

ACOUSTIC ECHO CANCELLATION USING A VARIABLE STEP SIZE AFFINE PROJECTION ALGORITHM

A DISSERTATION

*Submitted in partial fulfillment of the
requirements for the award of the degree*

of

MASTER OF TECHNOLOGY

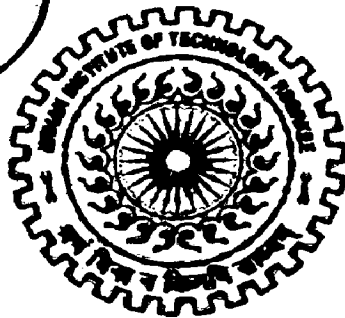
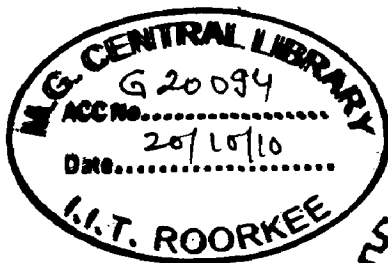
in

ELECTRONICS AND COMMUNICATION ENGINEERING

(With Specialization in Communication Systems)

By

SHARAN CORREA



DEPARTMENT OF ELECTRONICS AND COMPUTER ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE -247 667 (INDIA)
JUNE, 2010

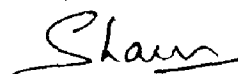
CANDIDATE'S DECLARATION

I hereby declare that the work presented in this dissertation report entitled, “**Acoustic Echo Cancellation Using a Variable Step Size Affine Projection Algorithm**” towards the partial fulfillment of the requirements for the award of degree of **Master of Technology in Electronics and Communication Engineering** with specialization in **Communication Systems**, submitted in the Department of Electronics and Computer Engineering, Indian Institute of Technology Roorkee, is an authentic record of my own work carried out during the period from July 2009 to June 2010, under the guidance of **Dr. S. K. Varma, Professor, Department of Electronics and Computer Engineering, Indian Institute of Technology Roorkee.**

The content of this dissertation has not been previously submitted for examination as part of any academic qualifications.

Date: 21-06-2010

Place: Roorkee



SHARAN CORREA

CERTIFICATE

This is to certify that the above statement made by the candidate is correct to the best of my knowledge and belief.

Date: 21-06-2010

Place: Roorkee



Dr. S. K. VARMA

Professor, E&CE Department

Indian Institute of Technology Roorkee

Roorkee-247667(India)

ACKNOWLEDGEMENTS

With great sense of pleasure and privilege, I take this opportunity to express my deepest sense of gratitude towards my supervisor and guide; **Dr. S. K. Varma** for his valuable suggestions, sagacious guidance, scholarly advice and insightful comments and constructive suggestions to improve the quality of the present work. His professionalism, accurate advice, suggestions and his ways of thinking inspired me, and this inspiration guides me in every moment of my life.

I would like to thank **Kevin D'souza** and **Kevin Saldana** , without whom I would have never made into this institution. I would like to thank **Atul, Parag, Rajesh** and **Shetty** for their company in all the night-outs, trips, stress-out parties, without whom it was impossible to spend two years in Roorkee.

I would also like to thank the Lab staff of Signal Processing Lab for their valuable support in completing my work.

Most of all I would like to thank my parents **Joseph** and **Carmine**, my sister, **Sharal** for their support throughout the numerous ups and downs that I have experienced. Finally, I would like to extend my gratitude to all those persons who directly or indirectly contributed towards this work.

Abstract

The rapid growth of technology in recent decades has changed the whole dimension of communications. Today people are more interested in hands-free communication. In such a situation, the use of a regular loudspeaker and a high-gain microphone, in place of a telephone receiver, might seem more appropriate. This would allow more than one person to participate in a conversation at the same time. Another advantage is that it would allow the person to have both hands free and to move freely in the room. However, the presence of a large acoustic coupling between the loudspeaker and microphone would produce a loud echo that would make conversation difficult. Furthermore, the acoustic system could become unstable, which would produce a loud howling sound to occur.

The solution to these problems is the elimination of the echo with an echo suppression or echo cancellation algorithm. Echo suppressor offers a simple but effective method to counter the echo problem. However, it possesses a main disadvantage as it supports only half-duplex communication. Half-duplex communication permits either calling or called person to speak at a time. This drawback lead to the development of echo cancellers. An important aspect of echo cancellers is that full-duplex communication can be maintained, which allows both speakers to talk at the same time.

Acoustic echo cancellation (AEC) provides one of the best solutions to control the acoustic echoes generated by the hands-free audio terminals. In this type of application, an adaptive filter provides an estimate of the echo. The primary requirement of an adaptive filter algorithm is to provide a high convergence rate and low misalignment. The performance of an adaptive filter algorithm is mainly dependent on the step-size parameter. When a fixed step-size is used, trade off has to be maintained on either high convergence rate or low misalignment, while both cannot be achieved simultaneously. To solve this problem a variable step-size parameter can be used, which keeps updating according to the system statistics.

This work mainly involves the detailed study of a variable-step size affine projection algorithm and its application to AEC. The performance of this algorithm is evaluated under various scenarios such as single-talk and double-talk.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iii
ABSTRACT	iv
LIST OF FIGURES	vii
LIST OF TABLES	x
LIST OF SYMBOLS/ABBREVIATIONS	xi
CHAPTER 1: INTRODUCTION.....	1
1.1 Acoustic Echo Cancellation	2
1.2 Problem Statement	3
1.3 Organization of the Report	3
CHAPTER 2: ECHO CANCELLATION	4
2.1 Hybrid/Electrical Echo	4
2.2 Acoustic Echo	5
2.3 Long Distance Connections Between Fixed Telephones	6
2.3.1 Echo Suppressors	7
2.4 Teleconference/Videoconference Communication Systems	8
2.5 Basic Echo Canceller	9
CHAPTER 3: AFFINE PROJECTION ALGORITHM.....	11
3.1 Affine Space	11
3.1.1 Affine Subspace	12
3.2 Solution for the Variable Convergence Speed of LMS and NLMS	12
3.3 Affine Projection Filter	15
3.3.1 Stability Analysis of Affine Projection Filter	17
CHAPTER 4: VARIABLE STEP SIZE AFFINE PROJECTION ALGORITHM	20
4.1 Single Talk Scenario	24

4.2 Double Talk Scenario	25
4.3 Under Modeling Scenario	25
4.4 Variable Regularized Affine Projection Algorithm.....	29
4.5 Double Talk Detectors	29
4.5.1 The Giegel Algorithm.....	30
4.5.2 Cross Correlation Method.....	31
4.5.3 Normalised Cross Correlation Method.....	31
CHAPTER 5: SIMULATION AND RESULTS	33
5.1 Exact Modeling Case	35
5.2 Under Modeling Case.....	45
CHAPTER 6: CONCLUSION	49
6.2 Future Scope.....	49
REFERENCES.....	50

LIST OF FIGURES

Figure No.	Title	Page No.
2.1	Hybrid echo	4
2.2	Sources of acoustic echo in a room	6
2.3	Simplified long distance connections	6
2.4	Echo suppressor at near-end talker B path	8
2.5	Acoustic echo cancellations in an enclosed environment	9
2.6	A basic echo canceller	10
3.1	Picture of an affine space	11
3.2	Block diagram of adaptive transversal filter	13
3.3(a)	Geometrical interpretation of the normalized LMS algorithm	14
3.3(b)	Geometrical interpretation of the affine projection filter	14
3.4	Multiple linear regression model	18
4.1	AEC configuration	20
5.1(a)	Typical room acoustic impulse response	34
5.1(b)	Far-end speech signal	34
5.1(c)	Near-end speech signal	35
5.2	APA with three different step sizes, and VSS-APA; input is an AR(1) process, SNR=40dB	36
5.3	APA with three different step sizes, and VSS-APA; input is an AR(1) process, SNR=20dB [18]	37

5.4	APA with three different step sizes, and VSS-APA; input is an AR(1) process, SNR=20dB	37
5.5	APA with three different step sizes, and VSS-APA; input is an AR(1) process, SNR=10dB [18]	38
5.6	APA with three different step sizes, and VSS-APA; input is a speech signal, SNR=20dB	38
5.7	Misalignment of APA with three different projection orders, $p=2, 4,$ and $8,$ input is an AR(1) process, SNR=40 dB	40
5.8	Misalignment of VSS-APA with four different projection orders, $p=1, 2,$ $4,$ and $8,$ input signal is speech, SNR=40 dB	40
5.9	Misalignment of VSS-APA with four different projection orders, $p=1, 2,$ $4,$ and $8,$ $L=N=512,$ SNR=20 dB [18]	41
5.10	Misalignments of APA with $\mu=0.2,$ VR-APA with $\zeta=1,$ and VSS-APA, the echo path changes at time 10 sec, SNR=40dB	41
5.11	Misalignments of APA with $\mu=0.2,$ VR-APA with $\zeta=1,$ and VSS-APA, the echo path changes at time 21 sec, $L=N=512,$ SNR=20dB [18]	42
5.12	Misalignments of APA with $\mu=0.2,$ VR-APA with $\zeta=1,$ and VSS-APA, the background noise variation at time 10 sec (SNR decreases from 40dB to 20dB)	43
5.13	Misalignments of APA with $\mu=0.2,$ VR-APA with $\zeta=1,$ and VSS-APA, the background noise variation at time 14 sec, for a period of 14sec (SNR decreases from 20dB to 10dB), $L=N=512$ [18]	43
5.14	14 Misalignments of APA with $\mu=0.2,$ VR-APA with $\zeta=1,$ and VSS-APA, double-talk scenario, without DTD, SNR=40dB	44
5.15	Misalignments of APA with $\mu=0.2,$ VR-APA with $\zeta=1,$ and VSS-APA, double-talk scenario, with Giegel DTD, SNR=40dB	45

5.16	Misalignments of APA with $\mu=0.2$, VR-APA with $\zeta=1$, and VSS-APA, single-talk scenario, $L=500$, $N=1000$, SNR=40dB	46
5.17	Misalignments of APA with $\mu=0.2$, VR-APA with $\zeta=1$, and VSS-APA, single-talk scenario, $L=512$, $N=1024$, SNR=20dB [18]	46
5.18	Misalignments of APA with $\mu=0.2$, VR-APA with $\zeta=1$, and VSS-APA, single-talk scenario, the echo path changes at time 10 sec, $L=500$, $N=1000$, SNR=40dB	47
5.19	Misalignments of APA with $\mu=0.2$, VR-APA with $\zeta=1$, and VSS-APA, single-talk scenario, noise level changes from 40 dB to 20dB, $L=500$, $N=1000$, SNR=40dB	48

LIST OF TABLES

Table 3.1	Summary of Affine Projection Adaptive Filter	18
Table 4.1	Summary of VSS-APA Algorithm	28
Table 5.1	Regularization Factors for APA and VSS-APA	39

LIST OF SYMBOLS/ABBREVIATIONS

d	Desired response
\mathbf{d}	Desired response vector
e	(a priori) Estimation error
\mathbf{e}	(a priori) Estimation error vector
h	Acoustic impulse response
$\hat{\mathbf{h}}$	Tap weight vector
u	Near-end speech
v	Near-end signal
\mathbf{v}	Near-end signal vector
x	Filter input
\mathbf{x}	Input vector
y	Echo signal
\hat{y}	Estimate of echo
AEC	Acoustic Echo Cancellation
APA	Affine Projection Algorithm
DTD	Double Talk Detector
LMS	Least Mean Square
NLMS	Normalised Least Mean Square
VR-APA	Variable Regularized Affine Projection Algorithm
VSS-APA	Variable Step Size Affine Projection Algorithm
VSS-NLMS	Variable Step Size Normalised Least Mean Square

Introduction

In this new age of global communications, wireless phone is regarded as essential communication tool and have a direct impact on people's day-to-day personal and business communications. As new network infrastructures are implemented and competition between wireless carriers increases, digital wireless subscribers are becoming ever more critical of the service and voice quality they receive from network providers. Subscriber demand for enhanced voice quality over wireless networks has driven a new and key technology termed echo cancellation, which can provide near wire line voice quality across a wireless network.

Today's subscribers use speech quality as a standard for assessing the overall quality of a network. Regardless of whether or not the subscriber's opinion is subjective, it is the key to maintain subscriber loyalty. For this reason, the effective removal of hybrid and acoustic echoes, which are inherent within the telecommunications network infrastructure, is the key to maintain and improve the perceived voice quality of a call. Ultimately, the search for improved voice quality has led to intensive research into the area of echo cancellation. By employing echo cancellation technology, the quality of speech can be improved significantly.

Echo is a phenomenon where a delayed and distorted version of an original sound or electrical signal is reflected back to the source. With rare exceptions, conversations take place in the presence of echoes. Echoes of our speech are heard as they are reflected from the floor, walls and other neighbouring objects. If a reflected wave arrives after a very short time of direct sound, it is considered as a spectral distortion or reverberation. However, when the leading edge of the reflected wave arrives a few tens of milliseconds after the direct sound, it is heard as a distinct echo [1].

In telecommunications networks there are two types of echo. One source for an echo is electrical and the other is acoustic [1]. The electrical echo is due to the impedance mismatch at the hybrids of a Public Switched Telephony Network (PSTN) exchange where the

subscriber two-wire lines are connected to four-wire lines. If a communication is simply between two fixed telephones, then only the electrical echo occurs. However, the development of hands-free teleconferencing systems gave rise to another kind of echo known as an acoustic echo. The acoustic echo is due to the coupling between the loudspeaker and microphone.

1.1 Acoustic Echo Cancellation

Acoustic echo cancellation (AEC) provides one of the best solutions to the control of acoustic echoes generated by hands-free audio terminals [2]-[4], which uses an “adaptive filter” for echo cancellation. An adaptive filter is a filter that learns and adjusts its coefficients (transfer function) according to an optimization criterion of an adaptive algorithm. In echo cancellation an adaptive filter learns the acoustic echo path between the terminals loudspeaker and microphone, i.e., the room impulse response. The output of the filter which provides the replica of the acoustic echo, is subtracted from the microphone signal to cancel the echo.

An adaptive filter encounters a number of challenging problems when used in real world AEC applications [4]. First, the echo path is extremely long (of the order of hundreds of milliseconds) and it may rapidly change at time during the connection. As a result, the adaptive filter works in an under-modeling situation, i.e., its length is smaller than the length of the acoustic impulse response. Hence, the residual echo caused by the part of the system that cannot be modelled acts like an additional noise and disturbs the overall performance. Second, the background noise which corrupts the microphone signal is nonstationary and could be time variant. Another important issue in AEC, is the presence of double-talk, i.e., talkers on both sides speak simultaneously. The double-talk can cause the adaptive filter to diverge, hence the presence of double-talk detector (DTD) becomes necessary [4]. The performance of an adaptive filter also depends on the properties of the input signal [5], since the input speech signal is nonstationary and highly correlated.

The problems addressed previously implies the need of some special requirements for the adaptive algorithms used for AEC, which are high convergence rate and good tracking capabilities but achieving low misalignment. Also, the algorithm should be robust against the microphone signal variations and double-talk. Finally, its computational complexity must be moderate.

Among the adaptive algorithms, least mean square (LMS) and normalised least mean square (NLMS) are the most popular algorithm for their simplicity and stability. However, when the input signal is highly correlated speech and a long-length adaptive filter is needed, the convergence speed of the LMS adaptive filter can deteriorate seriously [6]. To overcome this problem, the affine projection algorithm (APA) was proposed [7]. The improved performance of the APA is characterized by an updating-projection scheme of an adaptive filter on a p -dimensional data related subspace. The main drawback of APA is, its computational complexity, which can be overcome by using the 'fast' implementation of the affine projection filter [8],[9]. The stability and convergence speed of the NLMS and classical APA depends on the step-size parameter. The choice of this parameter reflects the trade-off between fast convergence on one hand and poor steady state misalignment on the other. These conflicting requirements can be achieved by using a variable step-size instead a fixed step-size and several variable step-size NLMS (VSS-NLMS) and variable step-size APA (VSS-APA) were developed [10]-[13].

1.2 Problem Statement

The objective of this work is to cancel the acoustic echo by means of an adaptive filter. An adaptive filter algorithm with fixed step-size has to compromise between high convergence rate and low misalignment. This work is concerned with the application of a Variable Step-Size Affine Projection Algorithm (VSS-APA) to cancel out the acoustic echo in an AEC application and comparison of its performance with fixed step-size APA, considering various real world scenarios like single-talk, double-talk and under-modeling conditions.

1.3 Organization of the Report

Chapter 2 covers the different types of echoes in telecommunication systems.

Chapter 3 discusses the classical affine projection algorithm and its advantages over the LMS and NLMS algorithms.

Chapter 4 presents the detailed study of a variable step-size parameter for APA and discusses various double-talk detectors.

Chapter 5 mainly concerned with the presentation and discussion of the results obtained through simulation.

Chapter 6 lists conclusions from the present work and outlines scope for further work in this area.

Echo Cancellation

This chapter deals with various types of echoes that are generated in telecommunication systems. As discussed in chapter one, there are two main types of echo, which are termed electrical, or hybrid, and acoustic. The structure of the basic echo canceller, used to eliminate the echo is also discussed in this chapter.

2.1 Hybrid/Electrical Echo

Hybrid echoes have been inherent within the telecommunications networks since the advent of the telephone. This echo is the result of impedance mismatches in the analog local loop. For example, this happens when mixed gauges of wires are used, or where there are unused taps and loading coils. In the Public Switched Telephone Network, (PSTN), by far the main source of electrical echo is the hybrid. This hybrid is a transformer located at a junction that connects the two-wire local loop coming from a subscriber's premise to the four-wire trunk at the local telephone exchange. The four-wire trunks connect the local exchange to the long distance exchange. This situation is illustrated in Figure 2.1.

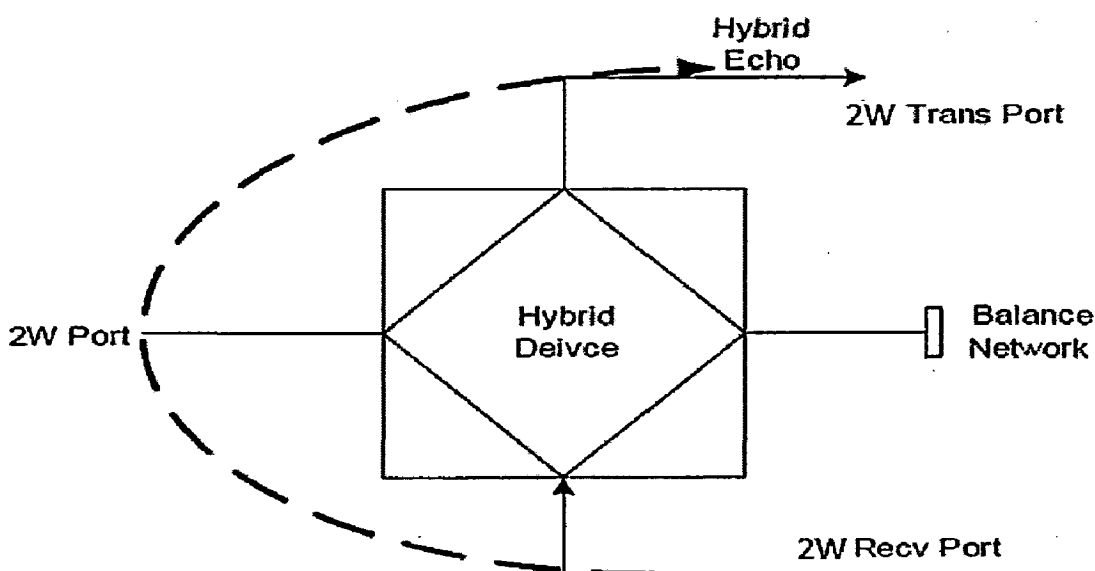


Figure 2.1 Hybrid echo.

The hybrid splits the two-wire local loop into two separate pairs of wires. One pair is used for the transmission path and the other for the receiver path. The hybrid passes on most of the signal. However, the impedance mismatch between the two-wire loop and the four-wire facility causes a small part of the received signal to “leak” back onto the transmission path. The speaker hears an echo because the far-end receives the signal and sends part of it back again. Electrical echo is definitely not a problem on local calls since the relatively short distances do not produce significant delays. However, the electrical echo must be controlled on long distance calls [14].

In the early years, when the public network was entirely circuit switched, the hybrid echo was the only significant source of echo. Since the locations of hybrids and most other causes of impedance differences in circuit switched networks were known, adequate echo control could be planned and provisioned. However, in today’s digital networks the points where two wires split into four wires is typically also the point where analog to digital conversion takes place. Regardless of whether the hybrid and analog to digital conversion is implemented in the same device or in two devices, the two to four wire conversions constitute an impedance mismatch and echoes are produced.

2.2 Acoustic Echo

The acoustic echo, which is also known as a “multipath echo”, is produced by poor voice coupling between the earpiece and microphone in handsets and hands-free devices. Further voice degradation is caused as voice-compressing and encoding/decoding devices process the voice paths within the handsets and in wireless networks. This results in returned echo signals with highly variable properties. When compounded with inherent digital transmission delays, call quality is greatly diminished for the wire-line caller.

Acoustic coupling is due to the reflection of the loudspeaker’s sound waves from walls, door, ceiling, windows and other objects back to the microphone. The result of the reflections is the creation of a multipath echo and multiple harmonics of echoes, which are transmitted back to the far-end and are heard by the talker as an echo unless eliminated. Adaptive cancellation of such acoustic echoes has become very important in hands-free communication systems such as teleconference or videoconference systems [14]. The multipath echo phenomenon is illustrated in Figure 2.2.

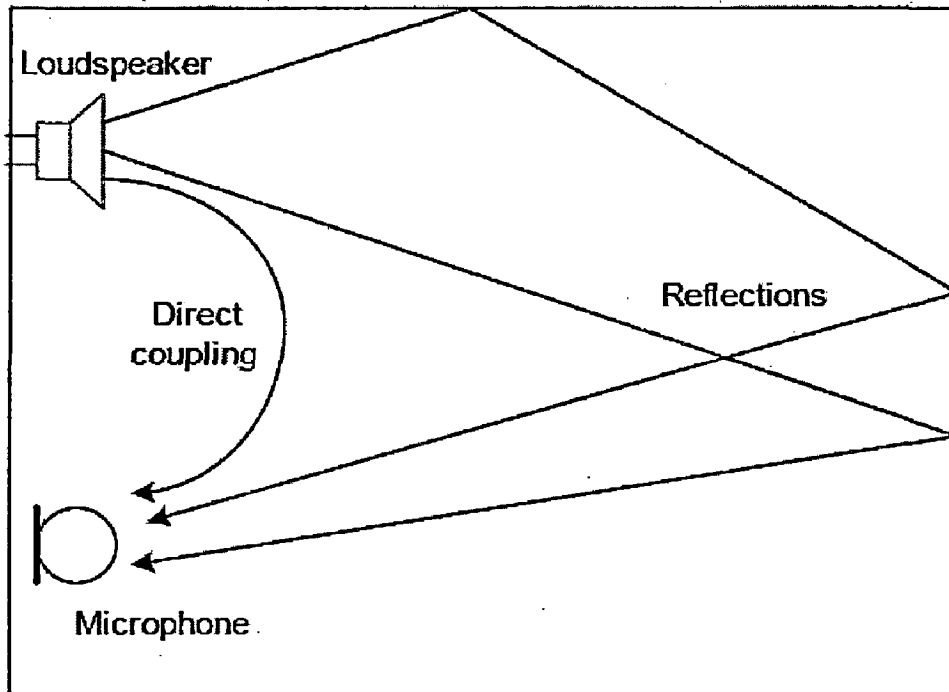


Figure 2.2 Sources of acoustic echo in a room.

In the following sections, the echo phenomena of two communication systems will be described. The communication systems are:

- Long-distance connections between fixed telephones
- Teleconference/videoconference systems

2.3 Long-distance connections between fixed telephones

A simple long-distance telephone connection is presented in Figure 2.3. This connection contains two-wire sections at the ends, the subscriber loops and possibly some portion of the local network. It also contains a four-wire section in the center, which is a carrier system for medium-range to long-range transmissions.

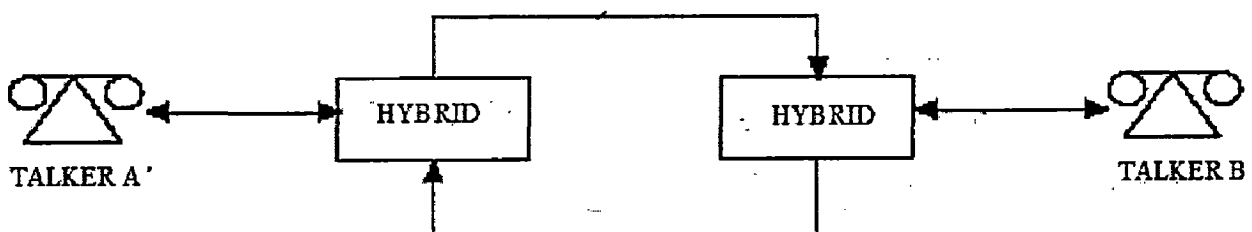


Figure 2.3 Simplified long distance connections.

Every conventional telephone in a given geographical area is connected to the local PSTN exchange by a two-wire line, called the subscriber loop, which carries a connection for both directions of transmission. Simply connecting the two subscriber loops at the local exchange sets up a local call. However, amplification of the speech signal becomes necessary when the distance between the two telephones exceeds about 56 km [14]. Therefore, a four-wire line is required, which segregates the two directions of transmission. A hybrid is used to convert from the two-wire to four-wire line and vice versa.

An echo can be decreased if the hybrid is perfectly balanced by impedance located at its four-wire portion. Unfortunately, this is not possible in practice since it requires knowledge of the two-wire impedance, which varies considerably over the population of subscriber loops. When the bridge is not perfectly balanced, impedance mismatch occurs. This causes some of the talker's signal energy to be reflected back as an echo. Adding an insertion loss to the four-wire portions of the connection can control the effects of echo. Such action is effective since the echo signals experience this loss two or three times while the talker's speech suffers this loss only once. However, on long-range connections the insertion loss can become very significant. Hence, it is not a favourable solution and other echo control techniques such as echo suppression must be used.

2.3.1 Echo Suppressors

Echo suppressors have been used since the introduction of long distance communication. This device basically takes advantage of the fact that people rarely talk simultaneously. The situation of two people talking simultaneously is termed "double-talk". The echo suppressor is also helped by the fact that during such double-talk poor transmission quality is less noticeable. Figure 2.4 illustrates how the echo suppressor dynamically controls the connection based on who is talking, which is decided by the speech and double talking detector. Double talking is detected if the level of the signal in path L1 is significantly lower than that in path L2. When the far-end talker A is speaking, the path used to transmit the near-end speech is opened so that the echo is prevented. Then, when the near-end talker B speaks, the same switch is closed and a symmetric one at the far-end talker A's path is opened. However, echo suppressors can clip speech sounds and introduce impairing interruption.

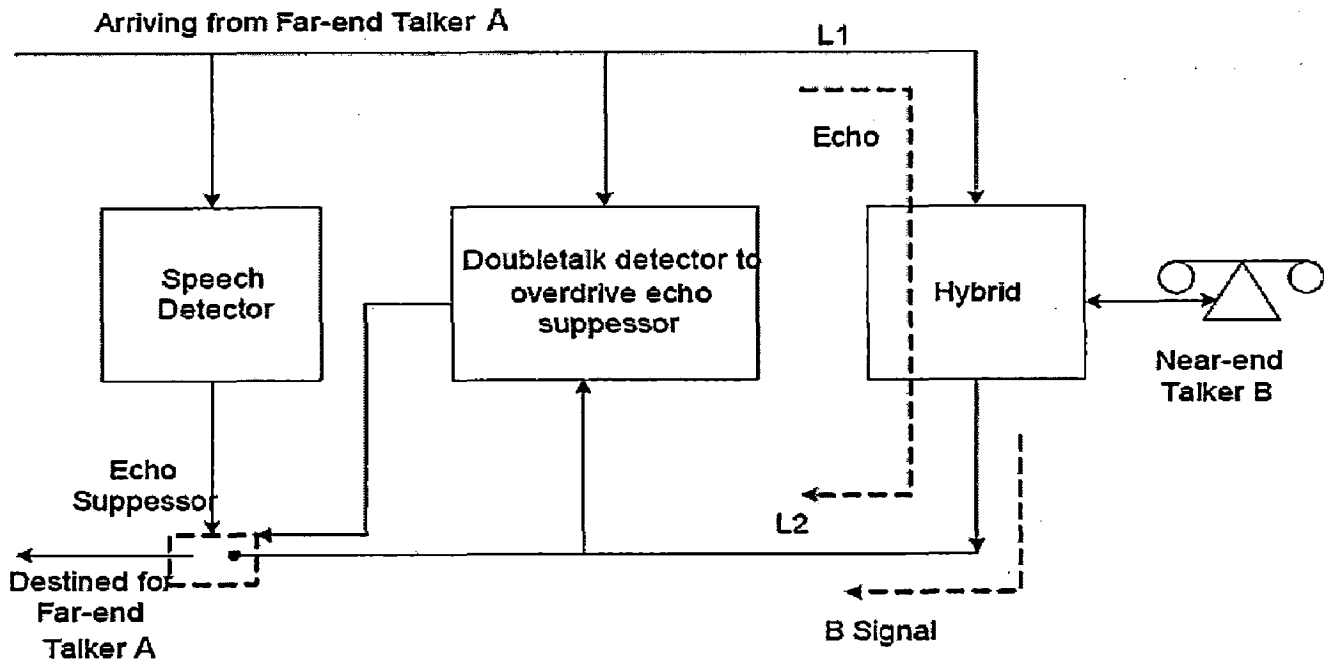


Figure 2.4 Echo suppressor at near-end talker B path.

For example, if talker B is initially listening to talker A but suddenly wants to talk, it is quite likely that the switch preventing talker A's echo from being transmitted will not close quickly enough. This will cause the far-end talker A to not be able to receive all the messages from the near-end talker B. This deletion is noticed by talker A, encouraging him/her to stop and wait for talker B to finish. The resulting confusion may stop the conversation entirely while each party waits for the other to say something [14]. Thus, an echo suppressor supports only half-duplex communication, which permits only one speaker to talk at a time. Therefore the best solution for removing echoes is to use echo cancellers, which provides a full-duplex communication.

2.4 Teleconference/Videoconference Communication Systems

When the telephone connection is between hands-free telephones or between two conference rooms, then an acoustic echo problem emerges that is due to the reflection of the loudspeaker's sound waves from the boundary surfaces and other objects back to the microphone. This acoustic echo can be removed using an adaptive filter as illustrated in Figure 2.5. The adaptive filter attempts to synthesize a model of the acoustic echo at its output.

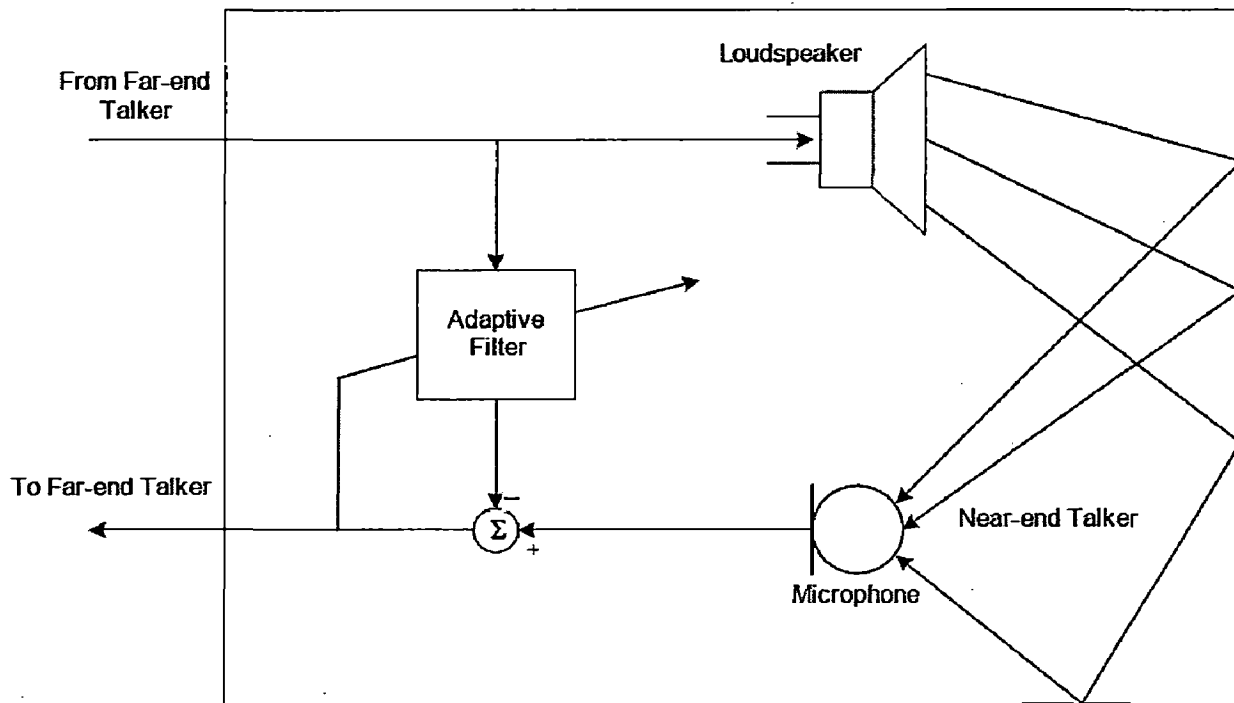


Figure 2.5 Acoustic echo cancellations in an enclosed environment.

Acoustic echo cancellation is a more challenging problem than the network echo cancellation for the following main reasons:

- The impulse response of the acoustic echo path is several times (usually between 100 to 500 msec.) longer than that of the network echo path.
- The characteristics of the acoustic echo path are non-stationary, as it may change with ambient temperature, pressure and also with the movement of objects and human bodies, while the network echo path is almost stationary.
- The acoustic echo path has a mixture of linear and nonlinear characteristics. The reflection of acoustic signals inside a room is almost linearly distorted. However, the contribution towards the nonlinearity is done by the loudspeaker.

2.5 Basic Echo Canceller

A basic echo canceller used to remove echo in telecommunication networks is presented in Figure 2.6. The echo canceller mimics the transfer function of the echo path in order to synthesize a replica of the echo. Then the echo canceller subtracts the synthesized replica from the combined echo and near-end speech or disturbance signal to obtain the near-end signal. However, the transfer function (echo path) is unknown in practice.

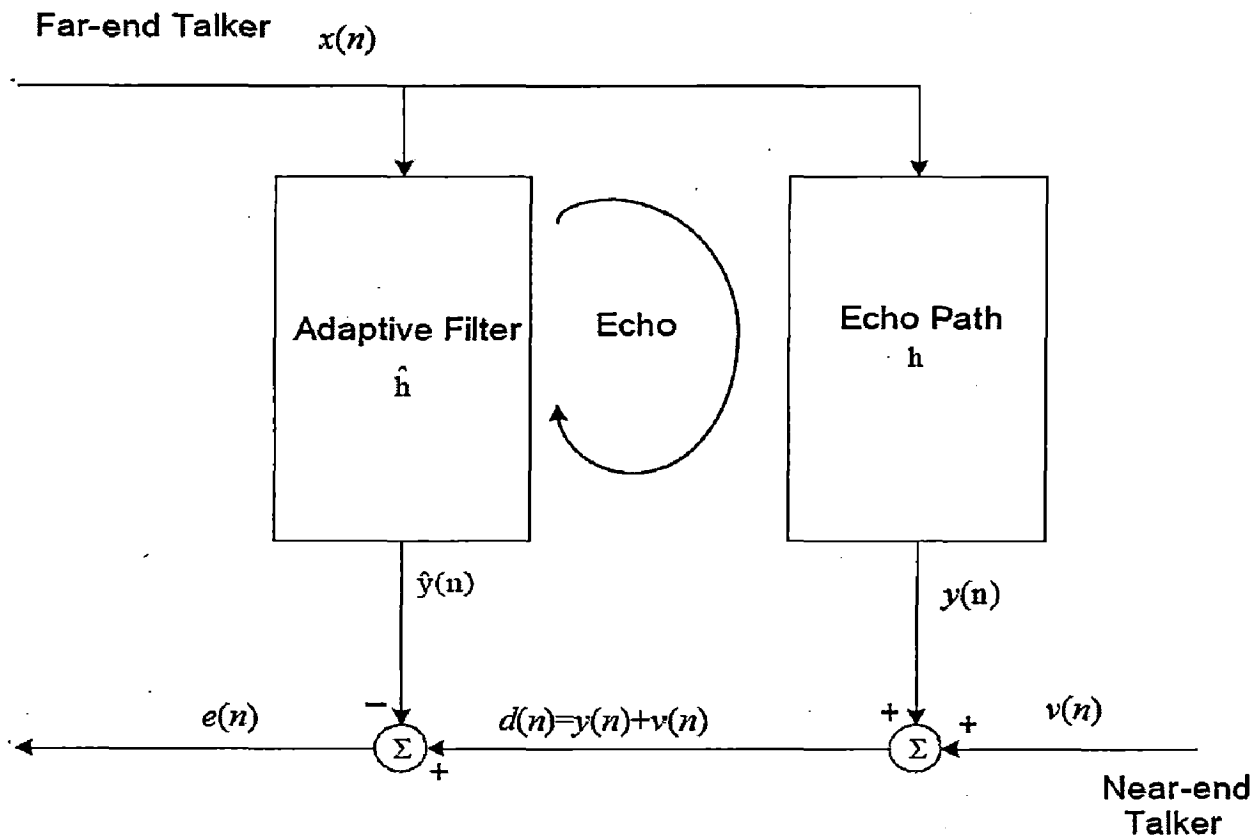


Figure 2.6 A basic echo canceller.

Therefore, it must be identified. This problem can be solved by using an adaptive filter that gradually matches its estimated impulse response, \hat{h} , to that of the impulse response of the actual echo path, h . This process is illustrated in Figure 2.6. The estimated echo, $\hat{y}(n)$, is generated by passing the reference input signal, $x(n)$, through the adaptive filter, $\hat{h}(n)$, that will ideally match the transfer function of the echo path, h . The echo signal, $y(n)$, is produced when $x(n)$ passes through the echo path. The echo $y(n)$ plus the near-end talker and/or disturbance signal, $v(n)$, constitute the desired response,

$$d(n) = y(n) + v(n) \quad (2.1)$$

for the adaptive canceller. The two signals $x(n)$ and $y(n)$ are correlated since the later is obtained by passing $x(n)$ through the echo path. The error signal $e(n)$ is given by

$$e(n) = d(n) - \hat{y}(n) \quad (2.2)$$

In the ideal case, $e(n) = v(n)$, i.e., when the adaptive filter produces the exact replica of the echo, which represents the case when the adaptive echo canceller is perfect.

Affine Projection Algorithm

The main motivation behind the discovery of affine projection algorithm was to provide an improvement in the rate of convergence over the normalised LMS algorithm. This chapter provides a detailed analysis of affine projection filter and its advantages over LMS and NLMS filters and also includes an introduction to the affine space.

3.1 Affine Space

Affine space is a generalization of Cartesian or Euclidian space. In an affine space, points can be subtracted to get vectors or a vector can be added to a point to get another point, but the addition of points is not allowed, as there is no distinguished point that serves as an origin. A coordinate system for the n -dimensional affine space R^n is determined by any basis of n vectors, which are not necessarily orthonormal. Therefore, the resulting axes are not necessarily mutually perpendicular nor have the same unit measure. In this sense, affine is a generalization of Cartesian or Euclidian space [15]. An affine space can most easily be defined in terms of a vector space over a field R^3 , by taking a set of points and a set of vectors over R^3 [16], as shown in Figure 3.1 below,

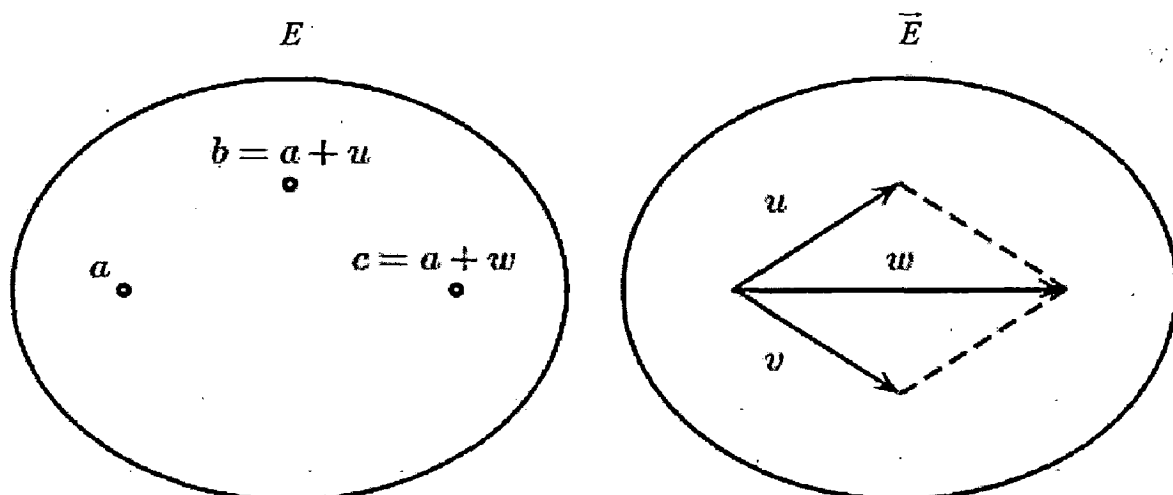


Figure 3.1 Picture of an affine space

where E is a set of points (with no structure) and \bar{E} is a vector space (of free vectors) acting on the set E . The vectors from \bar{E} act on the points in E , causing them to shift to a new location (point), i.e., the application of every vector $u \in \bar{E}$, is to move every point $a \in E$ to the point $a+u \in E$.

An affine space, is defined as a set $\langle E, \bar{E}, + \rangle$ consisting of a nonempty set E (of points), a vector space \bar{E} (of translations, or free vectors), and an action $+: E \times \bar{E} \rightarrow E$, satisfying the following conditions [16],

1. $a + 0 = a$, for every $a \in E$.
2. $(a + u) + v = a + (u + v)$, for every $a \in E$, and every $u, v \in \bar{E}$.
3. For any two points $a, b \in E$, there is a unique $u \in \bar{E}$ such that $a + u = b$.

The subtraction of points of an affine space is defined as follows:

The unique vector $u \in \bar{E}$ such that $a + u = b$ is denoted by $b - a$ (also \overline{ab} , or \overline{ab}). Thus $b - a$ is a unique vector in \bar{E} such that $a + (b - a) = b$. The dimension of the affine space $\langle E, \bar{E}, + \rangle$ is the dimension $\dim(\bar{E})$ of the vector space \bar{E} [16].

3.1.1 Affine Subspace

In linear algebra, a subspace can be characterized as a nonempty subset of a vector space closed under linear combinations. Similarly, an affine subspace is characterized as a subset of an affine space closed under affine combinations [16], and is defined as follows,

Given an affine space $\langle E, \bar{E}, + \rangle$, a subset V of E is an affine subspace, if for every family of weighted points $((a_i, \lambda_i))_{i \in I}$ in V such that $\sum_{i \in I} \lambda_i = 1$, the affine combination $\sum_{i \in I} \lambda_i a_i$ belongs to V , where $(a_i)_{i \in I}$ denotes a family of points in E and $(\lambda_i)_{i \in I}$ denotes a family of scalars.

3.2 Solution for the Variable Convergence Speed of LMS and NLMS

The affine projection algorithm is considered as a generalised case of normalised LMS algorithm.

The main reason behind its evolution was to overcome the drawback of NLMS, i.e., dependence of the convergence speed on the properties (autocorrelation function) of input signal.

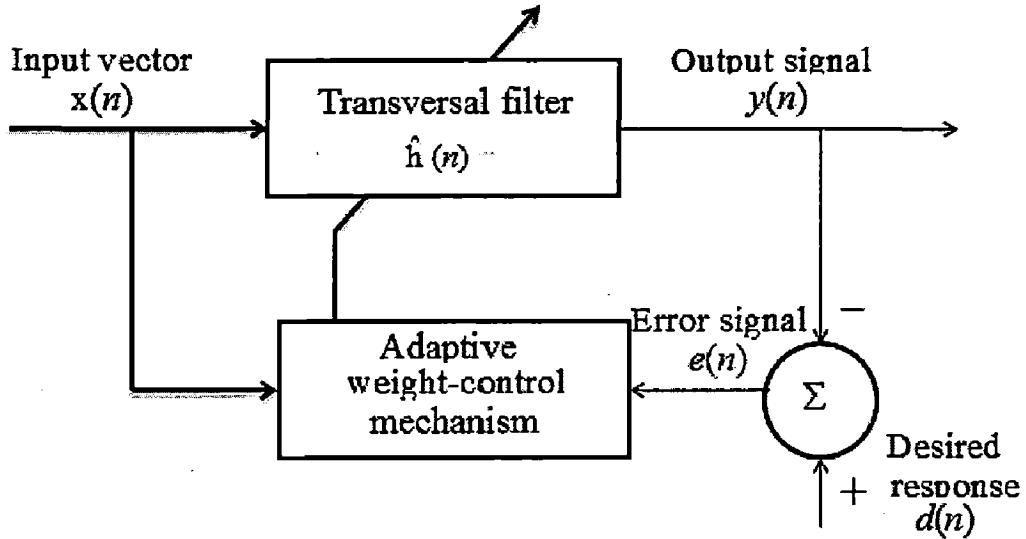


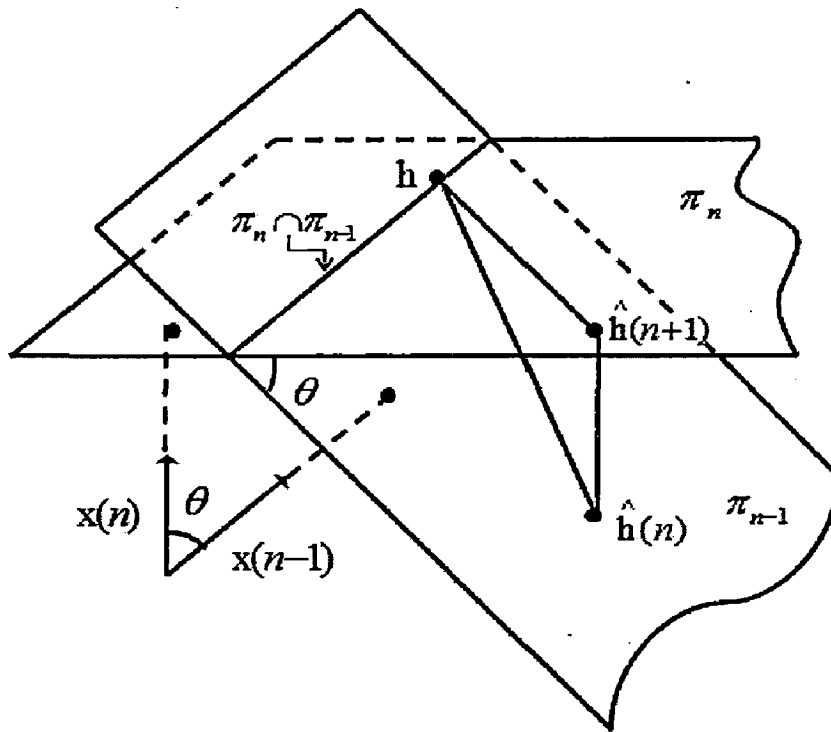
Figure 3.2 Block diagram of adaptive transversal filter.

Following the adaptive filter structure of Figure 3.2, the weight vector adjustment $\delta \hat{\mathbf{h}}(n+1)$ applied to a normalised LMS filter at iteration $n+1$, for a real valued data is given by [5],

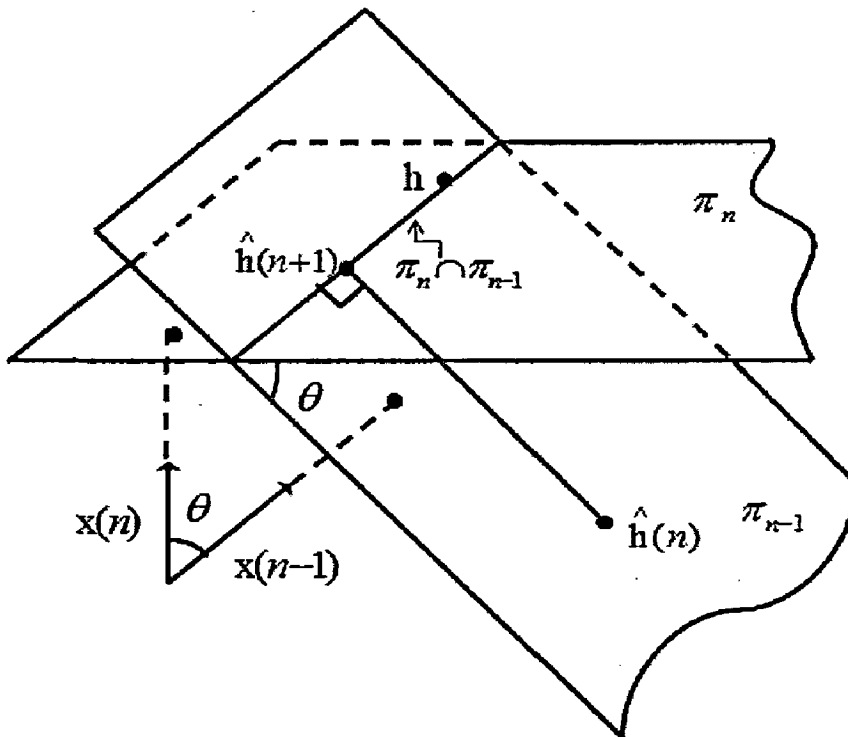
$$\begin{aligned} \delta \hat{\mathbf{h}}(n+1) &= \hat{\mathbf{h}}(n+1) - \hat{\mathbf{h}}(n) \\ &= \frac{\mu}{\|\mathbf{x}(n)\|^2} \mathbf{x}(n)e(n) \end{aligned} \quad (3.1)$$

where μ is the step-size constant.

The problem of variable convergence speed, associated with NLMS algorithm could be explained better considering the geometrical interpretation as shown in Figure 3.3(a) for $\mu=1$ and $L=3$, which consists the elements of two L - dimensional spaces, namely, the input data space and the weight space. Specifically, π_n represents the set of all weight vectors $\hat{\mathbf{h}}(n)$ that act on the input vector $\mathbf{x}(n)$ to produce the output $y(n)$, and similarly for π_{n-1} .



(a)



(b)

Figure 3.3 Geometrical interpretation of (a) the normalized LMS algorithm and (b) the affine projection adaptive algorithm.

The angle θ subtended between the hyperplanes π_n and π_{n-1} is the same as the angle between the input vector $x(n)$ and $x(n-1)$. The cosine of the angle θ between the vectors $x(n)$ and $x(n-1)$ is defined as,

$$\cos \theta = \frac{x^T(n)x(n-1)}{\|x(n)\| \cdot \|x(n-1)\|}. \quad (3.2)$$

The right side of Eq.(3.2) also represents the first- order sample autocorrelation function of the input signal $x(n)$. When angle θ is zero or 180° (i.e., when the input vectors $x(n)$ and $x(n-1)$ point in the same direction or opposite direction), the convergence speed is decreased, as the first- order sample autocorrelation function of $x(n)$ approaches 1 in absolute value. On the other hand, when angle θ is $\pm 90^\circ$ (i.e., when the input vectors $x(n)$ and $x(n-1)$ are orthogonal to each other), the convergence rate is increased.

To maintain the speed of convergence constant, independently of the angle θ between the input vectors $x(n)$ and $x(n-1)$, the affine projection filter is used [7], whose geometric interpretation for the case $\mu=1$ and $L=3$ is shown in Figure 3.3(b). In the same Figure $\pi_n \cap \pi_{n-1}$ denotes the intersection of the hyperplanes π_n and π_{n-1} . As compared to Figure 3.3(a), it can be observed in Figure 3.3(b), that in the weight space, the line joining $\hat{h}(n+1)$ to $\hat{h}(n)$ is normal to $\pi_n \cap \pi_{n-1}$ rather than π_n .

The intersection of π_n and π_{n-1} in Figure 3.3(b), not necessarily contain the origin of the L dimensional weight space. Thus $\pi_n \cap \pi_{n-1}$ is an affine subspace and hence the name affine projection algorithm [7].

3.3 Affine Projection Filter

The design of an affine projection filter is based on minimizing the squared Euclidean norm of the change in tap-weight vector [5],

$$\delta \hat{h}(n+1) = \hat{h}(n+1) - \hat{h}(n) \quad (3.3)$$

subjected to the set of p constraints, where p is known as the order of the APA filter,

$$d(n-k) = \hat{h}^H(n+1) x(n-k), \quad \text{for } k=0, 1, \dots, p-1 \quad (3.4)$$

Following the method of Lagrange multipliers [5], Eqs.(3.3) and (3.4) can be combined to set up the cost function for the affine projection filter as

$$J(n) = \|\hat{h}(n+1) - \hat{h}(n)\|^2 + \sum_{k=0}^{p-1} \text{Re}[\lambda_k^* (d(n-k) - \hat{h}^H(n+1) x(n-k))] \quad (3.5)$$

In the above equation λ_k are the Lagrange multipliers pertaining to the multiple constraints. For the convenience of presentation, the following definitions are introduced:

Let $X(n)$ be the p -by- L data matrix, whose Hermitian transpose is defined by

$$X^H(n) = [x(n), x(n-1), \dots, x(n-p+1)] \quad (3.6)$$

p -by-1 desired response vector, whose Hermitian transpose is defined as,

$$d^H(n) = [d(n), d(n-1), \dots, d(n-p+1)] \quad (3.7)$$

and p -by-1 Lagrange vector, whose Hermitian transpose is defined as,

$$\lambda^H = [\lambda_0, \lambda_1, \dots, \lambda_{p-1}]. \quad (3.8)$$

Using Eqs.(3.6), (3.7), and (3.8) into Eq.(3.5), the cost function can be written in a compact form as

$$J(n) = \|\hat{h}(n+1) - \hat{h}(n)\|^2 + \text{Re} [(d(n) - X(n) \hat{h}(n+1))^H \lambda]$$

Differentiating the cost function $J(n)$ with respect to complex valued weight vector $\hat{h}^*(n+1)$, using the rules of differentiation with respect to complex valued vector [5], results in

$$\frac{\partial J(n)}{\partial \hat{h}^*(n+1)} = 2 (\hat{h}(n+1) - \hat{h}(n)) - X^H(n) \lambda \quad (3.9)$$

Setting this derivative equal to zero and using Eq.(3.3),

$$\delta \hat{h}(n+1) = \frac{1}{2} X^H(n) \lambda \quad (3.10)$$

From Eqs.(3.4), (3.6), and (3.7) the desired response vector can be written as

$$d(n) = X(n) \hat{h}(n+1) \quad (3.11)$$

Premultiplying both sides of Eq.(3.10) by $X(n)$ and then using and then using Eqs.(3.3) and (3.11) to eliminate the update weight vector $\hat{h}(n+1)$, results in

$$\begin{aligned} X(n) \delta \hat{h}(n+1) &= \frac{1}{2} X(n) X^H(n) \lambda \\ \Rightarrow X(n) (\hat{h}(n+1) - \hat{h}(n)) &= \frac{1}{2} X(n) X^H(n) \lambda \end{aligned} \quad (3.12)$$

Using Eqs.(3.11) and (3.12), and rearranging terms, the desired response vector is

$$d(n) = X(n) \hat{h}(n) + \frac{1}{2} X(n) X^H(n) \lambda \quad (3.13)$$

The difference between $d(n)$ and $X(n) \hat{h}(n)$ at iteration n is known as *p-by-1 a priori* error vector, i.e.,

$$e(n) = d(n) - X(n) \hat{h}(n) \quad (3.14)$$

With the help of Eqs.(3.13) and (3.14), the solution for Lagrange vector λ is

$$\lambda = 2(X(n)X(n)^H)^{-1} e(n) \quad (3.15)$$

Substituting the value of λ from Eq.(3.15) into Eq.(3.10), results in the optimum change in the weight vector:

$$\delta \hat{h}(n+1) = X^H(n)(X(n)X^H(n))^{-1} e(n) \quad (3.16)$$

Introducing the step-size parameter μ into Eq.(3.16) for controlling the change in the weight vector from one iteration to the next, results in

$$\delta \hat{h}(n+1) = \mu X^H(n)(X(n)X^H(n))^{-1} e(n) \quad (3.17)$$

Using Eq.(3.3), the desired update equation for affine projection adaptive filter is

$$\hat{h}(n+1) = \hat{h}(n) + \mu X^H(n)(X(n)X^H(n))^{-1} e(n). \quad (3.18)$$

Table 3.1 provides the complete summary of affine projection adaptive filter.

3.3.1 Stability Analysis of the Affine Projection Filter

The physical mechanism responsible for generating the desired response $d(n)$, is governed by the multiple regression model [5] as shown in Figure 3.4.

Table 3.1 Summary of the Affine Projection Adaptive Filter [5]

Parameters : L = number of taps

μ = adaptation constant

p = projection order

Initialization: $\hat{h}(0) = 0$.

Data:

Given: $x(n)$ = L -by-1 tap input vector at time step n

$$= [x(n), x(n-1), \dots, x(n-L+1)]^H$$

$d(n)$ = desired step response at time step n

For $n = 0, 1, 2, \dots$

Compute: $X^H(n) = [x(n), x(n-1), \dots, x(n-p+1)]$

$$d^H(n) = [d(n), d(n-1), \dots, d(n-p+1)]$$

$$e(n) = d(n) - X(n) \hat{h}(n)$$

$$\hat{h}(n+1) = \hat{h}(n) + \mu X^H(n)(X(n)X^H(n))^{-1}e(n).$$

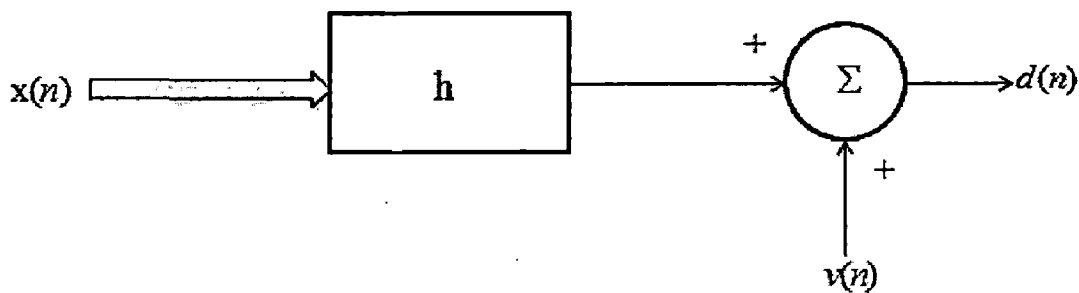


Figure 3.4 Multiple linear regression model.

The multiple linear regression model of Figure 3.4, is described by the formula

$$d(n) = \mathbf{h}^H \mathbf{x}(n) + v(n) \quad (3.19)$$

where \mathbf{h} is the model's unknown parameter vector and $v(n)$ is the additive disturbance. The tap-weight vector $\hat{\mathbf{h}}(n)$ computed by the affine projection filter is an estimate of \mathbf{h} . The mismatch between \mathbf{h} and $\hat{\mathbf{h}}(n)$ is measured by the *weight-error vector*,

$$\boldsymbol{\varepsilon}(n) = \mathbf{h} - \hat{\mathbf{h}}(n) \quad (3.20)$$

Subtracting Eq.(3.18) from \mathbf{h} , results in

$$\mathbf{h} - \hat{\mathbf{h}}(n+1) = \mathbf{h} - (\hat{\mathbf{h}}(n) + \mu \mathbf{X}^H(n)(\mathbf{X}(n)\mathbf{X}^H(n))^{-1} \mathbf{e}(n)) \quad (3.21)$$

Using Eq.(3.20), Eq.(3.21) can be written as,

$$\boldsymbol{\varepsilon}(n+1) = \boldsymbol{\varepsilon}(n) - \mu \mathbf{X}^H(n)(\mathbf{X}(n)\mathbf{X}^H(n))^{-1} \mathbf{e}(n) \quad (3.22)$$

The fundamental idea of an affine projection filter is to minimize the incremental change $\delta \hat{\mathbf{h}}(n+1)$ in tap-weight vector of the filter from iteration n to iteration $n+1$, i.e., to minimize the mean square deviation,

$$D(n) = E [\|\boldsymbol{\varepsilon}(n)\|^2] \quad (3.23)$$

Taking the squared Euclidean norms of both sides of Eq.(3.22), and then taking the expectations, results in,

$$E [\|\boldsymbol{\varepsilon}(n+1)\|^2] = E [\|\boldsymbol{\varepsilon}(n) - \mu \mathbf{X}^H(n)(\mathbf{X}(n)\mathbf{X}^H(n))^{-1} \mathbf{e}(n)\|^2]. \quad (3.24)$$

Rearranging and simplifying terms in Eq.(3.24)

$$\begin{aligned} D(n+1) - D(n) &= \mu^2 E [\mathbf{e}^H(n)(\mathbf{X}(n)\mathbf{X}^H(n))^{-1} \mathbf{e}(n)] \\ &\quad - 2 \mu E [\text{Re} (\boldsymbol{\xi}_x^H(n)(\mathbf{X}(n)\mathbf{X}^H(n))^{-1} \mathbf{e}(n))]. \end{aligned} \quad (3.25)$$

where $\boldsymbol{\xi}_x(n) = \mathbf{X}(n)(\mathbf{h} - \hat{\mathbf{h}}(n))$ is the *undisturbed error vector*.

From Eq.(3.25), it can be observed that the mean-square deviation $D(n)$ decreases monotonically with increasing n , and the affine projection filter is therefore stable in the mean-square error sense, provided that the step-size parameter satisfies the condition

$$0 < \mu < \frac{2E[\text{Re}(\boldsymbol{\xi}_x^H(n)(\mathbf{X}(n)\mathbf{X}^H(n))^{-1} \mathbf{e}(n))]}{E[\mathbf{e}^H(n)(\mathbf{X}(n)\mathbf{X}^H(n))^{-1} \mathbf{e}(n)]} \quad (3.22)$$

Variable Step-Size Affine Projection Algorithm

The variable step-size parameter for APA is derived in this section, considering the general AEC configuration as shown below. The goal of this scheme is to identify an unknown system (i.e., acoustic echo-path) using an adaptive filter.

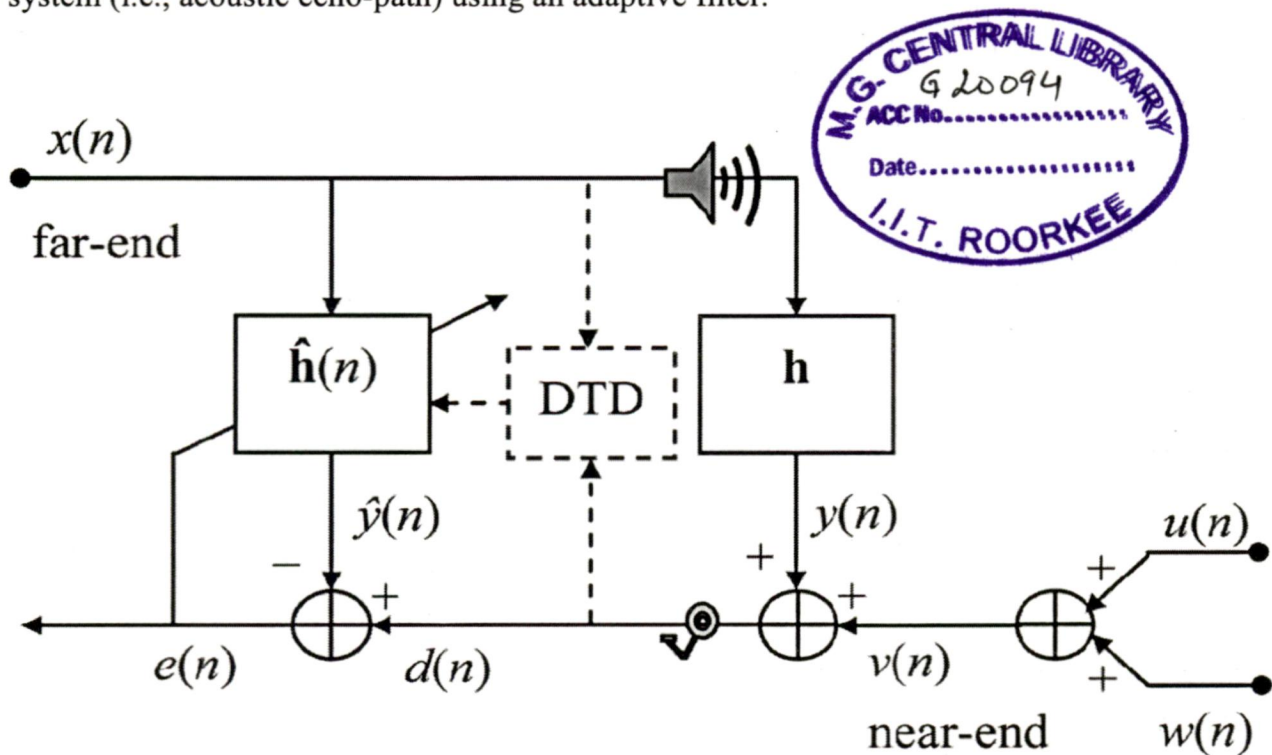


Figure 4.1 AEC configuration.

Both systems in the above configuration have finite impulse responses, defined by the real-valued vectors $\mathbf{h} = [h_0 h_1 \dots h_{N-1}]^T$ and $\hat{\mathbf{h}}(n) = [\hat{h}_0(n) \hat{h}_1(n) \dots \hat{h}_{L-1}(n)]^T$, where superscript T denotes transposition and n is the time index; N is the length of the echo path, while L is the length of the adaptive filter. The signal $x(n)$ is the far-end speech which goes through the acoustic impulse response \mathbf{h} , resulting the echo signal, $y(n)$.

This signal is picked up by the microphone together with the near-end signal $v(n)$, resulting the microphone signal $d(n)$. The near-end signal can contain both the background noise, $w(n)$, and the near-end speech, $u(n)$. The output of the adaptive filter $\hat{y}(n)$ provides a replica of the echo, which will be subtracted from the microphone signal. The DTD block controls the algorithm behaviour during double-talk.

To obtain the variable step-size parameter let us consider the following relations which define the classical APA, assuming the input signal to be real valued

$$e(n) = d(n) - X^T(n) \hat{h}(n-1) \quad (4.1)$$

$$\hat{h}(n) = \hat{h}(n-1) + \mu X(n) [X^T(n)X(n)]^{-1} e(n) \quad (4.2)$$

where $e(n)$ is the error signal, $d(n) = [d(n), d(n-1), \dots, d(n-p+1)]^T$ is the desired signal vector of length p , with p denoting the projection order.

The matrix $X(n) = [x(n), x(n-1), \dots, x(n-p+1)]$ is the input signal matrix,

where $x(n-l) = [x(n-l), x(n-l-1), \dots, x(n-l-L+1)]^T$ (with $l = 0, 1, \dots, p-1$) are the input signal vectors. The constant μ denotes the step-size parameter of the algorithm.

To make μ adaptive (variable), Eq.(4.2) is rewritten in a different form as

$$\hat{h}(n) = \hat{h}(n-1) + X(n) [X^T(n)X(n)]^{-1} \mu(n) e(n) \quad (4.3)$$

where $\mu(n) = \text{diag}\{\mu_0(n), \mu_1(n), \dots, \mu_{p-1}(n)\}$ is a $p \times p$ diagonal matrix.

Using the adaptive filter coefficients at time n , the *a posteriori* error vector can be defined as

$$\varepsilon(n) = d(n) - X^T(n) \hat{h}(n) \quad (4.4)$$

The error vector $e(n)$ from Eq.(4.1) plays the role of *a priori* error vector. Replacing Eq.(4.3) in Eq.(4.4) and taking Eq.(4.1) into account, it results that

$$\varepsilon(n) = [I_p - \mu(n)] e(n) \quad (4.5)$$

where I_p denotes a $p \times p$ identity matrix.

In consistence with the basic idea of the APA, it can be imposed to cancel *p a posteriori* errors, i.e., $\varepsilon(n) = \mathbf{0}_{p \times 1}$, where $\mathbf{0}_{p \times 1}$ denotes a column vector with all its elements equal to zeros. Since practically $\varepsilon(n) \neq \mathbf{0}_{p \times 1}$, it results from Eq.(4.5) that $\mu(n) = \mathbf{I}_p$. This corresponds to the classical APA update of Eq.(4.2), with the step-size $\mu=1$. In the absence of the near-end signal, i.e., $v(n)=0$, the scheme from Figure 4.1 is reduced to an ideal “system identification” configuration. In this case, the value of the step-size $\mu=1$ makes sense, because it leads to the best performance [7].

According to the adaptive filter theory [17], the AEC scheme from Figure 4.1 can be interpreted as a combination between two classes of adaptive system configurations. First, it represents the “system identification” configuration and the goal is to identify an unknown system (i.e., the acoustic echo path) with its output corrupted by an apparently “undesired” signal (i.e., the near-end signal). However, it also can be viewed as an “interference cancelling” configuration, aiming to recover an “useful” signal (i.e., the near-end signal) corrupted by an undesired perturbation (i.e., the acoustic echo); consequently, the “useful” signal should be recovered in the error signal of the adaptive filter. The main aim of the AEC is to remove the undesired echo picked by the microphone along with the near-end speech; the existence of the near-end signal cannot be omitted.

Therefore, a more reasonable condition is $\varepsilon(n) = v(n)$.

where the column vector $\mathbf{v}(n) = [v(n), v(n-1), \dots, v(n-p+1)]^T$ represents the near-end signal vector of length p . Taking Eq.(4.5) into account, it results that

$$\varepsilon_{l+1}(n) = [1 - \mu_l(n)]e_{l+1}(n) = v(n-l) \quad (4.6)$$

where the variables $\varepsilon_{l+1}(n)$ and $e_{l+1}(n)$ denote the $(l+1)$ th elements of the vectors $\varepsilon(n)$ and $e(n)$, with $l=0,1,\dots,p-1$. The goal is to find an expression for the step-size parameter $\mu_l(n)$ such that

$$E\{\varepsilon_{l+1}^2(n)\} = E\{v^2(n-l)\} \quad (4.7)$$

where $E\{\}$ denotes mathematical expectation. Squaring Eq.(4.6) and taking the expectations results in

$$[1 - \mu_l(n)]^2 E\{e_{l+1}^2(n)\} = E\{v^2(n-l)\} \quad (4.8)$$

By solving the quadratic Eq.(4.8), two solutions are obtained i.e.,

$$\mu_l(n) = 1 \pm \sqrt{\frac{E\{v^2(n-l)\}}{E\{e_{l+1}^2(n)\}}} \quad (4.9)$$

Following the analysis presented on convergence behaviour of APA from [18], which states that a value of the step-size between 0 and 1 is preferable over the one between 1 and 2 (even if both solutions are stable, but the former has less steady-state mean square error with the same convergence speed), it is reasonable to choose

$$\mu_l(n) = 1 - \sqrt{\frac{E\{v^2(n-l)\}}{E\{e_{l+1}^2(n)\}}} \quad (4.10)$$

In terms of the power estimates (4.10) can be expressed as

$$\mu_l(n) = 1 - \frac{\hat{\sigma}_v(n-l)}{\hat{\sigma}_{e_{l+1}}(n)} \quad (4.11)$$

The variable in the denominator can be computed in a recursive manner [18], i.e.,

$$\hat{\sigma}_{e_{l+1}}^2(n) = \lambda \hat{\sigma}_{e_{l+1}}^2(n-1) + (1-\lambda)e_{l+1}^2(n) \quad (4.12)$$

where λ is a weighting factor chosen as $\lambda = 1 - 1/(KL)$, with $K > 1$; the initial value is

$$\hat{\sigma}_{e_{l+1}}^2(0) = 0$$

The step size in Eq.(4.11) is derived assuming the case of exact modeling, i.e., the length of the adaptive filter and acoustic impulse response is assumed to be the same ($L=N$). In real world applications like AEC, the estimate of the $\hat{\sigma}_v(n-l)$ is not straightforward. The near-end signal $v(n)$ which could be background speech or/and near-end speech combined together with the acoustic echo results in the microphone signal, the only signal which is practically available. It is possible to express the power estimate of $v(n)$, in terms of the practically available signal, i.e., microphone signal $d(n)$.

The microphone signal at time index n can be expressed as

$$d(n) = y(n) + v(n) \quad (4.13)$$

where $y(n) = x^T(n)h$. Squaring Eq.(4.13) and taking the expectation of both the sides with the assumption that $y(n)$ and $v(n)$ are uncorrelated it results that $E\{d^2(n)\} = E\{y^2(n)\} + E\{v^2(n)\}$,

so that

$$E\{d^2(n)\} - E\{y^2(n)\} = E\{v^2(n)\} \quad (4.14)$$

If the adaptive filter converges to a certain degree, it is valid to consider that

$$E\{y^2(n)\} \cong E\{\hat{y}^2(n)\} \quad (4.15)$$

where $\hat{y}(n) = \mathbf{x}^T(n)\hat{\mathbf{h}}(n-1)$ is the estimate of the echo. Consequently Eq.(4.14) becomes

$$E\{v^2(n)\} \cong E\{d^2(n)\} - E\{\hat{y}^2(n)\} \quad (4.16)$$

or in terms of the power estimates

$$\hat{\sigma}_v^2(n) = \hat{\sigma}_d^2(n) - \hat{\sigma}_{\hat{y}}^2(n). \quad (4.17)$$

From Eq.(4.17) it is possible to express the power estimate of near-end signal in terms of the microphone signal and the estimate of the echo, so Eq.(4.11) becomes

$$\mu_l(n) = 1 - \frac{\sqrt{\hat{\sigma}_d^2(n-l) - \hat{\sigma}_{\hat{y}}^2(n-l)}}{\hat{\sigma}_{e_{nl}}(n)} \quad (4.18)$$

Several scenarios are considered based on the presence of only far-end signal or both far-end as well as near-end signals, as follows.

4.1 Single-Talk Scenario:

In the single-talk case, the near-end signal consists only the background noise, $w(n)$, i.e., $u(n)=0$. The ideal procedure is to estimate the power of $w(n)$ during silences, i.e., when the acoustic echo is zero and can be assumed constant, but estimate of noise power can also be obtained using the right-hand term in Eq.(4.17). This expression holds even if the level of the background noise changes, so that there is no need for the estimation of this parameter during silences. Under single-talk scenario Eq.(4.18) becomes

$$\mu_l(n) = 1 - \frac{\sqrt{\hat{\sigma}_d^2(n-l) - \hat{\sigma}_{\hat{y}}^2(n-l)}}{\hat{\sigma}_{e_{nl}}(n)} \quad (4.19)$$

We know that NLMS algorithm can be obtained from APA with a value of projection order $p=1$. For a value of $p=1$ in Eq.(4.19), the nonparametric VSS-NLMS (NPVSS-NLMS) algorithm proposed in [12] is obtained. For $p>1$, a VSS-APA can be obtained, by computing Eq.(4.19) for $l= 0,1,\dots,p-1$, then using a step-size matrix $\mu(n)$, and updating the filter coefficients according to Eq.(4.3).

4.2 Double-Talk Scenario:

In the double-talk case, both the far-end as well as near-end speech are present, i.e., the near-end signal consists of both background noise $w(n)$ and the near-end speech $u(n)$; so that $v(n)=w(n)+u(n)$. When the near-end speech is present the near-end signal power estimate can be expressed as $\hat{\sigma}_v(n) = \hat{\sigma}_w(n) + \hat{\sigma}_u(n)$ (assuming the near-end speech is uncorrelated with the background noise); the last parameter denotes the power estimate of the near-end speech. Accordingly, the right-hand term in Eq.(4.17) provides a power estimate of the near-end signal. Most importantly, this term depends only on the signals that are available within the AEC application, i.e., the microphone signal $d(n)$ and the output of the adaptive filter $\hat{y}(n)$. Therefore, from Eq.(4.17) the step size parameter under double-talk conditions can be written as

$$\mu_l(n) = 1 - \frac{\sqrt{\hat{\sigma}_d(n-l) - \hat{\sigma}_{\hat{y}}(n-l)}}{\hat{\sigma}_{e_{nl}}(n)} \quad (4.20)$$

The overall performance of the adaptive filter could be seriously affected during the double talk periods, up to divergence. Therefore, under such circumstances Eq.(4.20) becomes useless and the presence of DTD is must, in order to slow down or completely halt the adaptation process. Detailed description on DTD and various DTD algorithms is presented in section 4.5

4.3 Under-Modeling Scenario:

In both previous sections, the length of the adaptive filter and acoustic impulse response is assumed to be the same. In practice the acoustic impulse response is not static over time. As a result, the adaptive filter can work in an under-modeling situation, so that an under-modeling noise (i.e., the residual echo caused by the part of the system that cannot be modeled) appears. It can be interpreted as an additional noise that corrupts the near-end signal.

Since it is unavailable in practice, the power of the under-modeling noise cannot be estimated in a direct manner, and consequently, its contribution to the near-end signal power cannot be evaluated.

In practice the length of the acoustic impulse response is much longer than the length of the adaptive filter, so $L < N$. In this situation, the echo signal at time index n can be decomposed as

$$y(n) = y_L(n) + q(n) \quad (4.21)$$

where $y_L(n)$ represents the part of the acoustic echo that can be modeled by the adaptive filter. It can be written as

$$y_L(n) = \mathbf{x}^T(n) \mathbf{h}_L \quad (4.22)$$

where the vector $\mathbf{h}_L = [h_0, h_1, \dots, h_{L-1}]^T$ contains the L first coefficients of the echo path vector. The second term from the right-hand side of Eq.(4.21) is the under-modeling noise. This residual echo (which cannot be modeled by the adaptive filter) can be expressed as

$$q(n) = \mathbf{x}_{N-L}^T(n) \mathbf{h}_{N-L} \quad (4.23)$$

where $\mathbf{x}_{N-L}(n) = [x(n-L), x(n-L-1), \dots, x(n-N+1)]^T$ and $\mathbf{h}_{N-L} = [h_L, h_{L+1}, \dots, h_{N-1}]^T$ (i.e., the last $N-L$ coefficients of the echo path vector \mathbf{h}). The term from Eq.(4.23) acts like an additional noise for the adaptive process, so that (4.6) should be rewritten as

$$\varepsilon_{l+1}(n) = [1 - \mu_l(n)] e_{l+1}(n) = v(n-l) + q(n-l) \quad (4.24)$$

with $l=0, 1, \dots, p-1$. Squaring, then taking expectation on both the sides of Eq.(4.24) (considering the near-end signal and under-modeling noise to be uncorrelated) and solving for $\mu_l(n)$, results in

$$\mu_l(n) = 1 - \sqrt{\frac{E\{v^2(n-l)\} + E\{q^2(n-l)\}}{E\{e_{l+1}^2(n)\}}} \quad (4.25)$$

Unfortunately, expression Eq.(4.25) is useless in a real-world AEC application since it depends on some sequences that are unavailable, i.e., the near-end signal and the under-modeling noise. Considering the presence of under modeling noise into account the microphone signal is rewritten as

$$d(n) = y_L(n) + q(n) + v(n) \quad (4.26)$$

Squaring and then taking the expectations of both sides of Eq.(4.26), it results in

$$E\{d^2(n)\} = E\{y_L^2(n)\} + E\{q^2(n)\} + E\{v^2(n)\} \quad (4.27)$$

If the adaptive filter converges to a certain degree, it is valid to consider that

$$E\{y_L^2(n)\} \cong E\{\hat{y}^2(n)\} \quad (4.28)$$

Consequently

$$E\{q^2(n)\} + E\{v^2(n)\} = E\{d^2(n)\} - E\{\hat{y}^2(n)\} \quad (4.29)$$

Rewriting Eq.(4.25) in terms of the power estimates and taking Eq.(4.30) into account, the expression of the step-size parameter is

$$\mu_l(n) = 1 - \frac{\sqrt{\hat{\sigma}_d^2(n-l) - \hat{\sigma}_y^2(n-l)}}{\hat{\sigma}_{e_{n+1}}(n)} \quad (4.30)$$

For all the three cases of AEC, i.e., single-talk, double-talk and under-modeling scenarios the variable step-size parameters can be computed in a unified manner as in Eq.(4.30)

Practical Considerations:

In practice the update equation for VSS-APA is written as

$$\hat{\mathbf{h}}(n) = \hat{\mathbf{h}}(n-1) + \mathbf{X}(n)[\delta \mathbf{I}_p + \mathbf{X}^T(n)\mathbf{X}(n)]^{-1} \mu(n)\mathbf{e}(n) \quad (4.31)$$

where δ is a positive scalar known as the regularization factor and \mathbf{I}_p is the $p \times p$ identity matrix. The reason behind this regularization process is to prevent the problems associated with the inverse of the matrix $\mathbf{X}^T(n)\mathbf{X}(n)$, which could become ill-conditioned especially when highly correlated inputs (e.g., speech) are involved [18]. Considering the context of a “system identification” configuration, the value of the regularization factor depends on the level of the noise that corrupts the output of the system that has to be identified. A lower signal-to-noise ratio requires a higher value of the regularization factor. In the AEC context, different types of “noise” corrupt the output of the echo path, e.g., the background noise or/and the near-end speech; in addition, the under-modeling noise (if it is present) increases the overall level of “noise.”

Also, the value of the regularization factor depends on the value of the projection order of the algorithm. As the value of the projection order of the APA becomes larger, the condition number of the matrix $\mathbf{X}^T(n)\mathbf{X}(n)$ also grows; consequently, a higher value of δ is required.

A very small positive number ξ is added to the denominator in Eq.(4.30) to avoid division by zero. Hence the final step-sizes formula is

$$\mu_l(n) = 1 - \frac{\sqrt{\hat{\sigma}_d^2(n-l) - \hat{\sigma}_y^2(n-l)}}{\xi + \hat{\sigma}_{e_{l+1}}(n)} \quad (4.32)$$

The complete summary of VSS-APA is given in Table 4.1.

Table 4.1 Summary of VSS-APA Algorithm [18]

<p><i>Initialization:</i> $\hat{\mathbf{h}}(0) = \mathbf{0}_{L \times 1}$, $\hat{\sigma}_d(0) = 0$, $\hat{\sigma}_y^2(0) = 0$</p> <p>For $l=0, 1, \dots, p-1$</p> $\hat{\sigma}_{e_{l+1}}^2(0) = 0$ <p>For time index $n=1, 2, \dots$</p> $\mathbf{e}(n) = \mathbf{d}(n) - \mathbf{X}^T(n) \hat{\mathbf{h}}(n-1)$ $\hat{y}(n) = \mathbf{x}^T(n) \hat{\mathbf{h}}(n-1)$ $\hat{\sigma}_d^2(n) = \lambda \hat{\sigma}_d^2(n-1) + (1-\lambda) d^2(n)$ $\hat{\sigma}_y^2(n) = \lambda \hat{\sigma}_y^2(n-1) + (1-\lambda) \hat{y}^2(n)$ <p>For $l=0, \dots, p-1$</p> $\hat{\sigma}_{e_{l+1}}^2(n) = \lambda \hat{\sigma}_{e_{l+1}}^2(n-1) + (1-\lambda) e_{l+1}^2(n)$ $\mu_l(n) = 1 - \frac{\sqrt{\hat{\sigma}_d^2(n-l) - \hat{\sigma}_y^2(n-l)}}{\xi + \hat{\sigma}_{e_{l+1}}(n)}$ $\boldsymbol{\mu}(n) = \text{diag}\{\mu_0(n), \mu_1(n), \dots, \mu_{p-1}(n)\}$ $\hat{\mathbf{h}}(n) = \hat{\mathbf{h}}(n-1) + \mathbf{X}(n) [\delta \mathbf{I}_p + \mathbf{X}^T(n) \mathbf{X}(n)]^{-1} \boldsymbol{\mu}(n) \mathbf{e}(n)$
--

4.4 Variable Regularised APA (VR-APA)

The variable regularised APA proposed in [19] is included for comparison with VSS-APA in chapter 5. Due to the nature of this algorithm, it can also be considered as a VSS-APA. The VR-APA update is

$$\hat{\mathbf{h}}(n) = \hat{\mathbf{h}}(n-1) + \mathbf{X}(n)[\delta(n)\mathbf{I}_p + \mathbf{X}^T(n)\mathbf{X}(n)]^{-1}\mathbf{e}(n) \quad (4.33)$$

Its variable regularization factor is given by

$$\delta(n) = \min \left\{ \frac{L\sigma_x^2}{\zeta}, \frac{p\sigma_w^2\sigma_x^2L}{\hat{\sigma}_e^2(n) - p\sigma_w^2} \right\} \quad (4.34)$$

where ζ is a positive design parameter. The power of the background noise σ_w^2 and the power of the input signal σ_x^2 have to be known within the algorithm, while the term is $\hat{\sigma}_e^2(n)$ evaluated as

$$\hat{\sigma}_e^2(n) = \lambda\hat{\sigma}_e^2(n-1) + (1-\lambda)\|\mathbf{e}(n)\|^2 \quad (4.35)$$

where $\|\cdot\|$ denotes the l_2 norm and λ is a weighting parameter as in Eq.(4.12).

4.5 Double-Talk Detector

An important characteristic of a good echo canceller is its performance during double talk. The condition where both ends, the near-end and the far-end, are speaking is referred to as double talk. If the echo canceller does not detect a double talk condition properly the near end speech will cause the adaptive filter to diverge. Therefore, it is important to have a reliable double-talk detector. A DTD is used with an echo canceller to sense when the far-end speech is corrupted by the near-end speech. The role of this important function is to freeze adaptation of the model filter, $\hat{\mathbf{h}}(n)$, when the near-end speech, $u(n)$, is present in order to avoid divergence of the adaptive algorithm.

Almost all types of doubletalk detectors operate in the same manner. Therefore, the general procedure for handling double talk is described by the following four steps.

1. A detection statistic, ξ , is formed using available signals such as far-end signal $x(n)$, microphone signal $d(n)$ and error signal $e(n)$ and the estimated filter coefficients, $\hat{h}(n)$.
2. The detection statistic, ξ , is compared to a preset threshold, T , (a constant), and double talk is declared if $\xi < T$.
3. Once doubletalk is declared the detection is held for a minimum period of time T_{hold} , which is also termed as 'hangover time'[20]. While the detection is held, the filter adaptation is disabled.
4. If $\xi \geq T$ consecutively over a time T_{hold} the filter resumes adaptation while the comparison of ξ to T continues until $\xi < T$ again.

The hold time, T_{hold} , in steps 3 and 4 is essential to suppress detection dropouts due to the noisy behaviour of the detection statistic. Although there are some possible variations most of the DTD algorithms keep this basic form and only differ in how they form the detection statistic.

An optimum decision variable, ξ , for double talk detection should behave as follows:

- if $v = 0$ (doubletalk is not present), $\xi \geq T$
- if $v \neq 0$ (doubletalk is present), $\xi < T$

The threshold T must be a constant, independent of data. Moreover ξ must be insensitive to echo path variations when $v = 0$ [21].

In the following sections discussions of different DTD algorithms such as the Geigel algorithm, the cross- correlation method and the normalized cross-correlation method are presented.

4.5.1 The Giegel Algorithm

One simple algorithm due to A. A. Giegel [20] declares the presence of near-end speech whenever

$$\xi = \frac{\max\{|x(n)|, |x(n-1)|, \dots, |x(n-L+1)|\}}{|d(n)|} < T \quad (4.36)$$

where L is the length of the adaptive filter.

This detection scheme is based on a waveform level comparison between the microphone signal, $d(n)$, and the far-end speech, $x(n)$, assuming the near-end speech, $v(n)$, in the microphone signal will be stronger than the echo. The maximum of the L most recent samples of 'x' is chosen for the comparison due to uncertain delay in the echo path.

The threshold, T , is used to compensate for the energy level of the echo path response and is often set to 0.5, if the far-end to double-talk attenuation is assumed to be 6dB and to 0.71 if the attenuation is assumed to be 3dB[18].

4.5.2 Cross Correlation Method

This method uses the cross-correlation coefficient vector between far-end speech, $x(n)$, and microphone signal, $d(n)$, as a means for double talk detection. The cross-correlation coefficient vector between $x(n)$ and $d(n)$ is defined by

$$\begin{aligned} c_{xd} &= \frac{E\{x(n)d(n)\}}{\sqrt{E\{x^2(n)\}E\{d^2(n)\}}} \\ &= \frac{\Gamma_{xd}}{\sigma_x \sigma_d} \\ &= [c_{xd,0} \ c_{xd,1} \ \dots \ c_{xd,L-1}]^T \end{aligned} \quad (4.37)$$

where $E\{\cdot\}$ denotes the mathematical expectation and $c_{xd,i}$ is the cross-correlation coefficient between $x(n-i)$ and $d(n)$. The idea is to compare

$$\begin{aligned} \xi &= \|c_{xd}\| \\ &= \max |c_{xd,i}|, \ i = 0, 1, \dots, L-1 \end{aligned}$$

to a threshold level T . The decision rule is then very simple. If $\xi \geq T$, double talk is not present and if $\xi < T$, double talk is present.

The fundamental problem with this method is that the cross-correlation coefficient vectors are not well normalized. In general, it is assumed that $\xi \leq 1$. Therefore, if near-end signal $v(n) = 0$, it does not mean that $\xi = 1$ or any other known value. The value of ξ is not known in general. The amount of correlation will depend greatly on the statistics of the signal and of the echo path. As a result, the best value of T will vary from one experiment to another. There is no natural threshold level associated with the variable ξ when $v(n) = 0$. These complexities lead to another DTD algorithm, which is termed the normalized cross-correlation method. This method is simply a modification of the existing cross-correlation Method [21].

4.5.3 Normalised Cross Correlation Method

In this method a new normalized cross-correlation vector between a vector $x(n)$ (i.e. far-end speech) and a scalar $d(n)$ (i.e. microphone signal) is derived.

Suppose, the near-end signal $v(n) = 0$. In this case

$$\begin{aligned} \mathbf{R}_{dd} &= E\{d(n)d^T(n)\} \\ &= \mathbf{h}^T \mathbf{R}_{xx} \mathbf{h} \end{aligned} \quad (4.38)$$

where

$$\mathbf{R}_{xx} = E\{\mathbf{x}(n)\mathbf{x}^T(n)\} \quad (4.39)$$

Since

$$d(n) = \mathbf{h}^T \mathbf{x}(n) \quad (4.40)$$

$$\mathbf{R}_{xd} = \mathbf{R}_{xx} \mathbf{h} \quad (4.41)$$

which allows \mathbf{R}_{dd} to be written as

$$\mathbf{R}_{dd} = \mathbf{R}_{xd}^T \mathbf{R}_{xx}^{-1} \mathbf{R}_{xd} \quad (4.42)$$

In general for, for $v(n) \neq 0$,

$$\mathbf{R}_{dd} = \mathbf{R}_{xd}^T \mathbf{R}_{xx}^{-1} \mathbf{R}_{xd} + \mathbf{R}_{vv} \quad (4.43)$$

where $\mathbf{R}_{vv} = E\{v(n)v^T(n)\}$. The new decision variable is obtained by dividing Eq.(4.43) by \mathbf{R}_{dd} and extracting the square root, which yields

$$\xi = \sqrt{\mathbf{R}_{xd}^T \mathbf{R}_{xx}^{-1} \mathbf{R}_{xd} \mathbf{R}_{dd}^{-1}} \quad (4.44)$$

$$= \left\| c_{xd} \right\| \quad (4.45)$$

where $\left\| \cdot \right\|$ denotes the l_2 norm and $c_{xd} = \mathbf{R}_{xx}^{-1/2} \mathbf{R}_{xd} \mathbf{R}_{dd}^{-1/2}$ is the normalized cross-correlation vector between $\mathbf{x}(n)$ and $d(n)$. Substituting equation Eq.(4.41) and equation Eq.(4.43) into equation Eq.(4.44) produces the decision variable, which is given by

$$\xi = \frac{\sqrt{\mathbf{h}^T \mathbf{R}_{xx} \mathbf{h}}}{\sqrt{\mathbf{h}^T \mathbf{R}_{xx} \mathbf{h} + \sigma_v^2}} \quad (4.46)$$

where $\sigma_v^2 = \mathbf{R}_{vv} = E\{v(n)v^T(n)\}$. Eq.(4.47) shows that for $v = 0$; $\xi = 1$ and for $v \neq 0$; $\xi < 1$.

Also, from Eq.(4.47) it is clear that ξ is not sensitive to changes of echo path when

$v = 0$ [21].

Simulation and Results

This chapter presents the results obtained through computer simulation using MATLAB 7.4, for cancelling the acoustic echo using VSS-APA and comparison of the same algorithm with classical APA and VR-APA, in terms of the convergence rate and final misalignment under different scenarios like single-talk, double-talk and under-modelling.

The simulations were performed in an AEC context, as shown in Figure 4.1. The acoustic echo path was measured using an 8-kHz sampling rate. Its impulse response has 1000 coefficients [22] and is plotted in Figure 5.1(a), while the adaptive filter length is $L=500$.

The length of the acoustic impulse response is truncated to the first 500 coefficients for a first set of experiments performed in an exact modeling case. Then, the entire length of the acoustic impulse response is used for a second set of experiments performed in the under-modeling case. The far-end signal $x(n)$ is either an AR(1) process generated filtering a white Gaussian noise through a first-order $1/(1-0.95z^{-1})$ system, or a speech sequence as plotted in Figure 5.1(b). For the double-talk scenarios, the near-end speech is plotted in Figure 5.1(c). An independent white Gaussian noise signal is added to the echo signal, with 40-dB signal-to-noise ratio (SNR) for most of the experiments. The weighting factor $\lambda = 1-1/(KL)$ (for VR-APA and VSS-APA) is computed using $K=6$ [18]. The value of the parameter in the denominator of variable step-sizes in Eq.(4.32) is $\xi = 10^{-8}$. The performance is evaluated in terms of the normalized misalignment (in dB), defined as $20 \log_{10} \left(\frac{\|h - \hat{h}(n)\|}{\|h\|} \right)$.

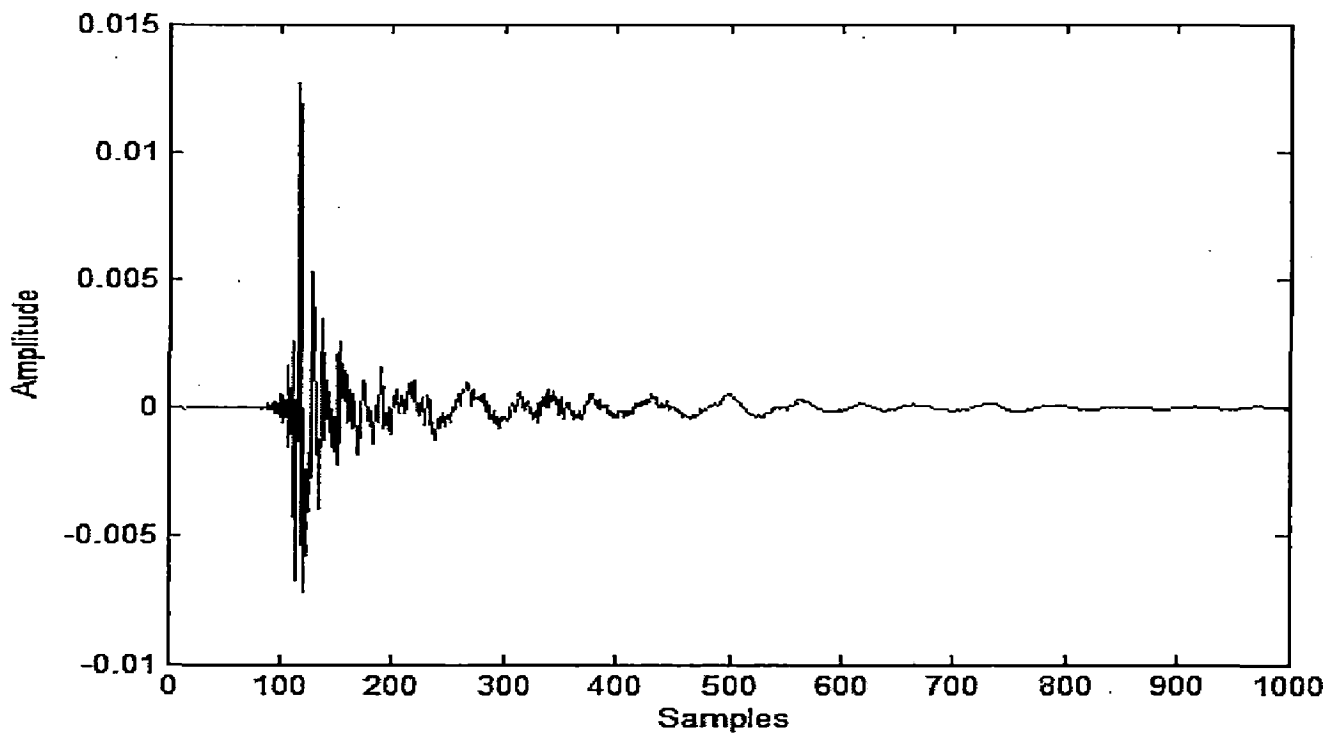


Figure 5.1(a) Typical room acoustic impulse response.

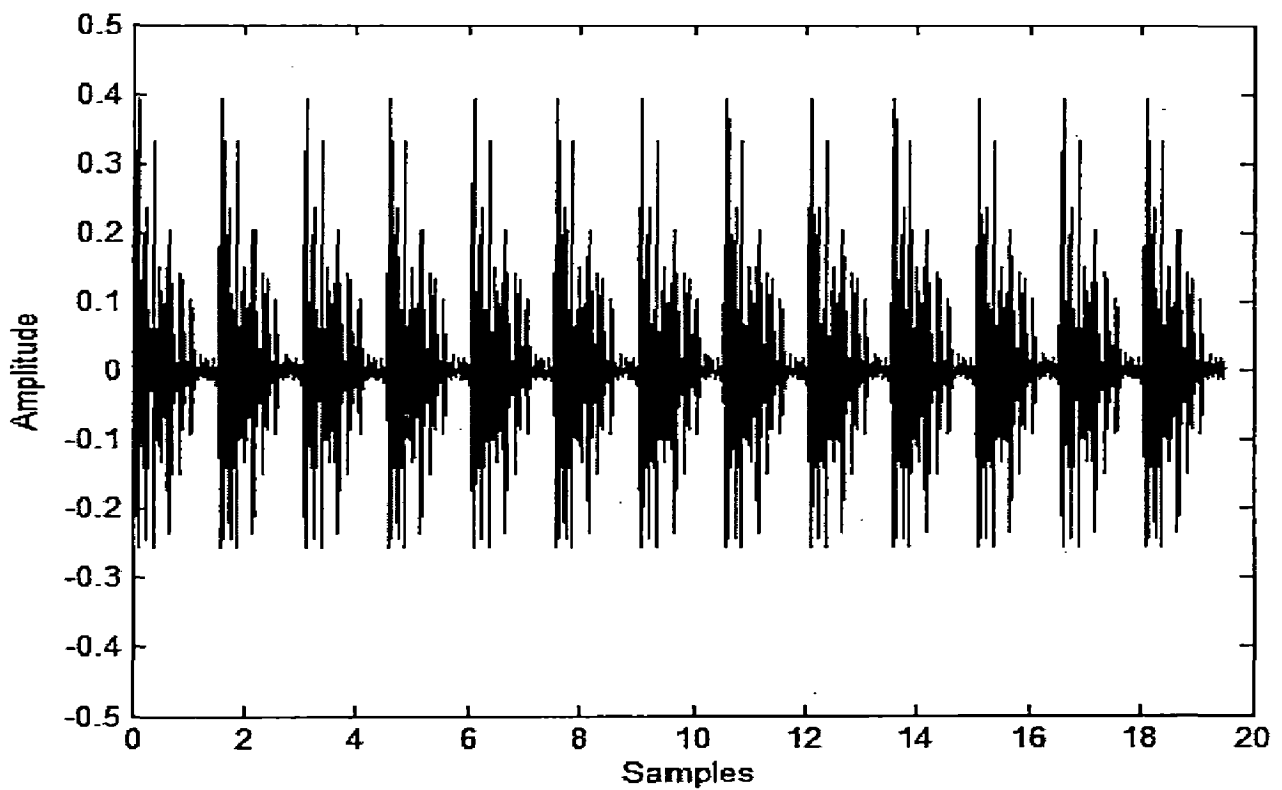


Figure 5.1(b) Far-end speech signal.

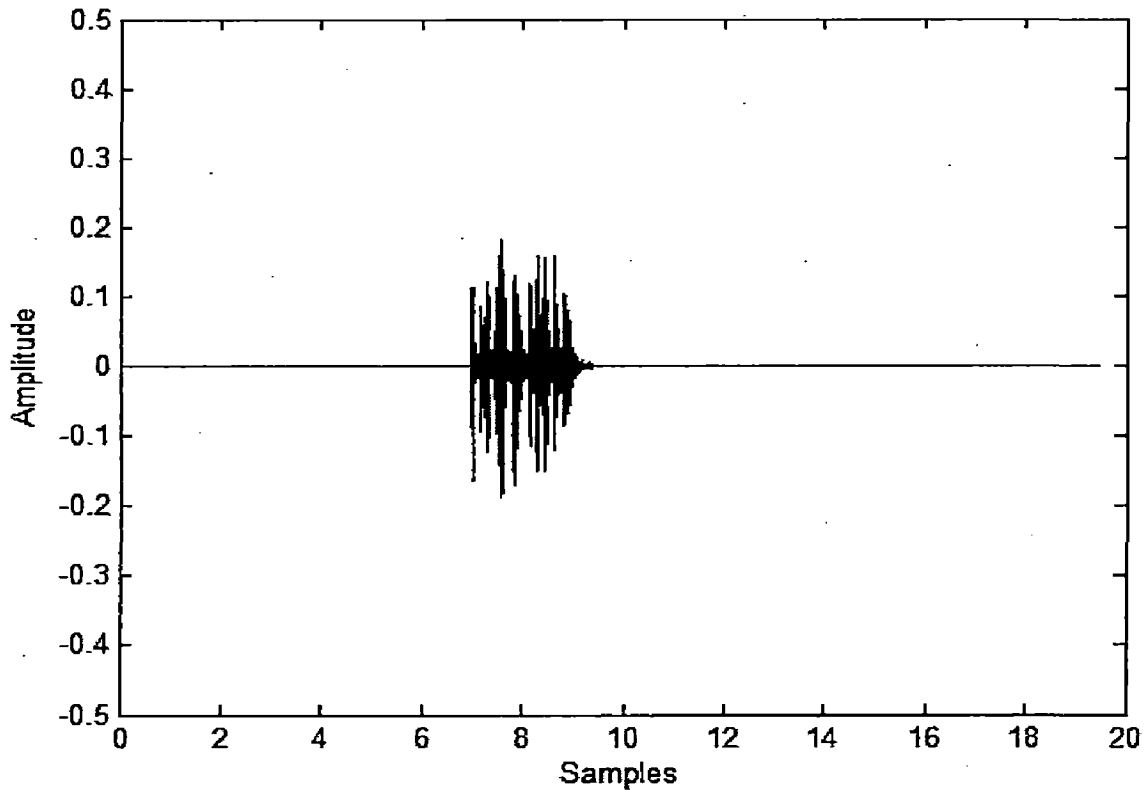


Figure 5.1(c) Near-end speech signal.

5.1 Exact Modeling Case

Simulations is carried out assuming the length of the acoustic impulse response to be the same as adaptive filter (i.e., $L=N=500$). The length of the acoustic impulse response in Figure 5.1(a) is truncated to the first 500 coefficients. The single-talk and double-talk scenarios are considered separately.

Single-Talk Scenario:

It is known that the overall behaviour of the APA depends on its step-size parameter μ . In Figures 5.2, 5.4 and 5.6, the misalignment curves for the APA with different values of the step-size are shown, as compared to the VSS-APA. The input signal is the AR (1) process for Figures 5.2 and 5.4; the speech sequence is used in Figure 5.6. The SNR is equal to 40 dB for Figure 5.2; SNR=20dB is used in Figures 5.4 and 5.6. The value of the projection order is $p=2$ and the regularization factor is $\delta = 50\sigma_x^2$ for all the algorithms. Since the requirements are for both high convergence rate and low misalignment, a compromise choice has to be made in the case of the APA.

From Figures 5.2, 5.4 and 5.6 it can be observed that the value of $\mu=1$ leads to the fastest convergence mode but offers a very high final misalignment. The value of $\mu=0.2$ offers a slower convergence rate but offers lower final misalignment as compared to the previous value. The VSS-APA has an initial convergence rate similar to the APA with the “middle” value (i.e., $\mu=0.2$) of the step-size, but it achieves a significant lower misalignment, which is close to the one obtained by the APA with the smallest step-size (i.e., $\mu=0.08$).

The results obtained in this section are verified with results of [18] as shown in Figures 5.3 and 5.5.

Misalignment of the algorithms in the single-talk case and exact modeling scenario, $L=N=500$, with projection order $p=2$ and $\delta = 50\sigma_x^2$ is shown in Figures 5.2, 5.4 and 5.6.

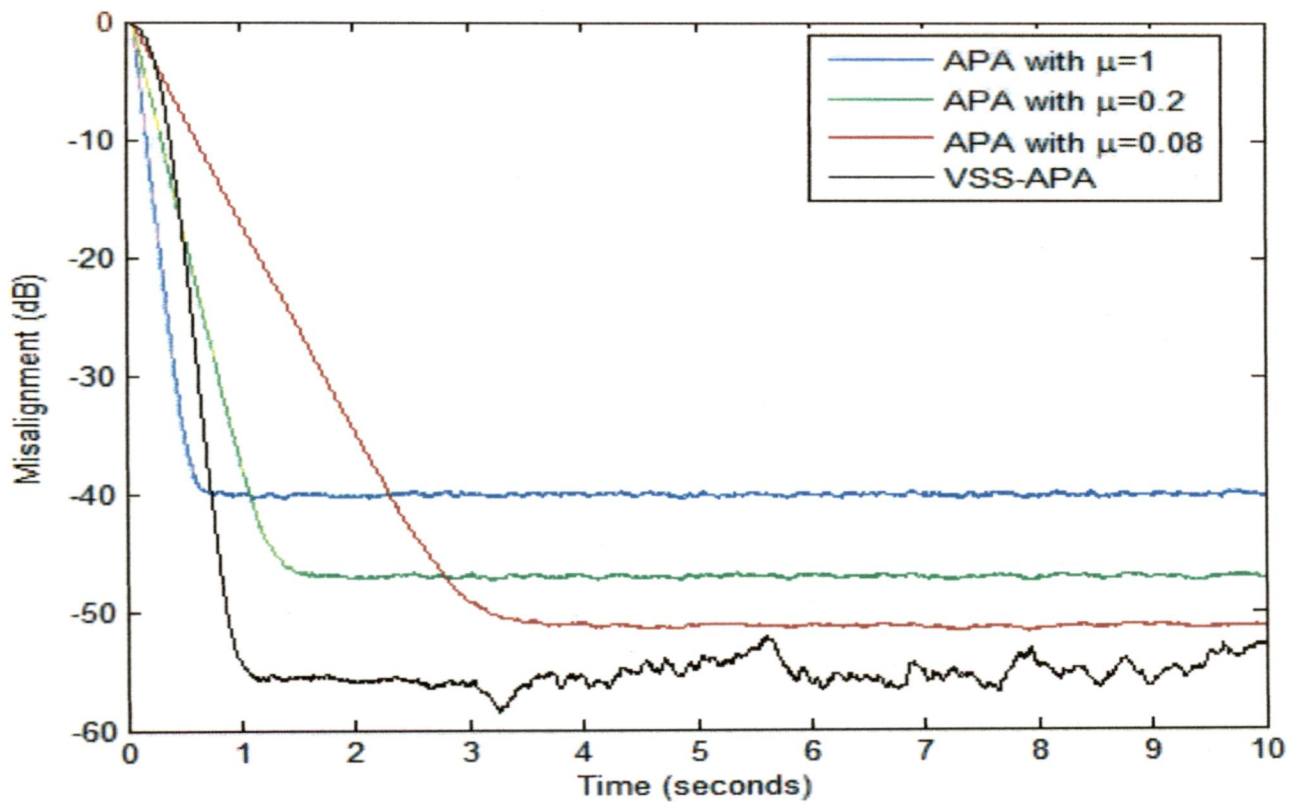


Figure 5.2 APA with three different step sizes ($\mu=1$, $\mu=0.2$, $\mu=0.08$), and VSS-APA; input is an AR(1) process, SNR=40dB.

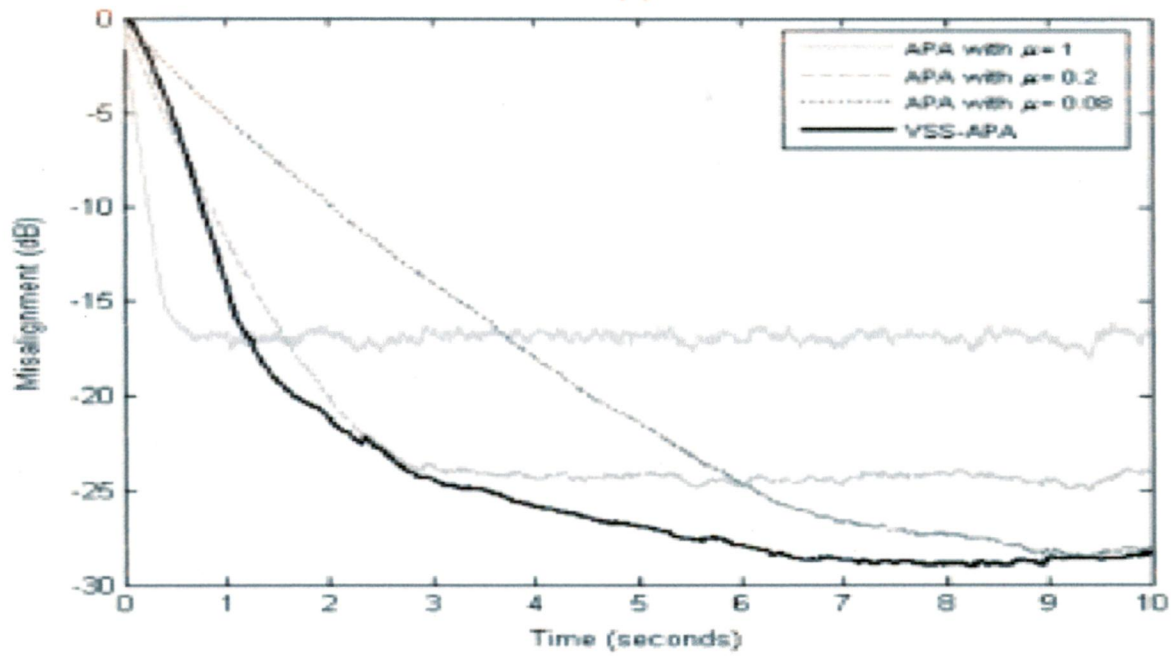


Figure 5.3 APA with three different step sizes ($\mu=1$, $\mu=0.2$, $\mu=0.08$), and VSS-APA; input is an AR(1) process, SNR=20dB [18].

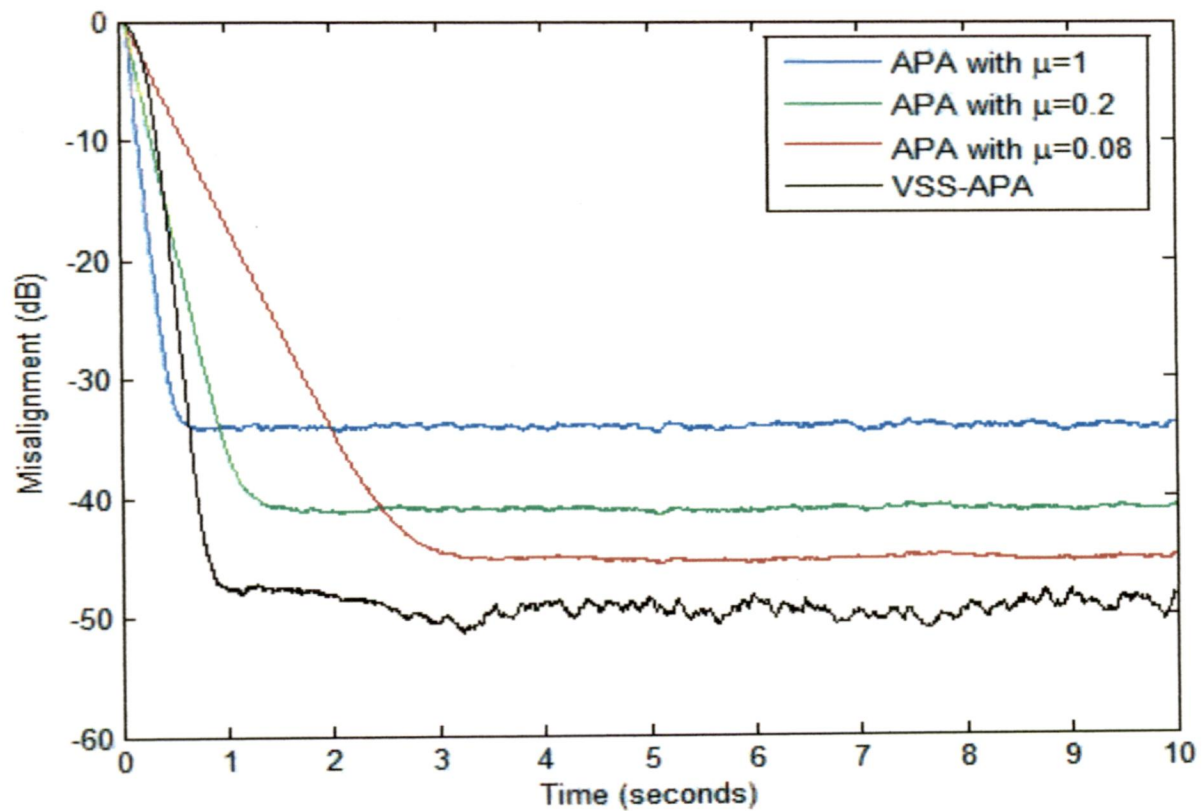


Figure 5.4 APA with three different step sizes ($\mu=1$, $\mu=0.2$, $\mu=0.08$), and VSS-APA; input is an AR(1) process, SNR=20dB.

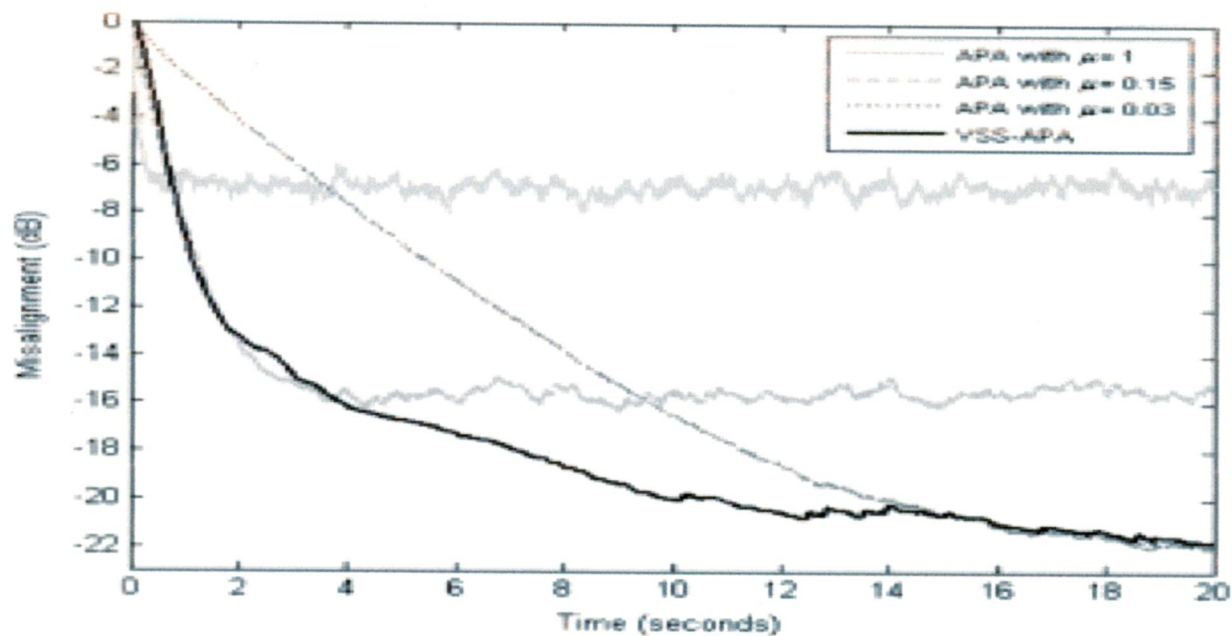


Figure 5.5 APA with three different step sizes ($\mu=1$, $\mu=0.15$, $\mu=0.03$), and VSS-APA; input is an AR(1) process, SNR=10dB [18].

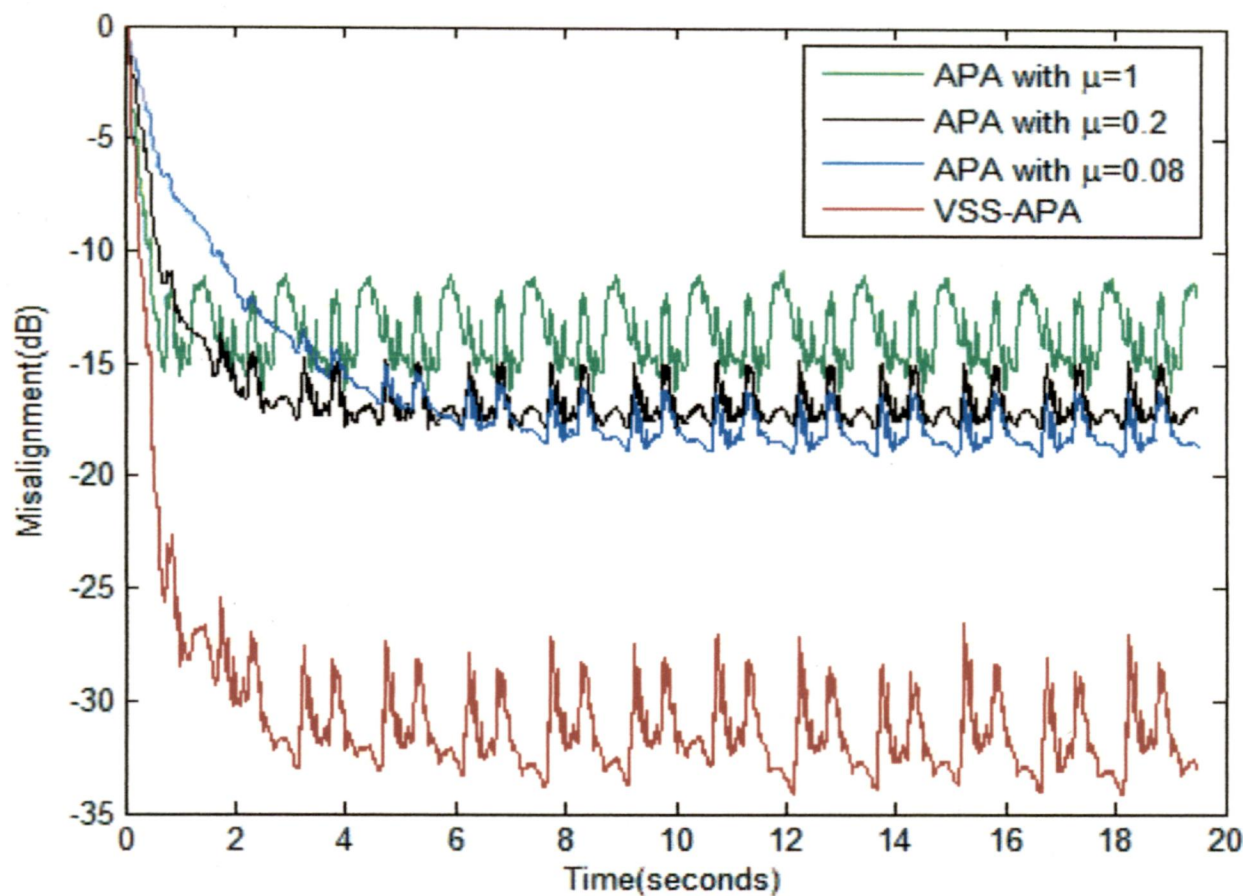


Figure 5.6 APA with three different step sizes ($\mu=1$, $\mu=0.2$, $\mu=0.08$), and VSS-APA; input is a speech signal, SNR=20dB.

Dependence of Convergence Rate on Projection Order p :

Simulations is carried out to evaluate the performance of APA and VSS-APA for different values of projection order p . Theoretically, for APA it is proven that the rate of convergence improves as the projection order is increased [7], which has been justified by the Figures 5.7 and 5.8, for $p= 2, 4$ and 8. Since, the regularization factor δ strongly depends on p , its choice for different values of p is given in Table 5.1 [18].

Table 5.1 Regularization Factors for APA and VSS-APA.

Projection order	Regularization factor
$p=1$	$\delta = 20\sigma_x^2$
$p=2$	$\delta = 50\sigma_x^2$
$p=4$	$\delta = 100\sigma_x^2$
$p=8$	$\delta = 200\sigma_x^2$

Change of the Acoustic Echo Path:

A possible scenario in AEC is the change of the acoustic echo path, i.e., acoustic impulse response is not constant; it may vary from time to time. The results of such an experiment are depicted in Figure 5.10, where the acoustic impulse response was shifted to the right by 12 samples after 10 s from the debut of the adaptive process. In Figure 5.10, the projection order for all the algorithms is $p=8$. The performance of the VSS-APA is compared with classical APA and VR-APA. The design parameter for VR-APA is set to $\zeta =1$. From Figure 5.10, it can be observed that the VSS-APA has better tracking capabilities over the other two; hence it is robust to the echo path variations.

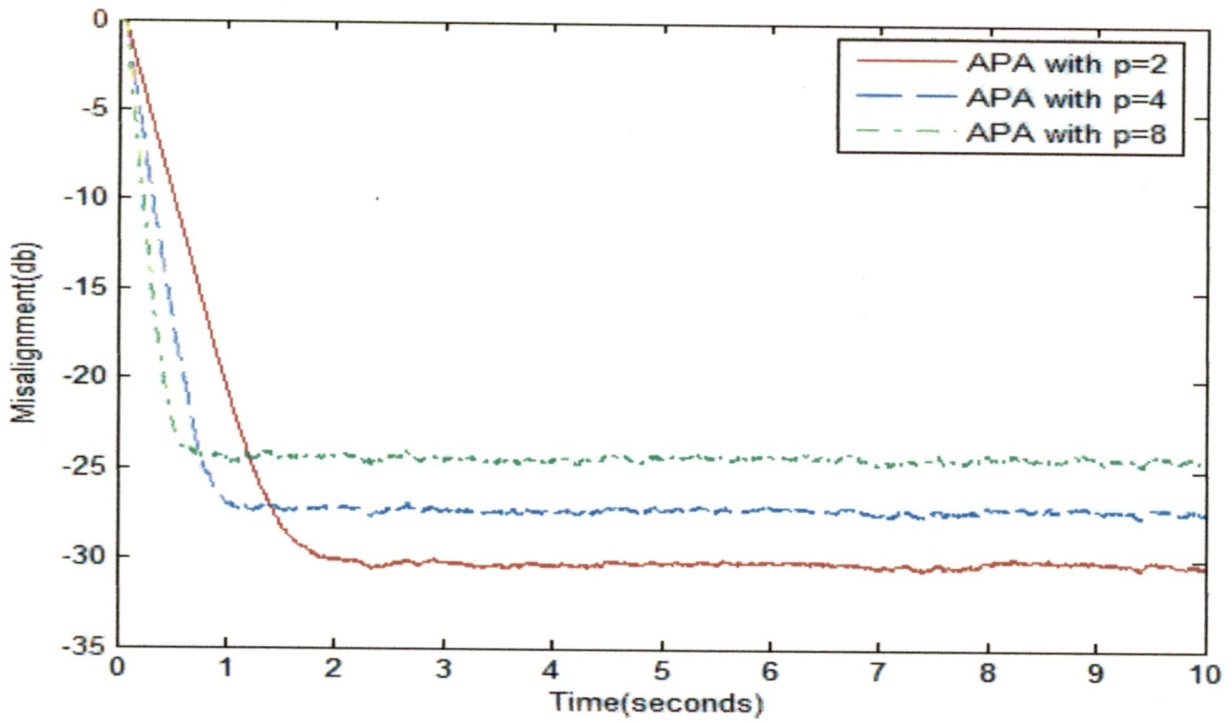


Figure 5.7 Misalignment of APA with three different projection orders, $p=2, 4,$ and $8,$ $\mu=0.2,$ δ is chosen from Table 5.1, input is an AR(1) process, SNR=40 dB.

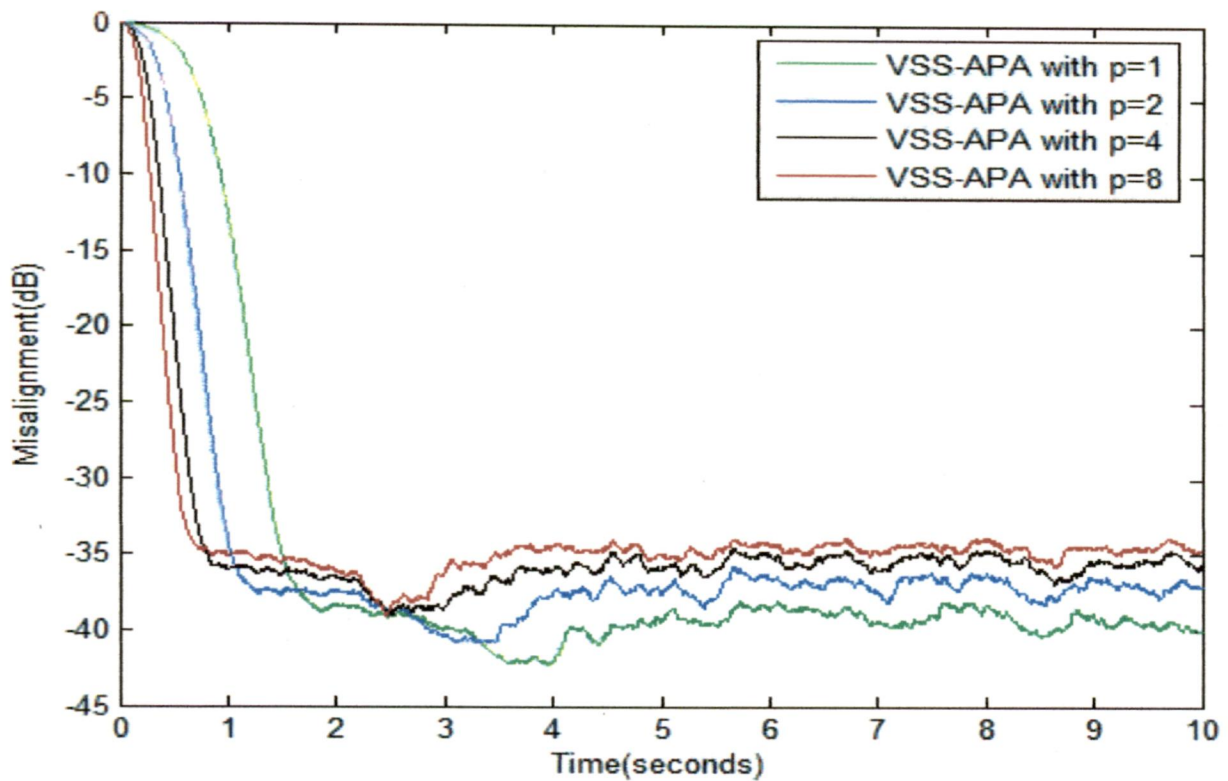


Figure 5.8 Misalignment of VSS-APA with four different projection orders, $p=1, 2, 4,$ and $8,$ δ is chosen from Table 5.1, input signal is speech, SNR=40 dB.

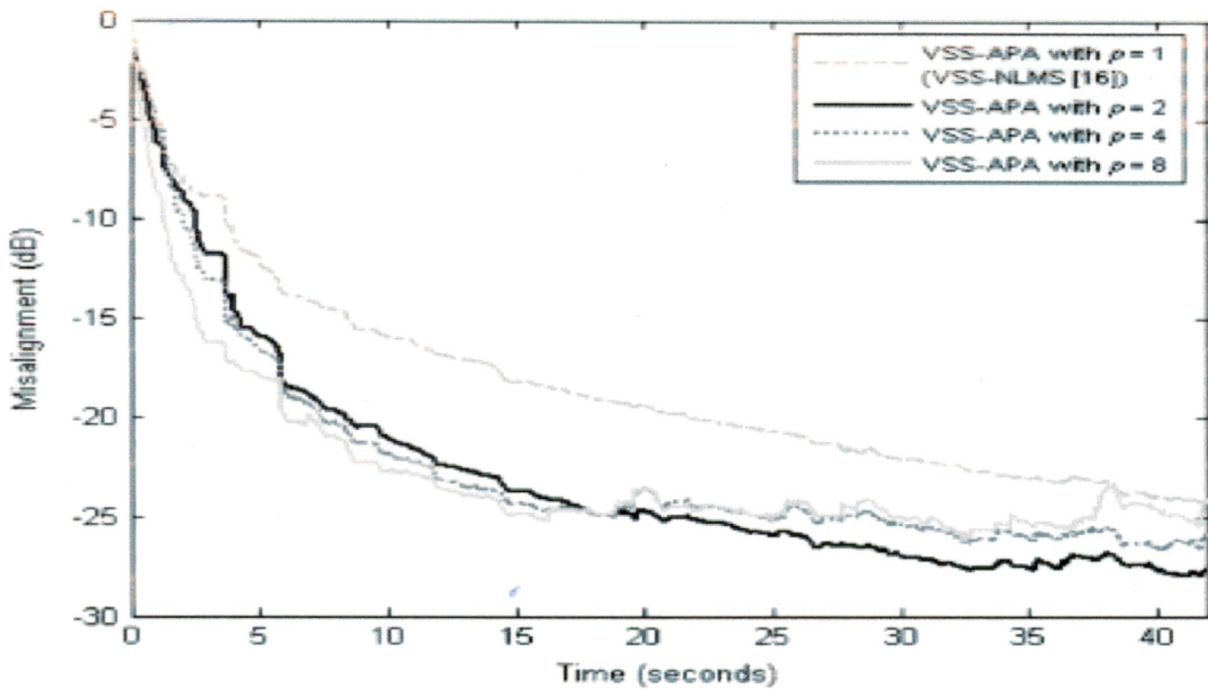


Figure 5.9 Misalignment of VSS-APA with four different projection orders, $p=1, 2, 4,$ and $8, L=N=512, \text{SNR}=20 \text{ dB}$ [18].

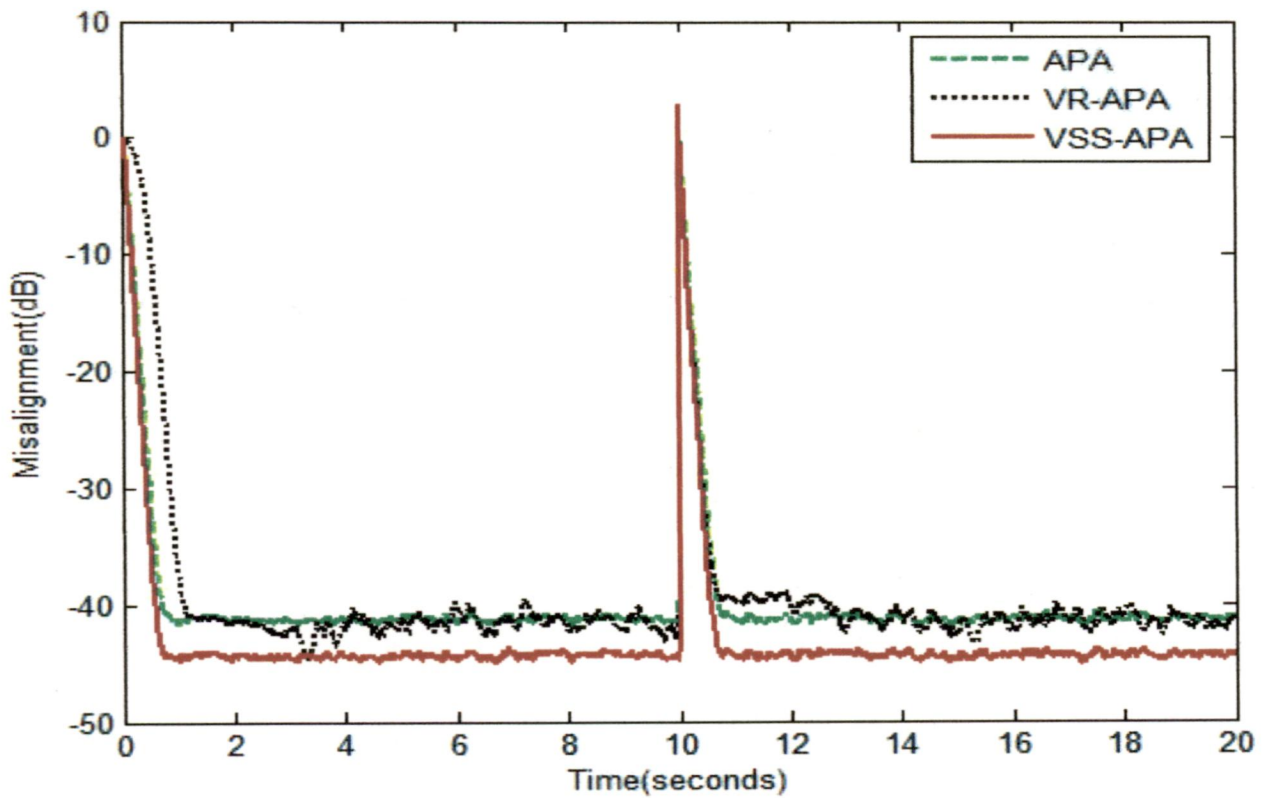


Figure 5.10 Misalignments of APA with $\mu=0.2$, VR-APA with $\zeta=1$, and VSS-APA, the echo path changes at time 10 sec, δ is chosen from Table 5.1(for APA and VSS-APA), $\text{SNR}=40\text{dB}$.

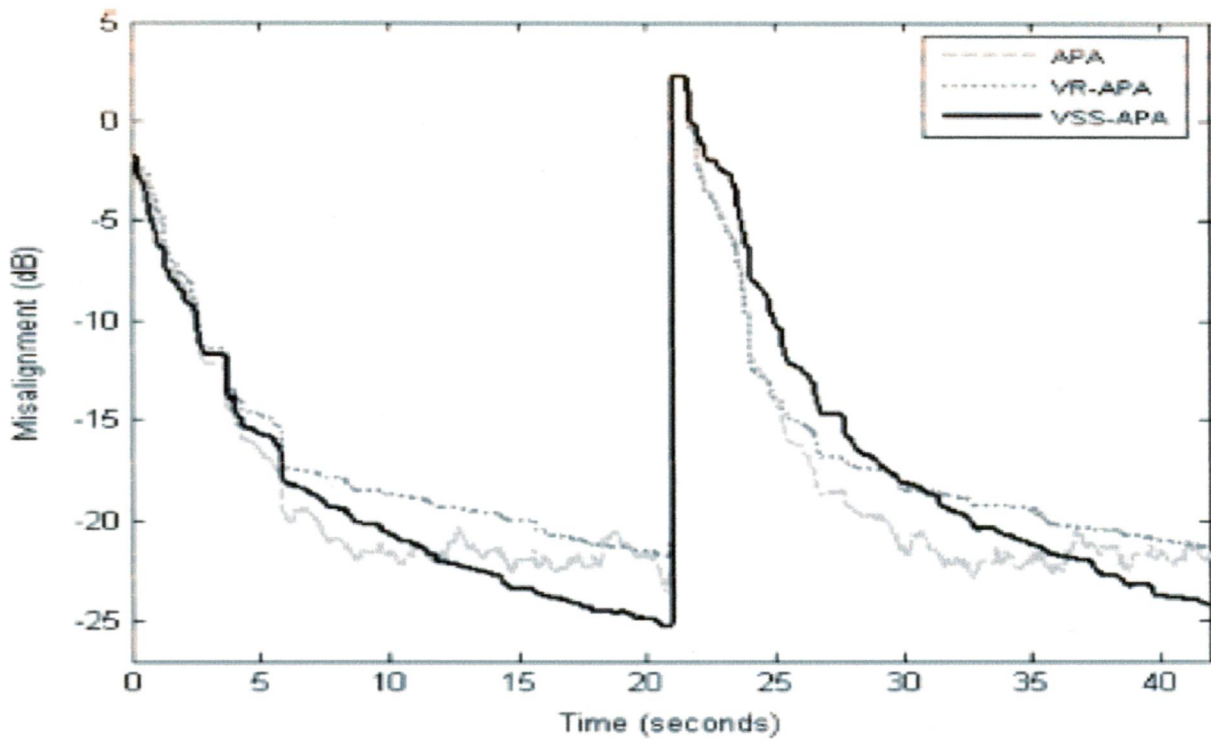


Figure 5.11 Misalignments of APA with $\mu=0.2$, VR-APA with $\zeta=1$, and VSS-APA, the echo path changes at time 21 sec, $L=N=512$, SNR=20dB [18].

Variations in the Background Noise Level:

The background noise can also vary in AEC, and consequently, its effects over the algorithms performance should be considered. The requirement of a good adaptive filter algorithm is, it should be robust against the background noise variations, which is an unavoidable situation in practice. In the experiment presented in Figure 5.12, the SNR decreases from 40 to 20 dB after 10 s from the debut of the adaptive process. While performing the simulations, it is assumed that the new background noise power estimate is not available for VR-APA. For all the algorithms, the projection order is $p=2$. From Figure 5.12 , it can be noticed that the VSS-APA is very robust against the background noise variation, while all the other algorithms are affected by this change in the acoustic environment.

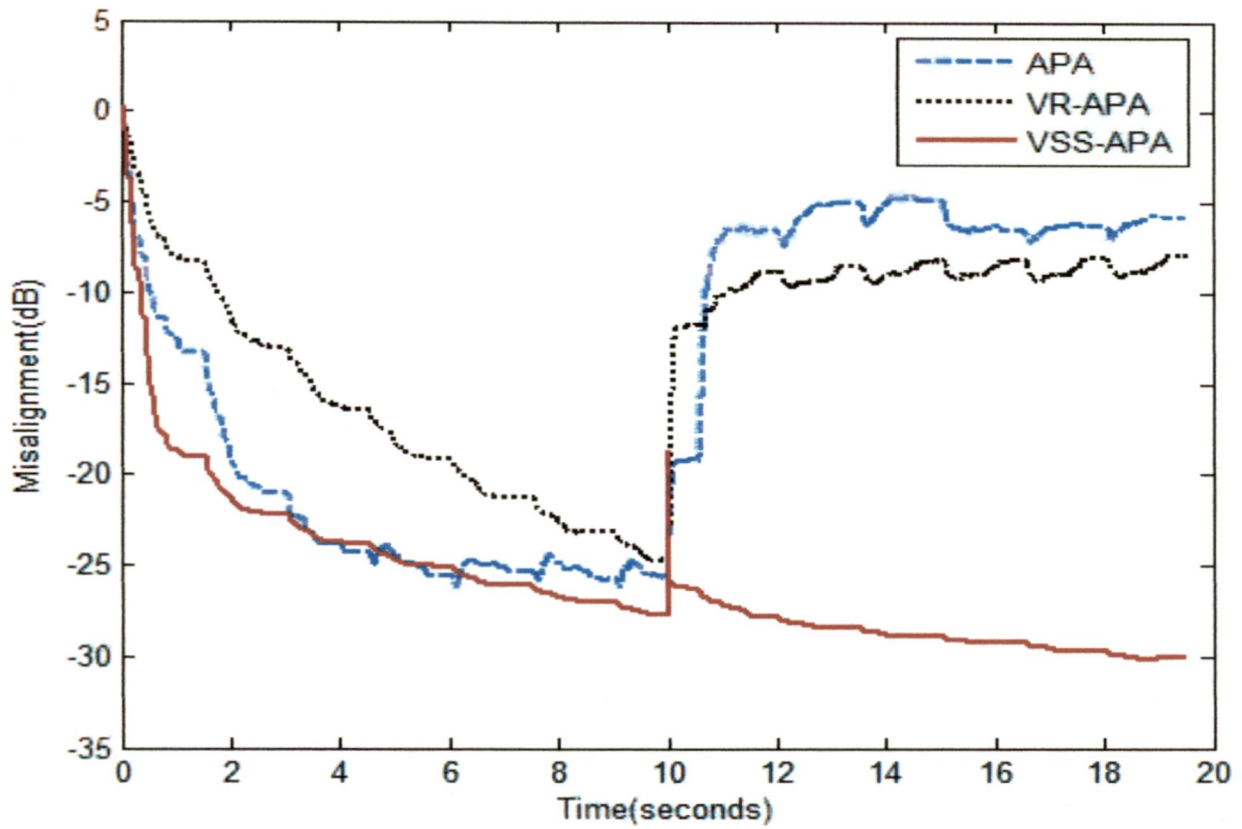


Figure 5.12 Misalignments of APA with $\mu=0.2$, VR-APA with $\zeta=1$, and VSS-APA, the background noise variation at time 10 sec (SNR decreases from 40dB to 20dB), $L=N=500$, δ is chosen from Table 5.1(for APA and VSS-APA).

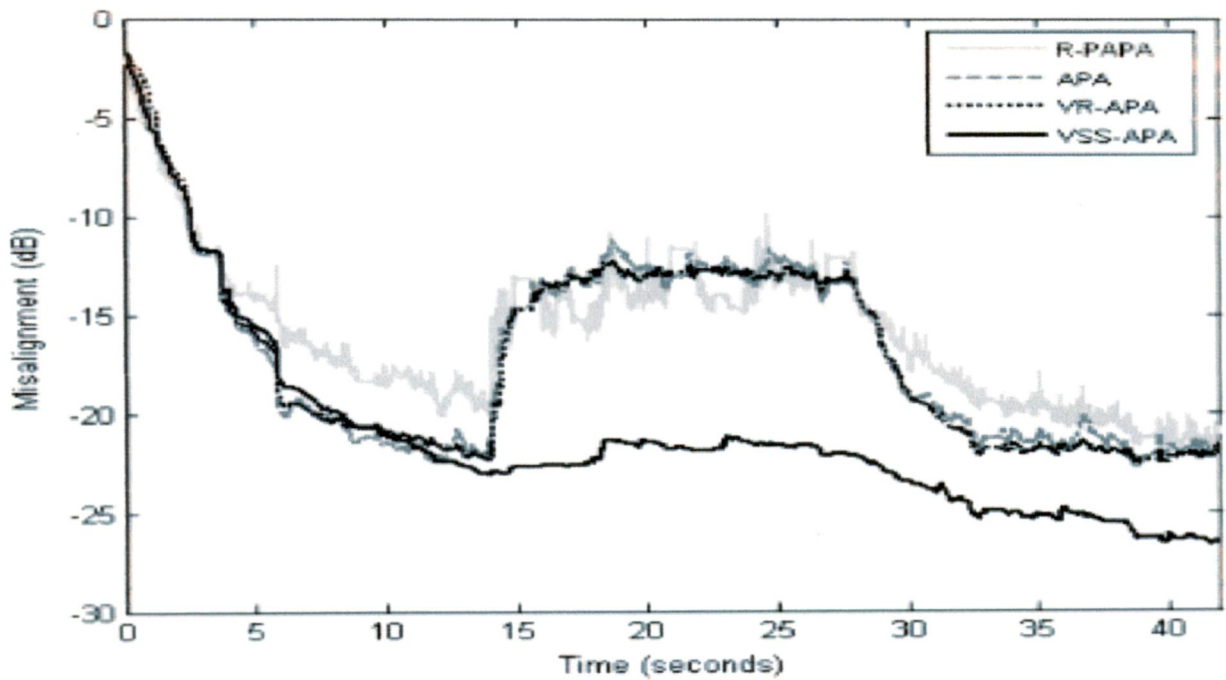
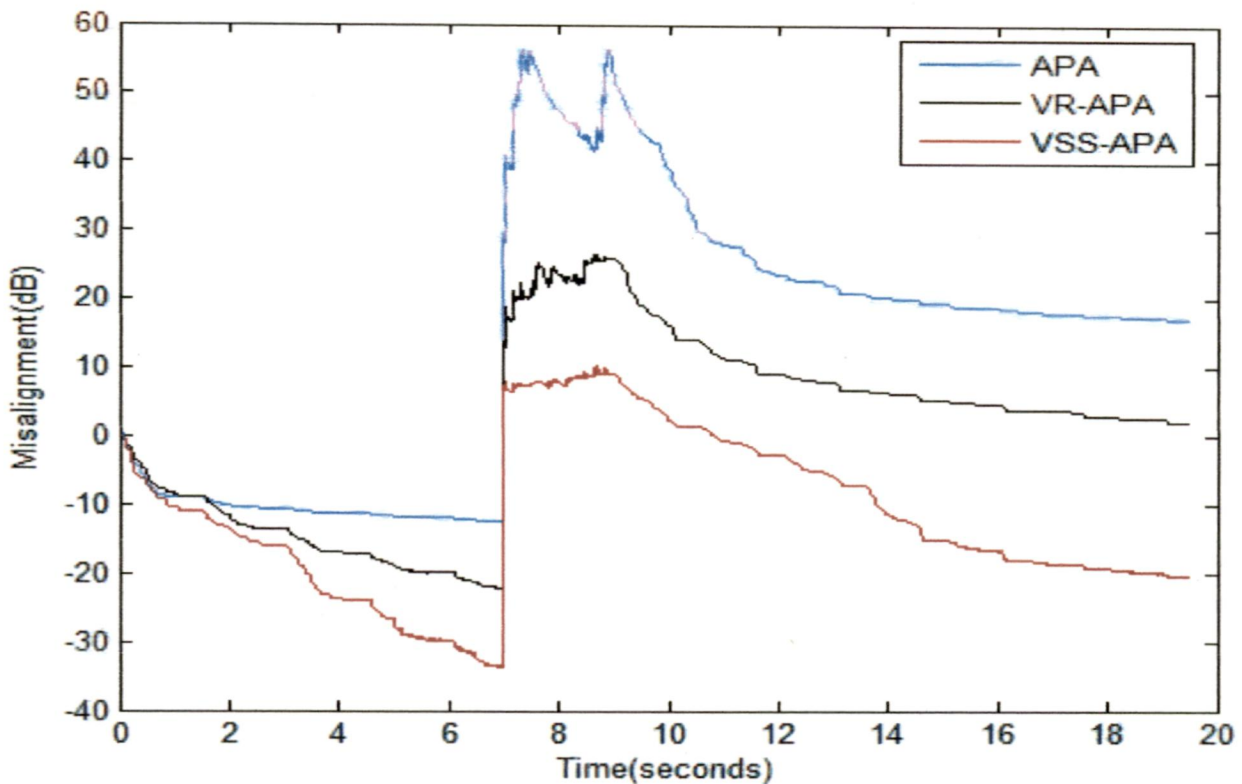


Figure 5.13 Misalignments of APA with $\mu=0.2$, VR-APA with $\zeta=1$, and VSS-APA, the background noise variation at time 14 sec, for a period of 14sec (SNR decreases from 20dB to 10dB), $L=N=512$ [18].

Double-Talk Scenario:

The most challenging situation in echo cancellation is the double-talk case. Such a scenario is considered in the simulations using the speech signals from Figures 5.14 and 5.15. In Figure 5.14, simulations is carried out to evaluate the performance of three algorithms; APA, VR-APA and VSS-APA in the presence of double talk and without using any DTD. From Figure 5.14, it can be noticed that the presence of double talk (i.e., presence of both far-end and near-end speech) causes the adaptive algorithms to diverge . The near end speech signal used in simulations (Figure 5.1(c)) extends over a period of 2sec (from 7-9sec), hence the algorithms diverges for the same period in Figure 5.14.

A simple solution to the double-talk is to use a Double-Talk Detector (DTD) to enhance the performance of the algorithms during double-talk periods. The simulation in Figure 5.14 is repeated using a Giegel DTD. Its settings are chosen assuming a 6-dB attenuation from far-end to double-talk, i.e., the threshold T is equal to 0.5 and the hangover time is set to 240 sec [18], i.e., once the double talk is detected the adaptation is halted for 240 future samples (0.03 sec at 8-kHz sampling rate). From Figure 5.15, it can be noticed that the performance is improved in the presence of a DTD and VSS-APA outperforms the other two.



Fig

ure 5.14 Misalignments of APA with $\mu=0.2$, VR-APA with $\zeta=1$, and VSS-APA, double-talk scenario, without DTD, SNR=40dB.

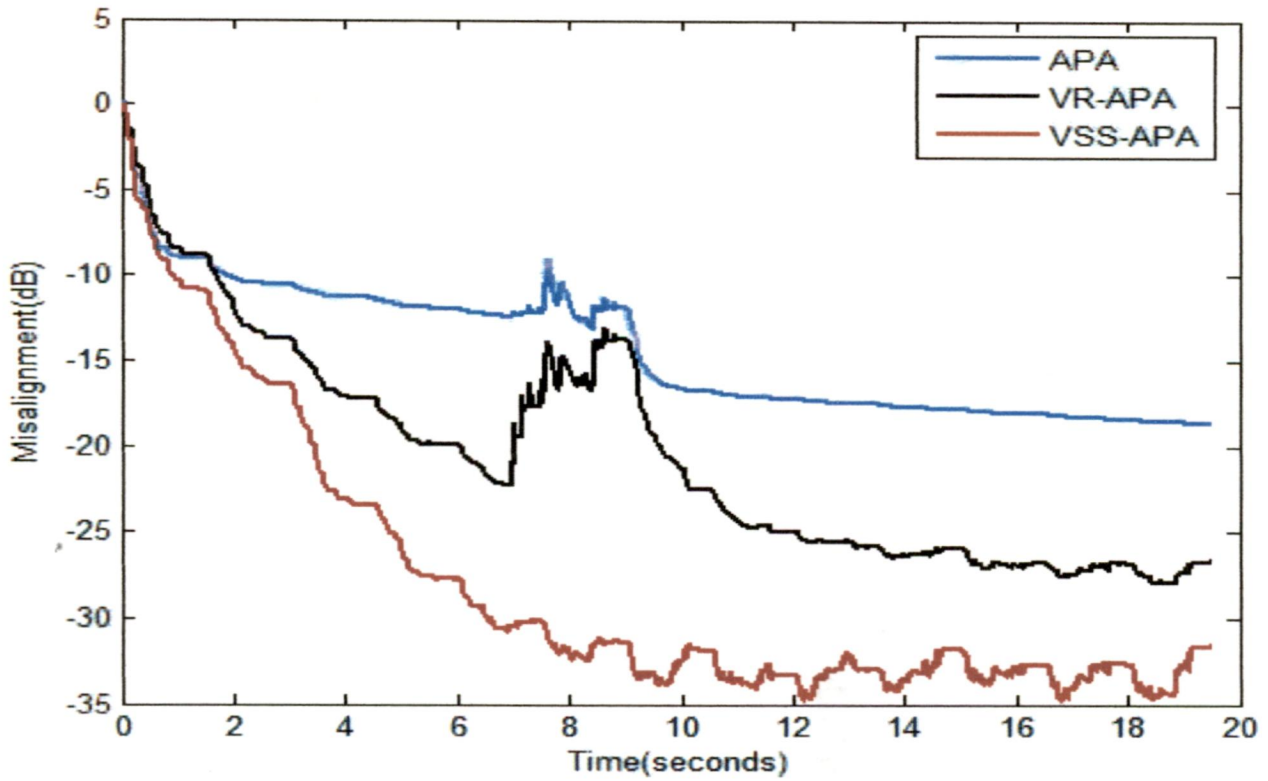


Figure 5.15 Misalignments of APA with $\mu=0.2$, VR-APA with $\zeta=1$, and VSS-APA, double-talk scenario, with Giegel DTD, SNR=40dB.

5.2 Under-Modeling Case

Simulations is performed in this section using the entire length of the acoustic impulse response from Figure 5.1(a), while the length of the adaptive filter remains the same ($L=500$, $N=1000$). In this case the expression of the misalignment is evaluated by padding the vector of the adaptive filter coefficients with $N-L$ zeros, i.e., $20 \log_{10} \left(\left\| \mathbf{h} - [\hat{\mathbf{h}}(n) \mathbf{0}_{N-L}^T]^T \right\| / \left\| \mathbf{h} \right\| \right)$.

Single-Talk Scenario:

In the first set of simulations, the performance of APA, VR-APA and VSS-APA is evaluated in a single-talk case, using a projection order $p = 2$; the results are presented in Figure 5.16, from which it is clear that the performance is affected by the presence of extra under-modeling noise.

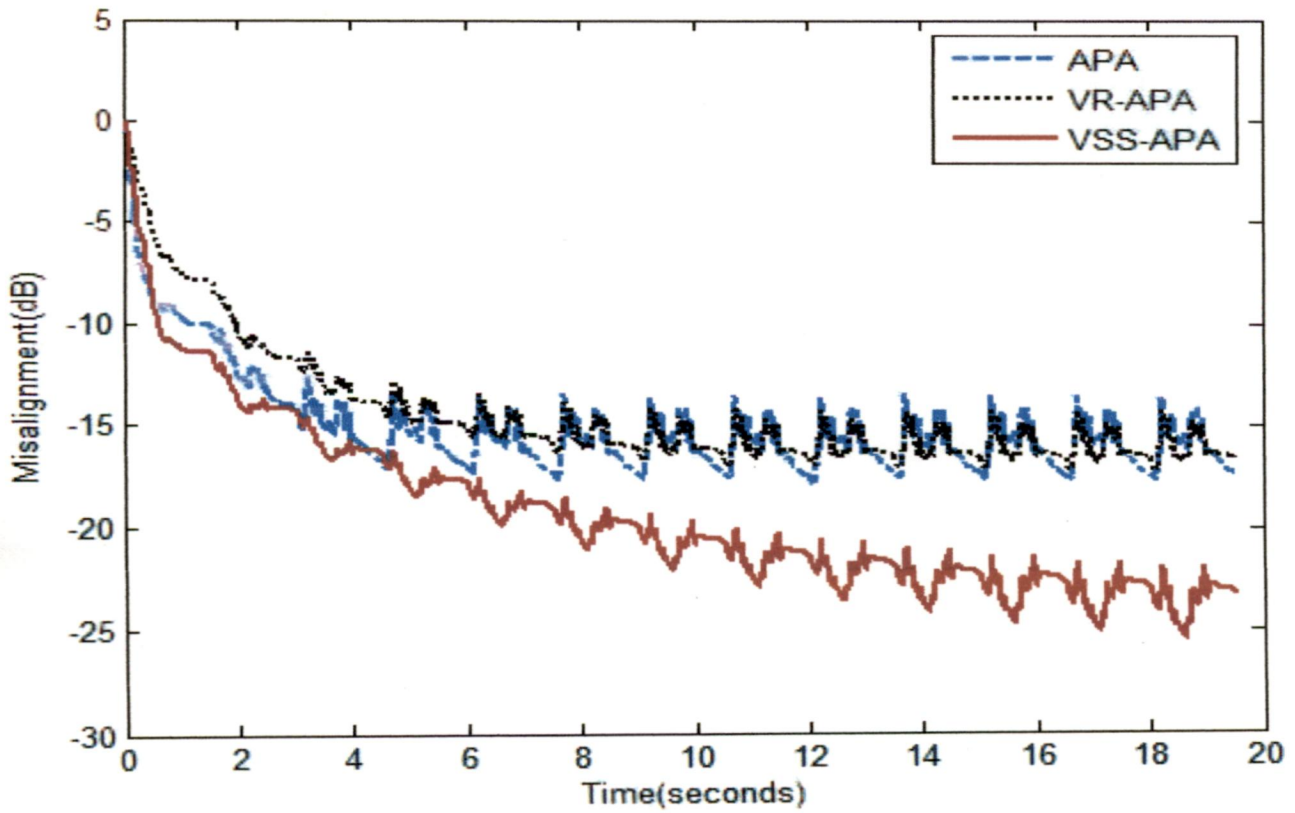


Figure 5.16 Misalignments of APA with $\mu=0.2$, VR-APA with $\zeta=1$, and VSS-APA, single-talk scenario, $L=500$, $N=1000$, SNR=40dB.

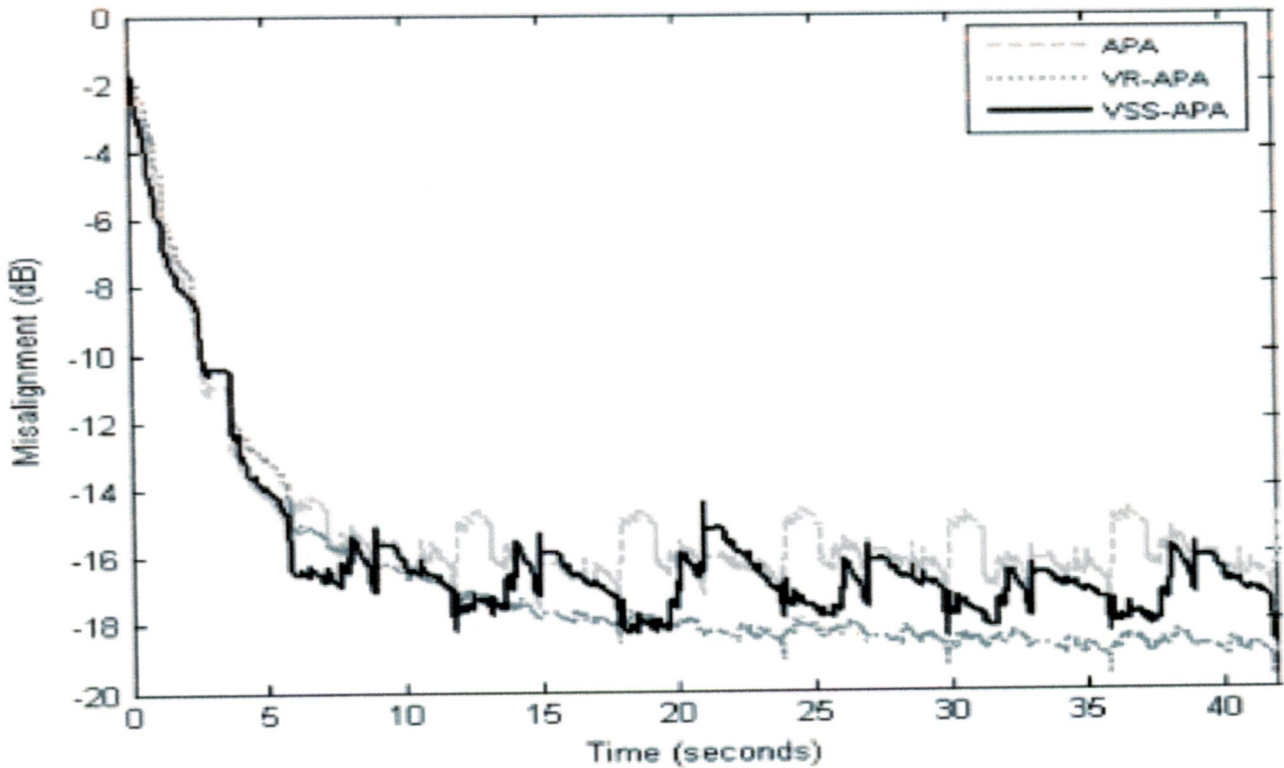


Figure 5.17 Misalignments of APA with $\mu=0.2$, VR-APA with $\zeta=1$, and VSS-APA, single-talk scenario, $L=512$, $N=1024$, SNR=20dB [18].

Conclusion

In this dissertation, a variable step-size affine projection algorithm suitable for AEC application has been discussed in detail. Based on the results obtained, following conclusions are drawn:

The trade-off between the high convergence rate and low misalignment associated with a fixed step-size adaptive filter algorithm was overcome using a variable step-size adaptive filter algorithm.

From the simulations performed in chapter 6, VSS-APA was found to be more robust to near-end signal variations like increase of the background noise or double-talk, and the effect of under-modeling noise on this algorithm was less as compared to classical APA.

6.1 Future Scope

The algorithm discussed in this thesis presents a solution for single channel acoustic echoes. However, most often in real life situations, multichannel sound is a major issue in telecommunication. For example, when there is a group of people in a teleconference environment and everybody is busy talking, laughing or just communicating with each other results in multichannel sound. Since there is just a single microphone the other end will hear just a highly incoherent monographic sound. In order to handle such situations in a better way the echo cancellation algorithm discussed during this research should be extended for the multichannel case.

REFERENCES

- [1] S. L. Gay and J. Benesty, "Acoustic Signal Processing for Telecommunication", chapter 2, Boston, MA: Kluwer, 2000.
- [2] J. Benesty and Y. Huang, "Adaptive Signal Processing—Applications to Real-World Problems", chapters 1 and 2, Berlin, Germany: Springer-Verlag, 2003.
- [3] C. Breining, P. Dreiseitel, E. Haensler, A. Mader, B. Nitsch, H. Puder, T. Schertler, G. Schmidt, and J. Tilp, "Acoustic echo control—An application of very high order adaptive filters," *IEEE Signal Process. Mag.*, vol. 16, no. 4, pp. 42–69, Jul. 1999.
- [4] J. Benesty, T. Gaensler, D. R. Morgan, M. M. Sondhi, and S. L. Gay, "Advances in Network and Acoustic Echo Cancellation", chapters 2 and 4, Berlin, Germany: Springer-Verlag, 2001.
- [5] S. Haykin, "Adaptive Filter Theory", chapters 1, 2, 4, 5 and 6, 4th ed. Upper Saddle River, NJ: Prentice-Hall, 2002.
- [6] A. H. Sayed, "Fundamentals of Adaptive Filtering", chapters 1 and 2, New York: Wiley, 2003.
- [7] K. Ozeki and T. Umeda, "An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties," *Electron. Commun. Jpn.*, vol. 67-A, no. 5, pp.19–27, May 1984.
- [8] S. L. Gay and S. Tavathia, "The fast affine projection algorithm," in *Proc. 1995 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Detroit, MI, vol. 5, pp. 3023–3026, May 1995.
- [9] Tanaka, S. Makino, and L. Kojima. "A block exact fast affine projection algorithm," *IEEE Trans. Speech and Audio Processing*, vol 7, pp.79-86,1999.
- [10] R.W. Harris, D.M. Chabries, and F.A. Bishop, "A variable step size adaptive filter algorithm," *IEEE Transactions on Acoustics, Speech, Signal Processing*, vol. 34, pp. 309–316, April 1986.
- [11] H.C. Chin, A.H. Sayed, and W.J. Song, "Variable step size NLMS and affine projection algorithms," *IEEE Signal Processing Letter*, vol. 11, pp. 132–135, Feb 2004.
- [12] J. Benesty, H. Rey, L. Rey Vega, and S. Tressens, "A nonparametric VSS NLMS algorithm," *IEEE Signal Process. Lett.*, vol. 13, no. 10, pp. 581–584, Oct. 2006.

- [13] C. Paleologu, S. Ciochinã, and J. Benesty, "Variable step-size NLMS algorithm for under-modeling acoustic echo cancellation," *IEEE Signal Process. Lett.*, vol. 15, no. 1, pp. 5–8, Jan. 2008.
- [14] Sadaoki Furui and M. Mohan Sondhi, "Advances in Speech Signal Processing", chapter 2, Marcel Dekker, Inc, 1992.
- [15] Cameron, Peter J. (1991), "Projective and polar spaces", QMW Maths Notes, 13, London: Queen Mary and Westfield College School of Mathematical Sciences.
- [16] Jean Gallier, "Geometric Methods and Applications", chapter 2, Springer.
- [17] Paulo S.R. Diniz, "Adaptive Filtering Algorithms and Practical Implementation", Chapters 2,3 and 4, Kluwer Academic Publishers, 1997.
- [18] C. Paleologu, J. Benesty, and S. Ciochina, "A variable step-size affine projection algorithm designed for acoustic echo cancellation," *IEEE Trans. Audio, Speech, Language Processing*, vol. 16, pp. 1466-1478, Nov. 2008.
- [19] H. Rey, L. Rey Vega, S. Tressens, and J. Benesty, "Variable explicit regularization in affine projection algorithm: Robustness issues and optimal choice," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 2096–2108, May 2007.
- [20] D. L. Duttweiler, "A twelve-channel digital echo canceler," *IEEE Trans. Commun.*, vol. 26, no. 5, pp. 647–653, May 1978.
- [21] J. Benesty, D. R. Morgan, and J. H. Cho, "A new class of doubletalk detectors based on cross-correlation," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 2, pp. 168–172, Mar. 2000.
- [22] www.comm.pub.ro/plant