a decision statistic based on the available signal. The adaptive filter then acts accordingly by adjusting its state.

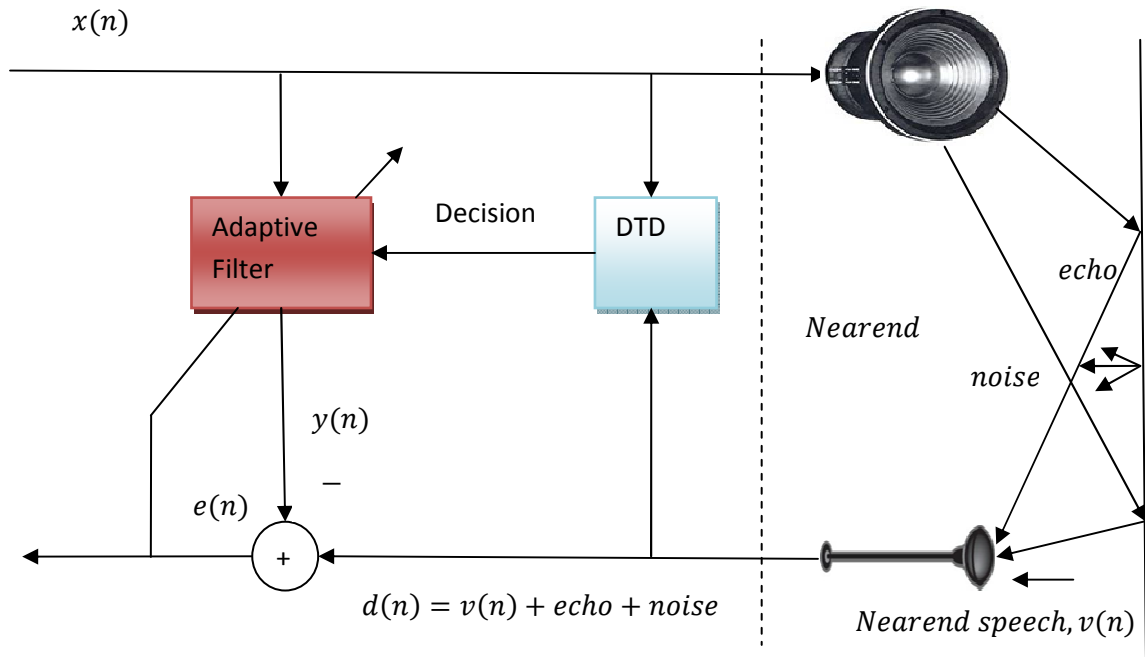


Figure 7.1: AEC with Double Talk Detector (DTD)

The Geigel algorithm [42] has been proven effective for line echo cancellers. However, it does not provide reliable performance when applied to AECs. Cross-correlation-based DTD techniques [29, 30] have been proposed, that appear to be suitable for AEC applications. However, the correlation-based criterion captures only the linear relationship between two random processes [30, 32, 29, and 31]. Although [33] derives an optimum log-likelihood ratio test (LRT), the Gaussian assumption of signals does not hold any more when nonlinearity is present. A new approach which can detect the DT situation for both linear and nonlinear relationships based on mutual information has been presented in [34] and it will be discussed and implemented here.

7.3. DTD Based on Mutual Information (MI)

Mutual information (MI) is in many ways the cornerstone of classic information theory, playing central roles in the analysis of both digital and analog communication systems [35]. The primary objective of DTD designs is to detect the presence of the near-end speech. In this section we will see how MI is suited for this task in AEC.

7.3.1. MI and Its Calculation

We start with the fundamentals of MI. Denote continuous-valued random variables x and y by the pair (x, y) . The entropy or uncertainty of the variable x is defined in terms of its probability density function (PDF), $f(x)$ is:

$$H(x) = -\int_x f(x) \log f(x) dx \dots \dots \dots 7.1$$

After having observed y , the uncertainty of x is given by the conditional entropy, defined in terms of the conditional PDF $f(x|y)$ and the joint PDF $f(x, y)$:

$$H(x|y) = -\int_x \int_y f(x, y) \log f(x|y) dy dx \dots \dots \dots 7.2$$

The MI between x and y is defined as in [35]

$$\begin{aligned} I(x: y) &= H(x) - H(x|y) \\ &= -\int_x f(x) \log f(x) dx + \int_x \int_y f(x, y) \log f(x|y) dy dx \dots \dots \dots 7.3 \end{aligned}$$

and measures the reduction in the uncertainty of x due to the knowledge of y .

Another view of MI is that it measures the degree to which x and y are not independent. With the identity $f(x, y) = f(x|y)f(y)$, the expression in (7.3) can be rewritten as

$$I(x: y) = \int_x \int_y f(x, y) \log \frac{f(x, y)}{f(x)f(y)} dy dx \dots \dots \dots 7.4$$

When x and y are statistically independent, $f(x, y) = f(x)f(y)$, and thus $I(x; y) = 0$. The value of $I(x, y)$ grows as x and y become more dependent. The more dependent x is on y , the more information one gains about x once y is known, and therefore the less uncertain x is when y is known.

The following properties hold for MI [35].

Property I: $0 \leq I(x; y) \leq \infty$

Property II: $I(x; y) = 0$ if and only if x and y are statistically independent.

Property III: $I(x; y) = \infty$ if and only if y is a function of x , i.e., $y = g(x)$ where $g(\cdot)$ is invertible.

Property IV: $I(x; y) = I(u; v)$ if the transformation $(x; y) \rightarrow (u; v)$ has the form $u = g(x), v = p(y)$ with $g(\cdot)$ and $p(\cdot)$ being one-to-one mappings.

In order to calculate the MI between x and y , we need to estimate the joint distribution $f(x; y)$. Histogram and Kernel methods are widely used to estimate MI but entail high computational complexity [36]. To reduce the complexity in these methods, we adopt a recent estimator that estimates entropy from the average distance to the k -nearest neighbors as in [37]. Consider a set of N input-output pairs, $z_i = (x_i; y_i); i = 1; \dots; N$, and the maximum norm [37]

$$\|z - z'\|_\infty = \max\{|x - x'|, |y - y'|\}, \dots \dots \dots 7.5$$

for a fixed positive integer k , we find $z_{k(i)} = (x_{k(i)}; y_{k(i)})$ as the k -th nearest neighbor of z_i according to the maximum norm.

Define the following distances

$$\epsilon_i/2 = \|z_i - z_{k(i)}\|_\infty \dots \dots \dots 7.6$$

$$\epsilon_i^x/2 = |x_i - x_{k(i)}| \dots \dots \dots 7.7$$

$$\epsilon_i^y/2 = |y_i - y_{k(i)}| \dots \dots \dots 7.8$$

where $\epsilon_i/2$ is the distance from z_i to its k -th neighbor. $\epsilon_i^x/2$ and $\epsilon_i^y/2$ are the distances between the same points projected onto x and y subspaces.

Let n_i^x and n_i^y be the numbers of sample points that satisfy

$$|x_i - x_j| \leq \epsilon_i^x/2 \text{ and } |y_i - y_j| \leq \epsilon_i^y/2 \text{ respectively.}$$

The estimator of the MI between x and y is then obtained:

$$\hat{I}(x; y) = \psi(k) - \frac{1}{k} - \frac{1}{N} \sum_{i=1}^N [\psi(n_i^x) + \psi(n_i^y)] + \psi(N) \dots \dots \dots 7.9$$

where $\psi(\cdot)$ is the Digamma function (see appendix A).

As can be noted, this MI estimator has a dependency on the value chosen for k (k -th nearest neighbor). As it is recommended in [38], for a tradeoff between variance and bias, in the examples, a mid-range value for k ($k = 6$) will be used.

Considering Figure 7.1 above, the DTD utilizes two signals- $x(n)$ which is the far-end speech and $d(n)$ which is the sum of the echo of far-end speaker, background noise and the near-end speech, $v(n)$ (if any)- to make the decision on whether the near-end speech $v(n)$ is present or not. The idea here is to form a decision statistic ξ and compare it with a preset threshold T . Once double-talk is declared, the AEC filter adaptation is disabled.

To design the decision statistics ξ , we first consider the linear case ($u = g(x) = \alpha x$, α is a non-zero constant). If $v(n)$ is absent then $d(n) = h(n) * g(x(n))$, i.e., $y(n)$ fully depends on $x(n)$. Hence, to determine whether the signal $v(n)$ is present or not is equivalent to measuring the degree of dependency between x and d . From Property III of MI, $I(x; d)$ achieves the maximum when x and d are fully dependent. If we treat sequences x and d as the realizations of random variables x and d , respectively, the presence of the near-end speech reduces $I(x; y)$. Therefore, the MI between x and d will be used as the detection statistic for the DTD.

$$\xi = I(x; y) \dots \dots \dots 7.10$$

and DT is declared if $\xi < T$ or

ST is declared otherwise;

When the echo path is nonlinear ($u = g(x)$ is an invertible nonlinear mapping of x), we know from property IV of MI that the MI between x and d is the same as that between u and d :

$$I(x; d) = I(g(x); d) = I(u; d) \dots \dots \dots 7.11$$

Therefore, the MI-based DTD still works in the presence of the memoryless nonlinearity, provided that the nonlinear mapping is one-to-one. It is worth pointing out that Shannon's mutual information is a classical measure of statistical dependence between random variables no matter whether the relationship between two random variables is linear or nonlinear. This makes MI based DTD robust to nonlinearities.

7.4. Simulation Setup

In this section we will simulate the DTD and AEC together to see the performance of the DTD. As mentioned before, the Mutual Information between the far-end speech, $x(n)$ and the desired signal, $d(n)$ will be considered for the DT detection. Here what's important is the determination of the threshold used to declare DT state. However, since there is no universal rule for the selection of the threshold and it also depends on the nature of the method implemented to detect the DT we will develop our own systematic approach to set the threshold. In the simulation first the algorithm is run with only ST situation and the resulting MI between $x(n)$ and $d(n)$ is considered as a threshold. Once the threshold is set the algorithm is run again with the near-end speech also applied.

- 'Melkamledet' music is considered as far-end speech signal, $x(n)$
- 'nearspeech' speech signal stored in Matlab is used as near-end speech signal, $v(n)$. It is applied at two different locations in order to verify the robustness of the DTD.

- The simulation is carried out for both linear acoustic echo and nonlinear acoustic echo

As a performance measure we will consider:-

- Algorithm misalignment error with DTD and without DTD
- The ability of the algorithm to pinpoint the DT region both in linear acoustic and nonlinear acoustic case.

7.5. Results

A. DT Region Detection and Algorithm Misalignment Error for Linear Acoustic Echo

i. Performance of DT Region Detection

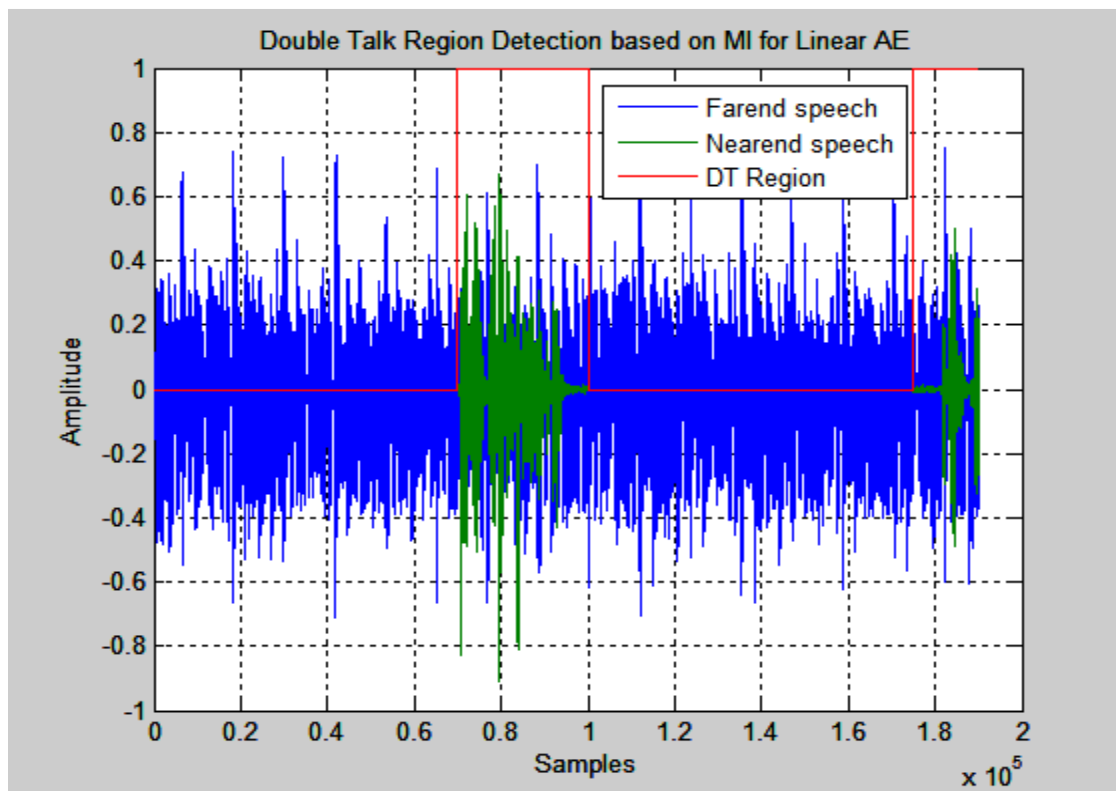


Figure 7.2: Double Talk Region Detection using MI for linear acoustic echo

In Figure 7.2 we have far-end speech (blue) which is the input to AEC, the near-end speech (green) and the red line which is the result of MI based DTD indicating the DT region. From the Figure we see that both of the DT regions are identified using the MI based DTD.

ii. Algorithm Misalignment Error without DTD

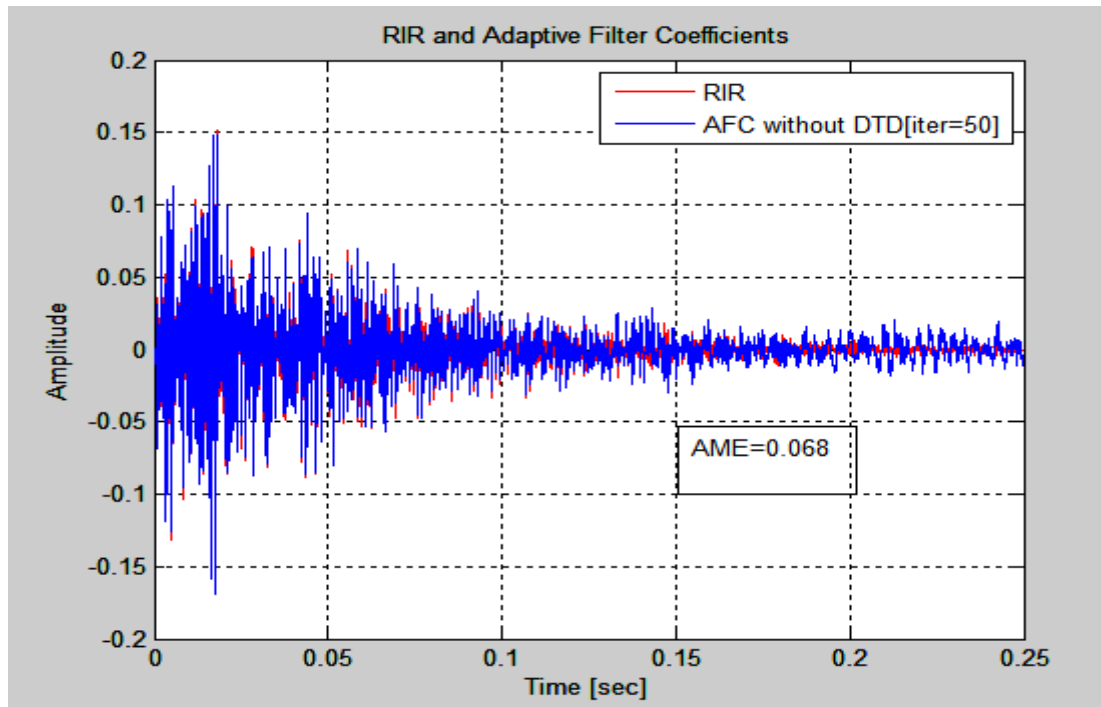


Figure 7.3: Algorithm misalignment error without DTD for linear AEC

iii. Algorithm misalignment error with MI based DTD

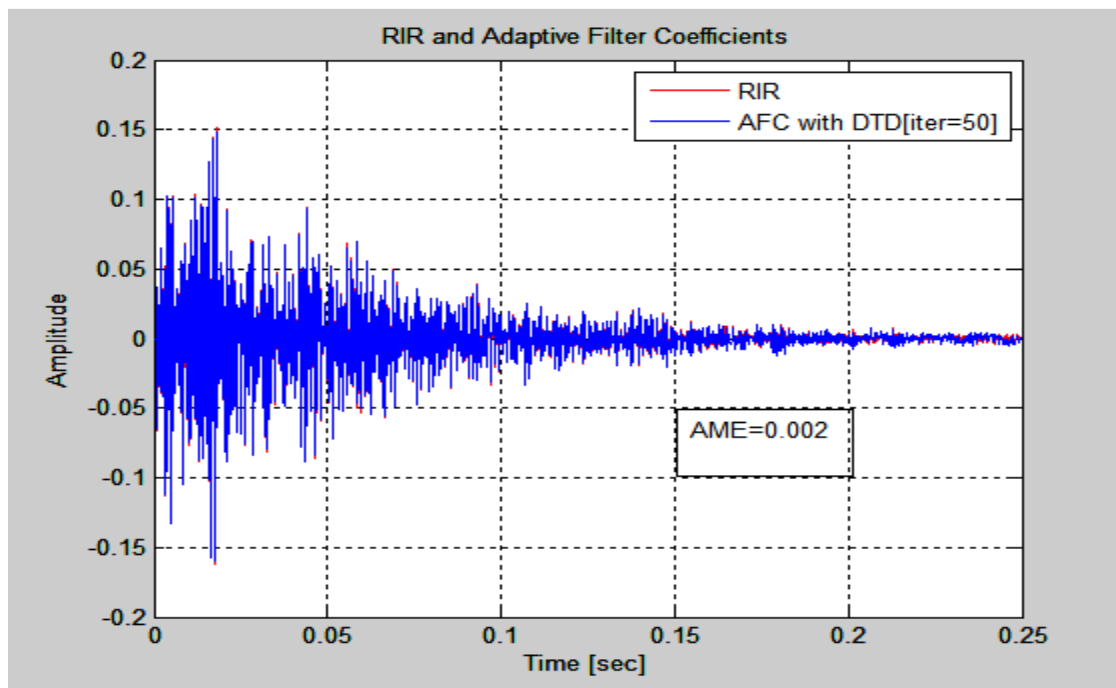


Figure 7.4: Algorithm misalignment error with MI based DTD for linear AEC

From Figure 7.3 [AME without DTD] and Figure 7.4 [AME with DTD] we see that the AME decreases from 0.068 to 0.002 that is we have 97.06% improvement which indicates we have achieved the desired improvement by using MI based DTD.

B. DT Region Detection and Algorithm Misalignment Error for nonlinear acoustic echo

i. DT Region Detection

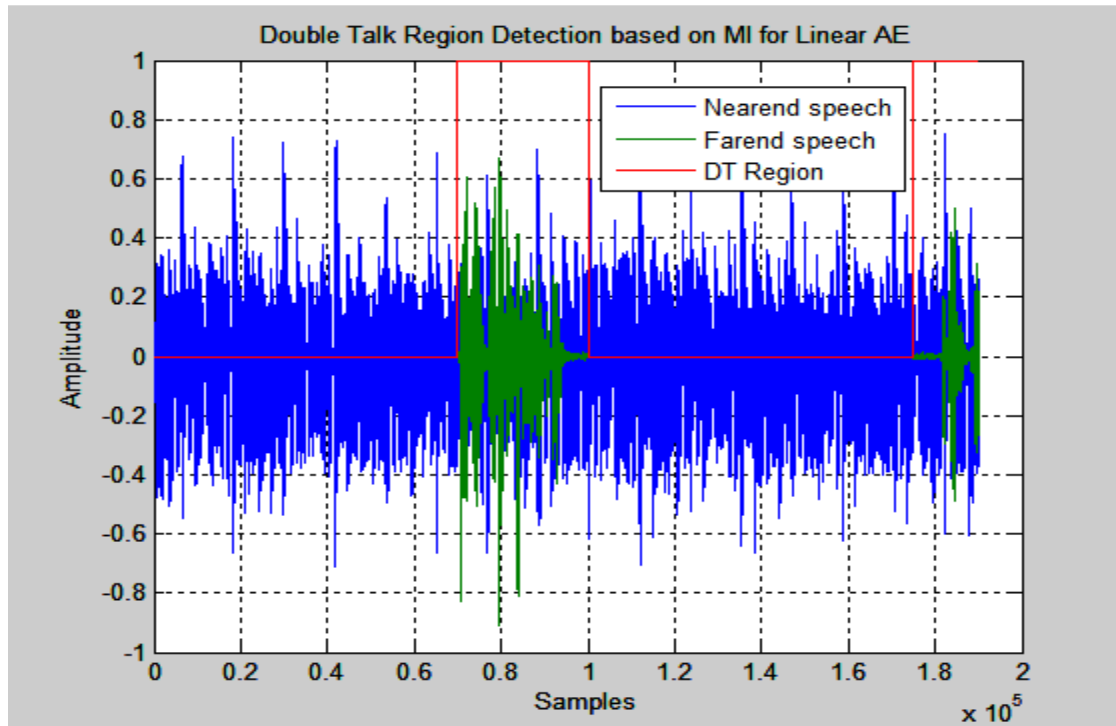


Figure 7.5: Double Talk Region Detection using MI for Nonlinear acoustic echo

In Figure 7.5 we have far-end speech (blue) which is the input to AEC, the near-end speech (green) and the red line which is the result of MI based DTD indicating the DT region. From the Figure we see that, the same as for the linear case, both of the DT regions are identified using the MI based DTD.

ii. Algorithm Misalignment

➤ Linear Kernel

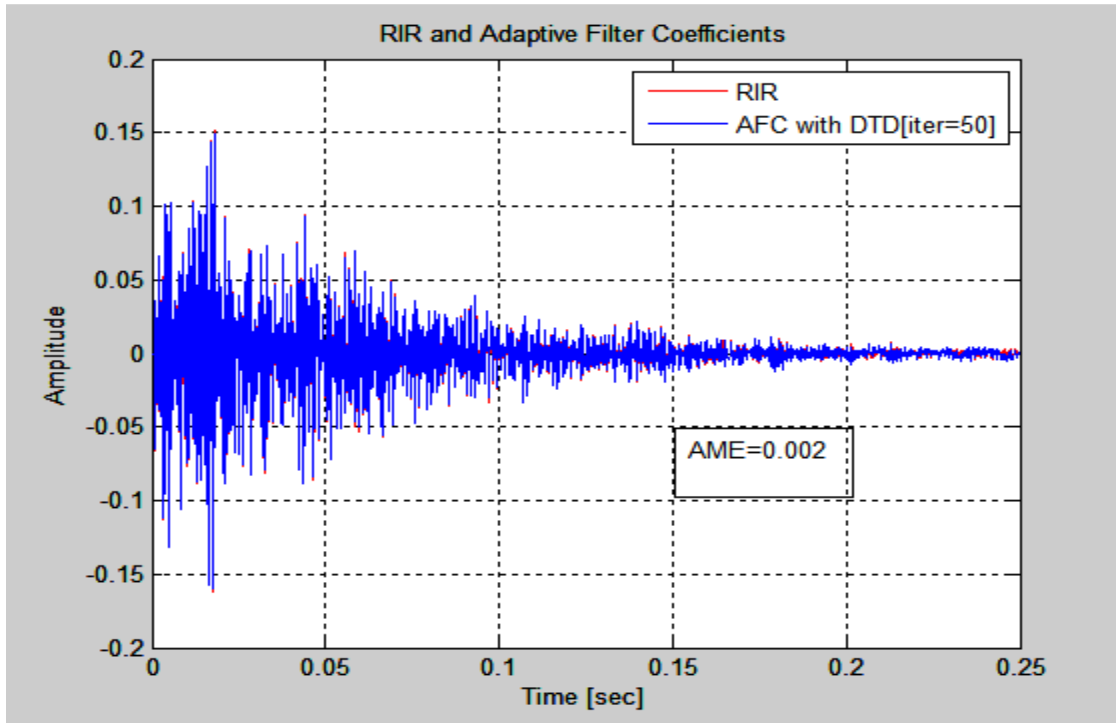


Figure 7.6: Algorithm misalignment error with MI based DTD for linear AEC

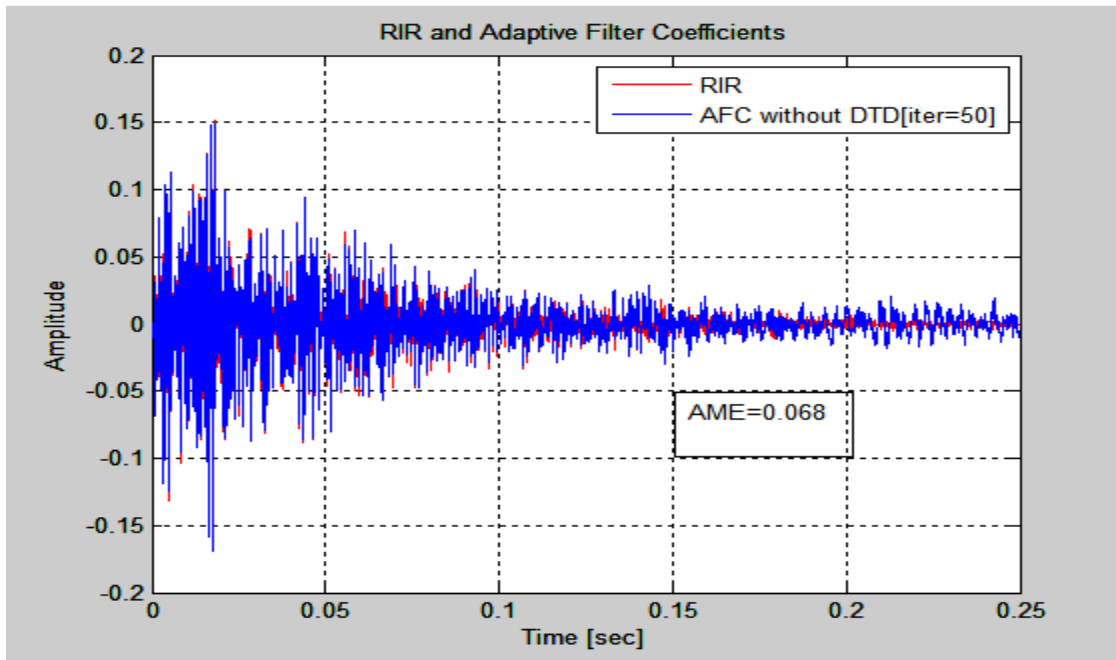


Figure 7.7: Algorithm misalignment error without DTD for linear AEC

➤ Quadratic Kernel

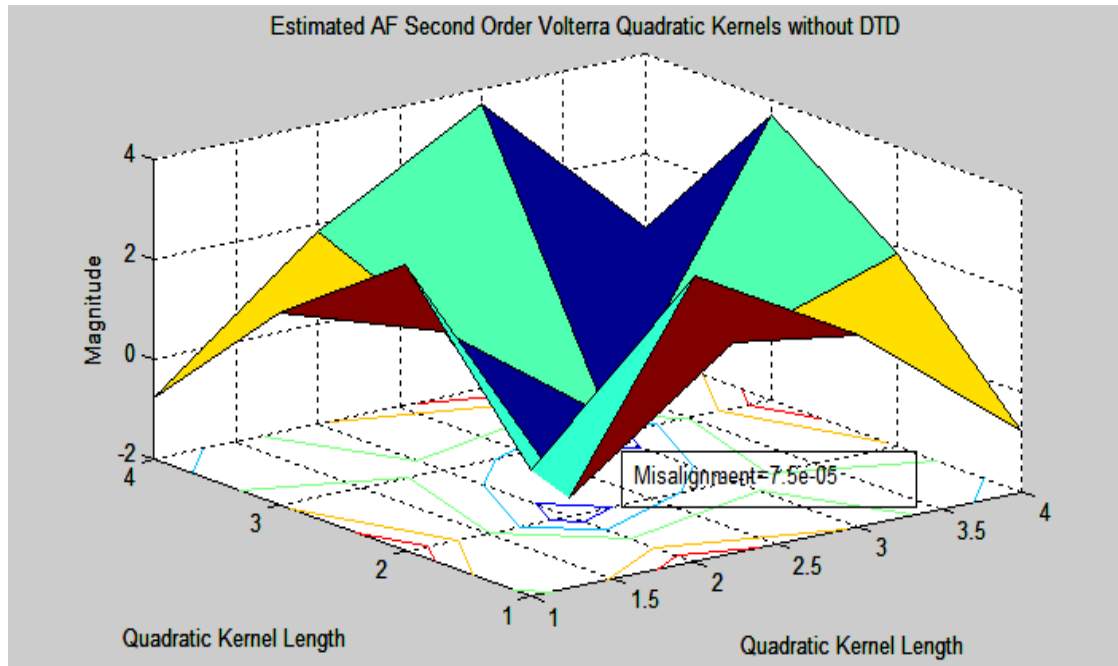


Figure 7.8: Algorithm misalignment error of quadratic kernel for nonlinear AEC with MI based DTD

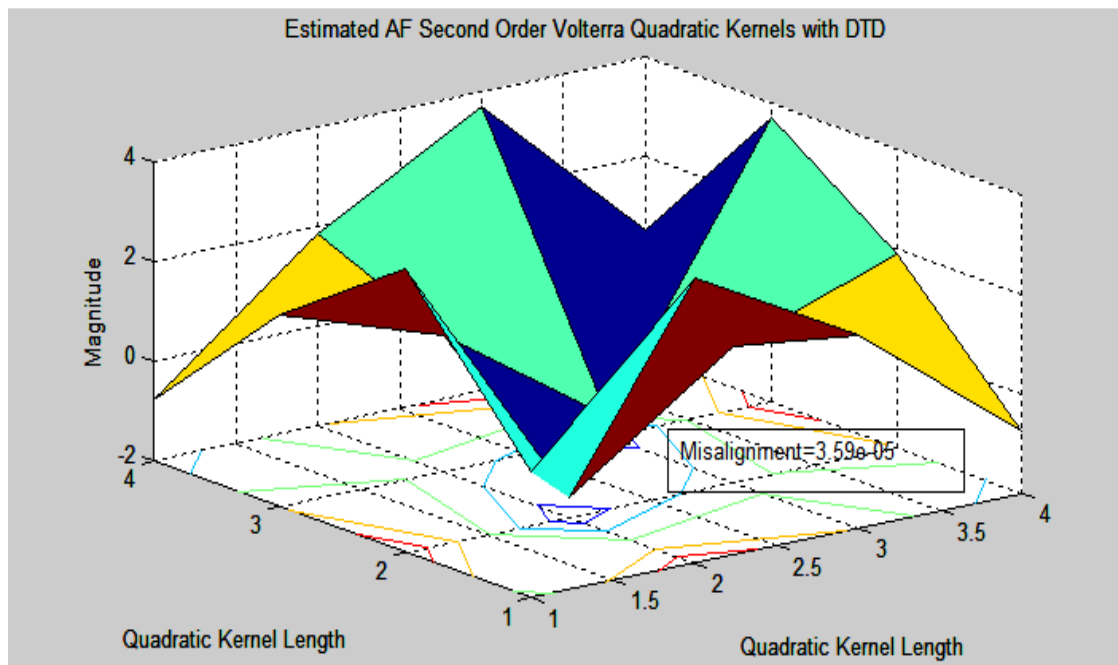


Figure 7.9: Algorithm misalignment error of quadratic kernel for nonlinear AEC without DTD

7.6. Discussion

In Figure 7.2 and Figure 7.5 we see that the double talk regions for both linear acoustic echo and nonlinear acoustic echoes are well detected by the MI based DTD which indicates the power of MI based approach for DTD design and also the results obtained are absolutely in agreement with the theory discussed in the foregoing sections.

The other important point that we see is in Figures 7.2 and 7.3 for linear acoustic echo case is that the algorithm misalignment error with MI based DTD is much lower than without it, i.e., the misalignment decreases from 0.068 to 0.002 that is with improvement of 97.06%. The same thing happens for the nonlinear acoustic echo as indicated in Figures 7.6 to 7.9. The AME for the linear kernel has the same amount of improvement as for the linear AEC case and the AME for the quadratic kernel decreases from $7.5e^{-5}$ to $3.59e^{-5}$. Therefore, the implementation of MI based DTD for AEC is very important as it has brought improvement in the echo cancelling capacity.

Chapter 8

Conclusions and Recommendations for Further Research

8.1. Conclusions

In this thesis we have seen that acoustic echoes are inevitable in modern hands-free communication and also they need to be removed for enhancing speech quality in communication systems. We have seen that convex combination of filters with different natures is of great interest and we have successfully applied this combination scheme for AEC in terms of linear and nonlinear filters. The results obtained also indicate that the combination scheme is the best way to tackle different echo scenarios, i.e., whether the echo is linear or nonlinear. The last but not the least is the use of DTD is highly essential to maintain the full duplex communication system by freezing the adaptation of the AEC. MI based DTDs are robust for both linear and nonlinear AEC systems.

8.2. Recommendations for Further Research

Even if we have seen that convex combination of adaptive linear FIR and nonlinear Volterra filters for AEC is robust to the different natures of the echo it has not come without any cost. In fact it has additional computational cost and memory requirement. Therefore one research area can be reducing this computational cost by implementing different kind of adaptive algorithms with low computational cost and/or implementing in frequency domain. Other research area of interest regarding this work is the hardware implementation of the convex combination scheme with DTD.

References

- [1] G. Shiers, ed., "The Telephone: A Historical Anthology. Historical studies in telecommunications", New York: Arno Press, 1977.
- [2] Saeed, V. V., "Advanced Digital Signal Processing and Noise Reduction," 2nd edition, New York: John Wiley & Sons Ltd, 2000.
- [3] International Telecommunication Union, "Echo Suppressors", ITU-T Recommendation G.164, Nov. 1988.
- [4] Christina, G., "Echo Cancellation Using the LMS Algorithm," *U.P.B. Sci. Bull., Series C*, Vol. 71, Iss. 4, 2009.
- [5] Amit, M., Vibha, A., & Gurpal, S., "RLS Algorithm for Acoustic Echo Cancellation," *Proceedings of 2nd National Conference on Challenges & Opportunities in Information Technology, COIT-2008*.
- [6] Stenger, A., Trautmann, L., & Rabenstein R., "Nonlinear Acoustic Echo cancellation with 2nd order volterra Filters," *IEEE Int. Conf. on Acoustics, Speech & Signal Processing (ICASSP), Phoenix, USA*, March 1999.
- [7] J. B. Seo, K. J., "Nonlinear Acoustic Echo Cancellation Using Volterra Filtering with a Variable Step-Size GS-PAP Algorithm," V57, 2009.
- [8] Azpicueta-Ruiz, L. A., Zeller, M., Arenas-Garcia, J., & Kellermann, W., "Novel Schemes for Nonlinear Acoustic Echo Cancellation Based on Filter Combinations," *IEEE ICASSP*, 2009.
- [9] Silva, M. T., & Nascimento, V. H., "Convex Combination Of Adaptive Filters With Different Tracking Capabilities," 2009 .
- [10] J. Arenas-Garcia, A. R. Figueiras-Vidal, and A. H. Sayed, "Mean-square performance of a convex combination of two adaptive filters," *IEEE Trans. Signal Process.*, vol. 54, pp. 1078–1090, 2006.
- [11] E. Hansler, "The hands-free telephone problem - an annotated bibliography", *Signal Processing*, vol. 27, pp. 259–271, June 1992.

-
- [12] A. Liavas and P. Regalia, "Acoustic echo cancellation: Do IIR models offer better modeling capabilities than their FIR counterparts," *IEEE Trans. Signal Processing*, vol. 46, pp. 2499–2504, Sept. 1998.
- [13] G. Glentis, K. Berberidis, and S. Theodoridis, "A unified view: Efficient least squares adaptive algorithms for FIR transversal filtering," *IEEE Signal Processing Magazine*, vol. 16, pp. 13–41, July 1999.
- [14] A. Gilloire and M. Vetterli, "Adaptive filtering in subbands with critical sampling: analysis, experiments, and application to acoustic echo cancellation," *IEEE Trans. Signal Processing*, vol. 40, pp. 1862–1875, Aug. 1992.
- [15] Q.-G. Liu, B. Champagne, and K. C. Ho, "Simple design of oversampled uniform DFT filter banks with applications to subband acoustic echo cancellation," *Signal Processing*, vol. 80, pp. 831–847, June 2000.
- [16] T. Aboulnasr and K. Mayyas, "Complexity reduction of the NLMS algorithm via selective coefficient update," *IEEE Trans. Signal Processing*, vol. 47, pp. 1421–1424, May 1999.
- [17] S. L. Gay, "An efficient, fast converging adaptive filter for network echo cancellation," in *Proc. Asilomar Conf. Signals, Systems, Computers*, vol. 1, (Pacific Grove, USA), pp. 394–398, Nov. 1998.
- [18] K. Ozeki and T. Umeda, "An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties," *Electronics and Communications in Japan*, vol. 67-A, no. 5, pp. 19–27, 1984.
- [19] S. L. Gay and S. Tavathia, "The fast affine projection algorithm," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP'95)*, (Detroit, USA), pp. 3023–3026, May 1995.
- [20] Jont B. Allen and David A. Berkley, "Image method for efficiently simulating small-room acoustics," *Journal of Acoustics Society of America*, vol. 65 no. 4, 1978.
- [21] V. J. Mathews and G. L. Sicuranza, "Polynomial Signal Processing," *John Wiley & Sons Inc.*, 2000.
- [22] A. Stenger, L. Trautmann, and R. Rabenstein, "Nonlinear acoustic echo cancellation with second order adaptive Volterra filters," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Phoenix, AZ, vol. 2, Mar. 1999, pp. 877-880.

-
- [23] International Telecommunication Union, "General characteristics of international telephones connections and international telephone circuits – acoustic echo controllers," *ITU-T Recommendation G.167*, 1993.
- [24] M. Martinez-Ramon, et al., "An adaptive combination of adaptive filters for plant identification," in *Proc. 14th Intl. Conf. Digital Signal Process*, 2002.
- [25] M. T. M. Silva and V. H. Nascimento, "Improving the Tracking Capability of Adaptive Filters via Convex Combination," *IEEE Trans. Signal Process.*, vol. 56, pp. 3137–3149, 2008.
- [26] L.A. Azpicueta-Ruiz, A.R. Figueiras-Vidal and J. Arenas-Garcia, "A Normalized Adaptation Scheme for the Convex Combination of Two Adaptive Filters," in *Proc. IEEE ICASSP'08*, pp. 3301–3304, 2008.
- [27] Chao, J., Tsujii, S., "A stable and distortion-free echo and howling canceller," *International Conference on Acoustics, Speech, and Signal Processing (ICASSP-88)*, pp. 1620–1623, 1998.
- [28] J.H. Cho, D.R. Morgan, J. Benesty, "An objective technique for evaluating doubletalk detectors in acoustic cancellers," *IEEE Trans. Speech Audio Process*, (November 1999) 718–724.
- [29] Benesty, J., Morgan, D. R., and Cho, J. H., "A new class of doubletalk detector based on cross-correlation", *IEEE Trans. on Speech and Audio Processing*, vol. 8, pp. 168-172, Mar. 2000.
- [30] Ye, H. and Wu, B.-X., "A new double-talk detection algorithm based on the orthogonality theorem," *IEEE Trans. on Communications*, vol. 39, pp. 1542-1545, Nov.1991.
- [31] GÄansler, T. and Benesty, J., "The fast normalized cross-correlation double-talk Detector", *Signal Processing*, vol. 86, pp. 1124-1139, June 2006.
- [32] GÄansler, T., Hansson, M., Ivarson, C.-J., and Salomonsson, G., "A double talk detector based on coherence", *IEEE Trans. on Communications*, vol. 44, pp. 1421-1427, Sept. 1996.
- [33] Bershad, N. J. and Tourneret, J.-Y., "Echo cancellation-a likelihood ratio test for double-talk versus channel change", *IEEE Trans. on Signal Processing*, vol. 54, pp. 4572-4581, Dec. 2006.
- [34] Shi, K., Ma, X., and Zhou, G. T., "A mutual information based double talk detector for nonlinear systems", in *Proc. IEEE Conference on Information Sciences and Systems*, (Princeton, NJ), Mar. 2008.
- [35] Cover, T. and Thomas, J., "Elements of Information Theory", *Wiley New York*, 1991.

- [36] Scott, D., "Multivariate Density Estimation: Theory, Practice, and Visualization", *Wiley-Interscience*, 1992.
- [37] Kraskov, A., Stogbauer, H., and Grassberger, P., "Estimating mutual information", *Physical Review*, vol. 69, p. 66138, 2004.
- [38] S. Harald, K. Alexander, A.A. Sergey, G. Peter, "Least dependent component analysis based on mutual information", *Phys. Rev.*, E 70, 2004, 066123.
- [39] Ronald E. Walpole et.al, "Probability and Statistics for Engineers and Scientists", *Prentice Hall*, 8th ed. 2007.
- [40] Alan Jeffrey, "Advanced Engineering Mathematics," *Harcourt/Academic Press*, 2002.
- [41] M. Tohyama and T. Koike, "Fundamentals of Acoustic Signal Processing," *San Diego:Academic Press*, 1998.
- [42] Duttweiler, D. L., "A twelve-channel digital echo canceler," *IEEE Trans. on Communications*, vol. COMM-26, pp. 647-653, May 1978.
- [43] S. Haykin, "Adaptive Filter Theory. Upper Saddle River:" *Prentice-Hall*, 4th ed., 2002.
- [44] C. Breining, P. Dreiseitel, E. Hansler, A. Mader, B. Nitsch, H. Puder, T. Schertler, G. Schmidt, and J. Tilp., Acoustic echo control, *IEEE Signal Processing Magazine*, 16(4):42-69, July 1999.

Appendix

DIGAMMA FUNCTION

The function $\psi(x)$, called the **psi function** or the **digamma function**, is defined as [40]

$$\psi(x) = \frac{d}{dx} (\ln \Gamma(x)) \dots \dots \dots 1$$

Where

$$\Gamma(x) = \int_0^\infty e^{-t} t^{x-1} dt \text{ for } x > 0 \dots \dots \dots 2$$

the $\Gamma(x)$ function also satisfies the recursion

$$\Gamma(x + 1) = x\Gamma(x) \text{ for } x > 0 \dots \dots \dots 3$$

and that when x is a positive integer n the gamma function reduces to

$$\Gamma(n + 1) = n! \dots \dots \dots 4$$

Thus, for any real $x > 0$, the function $\Gamma(x)$ interpolates continuously between successive values of $n!$, and so generalizes the factorial function to nonintegral values of n . For obvious reasons the gamma function is sometimes called the **factorial function**.

In probabilistic and stochastic model [39] the gamma function is used to represent the probability distribution of a random variable X .

The continuous random variable X has a gamma distribution, with parameters α and β , if its density function is given by

$$f(x; \alpha, \beta) = \begin{cases} \frac{1}{\beta\Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}}, & x > 0 \\ 0, & \text{otherwise} \end{cases} \dots \dots \dots 5$$

where $\alpha > 0$ and $\beta > 0$.