NEAR EAST UNIVERSITY
LIBRARY
1988 - LEFKOŞA

48 27 08 45 53
NEU

NEAR EAST UNIVERSITY
LIBRARY
1988 - LEFKOŞA

NEAR EAST UNIVERSITY

Faculty of Engineering

Department of Electrical and Electronic Engineering

ADAPTIVE ECHO CANCELLATION

Graduation Project
EE- 400

Student: Ashraf Safi (970935)

Supervisor: Prof. Dr. Fakhreddin Mamedov

Lefkoşa - 2000

DG

ACKNOWLEGDMENT

- This project present highly technical subject matter of the technique and the utilization of Adaptive echo cancellation.

This project was not possible to be prepared without the guidance and the support of my supervisor Prof. Dr. Fakherddin Mamedov, the chairman of Electrical and Electronic Engineering Department.

I am indebted to him for his complete support and showing me the guidance throughout all the stages of the preparation, and providing his constructive comments.

So I would like to take this opportunity to thank Prof. Dr. Fakerddin for every single help and support, not just throughout this project but also through the courses which he provides to the students in The Department of Electrical and Electronic Engineering, because through these courses I have gained a lot of knowledge which helped me in the preparation of this project.

ABSTRACT

The contamination of a desired signal by unwanted and unpredictable noise is a problem often encountered in digital communication and control system. When a spectral overlap between signal and noise occurs or band occupied by the noise is unknown or varies with time .it is necessary to design filter adapted to changing of the signal characteristics.

One of the important problems in long distance communication system over a telephone channel is removing residual and far end echo signals.

Aim of this thesis is the analysis and interpretation of the adaptive noise canceller based on LMS algorithm for removing hybrid and acoustic echo noises within the digital environment for improving voice quality of the long distance telecommunication systems.

CONTENTS

CHAPTER 1

Introduction to Adaptive Echo Cancellation

Overview:

The performance limits of adaptive echo cancellation techniques are investigated. In particular we analyze the effects of signal characteristics such as auto and cross correlation on the achievable echo suppression. Techniques to enhance signal characteristics such as to improve both the learning ability and the steady state echo suppression quality are identified. A nice feature of our work is that it links in a natural way the complexity of the learning task (via the dimension of the adapted parameter vector), the available information (via the signal characteristics) to the achievable echo suppression quality. A number of papers are currently in a review process for journal publication

1.1 Definition

Wireless phones are increasingly being regarded as essential communications tools, dramatically impacting how people approach 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. A key technology to provide near-wire line voice quality across a wireless carrier's network is echo cancellation.

Subscribers use speech quality as the benchmark for assessing the overall quality of a network. Regardless of whether or not this is a subjective judgment; it is the key to maintaining subscriber loyalty. For this reason, the effective removal of hybrid and acoustic echo inherent within the digital cellular infrastructure is the key to maintaining and improving perceived voice quality on a call. This has led to intensive research into the area of echo cancellation, with the aim of providing solutions that can reduce background noise and remove hybrid and acoustic echo before any transcoder processing. By employing this technology, the overall efficiency of the coding can be enhanced, significantly improving

the quality of speech. This tutorial discusses the nature of echo and how echo cancellation is helpful in making mobile calls meet acceptable quality standards.

1.2 History of Echo Cancellation

The late 1950s marked the birth of echo control in the telecommunications industry with the development of the first echo-suppression devices. These systems, first employed to manage echo generated primarily in satellite circuits, were essentially voice-activated switches that transmitted a voice path and then turned off to block any echo signal. Although echo suppressers reduced echo caused by transmission problems in the network, they also resulted in choppy first syllables and artificial volume adjustment. In addition, they eliminated double-talk capabilities, greatly reducing the ability to achieve natural conversations.

Echo-cancellation theory was developed in the early 1960s by AT&T Bell Labs, followed by the introduction of the first echo-cancellation system in the late 1960s by COMSAT TeleSystems (previously a division of COMSAT Laboratories). COMSAT designed the first analog echo canceller systems to demonstrate the feasibility and performance of satellite communications networks. Based on analog processes, these early echo-cancellation systems were implemented across satellite communications networks to demonstrate the network's performance for long-distance, cross-continental telephony. These systems were not commercially viable, however, because of their size and manufacturing costs.

In the late 1970s, COMSAT TeleSystems developed and sold the first commercial analog echo cancellers, which were mainly digital devices with an analog interface to the network. The semiconductor revolution of the early 1980s marked the switch from analog to digital telecommunications networks. More sophisticated digital interface, multichannel echo-canceller systems were also developed to address new echo problems associated with long-distance digital telephony systems. Based on application-specific integrated circuit (ASIC) technology, these new echo canceller's utilized high-speed digital signal-processing techniques to model and subtract the echo from the echo return path. The result was a new digital echo-cancellation technique that outperformed existing suppression-based techniques, creating improved network performance.

The 1990s have witnessed explosive growth in the wireless telecommunications industry, resulting from deregulation that has brought to market new analog and digital wireless handsets, numerous network carriers, and new digital network infrastructures such as TDMA, CDMA, and GSM. According to the Cellular Telecommunications Industry Association (CTIA), new subscribers are driving the growth of the wireless market at an annual rate of 40 percent. With wireless telephony being widely implemented and competition increasing as new wireless carriers enter the market, superior voice transmission quality and customer service have now become key determining factors for subscribers evaluating a carrier's network. Understanding and overcoming the inherent echo problems associated with digital cellular networks will enable network operators and Telco's to offer subscribers the network performance and voice quality they are demanding today.

1.3 Types of Echo

1.3.1 Acoustic Echo

Acoustic echo is generated with analog and digital handsets, with the degree of echo related to the type and quality of equipment used. This form of 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 encoding/decoding devices (decoders) 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 wireless caller.

Acoustic echo was first encountered with the early video/audio conferencing studios and now also occurs in typical mobile situations, such as when people are driving their cars. In this situation, sound from a loudspeaker is heard by a listener, as intended. However, this same sound also is picked up by the microphone, both directly and indirectly, after bouncing off the roof, windows, and seats of the car. The result of this reflection is the creation of multipath echo and multiple harmonics of echo, which, unless eliminated, are

transmitted back to the distant end and are heard by the talker as echo. Predominant use of hands-free telephones in the office has exacerbated the acoustic echo problem.

Acoustic echo cancellation is required in order to provide full duplex, fully interruptible speech. The acoustic echo canceller functions by modeling the speech being passed to the loudspeaker and removing any echoes picked up by the microphone. This type of operation necessitates a much more complex unit than is used in telephony in order to remove the many acoustic (multipath) echoes generated with each syllable of speech. The tail circuit requirement, or the amount of time the canceller has to hold the power.

1.3.2 Hybrid Echo

Hybrid echo is the primary source of echo generated from the public-switched telephone network (PSTN). This electrically generated echo is created as voice signals are transmitted across the network via the hybrid connection at the two-wire/four-wire PSTN conversion points, reflecting electrical energy back to the speaker from the four-wire circuit.

Hybrid echo has been around almost since the advent of the telephone itself. The signal path between two telephones, involving a call other than a local one, requires amplification using a four-wire circuit. Although not a factor in itself on digital cellular networks, hybrid echo becomes a problem in PSTN–originated calls. The cost and cabling required rules out the idea of running a four-wire circuit out to the subscriber's premise from the local exchange. For this reason, an alternative solution had to be found. Hence, the four-wire trunk circuits were converted to two-wire local cabling, using a device called a "hybrid" (see Figure 1.3.1).

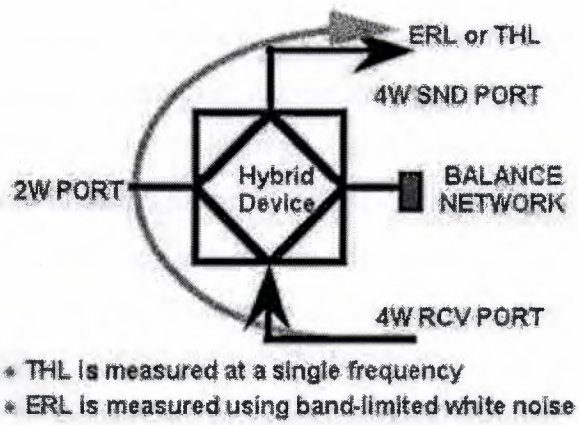


Figure 1.3.1. Hybrid Echo

Unfortunately, the hybrid is by nature a leaky device. As voice signals pass from the four-wire to the two-wire portion of the network, the energy in the four-wire section is reflected back on itself, creating the echoed speech. Provided that the total round-trip delay occurs within just a few milliseconds (i.e., within 28 ms), it generates a sense that the call is live by adding side tone, which makes a positive contribution to the quality of the call.

In cases where the total network delay exceeds 36 ms, however, the positive benefits disappear, and intrusive echo results. The actual amount of signal that is reflected back depends on how well the balance circuit of the hybrid matches the two-wire line. In the vast majority of cases, the match is poor, resulting in a considerable level of signal reflecting back. This is measured as echo return loss (ERL). The higher the ERL, the lower the reflected signal back to the talker, and vice versa.

1.4 Causes of Echo

Acoustic echo apart, background noise is generated through the network when analog and digital phones are operated in hands-free mode. As additional sounds are directly and indirectly picked up by the microphone, multipath audio is created and transmitted back to the talker. The surrounding noise, whether in an automobile or in a crowded, public environment, passes through the digital cellular decoder causing distorted speech for the wireline caller.

Digital processing delays and speech-compression techniques further contribute to echo generation and degraded voice quality in wireless networks. Delays are encountered as signals are processed through various routes within the networks, including copper wire, fiber optic lines, microwave connections, international gateways, and satellite transmission. This is especially true with mixed technology digital networks, where calls are processed across numerous network infrastructures.

Echo-control systems are required in all networks that produce one-way time delays greater than 16 ms. In today's digital wireless networks, voice paths are processed at two points in the network within the mobile handset and at the radio frequency (RF) interface of the network. As calls are processed through vocoders in the network, speech

processing delays ranging from 80 ms to 100 ms are introduced, resulting in an unacceptable total end-to-end delay of 160 ms to 200 ms. As a result, echo cancellation devices are required within the wireless network to eliminate the hybrid and acoustic echoes in a digital wireless call.

1.5 The Combined Problem on Digital Cellular Networks

To deal with hybrid echo created by vocoder processing delays, it is mandatory for digital cellular mobile calls to have a group echo canceller installed–even for local calls. As a result, all calls on to the PSTN must pass through an echo canceller to remove what would otherwise be a noticeable and annoying echo, as shown in Figure 1.5

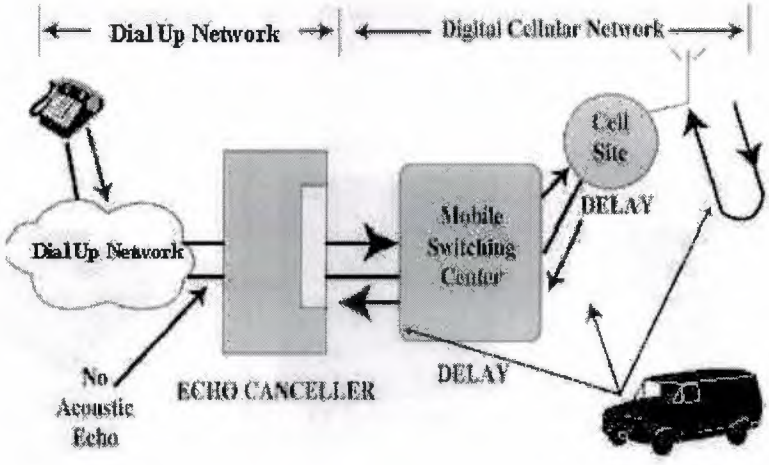


Figure 1.5, Digital Cellular Network

For example, consider a digital cellular mobile user who makes a call to the PSTN without an echo canceller in place. The user would hear his or her own speech being echoed back 180 ms or more later, even if the called person is in the same locality. The mobile user will either be using a hands-free system installed in his or her vehicle or a hand portable. In either case, these units will involve the occurrence of direct and indirect coupling between the microphone and the speaker, creating acoustic echo. In this situation, however, it is the PSTN user who suffers by experiencing poor speech quality. Hence, the echo canceller installed in the digital cellular network must be capable of handling both sources of echoes.

1.6 Process of Echo Cancellation

In modern telephone networks, echo cancellers are typically positioned in the digital circuit, as shown in Figure 1.6. The process of canceling echo involves two steps. First, as the call is set up, the echo canceller employs a digital adaptive filter to set up a model or characterization of the voice signal and echo passing through the echo canceller. As a voice path passes back through the cancellation system, the echo canceller compares the signal and the model to cancel existing echo dynamically. This process removes more than 80 to 90 percent of the echo across the network. The second process utilizes a non-linear processor (NLP) to eliminate the remaining residual echo by attenuating the signal below the noise floor.

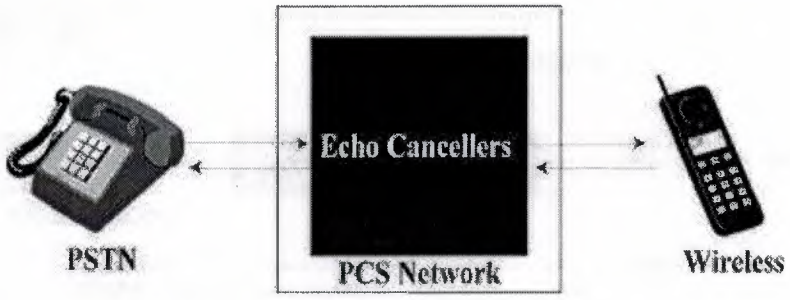


Figure 1-7.1. Typical Location of Echo Cancellers

Today's digital cellular network technologies, namely TDMA, CDMA, and GSM, require significantly more processing power to transmit signal paths through the channels. As these technologies become even more sophisticated, echo control will be more complex. Echo cancellers designed with standard digital signal processors (DSPs), which share processing time in a circuit within a channel or across channels, provide a maximum of only 128 ms of cancellation and are unable to cancel acoustic echo. With network delays occurring in excess of 160 ms in today's mixed-signal network infrastructures, a more powerful, application-specific echo-cancellation technology is required to control echo across wireless networks effectively

1.7 Controlling Acoustic Echo

In echo cancellation, complex algorithmic procedures are used to compute speech models. This involves generating the sum from reflected echoes of the original speech, and then subtracting this from any signal the microphone picks up. The result is the purified speech of the person talking. The format of this echo prediction must be learned by the echo canceller in a process known as adaptation. It might be said that the parameters learned from the adaptation process generate the prediction of the echo signal, which then forms an audio picture of the room in which the microphone is located. Figure 1-8.1, shows the basic operation of an echo canceller in a conference room type of situation.

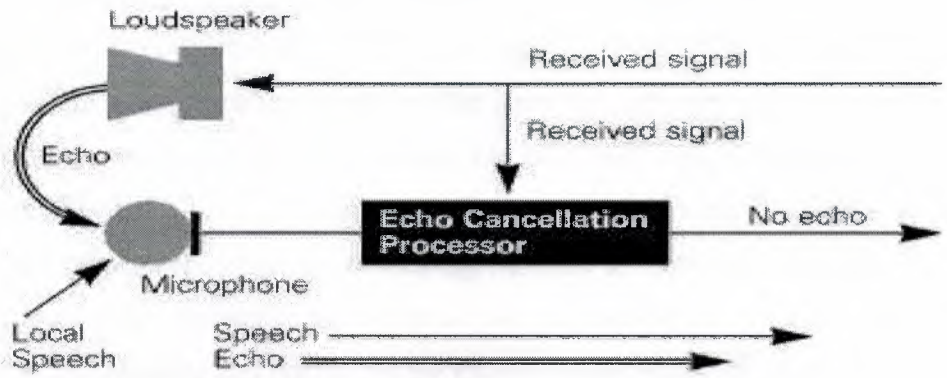


Figure 1-8.1. Operation of an Acoustic Echo Canceller

During the conversation period, this audio picture constantly alters, and, in turn, the canceller must adapt continually. The time required for the echo canceller to fully learn the acoustic picture of the room is called the convergence time. The best convergence time recorded is 50 ms, and any increase in this number results in syllables of echo being detected.

Other important performance criteria involve the acoustic echo canceller's ability to handle acoustic tail circuit delay. This is the time span of the acoustic picture and roughly represents the delay in time for the last significant echo to arrive at the microphone. The optimum requirement for this is currently set at 270 ms–any time below

this could result in echoes being received by the microphone outside the ability of the echo canceller to remove them, and hence in participants hearing the echoes.

Another important factor is acoustic echo return loss enhancement (AERLE). This is the amount of attenuation which is applied to the echo signal in the process of echo cancellation–i.e., if no attenuation is applied, full echo will be heard. A value of 65 dB is the minimum requirement with the non-linear processor enabled, based on an input level of -10 dBm white noise electrical and 6 dB of echo return loss (ERL).

The canceller's performance also relies heavily on the efficiency of a device called the center clipper, or non-linear processor. This needs to be adaptive and has a direct bearing on the level of AERLE that can be achieved.

1.8 Controlling Complex Echo in a Wireless Digital Network

Although acoustic echo is present in every hands-free mobile call, the amount of echo depends on the particular handset design and model that the mobile user has. On the market are a few excellent handsets that limit the echo present, but, due to strong price pressures, most handsets do not control the echo very well at all–in fact, some phones on the market have been determined to have a terminal compiling loss of 24 dB. Echo becomes a problem when the processing inherent to the digital wireless network adds an additional delay (typically in excess of 180 ms round-trip). This combination makes for totally unacceptable call quality for the fixed network customer, as shown in Figure 1.8.

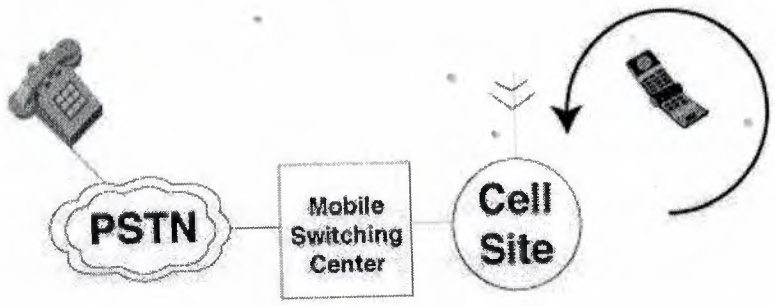


Figure 1.8. Acoustic Echo in a Mobile Environment

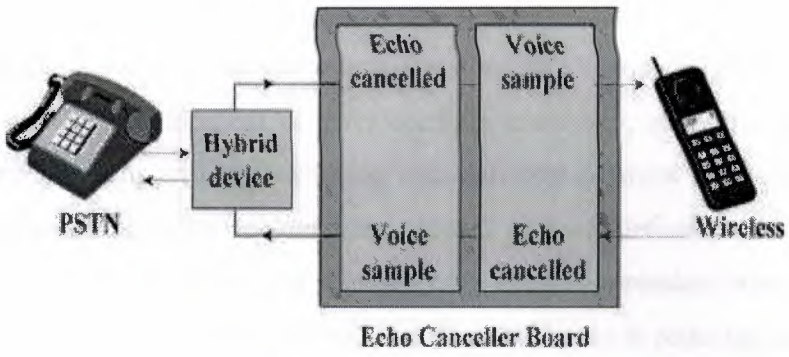


Figure 1.8.1, Bi-directional Echo Cancellation

This back-to-back configuration ensures a high audio quality for both PSTN and mobile customers. In addition, the echo canceller's software configuration is designed to provide a detailed analysis of background noises, including acoustic echo from the mobile user's end. Some echo cancellers incorporate a user-settable network delay, which enables network operators to fine-tune the echo control to suit their parameters via a menu option on the canceller's hand-held terminal or on the network management system (NMS).

1.9 Room for Improvement in the Handset

Applying effective echo control via the echo-cancellation platform is one way of improving the overall call clarity on digital cellular networks. Another derives from improvements that must be made within the handset or terminal itself. There also is considerable room to enhance the network itself, focusing principally on decoder development.

Recent headlines have charted the ongoing commercial battles regarding which digital technologies will eventually emerge as the winners, as equipment manufacturers fight it out. However, this public battle will soon be overshadowed by another battle concerning handsets. At present, there are four major players in the digital cordless market. Europe has cordless telephony (CT2) and digital European cordless telephony (DECT), while Japan has the personal handy-phone system (PHS) and the United States has personal communications services (PCS).

Connecting directly into the plain old telephone system, CT2 was one of the first digital Technologies to provide low-cost mobile phones. Although the technology worked Well, it had a fundamental problem: it could not handle cell handovers. DECT and GSM have overcome this problem and will eventually dominate European cellular services.

During the development of early cordless telephony, attention was paid to basic and enhanced functions and inter-working with different network architectures. While the early generation of handsets looked very elegant and aesthetically pleasing, very little attention was paid to designing the handset with echo suppression/cancellation in mind. The result was that they looked good but were extremely poor at reducing acoustic echo.

In the setting of standards for GSM and PCS, handset design and the impact of different design approaches on call quality were researched. As a result, recommendations stated a range of parameters, including side-tone tolerance and echo return loss performance. With the resultant advent of new recommendations with much tighter requirements for handsets, there is a call for greatly improved designs to be implemented. This, complemented by ongoing improvements in network technology and echo cancellation techniques, will bring digital wireless telephony much closer to matching wireless quality.

1.10 Echo Cancellation System For Radio Telephony

The customer has developed technology for major manufacturers in the telecommunications industry. This has included a host of popular electronic devices, from cellular and cordless telephones, to computers and digital television products.
Microsystems Engineering has worked extensively with the customer over a number of years developing echo cancellation systems for radiotelephony projects. The project described here is an example of a recent echo cancellation system designed and implemented by Microsystems Engineering.

Many radiotelephony devices require additional echo control to be incorporated into the system. The main reason for this is the group delay usually imposed by the radio link protocol. Sources of echo that would not normally be noticeable to the user become annoying due to the 10 - 20ms round trip delay that often exists between the portable part and the fixed part of the system.

The diagram below illustrates the elements of a typical radiotelephony system that relate to echo and its control.

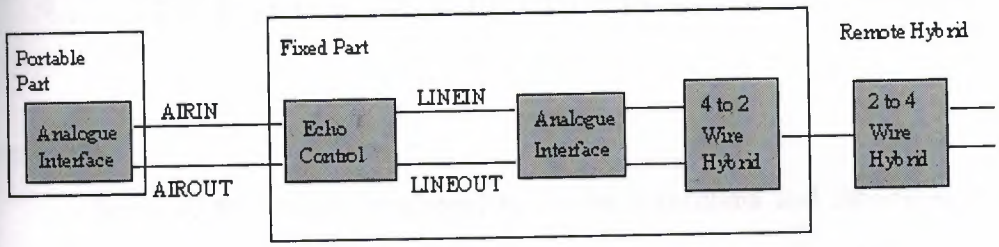


Figure 1.10, typical radiotelephony system

The echo control part of the system would generally reside at the base station (in the Fixed Part of the above diagram) and would be responsible for three kinds of echo:

- Coupling at the portable part resulting in echo of the AIROUT signals back into the AIRIN signal. This is generally of the order of 20-21ms. For the PSTN equipment to suppress this an artificial echo signal needs to be added by the system.
- Reflection at the fixed part 4-wire to 2-wire hybrid resulting in short delay echo of LINEOUT signals in the LINEIN signal (0- 4ms). This is cancelled using an adaptive FIR algorithm.
- Reflection at the exchange 2-wire to 4-wire hybrid resulting in long delay echo of LINEOUT signals in the LINEIN signal. This can be between 0 and 70ms. This is reduced to acceptable levels using a soft suppressor algorithm.

CHAPTER 2

Acoustic Echo Cancellation - Recent Developments

Overview:

Some of the recent developments in the algorithms and structures of acoustic echo cancellers (AECs) are reviewed in this paper. Most existing echo cancellers are designed with adaptive transversal FIR digital filters, and based on variants of the least mean square (LMS) and the least square (LS) algorithms. Thus, the concepts of conventional LMS and LS algorithms based on transversal FIR filters are first recalled. Then, other methods are presented that can be used to overcome the inherent problems of slow convergence in LMS algorithm and high computation in LS algorithm. In general, all AECs can be classified into two large groups, namely the full band and sub band AECs. The direct method of full band AECs is first presented which includes the excited normalized LMS, the exponentially weighted step-size normalized LMS, the fast Newton transversal filter and the IIR gradient instrumental variable algorithms. The special structure of sub band AECs is attracting current research activities. The performance of these AECs depends on the structures of the echo canceller and the adopted adaptive algorithms. The synthesis-dependent, synthesis independent and various delay less structures are presented. These structures use variations of the LMS and LS algorithm as the adaptive algorithms. A modified version of the recently introduced fast affined projection algorithm is also discussed.

2.1 Introduction

In a telephone connection between one or more hands-free telephones, a feedback-coupling path is set up between the loudspeaker and microphone at each end. This acoustic coupling is due to the reflection of loud-speaker's sound from walls, floor, ceiling, windows and other objects back to the microphone 1 . Adaptive can the author is with the Signal Processing Group, Dept. of Applied Electronics, Charmers University of Technology, Gothenburg, Sweden. The coupling can also due to the direct path from

loudspeaker to microphone .Collation of this acoustic echo is becoming very crucial in hands-free telephony Applications. For example, during the hands-free operation of a cellular telephone that is used in an enclosed environment, i.e. in a vehicle or room, and in teleconference or videoconference meetings.

The effects of an echo depend on the time de-lay between the incident and the reflected waves, the strength of the reflected waves and the number of paths through which the waves are reflected. If the time de-lay is not long, the acoustic echo can be perceived as soft reverberation, which adds artistic quality in concert hall. However, echo arriving a few tens of milliseconds or more after the direct sound will be highly undesirable. Such a long delay can be caused by program time over long distances, the coding of the transmitted signals or the end acoustic echo path itself. The cancellation of acoustic echo differs from the cancellation of telephone network line echo due to the different nature of the echo paths. Acoustic echo cancellation is far more challenging than line echo cancellation for a number of reasons:

The duration of the impulse response of the acoustic Echo path is usually several times longer (100 to 400msec) which implies that impracticably large transversal FIR filters with thousands of taps are required.

The characteristic of the echo path is more non-stationary, e.g. due to opening or closing of a door or a moving person, while the line echo path is almost stationary once a call connection is established.

Acoustic echo is due to reflection of the signal from a multitude of different paths, e.g. off the walls, the floor, the ceiling, the windows, etc. The echo path is not well approximated by FIR or IIR linear filters because it has a mixture of Perhaps, modeling the echo path as a recursive IIR filter can reduce the number of filter coefficients. Linear and non-linear characteristics. The reflection of acoustic signals inside a room is almost linearly distorted, but the loudspeaker introduces non-linearity's. The main causes of this non-linearity are the suspension non-linearity, which affects distortion at low frequency and the in homo-geniality of flux density, which produces non-linear distortion at large output signal levels.

Due to these reasons, the AECs require more computing power to compensate for the length of impulse response and to obtain a faster converging algorithm. One of the

problems of effective AECs design is the event of double-talking. The situation that must be avoided is interpreting the near-end signal, *x(n)* as part of the true error signal, as shown in Fig. 2, which results in making large corrections to the estimated echo path in a doomed-to-failure attempt to cancel it. In order to avoid this possibility, the tap weights must not be updated as soon as double-talking is detected. The design of a good double-talking detector is difficult. The AECs may diverge due to midsection or the time delay before the decision is made. Even with the assumption of a fast-acting detector, there is still a possibility of changes occurring in the acoustic echo path during the time that the canceller is frozen, which leads to in-creased unconcealed echo. The adaptive filter structures and common algorithms are briefly discussed in Sec 2.2. In Sec 2.3, the direct method of full band AECs is presented. The ad-vantage of the full band method is that no special design has to be made on the structure of the canceller. However, the disadvantages are high computation and slow convergence. In Sec.2.4, the sub band AECs is discussed. The basic idea is to divide the signal into a few frequency sub bands and decimate the sampling rate before applying the sub band signals to an adaptive filter. This method yields a faster convergence and a reduction in computational burden.

2.2 Adaptive Filter Structure And Algorithm

The selection of the adaptive filter structure and algorithm for the adaptation will effect the echo canceller's accuracy in estimating the echo path and the speed to adapt to its variation. Naturally, the choice of a filter's structure has a deep effect on the operation of the selected algorithm as a whole. Fig.2.2 shows the general configuration of an adaptive filter, which can be designed for the AEC application. The adaptive filter has two main parts: a filter, whose structure is designed to perform a desired processing function and an adaptive algorithm, for adjusting the coefficients of that filter to improve its performance. The output, $\hat{d}(n)$ is a weighted incoming signal, *y(n)* in a digital adaptive filter. The adaptive algorithm adjusts the weights in the filter to minimize the error, *e(n)* between $\hat{d}(n)$ and *d(n)+(n)*, where *d(y)* is a desired output from the unknown system and *n(n)* is a corrupting ambient noise.

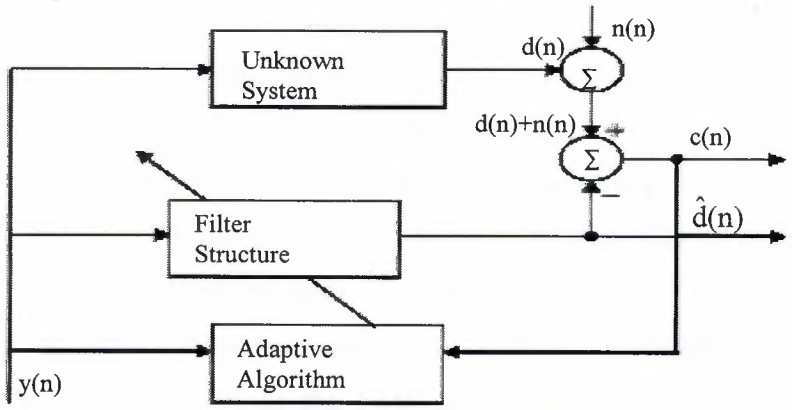


Figure 2.2: General adaptive filter configuration $\hat{d}(n)$

2.2.1. Filter Structures

In practice, a *finite impulse response* (FIR) transversal filter structure is often used because the convergence property of its coefficients to the optimum value is well proven. The major drawback is that as the echo path delay is increased, the number of taps increases proportionately and the convergence speed decreases. The echo canceller may also be an *infinite impulse response* (IIR) filter. The main advantage of an IIR is that a long delay echo can be synthesized by a relatively small number of filter coefficients due to the presence of a feedback loop. However, this feedback loop presents the problem of instability.

2.2.2. Adaptation Algorithm

There are two important basic categories of algorithms for AECs, i.e. the *least mean square* (LMS) and the *least square* (LS) algorithms.

2.2.2.1. LMS algorithm

The approach to examine LMS algorithm is to start with the concept of Wiener filters follows by the method of steepest descent before these concepts are used to derive the conventional LMS algorithm. The cost function can be defined as the mean-squared error

$$J=E[e(n)^2] \qquad [eq\ 1]$$

$$e(n) = d(n)-\hat{d}(n) \qquad [eq\ 2]$$

Where $n(n)$ is assumed to be negligibly small and E denotes the statistical expectation operator. For the cost function J to attain its minimum value, all the elements of the gradient vector ∇J must be simultaneously equal to zero. Under this set of conditions, the filter is said to be optimum in the *mean-squared-error sense*, which produces the minimum mean-squared error J_{min}. At this point the tap weight vector assumes its optimum value W_{opt} that satisfies the Weiner-Hopf equation. The *method of steepest descent* is basic to the understanding of other adaptive algorithms in which gradient-based adaptation is implemented in practice, such as the LMS algorithm. Using $\hat{r}(n) = w(n)^T y(n)$ for a transversal FIR filter, we will obtain

$$\nabla J(n) = -2P + 2Rw(n) \qquad [eq\ 3]$$

where $P=e[y(n)\ r(n)]$, $R=E[y(n)\ y(n)^T]$, superscript T denotes matrix transpose operation and $w(n)$ denotes the value of the tap weight vector, $w=(w_0, w_1,.....,w_{M-1})^T$ at time n. Since $w(n)$ varies with time n, $J(n)$ also varies in a corresponding fashion and this signifies that the estimation error $e(n)$ is non-stationary. The dependence of $J(n)$ on $w(n)$ is visualized as the error-performance surface of the adaptive filter. The adaptive process has the task of continually seeking the minimum point of the surface, where the tap weights take on the optimum value W_{opt}. According to the method of steepest gradient, the updated value of tap weight at time $n+1$ is computed by using the simple recursive relation

$$w(n+1)=w(n)+1/2\ \mu\ (-\nabla J(n)) \qquad [eq\ 4]$$

where μ is a positive real-valued constant. The factor is used to cancel the factor 2 in (3). The equation also shows that successive corrections to the tap weight vector is in the direction of the negative gradient vector which should eventually lead to J_{min} at which point the tap weights equal W_{opt}. Substituting (3) into (4), we will get a simple recursive formula

$$w(n+1) = w(n) + \delta w(n) \qquad [eq\ 5]$$

$$= w(n) + \mu(P - Rw(n)) \qquad [eq\ 6]$$

$$= w(n) + \mu E[y(n)\, e(n)] \qquad [eq\ 7]$$

According to (5) - (7), the correction $\delta w(n)$ is applied to the tap weight vector at time $n+1$. Thus, μ can be referred as the *step-size parameter* that controls the incremental correction

applied to the tap weight vector as we proceed from one iteration cycle to the next. The theory of the LMS algorithm was first introduced in 1960 for adaptive switching by its originators, Windrow and Hoff [28]. The exact measurement of gradient vector requires prior knowledge of vector Panned matrix R that is not possible in reality. Thus, the gradient vector can only be estimated from the available data. Here we have used a hat over some symbols to distinguish them from the values obtained using the steepest descent algorithm. First, Panned R are estimated by using *instantaneous estimates* that are based on sample values of the Tap *y(n)* and *r(n)*

$$\hat{R}(n) = y(n)\, y(n)^T \qquad \textit{[eq 8]}$$

$$\hat{P}(n) = y(n)\, r(n) \qquad \textit{[eq 9]}$$

correspondingly, the new recursive relation is

$$\hat{w}(n+1) = \hat{w}(n) + \mu y(n)(r(n) - \hat{w}(n)^T\, y(n)) \qquad \textit{[eq 10]}$$

$$= \hat{w}(n) + \mu y(n)\, e(n) \qquad \textit{[eq 11]}$$

Comparing with the method of steepest descent, we see that the expectation operator E [.] is missing in the LMS algorithm. Accordingly, the recursive computation of each tap weight in the LMS algorithm suffers from a *gradient noise*, which causes $\hat{w}(n)$ to move randomly around the minimum point of the error-performance surface rather than terminating on the Wiener solution W_{opt} as before.

Since the LMS algorithm involves feedback in its operation, an issue of stability is raised. In this case, a meaningful criterion is to require

$$J(n) \rightarrow J(\infty) \text{ as } n \rightarrow \infty \qquad \textit{[eq 12]}$$

Where *J(n)* is the mean-squared error produced by the LMS algorithm at time n and its final value, $J(\infty)$ is a constant. An algorithm that satisfies this requirement is said to be *convergent in the mean square*.

2.2.2.2. LS algorithm

As mentioned earlier, the LMS algorithm estimates the statistical expectation operator E [.] by using the instantaneous values as in (8)-(11). The least-square (LS) algorithm avoids such estimation because it is deterministic in approach. Specifically, it

minimizes a cost function that consists of the sum of error squares by choosing optimally the tap weights ($w_0, w_1, \ldots\ldots, w_{M-1}$) of the transversal filter:

$$J(w) = \sum_{i=i_1}^{i_2} e(i)^2 \qquad \textit{[eq 13]}$$

$$= \sum_{i=i_1}^{i_2} \left(r(i) - \hat{r}(i)\right)^2 \qquad \textit{[eq 14]}$$

where i_2 and i_2 define the index limits at which the error minimization occurs. The values assigned to these limits depend on the type of data windowing methods employed. It can be covariance, autocorrelation, and pre windowing or post windowing method. The discussion is limited to the pre windowing method which assumes the input data prior to $i=0$ are zero, but make no assumption about the data after $i=N-1$. Thus, $i_1=$ and $i_2=N-1$. For minimization, the tap weights $w_0, w_1, \ldots\ldots, w_{M-1}$ are held constant during the interval $i_1 \leq i \leq i_2$. The filter resulting from the minimization is termed as a linear LS filter. From (14) and using $\hat{r}(i) = w^T y(i)$, the linear LS estimator can be found by minimizing

$$J(w) = \sum_{i=0}^{N-1} \left(r(i) - w^T y(i)\right)^2 \qquad \textit{[eq 15]}$$

$$= (r - Hw)^T (r - Hw) \qquad \textit{[eq 16]}$$

The matrix H is a known MxN matrix of full rank M and it is referred as the input sample observation matrix:

$$H = \begin{pmatrix} y(0) & 0 & 0 \\ y(1) & y(0) & 0 \\ y(2) & y(1) & 0 \\ . & . & . \\ . & . & . \\ y(N-1) & y(N-1) & y(N-1) \end{pmatrix} \qquad \textit{[eq 17]}$$

And the vector $r=(r(0), r(1), r(2), \ldots\ldots, r(N-1))^2$ is the $Nx1$ vector of the echo signal. The minimization is easily accomplished by setting the gradient to zero, as eq.18

$\nabla J(w) = -2H^T r + 2H^T Hw = 0$ *[eq 18]*

and yields the LS estimator, as eq.19

$$\hat{w} = (H^T H)^{-1} H^T r \qquad [eq\ 19]$$

The equation $H^T Hw = H^T r$ to be solved for $\hat{w}$ is termed the normal equation. The assumed full rank of H guarantees the inevitability of $H^T H$. The inversion requires $O(M^3)$ computations. However, since the filter structure is transversal FIR, then the matrix H is to elite. This property indicates that the inversion can be performed in $O(M^2)$ operations [18].

2.3 Full band Echo Cancellation

Fig.2.3 shows the use of the adaptive filter as an AEC, which attempts to synthesize a replica of the acoustic feedback at its output. The signal *y(n)* is the farend signal, *d(n)* is the desired signal, *n(n)* is the ambient noise which is assumed to be negligibly small, *x(n)* is the near-end signal, *r(n)* is the end acoustic echo, $\hat{r}(n)$ is the synthesized echo $\hat{r}(n)$ from the adaptive filter, e(n) is the echo-free error signal and $\delta\hat{w}(n)$ is the estimated tap-weights correction vector for the adaptive FIR filter. The following mathematical notation conventions will be used in this section unless otherwise stated.

M Number of tap-weights in an adaptive filter $\hat{w}(n)$. Tap-weights vector *(Mx1) y(n)* Input signal vector *(Mx1)*. The LMS belongs to the stochastic gradient type algorithm family, which has both the low computational complexity of *O(M)* and slow convergence rate property when the input excitation signal is highly correlated, e.g. speech signal. Although the LS has faster

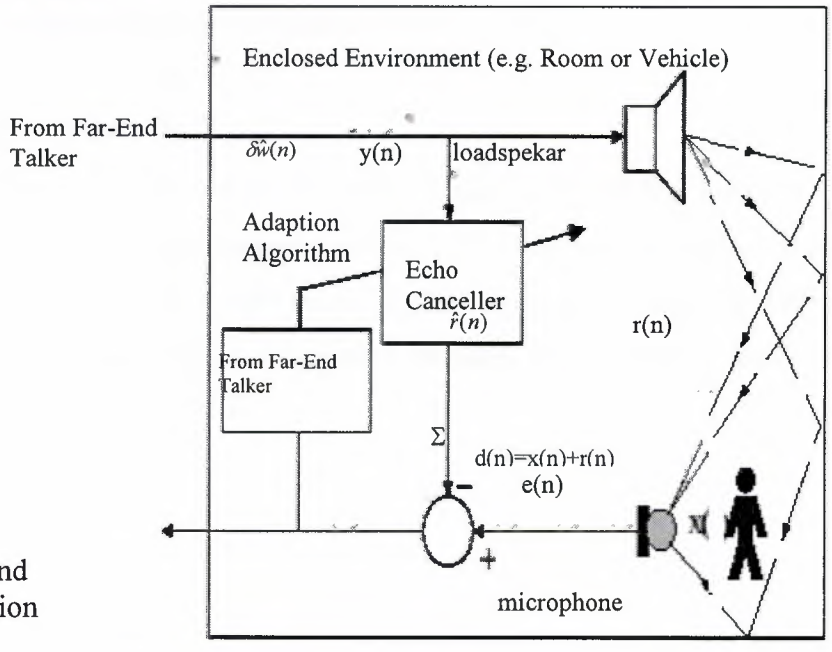


Fig2.3, General full band AEC configuration

Convergence rate, which is invariant to the eigen value spread of the input signal, it is still infeasible for long impulse responses due to the complex computation of $O(M^3)$. Thus, other more powerful algorithms must be used to compensate these drawbacks. Many of these algorithms can be classified as being between the LMS and LS algorithms in terms of their convergence rate and computational load. In this section, some of the variants of the two algorithms are discussed for the full-band AEC's application.

2.3.1 Excited NLMS

In (11), the correction $\mu y(n)\ e(n)$ is directly proportional to $y(n)$. Thus, when $y(n)$ is large, the LMS algorithm experiences a *gradient noise amplification* problem. Another drawback is that the convergence rate fluctuates considerably if the condition number of input signal correlation matrix R is large. One of the sources of such variability in eigen values is the change of input signal level. The *normalized LMS* (NLMS) algorithm is used to overcome these two problems where (11) is changed to

$$\hat{w}(n+1) = \hat{w}(n) + \frac{\hat{\mu}}{\|y(n)\|^2} y(n)e(n) \qquad \textit{[eq 20]}$$

where $\hat{\mu}$ is a positive real scaling constant. Unfortunately, the problem is not very well solved by the NLMS algorithm. An alternative method is used in [1], which pre whitens the input signal by applying perfect sequences periodically. Perfect sequences, $\hat{p}(n)$ of period M (same as the adaptive filter length) are characterized by their periodic autoscore-relation function, which disappear for all out-of-phase values:

$$R\overline{pp}(n) = \begin{cases} 1 & \text{for } i \bmod N = 0 \\ 0 & \text{otherwise} \end{cases} \qquad \textit{[eq 21]}$$

Perfect sequences are the optimal excitation signal of the NLMS algorithm because they are orthogonal in the M dimensional vector space. The new algorithm is called the *excited LMS* (ELMS) algorithm as shown in Fig.2.3.1, which combines the conventional NLMS algorithm and an optimal adaptation method using perfect sequences. The sequences are amplified by a factor K.

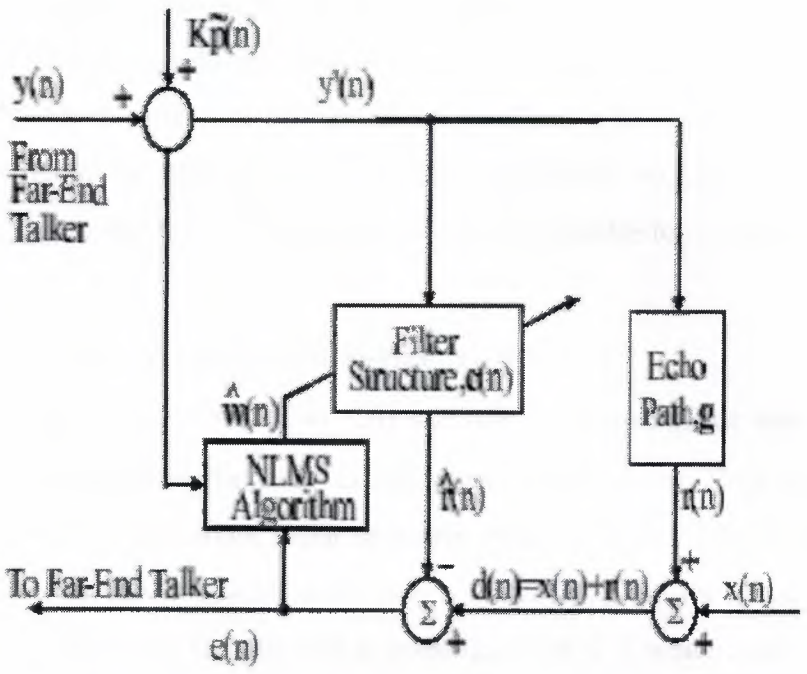


Figure 2.3.1: Application of excited LMS algorithm in AEC by Antweiler ET. Al.

In this optimal adaptation method, the far-end signal, *y(n)* causes interference while in the NLMS algorithm, it is the information signal. Thus, the integration of both methods into the ELMS algorithm combines the different effects of *y(n)*. The superposition of *y(n)* and $\hat{p}(n)$ can be regarded as a pre whitening technique where $\hat{p}(n)$ fills the gaps of the spectrum of *y(n)*. The system distance is measured by:

$$D(n)_{db} = 10\log\frac{\|g - c(n)\|^2}{\|g\|^2} \qquad \textit{[eq 22]}$$

and the power ratio by:

$$10\log\frac{E\left[y^2(n)\right]}{K^2\hat{p}^2(n)} \qquad \textit{[eq 23]}$$

where *g* and *c(n)* are the impulse response of the echo path and AEC respectively. The simulation shows that a lower power ratio yields the desirable lower system distance. A power ratio of as high as 40dB can provide sufficient improvement. An informal subjective listening test proved that at this power level, the perfect sequences are hardly audible to the near-end talker. Thus, the disturbing effect due to the superposition of $\hat{p}(n)$ can be kept

sufficiently small if the dynamic range of the acoustic path (includes AD and DA conversion) is sufficiently large. In [2], two orthogonal zing techniques, i.e. the ELMS algorithm plus the linear prediction method, is used to further improve the performance of the AECs. The linear predictors prewritten the *y(n)* and *e(n)* signal for the adaptation process. In addition, the predictors perform spectral shaping on $\hat{p}(n)$ according to the speech signal, *y(n)* so that the auxiliary signal $\hat{p}(n)$ is less audible to the near-end talker.

2.3.2 Exponentially Weighted Step-size NLMS

The step-size parameter, $\hat{\mu}$ in (20) controls the convergence rate of the filter coefficients and determines the final excess mean-squared error, $J(\infty)$ of the Wiener solution. Therefore, a time-variant scalar or matrix step-size method can be used to obtain fast convergence in the transient state and a small $J(\infty)$ in the steady state. The characteristic of a room impulse response is investigated in [17], which shows that impulse responses attenuate exponentially and the variation 3 of these impulse responses also attenuates by the same exponential ratio, γ. Using this knowledge in the conventional NLMS, a new algorithm called the *exponentially weighted step-size NLMS* (ESNLMS) is proposed. The algorithm updates the coefficients with large errors in large steps and those with small errors in small steps. For this purpose, a step-size matrix A with diagonal form is introduced:

$$\begin{pmatrix} \alpha_1\alpha_1 & 0 \\ 0 & \alpha_1 \end{pmatrix} \qquad \textit{[eq 24]}$$

where $\alpha_i=\alpha_0\,\gamma^{i-1}(i=1,\ldots..,1)$ and $0<\gamma<1$.

It is the difference between two differently measured impulse responses in a same room. Elements α_i decrease exponentially from α_1 to α_2 with the same ratio γ as the room impulse response and they are time-invariant. The ratio γ can be derived from the reverberation time which is determined by the acoustical property of the room. It can be shown that:

$$\gamma = \exp\left(-6.9\frac{T_s}{T}\right) \qquad \textit{[eq 25]}$$

where -6.9 is $In\ 10^{-3}$, T_s is the sampling interval and T_r is the reverberation time of the room. The algorithm is expressed as:

$$\hat{w}(n+1) = \hat{w}(n) + A\frac{e(n)}{\|y(n)\|^2}y(n) \qquad [eq\ 26]$$

The real-time experiments in a room based on multiple DSP chips implementation show that the ESNLMS algorithm is three times faster for a white noise input signal and twice faster for speech. The algorithm also has the same computational load of 2L as the conventional NLMS.

2.3.3 Fast Newton Transversal Filter

One of the alternative algorithms to LS algorithm is the recursive least-squares (RLS) algorithm, which obtains an optimum estimate of filter tap weight recursively sample-by-sample. The RLS algorithm can also be deduced in the exact form directly from the covariance Kalman filtering by using the state-space model that matches the RLS problem [22]. Thus, the RLS algorithm can be viewed as a special case of Kalman filter, i.e. the deterministic approach of Kalman filter. The RLS has faster convergence rate and a better minimum mean squared error performance than the LMS algorithm. Although its requirement of $O(M^2)$ operations per iteration is less than the $O(M^3)$ computations as in the LS algorithm, it is still much more than the LMS algorithm which requires only $O(2M)$ operations per iteration. A variant of RLS which has computation complexity comparable to LMS (i.e. increases linearly with the number of adjustable parameters, M) is known as *fast RLS* (FRLS). Similar to the LS algorithm, the computation requirements of RLS can be reduced if the filter structure is a transversal FIR. One example of such algorithm, which uses adaptive transversal filter structure, is called the *fast transversal filter* (FTF) algorithm [8]. The FRLS algorithm family is a potential solution to the high computation complexity of RLS by reducing the computation to an order of $O(M)$. However, bits practical use is prevented in the past because of divergence due to the numerical error accumulation in its linear prediction parameters. RLS scheme is a special member of the New-ton algorithm family. A Newton algorithm requires the stochastic estimation of the Hussein matrix, which in fact is the correlation matrix of the input signal. Therefore, a *fast Newton transversal filter* (FNTF) method that is developed in [21] is actually a complexity reduced FRLS

algorithm. It also provides possible tradeoffs between complexity and performance by appropriately choosing the prediction order versus the adaptive filter length. The established RLS algorithm has the following set of recursive equations:

$$\hat{w}(n) = \hat{w}(n+1) + C(n)e(n) \qquad \textit{[eq 27]}$$

$$e(n) = d(n) - \hat{d}(n) \qquad \textit{[eq 28]}$$

$$\hat{d}(n) = \hat{w}^T(n-1)y(n) \qquad \textit{[eq 29]}$$

$$c(n) = -R^{-1}(n)\, y(n) \qquad \textit{[eq 30]}$$

where $R(n)$ is the MxM covariance matrix of the input signal. The update of the gain vector, $C(n)$ in the RLS requires the update of the inverse covariance matrix. In FRLS, this is achieved by using LS optimal forward and backward predictors of the input signal. As a result, the complexity of the FRLS is reduced to 5M . The calculated complexity assumes that the input predictors have the same order as M . However, this is unnecessary when the input signal is speech. The predictable part of the input series can be extracted with predictors of much lower order, MxM. Thus, the estimate of the covariance matrix of order NxN can be now be extrapolated from a lower order estimate. The complexity of this new algorithm falls down to $12N+2M$. Feeding back the numerical errors of the unstable variables of the algorithm, i.e. basically the backward prediction variables solve the numerical instability.

2.3.4 Adaptive IIR Gradient Instrumental Variable

The advantages of IIR filters over FIR filters are: They require less computation per iteration for the same performance level. they can match the poles and zeroes of physical systems 4 , whereas the FIR filters can only approximate them. These advantages also exist in the adaptive IIR filters. However, unlike the adaptive FIR filters, they may not have unmoral error surfaces. In addition, it is also difficult to maintain stability during adaptation. One of the early works in developing adaptive IIR filtering algorithm for the use in echo cancellation is [9]. The algorithm is known as profiteering (PF) algorithm, and its adaptive filtering mode is shown in Fig. 2.3.4. It uses the direct form of IIR filter structure. In the diagram, $A(z^{-1}) = 1 - \sum_{i=1}^{\hat{n}_a} \hat{a}_i(n) z^{-1}$ and $B(z^{-1}) = 1 - \sum_{i=0}^{\hat{n}_b} \hat{b}_i(n) z^{-1}$ are the

adjustable denominator and numerator functions respectively for the adaptive IIR filter. Although the implementation of the algorithm is to cancel telephone network echo, the work presents a good fundamental understanding of the general problems and possible solutions in using adaptive IIR filters in echo cancellation.

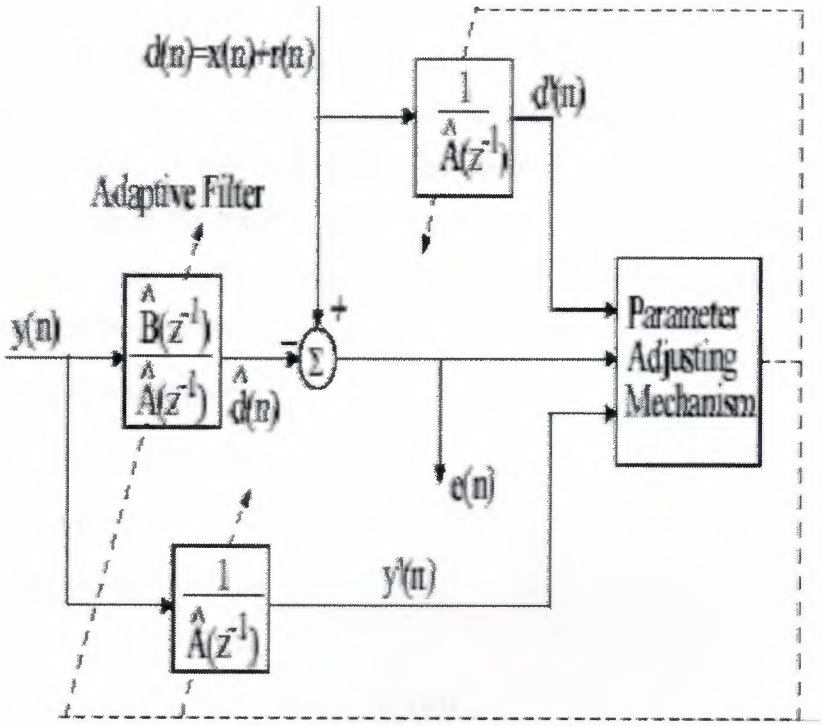


Figure 2.3.4: Adaptive filtering mode structure of the PS algorithm by Fan and Jenkins

The simulation in [9] shows that the stability of the canceller can be achieved by using a sufficiently small constant gain without incorporating a monitoring device. However, the price is a slower convergence. The IIR canceller also achieves a much higher *echo return loss enhancement* (ERLE) for the same amount 4 A room transfer function is better modelled with poles and zeroes [20]. 5 It depends on the structures that model the noise, i.e. equation error or output-error model structures. 7of computation if the order selection is corrects (sufficient order). In the presence of white Gaussian ambience or measurement noise, the improvement is considerably degraded. In practice, the order of the echo path is unknown or varies for different situation, therefore a reduced order case may occur. It is shown that in this case where the order of the IIR canceller is less than the order of the echo

path, it performs worse than a FIR canceller. It is clear from the results that correct order selection is important for adaptive IIR filtering. In [6], a new adaptive IIR-based gradient instrumental variable (IV) echo canceller (GIVE) is presented. The GIVE algorithm is implemented on both the *series-parallel* and *parallel* structures. As the output error method in adaptive filtering does not guarantee to converge from any initial point to the global minimum, the equation error method is used. The GIVE algorithm is capable of updating the filter's coefficients even in double-talking periods, and is guaranteed to converge to the unique global minimum. It can also avoid the need to invert non-symmetrical cross-correlation matrices, as in the traditional IV method. This gives the ad-vantage of having robust numerical stability and a computational complexity that is comparable to the equation error IIR LMS algorithm. The algorithm is as follows:

$$\hat{B}(n) = \hat{B}(n-1) + \mu e(n) y(n) \qquad [eq\ 31]$$

$$e(n) = d(n) - \psi^T(n)\theta(n-1) \qquad [eq\ 32]$$

$$d(n) = \theta^T \psi + \partial^T x(n) \qquad [eq\ 32]$$

with the following vector definition:

$$\theta = (a_1, \ldots\ldots, a_p, b_0, \ldots\ldots\ldots, b_q)^T \qquad [eq\ 34]$$

$$\hat{\theta}(n) = (\hat{a}_1(n), \ldots\ldots, \hat{a}_p(n), \hat{b}_0(n), \ldots\ldots, \hat{b}_q(n))^T \qquad [eq\ 35]$$

$$\psi(n) = (d(n-1), \ldots\ldots, d(n-p), y(n), \ldots\ldots, y(n-2))^T \qquad [eq\ 36]$$

$$\xi(n) = (v(n-1), \ldots\ldots, v(n-p), y(n), \ldots\ldots, y(n-2))^T \qquad [eq37]$$

$$x^T(n) = (x(n), \ldots\ldots\ldots, x(n-p)) \qquad [eq\ 38]$$

$$\partial^T = (a_1, \ldots\ldots\ldots, a_p) \qquad [eq\ 39]$$

Where θ is the parameter vector, $\hat{\theta}(n)$ is the estimated parameter vector at time n, $\psi(n)$ is the input-output vector, $v(n)$ is the IV signal and $\xi(n)$ is the information vector. If the algorithm converges, $E[e(n)\xi(n)]=0$ from (31) and by using (32), the extended normal equation is obtained:

$$E[y(n)\xi(n)] = E[\xi(n)\psi^T(n)]\hat{\theta}(n) \qquad [eq\ 40]$$

If $E[\xi(n)\psi^T(n)]$ is nonsingular, (40) implies that once it converges, it is guaranteed to converge to the unique global minimum. Since $\hat{\theta}(n)$ has converged to θ, $e(n)$ will converge to $A(z^{-1})x(n)$ during double talking. During single talking $(x(n)=0)$, $e(n)$ con-verges to zero. The echo replica is used as the IV signal, i.e. $v(n)=\hat{r}(n)=(\hat{B}(z^{-1})/\hat{A}(z^{-1}))v(n)$ where $\hat{A}(z^{-1})$ is obtained by stabilizing the transfer function estimate of $\hat{A}(z^{-1})$. Fig. 5 shows the implementation of GIVE algorithm:

Parallel structure: The echo replica is used as the IV signal and to cancel the echo directly.

Series-parallel GIVE structure: The equation error, $e(n)=\hat{A}(z^{-1})x(n)$ is fed into an equalizer $1/\hat{A}(z^{-1})$ to eliminate distortion. The echo replica, $\hat{r}(n)$ is used as the IV signal, but does not directly contribute to echo cancellation.

2.4 Sub band Echo Cancellation

The following mathematical notation conventions will be used in this section unless otherwise stated.

K Order of FIR filter in the analysis and synthesis bank for each sub band Expansion factor

L Number of tap-weights in an adaptive filter for each suburban

$\overline{M}$ Number of tap-weights in an adaptive filter for a full band AEC

M Number of sub bands

N Decimation factor

$\hat{w}_i(m)$ Estimated complex tap-weights vector $\overline{M}x1$ of the its adaptive filter at the decimated time m, where $i=0,1,.....N$

$e_i(m)$ its complex echo-free error signal

$y_i(m)$ its complex input signal vector

$y^x(m)$ complex conjugate input signal vector

$d_i(m)$ desired signal

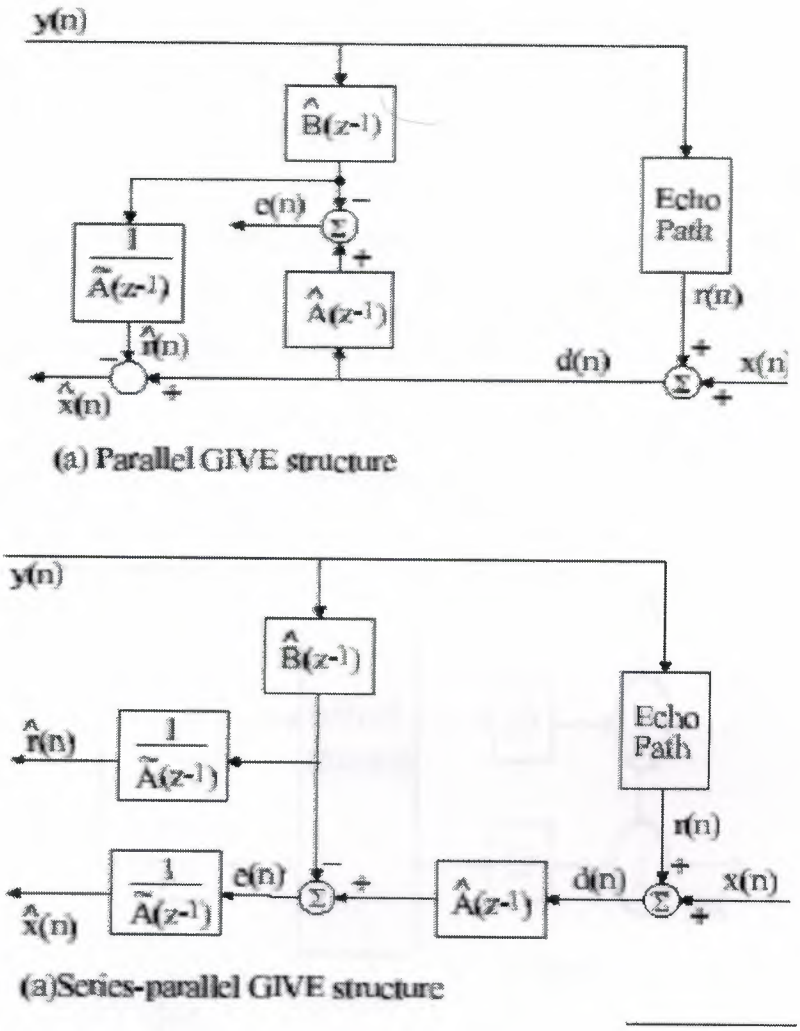


Figure 2.4, GIVE algorithm structures by Chao, et. al.

Sub band processing is an important application of MultiMate signal processing. It is based on the idea of dividing the frequency range of an input signal into segments (subtends). Each sub band is processed independently as required by a specific application and in this case, it is the echo cancellation. If necessary, the subtends are recombined after processing, to form an output signal whose bandwidth occupies the same entire frequency range. The division of a signal into subtends is done by using a band-pass (BP) filter bank as shown in Fig. 6. This filter bank is also known as an *analysis* bank. The output of the analysis bank is fed into a sub band processing system. The reconstruction of the signal after processing is done by another filter bank called a *synthesis* bank. If the signal is divided into subtends

[illegible] have the same width (i.e. $2\pi/N$), the filter bank is called *uniform*. The filter [illegible] in Fig. 6 does not use the property of the sub band signals, i.e. the signals [illegible] by a factor of up to N without being aliases. The decimation operation on [illegible] signal will reduce the required computation operations at each Kth-order FIR [illegible] analysis bank. This advantage is hard to be realized on IIR filters. The [illegible] number of multiply-add operations is about K/R per input data point at each [illegible] Likewise, the combination of expansion followed by band-pass filters at the synthesis [illegible] also reduce computation. Then, the necessary multiply-add operation is about K/L [illegible] data point.

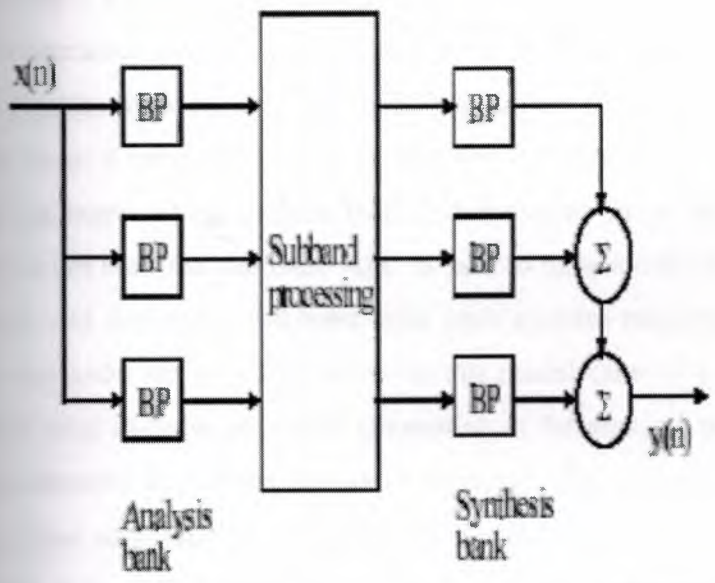


Figure 2.4.1, Application of filter banks in sub band processing

[illegible] shows the modified filter bank. If $R=N$ it is called a *maximally decimated filter* [illegible] uniform filter banks. The processed sub band signal should be expanded by the [illegible] decimation factor, i.e. $L=R$ unless sampling-rate conversion is required. Maximally [illegible] filter banks have the most computational savings because the sub band signals [illegible] lowest possible rate.

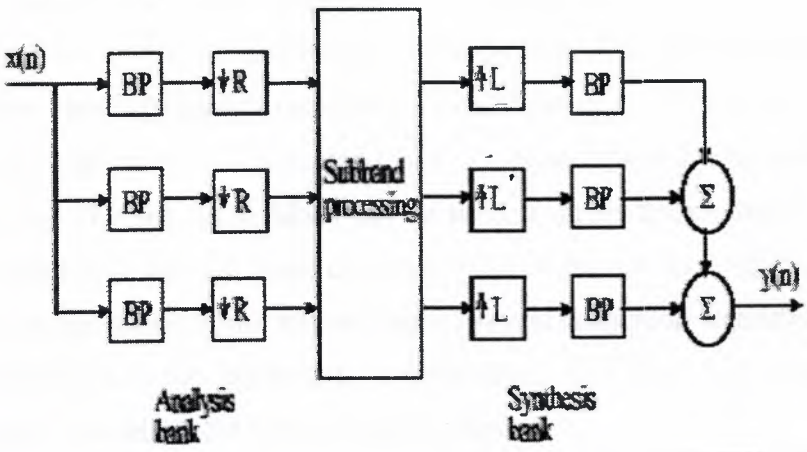


Figure 2.4.2, Sub band processing with decimated filter bank

This benefit however assumes an ideal situation where the analysis bank produces aliasing-free sub band signals. This requires the band-pass filters to have in-finite stop band attenuation. The factor R must also be chosen such that the pass band and transition region of the signals at the output of the analysis bank do not overlap in the frequency domain. If these requirements are met, the sub band AEC is said to have modeled a version of ideal band-pass filtered and decimated full-band echo path impulse response. The solution is called the *frequency subtends solution*. However, this model cannot be realized in a strict sense. Firstly, the ideal in-finite stop band attenuation at the analysis and synthesis banks can only be approximated in practice. Secondly, filtering a finite impulse response with an ideal band-pass filter will yield an unrealizable impulse response with $-\infty$ to $+\infty$ time domain. Therefore, it is prefer-able to decimate and expand by a smaller factor than N The other solution to be discussed is the *fast convolution* solution, which utilizes a uniform DFT filter bank as shown in Fig. 2.4.3,. The operation is to decimate the delayed outputs by $R=N$ before feeding them into the matrix $\overline{F}N$ (conjugate DFT). The consecutive input vectors to matrix consist of consecutive non-overlapping segments of length N of the input signal. Now, the DFT is performed once every N point in each subtends. The synthesis bank is treated in a similar manner: the output vector from F_N (DFT) is expanded by $L=N$ and passed through a delay chain. The operation of linear convolution between the decimated input vector and $\overline{F}_N$ matrix can be regarded as a linear convolution between the vector and

a F_N thrower FIR filter. There is N such filters in the bank and each one of them is a shifted replica of the first sub-band's prototype filter. This is achieved by right shifting the prototype filter's frequency response by $2\pi m/N$ $(m=0, 1...N)$ in the frequency domain. The design of sub band AECs is based on these fundamental characteristics of sub band processing. The sub band AECs can be used to solve the problems associated with the computation load and the speed of convergence of the full-band AECs. Their performances depend on the choice of the sub band structure and adaptation algorithm. Some examples of recent advances in the design will be discussed in Sec. 2.4.1. The main advantages of sub band echo chancellors are summarized as below:

Each sub band signal in the uniform analysis and:

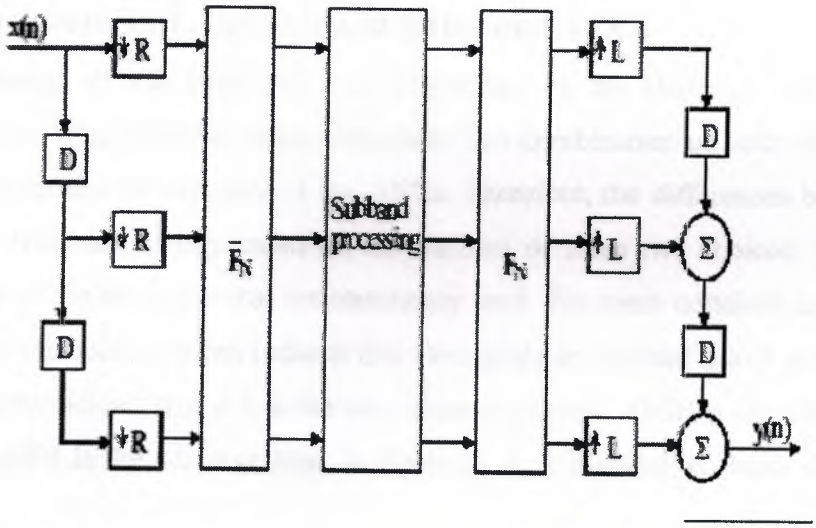


Figure 2.4.3: Sub band processing with maximally decimated uniform DFT filter bank

Synthesis bank is assumed to be decimated and expanded respectively by the same factor R. The impulse response length of each sub band is assumed to be the same in time with the full band impulse response. Then, the amount of tap-weights, $\overline{M}$ in an adaptive FIR filter for each sub band is reduced by a factor R because of the decimation, i.e. $\overline{M} = M / R$. In addition, the filtering and adaptation is now performed at a reduced sampling rate, resulting in the total reduction of computation per second equals to I/R^2 of the flubbed echo cancellers. Taking into account the N subtends, the reduction factor is approximately R^2/N.

The LMS algorithm is a good candidate for full-band AECs because of its low computation load. However, it has the problem with convergence speed because of the required long filter length and the spread of eigen values in the input speech signal. The use of LMS algorithm in sub band AECs does not face the same problems. The convergence rate can be increased because of the reduction in the FIR filter length for each sub band. A more important factor is that the signal's spectrum in each sub band is expected to be flatter than the flubbed signal. This improves the convergence speed because of the decrease in eigen value spread. How-ever, it is worth to note that the finite stop band attenuation and the transition band of each band-pass filter can create some small eigen values at the edges of each sub band's spectrum.

2.4.1The Structure and Algorithm of Sub-band AECs

The design of sub band AECs is determined by the choice of the adaptation algorithm and the echo canceller model structure. The combination of both choices has a profound effect on the performance of the AECs. Therefore, the differences between one design to the others are mainly based on the variants of these two choices. Fig. 2.4.1.1 illustrates some of the structures that are commonly used. The same notations as in Fig. 2.3 are used, except the bold notations indicate that the signals are divided into N subtends. The letter in the boxes indicate that AA is the adaptation algorithm, AFSB is the adaptive filter in sub band, AFFB is the adaptive filter in flubbing, A is the analysis bank and S is the synthesis bank.

2.4.1.1 Synthesis-dependent Structure

The synthesis-dependent structure is shown in Fig. 2.4.1.1, The disadvantage of this structure is that the error signal is fed back to the adaptation algorithm after a delay in the synthesis bank. This will effect the convergence speed, especially in a rapidly changing echo path. An-other shortcoming is that the adaptation of the N parallel adaptive filters is based on a error signal, $e(n)$. Thus, the components of the error signal outside the frequency range of each filter act as a noise on the adaptation process.

shows that if *T=4096,N=32* and $\overline{T} = 128$, the delay is only 8msec at a 16kHz sampling frequency.

2.4.1.2 Synthesis-independent Structure

The synthesis-independent class of structure in Fig. 2.4.1.2 is more remarkable in the application of echo cancellation because the sub band concept is more directly applicable. In addition, the problem of delay in the adaptation loop as in Sec. 2.4..1.1 is avoided:

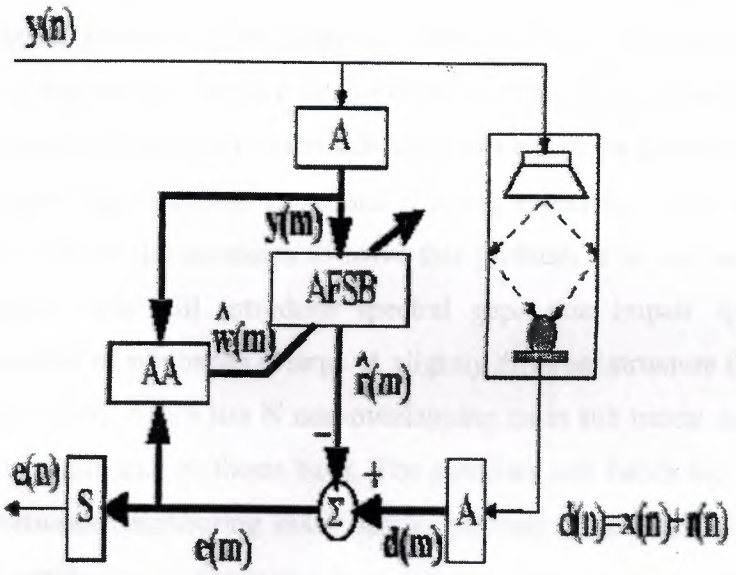


Figure 2.4.1.2, Synthesis-independent AEC

In [12], a *fast recursive LS* (Flotation is proposed for this structure. The simulation Uses *N=16* sub bands, *R=12* decimation factor, $\overline{M} = 128 + h -$ order the-order adaptive filter at each sub band and polyphone decomposition of the prototype low pass FIR filter at the filter banks. The algorithm which is a modified version of the FTF algorithm in [8] is as follows:

$$\hat{w}_i(m) = \hat{w}_L(m-1) + \mu_i(m)R_i^{-1}(m)y_i(m)e_i(m) \qquad \textit{[eq 41]}$$

$$e_i(m) = d_i(m) - \hat{w}_i^T(m-1)y_i(m) \qquad \textit{[eq 42]}$$

where $\mu_i(m)$ is the its step-size which can be used to improve the behavior of the FRLS algorithms and $R_i(m)$ is the its deterministic input signal correlation matrix. The Kalman gain, $K_i(m)=R_i^{-1}(m)y_i(m)$ can be efficiently calculated by the FTF algorithm of [8] with a complexity of $O(\overline{M})$. Therefore, the computation complexity is only in the order of a full band LMS AEC and the initial convergence speed is com-parable to a full band RLS AEC. Instability is a problem of FRLS algorithm due to the possible singularity of $R_i(m)$ and the numerical inaccuracies. One method to overcome the numerical inaccuracies is proposed by periodic re initialization of the internal state variables in the algorithm for a short period of time. The combination of analysis and synthesis banks that yield aliasing free output signal is said to have *perfect reconstruction* property. However, the spectral con-tents of the decimated output are not entirely disjoint due to the non-ideal nature of the analysis bank. Thus, the use of adaptive filters within each sub band will effect the perfect reconstruction property and may cause high levels of inter-band aliasing, especially if the filter banks are critically decimated. One of the measures to solve this problem is to use non-overlapping filter banks. However, this will introduce spectral gaps that impair speech quality, particularly if the number of sub bands is large. A slightly different structure than the one in Fig. 10 is presented in [25], which has N non-overlapping main sub bands and N auxiliary sub-bands in each analysis and synthesis bank. The auxiliary sub bands are used to cover the spectral gaps between neighboring main bands and they have narrower bandwidths. Thus, the auxiliary bands can be decimated by a factor as high as 2N to reduce the extra computation cost. Both the main and auxiliary sub bands are developed from the uniform DFT filter banks. The sub-bands adaptive filters use the complex NLMS algorithm for the adaptation and a complex thin lattice structure to whiten the error signal. The thin lattice is a simplified version of the conventional gradient lattice where the required computation is about one third of the conventional type. Another solution to the aliasing problem is to use the polyphone IIR filters, which have sharp transition band and high stop band attenuation. They appear to be an attractive substitution to the FIR polyphone filters in the filter banks. The *all pass polyphone network* (APN) implementation in [26] is a polyphone configuration of the prototype low pass IIR filter. The most appropriate sub band decomposition based on APN is N=2subbands. Thus, a tree-structured filter bank is used.

The APN gives perfect amplitude reconstruction but with a non-ideal phase reconstruction due to the non-linear phase response of the all pass sections. However, this in not critical in AECs because the phase distortion is normally unnoticeable by the human auditory system. The finite precision implementation of the APN is also discussed. The APN is defined by substituting

$$H_i(z^{-1})0\prod_{J=6}^{P_i}\frac{\alpha_{i,j}+z^{-1}}{1+\alpha_{i,j}z^{-1}} \qquad \textit{[eq 43]}$$

$$H(z^{-1})=\sum_{i=0}^{N-1}2^{-i}H_i(z^{-N}) \qquad \textit{[eq 44]}$$

where P_i is the number of all pass taps at the its-phase, α_{ij} is the all pass coefficients and $H_i(z^{-N})$ is the IIR polyphone component of the prototype low pass filter $H(z^{-1})$.

2.4.1.3 Delay less Structure

All the structures that have been discussed so far intro-duce a delay into the transmission path from the near-end talker to the far-end talker. It is due to the group delay of the cascaded analysis and synthesis banks. Examples of the common types of configuration used in these banks are the *quadrate mirror filter* (QMF) and DFT, which are usually a linear phase FIR filter. The FIR filter of order K produces a delay of K/2 samples. Therefore, the total delay due to the filter banks is in the order of the FIR filter. If more computation savings are required, the signal must be divided into more than N sub bands to increase the decimation factor, R . Thus, the bandwidth of each sub band will be narrower and this will need a higher order of FIR filter at the banks. Subsequently, more delay in the transmission path is introduced. To solve this problem, several new structures and algorithms are introduced as follows. A new structure with zero delay as in Fig. 2.4.1.3 is proposed in [7]. In this structure, a compensating adaptive filter (AFFB) is used. It yields a full band synthesized echo, $\hat{\hat{r}}(n)$ at the original sampling frequency. The analysis and synthesis bank is removed from the transmission path. Assuming that the AFSB and end echo path have identical phase characteristics, then $\hat{r}$ arrives K samples later than the echo, $r(n)$ itself due to the group delay of the analysis and synthesis bank. Due to the delay, the first K taps of the echo impulse response are not identified. The AFFB is used to

compensate that delay, which has K taps if the analysis and synthesis bank have linear phase characteristic. The simulation in [7] uses a simple weighted overlap-add (weighted window and FFT/IFFT) method for the filter banks, a complex $\overline{M}$ the-order adaptive filter for each sub band, $N=128$ sub bands, $R=32$, decimation factor and a modified NLMS adaptation algorithm as follows:

$$\hat{w}_i(m+1) = \hat{w}_i(m)2\mu_i(m)e_i(m)y_i^x(m-\delta) \quad \textit{[eq 45]}$$

$$\mu_i(m) = \frac{\overline{M}}{P_i(m)+P_0} \quad \textit{[eq 46]}$$

$$P_i(m) = 1 - \sqrt{M}P_i(m-1) + \sqrt{M}y_i(m)^2 \quad \textit{[eq 47]}$$

where δ is the delay due to the filter banks, $\mu_i\ (m)$ is The ith normalized step-size, $\overline{\mu}$ is a small constant of $0 < \overline{\mu} < 1$, P_0 is a small constant to prevent large $\mu_i(m)$ adjustment when the input signal, $y_i(m)$ is very small and $P_i(m)$ is the estimated total input power. In [19], two new types of structures are presented which avoid signal path delay and retain the computational and convergence speed advantage of sub band processing. The technique is to compute the adaptive weights in sub bands (time domain) and collectively transform the weights into an equivalent set of full band filter coefficients. Fig. 2.4.1.3.1 shows the delay less sub band AEC of a close loop type where the error signal is fed back to the sub band error filter bank. The following complex LMS algorithm is used:

$$\hat{w}_i(m+1) = \hat{w}_i(m) + \mu e_i(n)y_i^x(m) \quad \textit{[eq 48]}$$

$$e(n) = d(n) - \hat{w}^T(n)y(n) \quad \textit{[eq 49]}$$

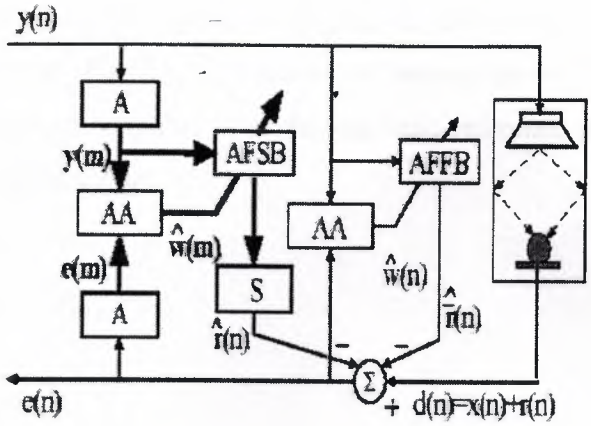


Figure 2.4.1.3, Zero-delayed sub band AEC by Chen, et. al. (simplified version)

Where μ is a constant step-size. Note that (49) is the same error equation as in a full band AEC.

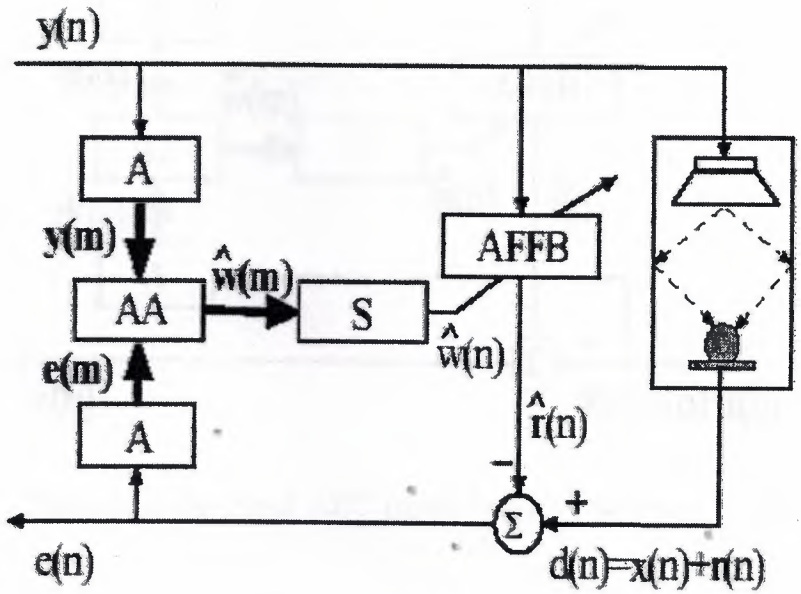


Figure 2.4.1.3.1, Delay less sub band AEC (close loop) by Morgan, et. Al (simplified version)

in the block AA, the estimated adaptive weights of each sub band are transformed (FFT) into the frequency domain and appropriately stacked. In block S, the point array is inverse-transformed (IFFT) to obtain the full band adaptive weights. The shortcoming of this

structure is that a delay is introduced into the weight update loop, which will limit the convergence rate of the LMS algorithm. A modified structure is developed by computing the error and subsequently the adaptive weights in sub bands, similarly to Sec.2.4.1.2, before the FFT operation. Fig. 2.4.1.3.2 shows the second structure. In this case, the LMS algorithm also computes a convolution of the sub band reference signal and tap-weights to obtain a local error signal, $e_i(m)$

$$\hat{d}_i(m) = \hat{w}_i^T(m) y_i(m) \qquad [eq\ 50]$$

$$e_i(m) = d_i(m) - \hat{d}_i(m) \qquad [eq\ 51]$$

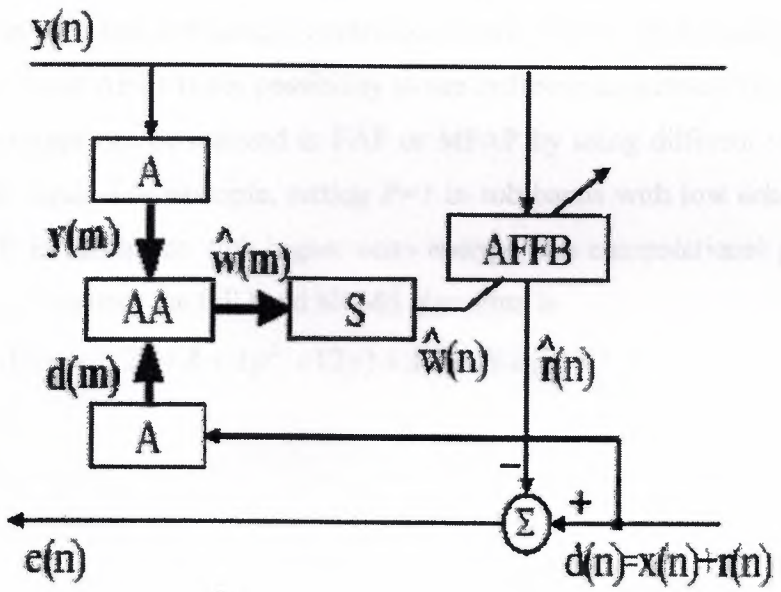


Figure 2.4.1.3.2, Delay less sub band AEC (open loop) by Morgan, et. Al (simplified version)

The structure is called the open loop type because the full band error is not fed back to the sub band weights calculation. The implementation uses the polyphone FFT decomposition technique to design the analysis bank, which yields N contiguous single side-band band pass filters. The output of these filters are decimated by a factor of $R=N/2$. For the $M=512$

and $N=32$ simulation example in [19], the computation load is estimated to be only one third of a conventional full band design.

2.4.1.4. Affined Projection algorithm .

In the *affined projection algorithm* (APA), the weight vector update is obtained from a projection on an affined subspace with dimension J-P, where J is the length of a filter and Pis an arbitrary integer. Later, a faster version of APA is introduced in [11], known as *fast affined Projection* (FAP). The FAP however employs the sliding window FRLS algorithm that has inherent instability due to the finite numerical precision in computation. To overcome the numerical instability, a new algorithm is introduced in [16]. It is known as *modified fast affined projection* (MFAP) algorithm that uses the conventional RLS algorithm to calculate the sample covariance matrix of the input signal. One distinct advantage of sub band AECs is the possibility to use different algorithms for different sub-bands. This advantage can be realized in FAP or MFAP by using different values of P in the different sub bands. For example, setting $P=1$ in sub bands with low echo energy and larger value of P in sub bands with higher echo energy. The computational gain of a sub band MFAP algorithm over the full band NLMS algorithm is

$$G = 2MR/\left(2(n-1)(2m/R+3p^2+12p)+3(k+N\log_2 N\right) \qquad \textit{[eq 52]}$$

CHAPTER 3
THE FUNDAMENTAL PROBLEMS AND SOLUTIONS TO ECHO CANCELLATION

Overview:

Communication applications are discussed. The applications that yield line echo are the Long-distance calls between ordinary fixed telephones and the digital data transmission on subscribers' loop. The application of calls between a cellular to a fixed telephone can produce either line echo or both line and acoustic echoes depending on whether the hands-free operation on the cellular is used. Tele-conferencing/ videoconferencing application causes acoustic feedback coupling between the loudspeaker and microphone, and thus creates the acoustic echo. To remove the line and acoustic echoes successfully requires the use of adaptive echo cancellers. These devices have better performance than the non-adaptive echo suppressors. There are several problems associated with the design of effective echo cancellers, i.e. divergence due to double-talking or silent far-end signal and residual echoes. Most existing echo cancellers are designed with adaptive transversal finite impulse response (FIR) digital filters, and based on variations of the least mean square (LMS) and least square (LS) algorithms. Therefore, the concepts of conventional LMS and LS algorithms for the use in echo cancellers are discussed. Other methods are also recommended that can overcome the inherent problems of slow convergence in the LMS algorithm and high computation in the LS algorithm.

3.1 Introduction

Echo is a phenomenon in which a delayed and distorted version of an original sound or electrical signal is reflected back to the source. There are two types of echo, namely line and acoustic Echoes. Telephone line echo the author is with the Signal Processing Group, Dept. of Applied Electronics, Chalmers University of Technology, Gothenburg, Sweden. Results from impedance mismatch at the telephone ex-change hybrids where the subscriber's two-wire line is connected to a four-wire line. If the communication is just between two handsets, then only line echo will occur. However, if the telephone connection is between one or more hands-free telephones, a feedback

path is set up between the loudspeaker and microphone at each end. This acoustic coupling is due to the reflection of loudspeaker's sound from walls, floor, ceiling, windows and other objects back to the microphone 1. Adaptive cancellation of this acoustic echo has become very important in hands-free telephony or teleconference communication system. The effects of an echo depend on the time delay between the incident and the reflected waves, the strength of the reflected waves and the number of paths through which the waves are reflected. If the time delay is not long, the acoustic echo can be perceived as soft reverberation, which adds artistic quality for example in a concert hall. However, echo arriving a few tens of milliseconds or more after the direct sound will be highly undesirable because long delayed echo is irritating. Likewise in line echo, the short delayed echo cannot be distinguished from the normal side-tone of the telephone, which is intentionally inserted to make the telephone communication channel sound "alive", and a round trip delay of more than 40msec will cause significant disturbances to the talker. Such a long delay

Is caused by the propagation time over long distances and/or the digital encoding of the Transmitted signals. In digital cellular systems, the one-way transmission delay is about 100ms when blocks of speech samples are transmitted in wireless, and the speech and channel coding methods used in radio communication cause this delay. It is worse in geostationary satellite links, which have a round trip delay of about 520msec. The International Telecommunical. The coupling can also due to the direct path from the loudspeaker to microphone.

Union-Telecommunication Standardization Section (U-TSS) recommends the use of echo cancellers for calls with round-trip delay that is above 50msec. In digital cellular communications, these devices are normally located at the mobile switching center (MSC), while in long distance telephony path, they are usually

Located in an international switching center (ISC). This report describes the echo phenomena and the general methods of removing the echo in a long-distance International call between ordinary fixed telephones, in a full-duplex data transmission between voice-band modems, in a national call between a fixed telephone and a cellular telephone and in teleconference/ videoconference communication systems.

3.2 Long-Distance International Calls Between Ordinary Fixed Telephones

A simplified long-distance telephone connection is shown in Figure 3.2. This connection contains two-wire sections on the ends (the subscriber loops and possibly some portion of the local network), and a four-wire section in the center (carrier systems for medium to long-haul transmission).

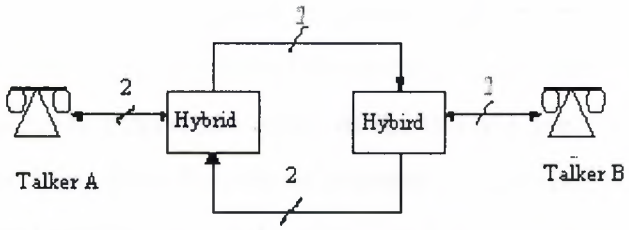


Figure 3.2, A long-distance connection.

Every telephone in a given geographical area is connected to the local exchange by a two-wire line, called the subscriber's loop, which carries connection for both directions of transmission. A local call is established simply by connecting the two subscribers' loops at the local exchange. However, repeaters are used to amplify the speech signal when the distance between the two telephones exceeds 50 km. Thus, a four-wire line is required which segregates the two directions of transmission on two different transmission paths. A hybrid is used to convert from the two-wire to four-wire line and vice versa as shown in Fig. 3.2.1 and it is basically a bridge network.

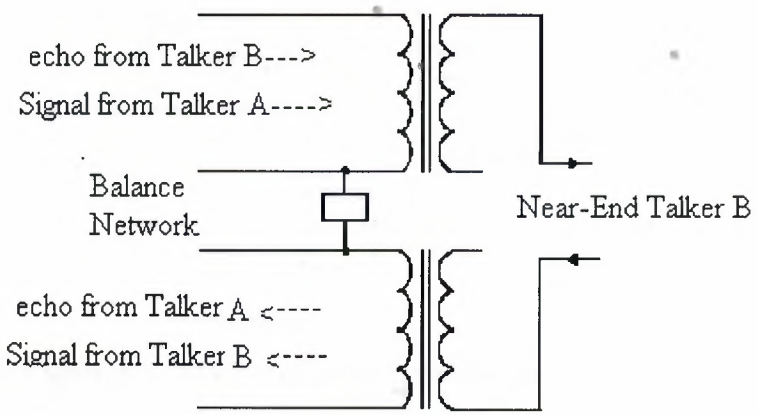


Figure 3.2.1,Two-wire to four-wire hybrid.

An echo can be decreased if the hybrid has significant loss between its two four-wire ports. To achieve this large loss will require the hybrid to be perfectly balanced by impedance located at its four-wire portion. Unfortunately, this is impossible in practice because it requires the knowledge of the two-wire impedance, which varies considerably over the population of sub-scriber loops. When the bridge is not perfectly balanced, impedance mismatch occurs and this causes some of the talker's signal energy to be reflected back as echo. The crucial talker path, as shown in Fig. 3a, requires that the hybrid does not have a lot of attenuation between its two-wire and either four-wire port. There are two types of echo as shown in Fig. 3b and 3c. Talker echo results in the talker hearing a delayed version of his or her own speech, while in listener echo it is the listener who hears a delayed version of the talker's speech. When there is insignificant transmission delay, this phenomena presents no problem, and, in any case, the talker or listener already heard the "side tone" of his or her own speech via the telephone instrument. The effects of echo can be controlled by adding an insertion loss to the four-wire portions of the connection, since the echo signals experience this loss two or three times (for talker and listener echo respectively) while the talker speech suffers this loss only once. However, on long connections, this loss can become very significant and as a result it is not an optimum solution and other echo control techniques must be used.

3.3 Echo Suppressor

Echo suppressors have been used since the introduction of long-distance communication

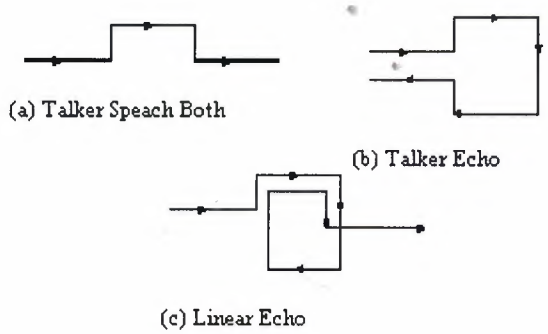


Figure 3.3, Sources of echo in the telephone network

Vantage of the fact that people seldom talk simultaneously. It is also helped by the fact that during such double-talking, poor transmission quality is less noticeable. As shown in Fig.3.4, 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 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 pre-vented. Then, when talker B speaks or in double talking Case, the same switch is closed and a symmetric one at the far-end speech path is opened. However, echo suppressors can clip speech sounds and introduce impairing interruptions. For example, if the near-end talker is initially listening to the far-end but suddenly wants to talk, it is quite likely that the switch preventing his or her speech from being transmitted will not close quickly enough, and the far-end talker may not receive the entire message. This distortion is more pronounced in long transmission with a round-trip delay of more than 200 ms. Due to the long delay, a quick response by talker B may cause suppression of something said by talker A at a later time. Talker B, encouraging him/her to stop and wait for talker A to get
Through notices this deletion. The resulting confusion may stop the conversation entirely while each party waits for the other to say something.

3.4 Adaptive Echo Cancelled

An alternative solution to remove echo is to use an echo cancelled as shown in Fig. 3.4.1. The echo canceller mimics the transfer function of the echo path to synthesize a replica of the echo, and then subtracts that replica from

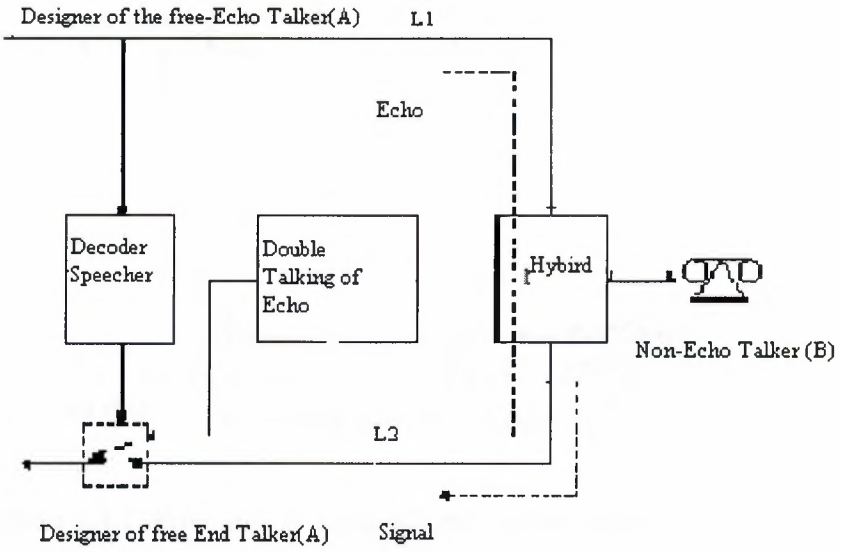


Figure 3.4, Echo suppressor at near-end talker B path.

It passes talker A's signal and blocks his/her echo with the open switch. The combined echo and near-end speech signal to obtain the near-end speech signal alone. However, the transfer function is unknown in practice and so it must be identified. The solution to this problem is to use an adaptive filter that gradually matches its impulse response to the impulse response of the actual echo path, as shown in Fig. 3.4.2, The echo path is highly variable, depending on the distance to the hybrid, the characteristics of the two-wire circuit, etc. These variations are taken care of by the adaptive control loop built into

The canceller. The canceller in Fig. 3.4.1 - 3.4.3 is for one direction of transmission only (from talker A to talker B).

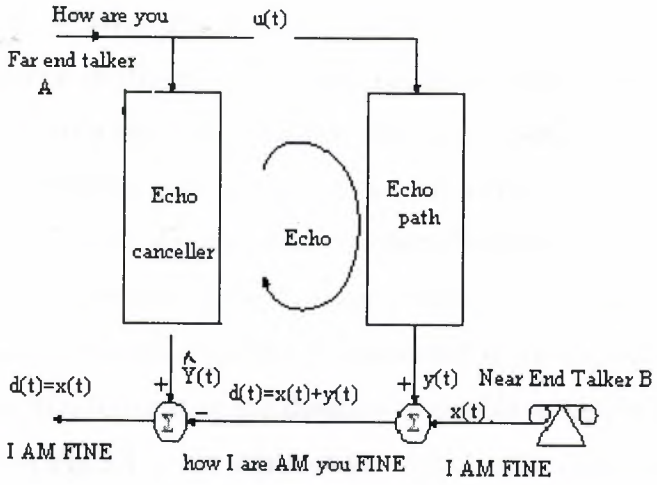


Figure 3.4.1, Principle of a non-adaptive echo canceller

The adaptive canceller in Fig. 3.4.2 - 3.4.3 is placed at the four-wire path near the origin of the echo. The synthetic Echo, $\hat{r}(n)$ is generated by passing the reference input signal, *y (n)* from the far-end talker through the adaptive filter that will ideally match the transfer function of the echo path. The echo signal, *r (n)* is produced when *y (n)* passes through the hybrid. The signal, *r (n)* plus the near-end talker signal, *x (n)* constitutes the "desired" response for the adaptive canceller. The two signals, *y (n)* and *r (n)* are correlated because the later is obtained by passing *y (n)* through the echo channel. The synthetic echo is subtracted from the desired response *r (n)+x (n)* to yield the canceller error signal,

$$e(n) = r(n) - \hat{r}(n) + x(n) \qquad [eq\ 1]$$

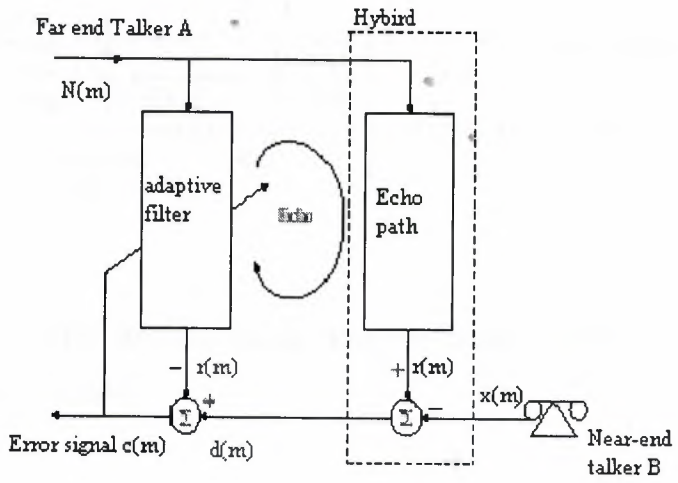


Figure 3.4.2, General configuration of an adaptive echo canceller

Similar to the echo suppressors, adaptive echo cancellers also face the problem of double-talking. The situation that must be avoided is interpreting *x(m)* as part of the true error signal, as shown in (figure 3.1), which results in making large corrections to the estimated echo path in a doomed-to-failure attempt to cancel it. In order to avoid this possibility, the tap weights must not be up-dated as soon as double-talking is detected as shown in Fig 3.4.3. The design of a good double-talking detector is difficult. Even with the assumption of a fast-acting detector, there is still a possibility of changes occurring in the echo channel during the time that the canceller is frozen, which leads to increase unconcealed echo. But, fortunately the duration of double-talking is usually short. In the system, as shown in Fig.3.4.3, the effect of the speech/echo misclassification is that the echo is sub optimally cancelled. This is more acceptable than in the case of echo suppressor that removes part of the speech signal during misclassification.

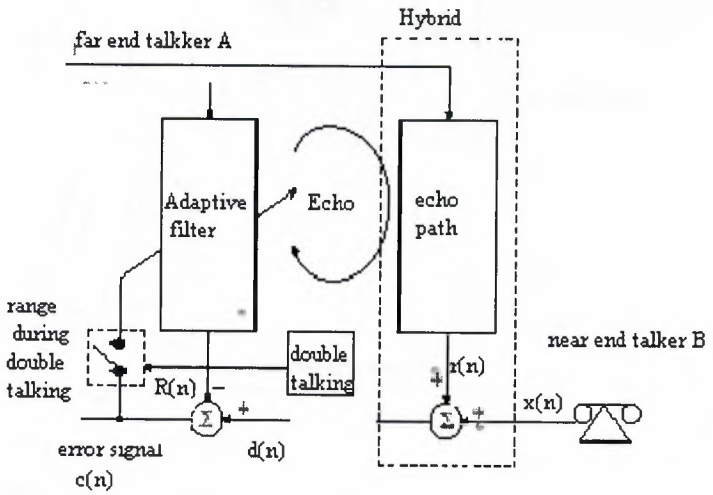


Figure 3.4.3,the double-talking detector stops adjustment when the near-end talker is active

To add to the problems of effective echo canceller de-sign, it sometimes occur that no far-end signal is present and even an echo canceller which is working well will leave

some residual unconcealed echo. In the former case, the adaptation is generally halted once the signal is estimated to be insignificant and in the latter case, a non-linear processor is used to remove the residual echo [8]. The presence of residual echo or the limitations on the achievable cancellation ratio are imposed by the presence of additive noise, nonlinear distortion, echo dispersion beyond the length of the transversal filter and digital resolution constraints. The working mechanism of the non-linear processor is to block this small-unwanted signal if the signal magnitude is lower than a certain (small) threshold value during single talking. However, the non-linear processor will only distort and not block the near-end signal during double-talking. The distortion is generally unnoticeable and so the processor does not have to be removed during double-talking. In practice, it is desirable to cancel the echoes in both directions of the trunk. Therefore, two adaptive echo cancellers are used as shown in Fig. 3.4.4. One of the cancellers removes the echo from each end of the connection. The near-end talker for one canceller is the far-end talker of the other. The requirements of an echo canceller are influenced by the following transmission characteristics: í *One way transmission delay*: The required

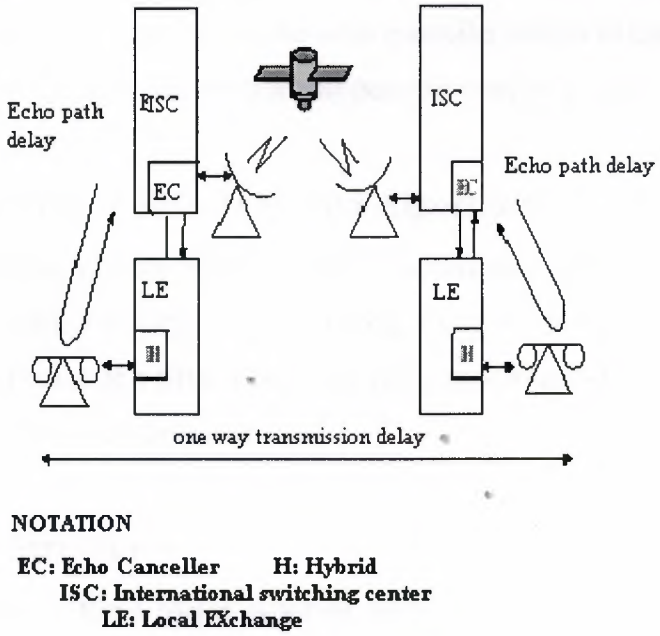


Figure 3.4.4, Location of echo canceller

Echo return loss (TERL) for a connection is determined by the one-way transmission delay (de- picted in Fig. 3.4.4). This loss is defined as the total level loss between the talker's mouth and his ear. ITU-T Recommendation G.131 sets the minimum value of

NEAR EAST UNIVERSITY LIBRARY LEFKOŞA

TERL at the range of 46-54dB. İ *Echo path delay*: The number of coefficients needed in the adaptive filter depends on the echo path delay and the length of the impulse response of the hybrid, which both are relatively short. The echo delay is defined as two times the delay from the canceller to the hybrid as shown in Fig. 3.4.4. For the adaptive echo canceller to operate properly the impulse response of the adaptive filter should have a length greater than the combined effect of the hybrid's impulse response length and echo path de- lay. Let Ts be the sampling period of the digitized speech signal, *M* be the number of adjustable coefficients in an adaptive finite impulse response filter, and τ be the combined effect to be accommodated. Therefore,

$$MT_s > \tau \qquad [eq\ 2]$$

The value of Ts is 125μs in the telephone network, and if τ ms, then we must choose M *>240* taps for a satisfactory performance. *Type of transmission and end-user equipment*: Non-linearity's in the echo path will affect the performance of the echo canceller. Devices such as bit-rate coders and end-user equipment with acoustic cross talk can also cause non-linearity. The TERL requirement must be met even when this factor is considered. Naturally, the degree of impairment will vary from connection to connection. Thus, it is crucial that the echo canceller adapts to the specific situation on a per call basis in order to achieve the best possible speech quality.

3.5 Adaptive Filter Structure And Algorithm

The selection of the adaptive filter structure and adaptation algorithm will affect the echo canceller's accuracy of estimating the echo path and the adaptation speed. Naturally, the choice of a filter's structure has a pro- found effect on the operation of the selected algorithm as a whole.

3.5.1 Filter Structures

There are three major types of *finite impulse response* (FIR) filter, namely transversal filter, lattice predictor and systolic array. The transversal filter is shown in Fig. 3.5, which is also known as tapped-delay line filter. In practice, FIR transversal filter structure is used because the convergence property of its coefficients to the optimum value is well proven. The important drawback is that as the echo path delay is increased, the number of taps increases proportionally and the convergence speed

decreases. A lattice predictor is modular in structure, where it consists of a number of individual stages, each in the appearance of a lattice. The weights are the filter coefficients and are adapted in a similar way as in a FIR transversal filter. The weighted sum of signals obtained at each stage of the lattice gives the echo replica. It is used usually to whiten the input speech signal so that rapid convergence is obtained. A systolic array represents a *parallel computing* network suitable for mapping a number of important linear algebra computations, such a matrix multiplication, triangularization and back substitution. It is well suited for implementing complex signal processing algorithms; the echo canceller may also be an infinite impulse response (IIR) filter. The main advantage of an IIR is that a long delayed echo can be synthesized by a relatively small number of filter coefficients due to the presence of a feedback loop. However, this feedback loop presents the potential problem of instability.

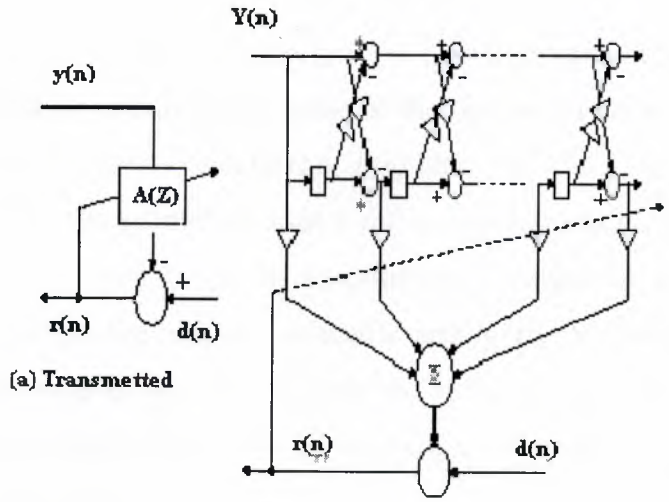


Figure 3.5.1: Adaptive digital filter structures

3.5.2 Adaptation Algorithm

There are two basic categories of algorithms for echo cancellers, i.e. the *least mean square* (LMS) and the *least square* (LS) algorithms. **LMS algorithm** The approach to examine LMS algorithm is to start with the concept of Wiener filters, follows by the method of steepest descent before these concepts are used to derive the conventional LMS algorithm. The discussion in this section assumes an ideal case where the echo path is almost stationary with no added noise component and the

double-talking situation does not exists. If $x(n)=0$ in (1), the cost function can be defined as the mean-squared error

$$J = E\left[e(n)^2\right] \qquad [Eq\ 3]$$

Where $e(n) = r(n) - \hat{r}(n)$ and E denotes the statistical expectation operator. For the cost function J to attain its minimum value, all the elements of the gradient vector ∇J (must be simultaneously equal to zero. Under this set of conditions, the filter is said to be optimum in the *mean squared-error sense*, which produces the minimum mean squared error 7 At this point the tap weight vector assumes its optimum value W_{opt} that satisfies the WeinerHopf equation. The *method of steepest descent* is basic to the understanding of other adaptive algorithms in which gradientbased adaptation is implemented in practice, such as the LMS algorithm.

Using $\hat{r}(n) = w(n)^T y(n)$ for a transversal FIR filter, we will obtain

$$\nabla J(n) = -p + 2Rw(n) \qquad [eq\ 4]$$

Where $P=E[y(n)\ r(n)]$, $R=E[y(n)\ y(n)^T]$ superscript τ denotes matrix transpose operation and W *(n)* denotes the value of the tap weight vector, $W=(w_{0,\ w1,\ldots,}w_{m-1})^T$ at time *n*. Since *W (n)* varies with time n, *J (n)* also varies in a corresponding fashion and this signifies that the estimation error e *(n)* is non-stationary. The dependence of J *(n)* on W *(n)* is visualized as the error-performance surface of the adaptive filter. The adaptive process has the task of continually seeking the minimum point of the surface, where the tap weights take on the optimum value W_{opt} According to the method of steepest gradient, the updated value of tap weight at time $n+1$ is computed by using the simple recursive relation

$$W(n+1) = W(n) + \frac{1}{2}\mu(-\nabla J(n)) \qquad [eq\ 5]$$

Where μ is a positive real-valued constant. The factor 1/2 is used to cancel the factor 2 in (4). The equation also shows that successive corrections to the tap weight vector is in the direction of the negative gradient vector which should eventually lead to J_{min} at which point the tap weights equal W_{opt}.

Substituting (4) into (5), we will get a simple recursive formula

$$W(n+1) = W(n) + \delta W(n) \qquad [eq\ 6]$$

$$= W(n) + \mu(P \cdot Rw(n)) \qquad [eq\ 7]$$

$$= W(n) + \mu E\left[y(n)e(n)\right] \qquad [eq\ 8]$$

According to (6) - (8), the correction $\delta W(n)$ is applied to the tap weight vector at time $n+1$. Thus, μ can be referred as the *step-size parameter* that controls the incremental correction applied to the tap weight vector as we proceed from one iteration cycle to the next. The theory of the LMS algorithm was first introduced in 1960 for adaptive switching by its originators, Windrow and Hoff [9]. Fig. 3.5.2 illustrates the block diagrams of a LMS adaptive transversal FIR filter that can represent the adaptive filter block of the echo canceller as shown earlier in Fig. 3.4.2. The exact measurement of gradient vector requires prior knowledge of vector P and matrix R which is not possible in reality. Thus, the gradient vector can only be estimated from the available data. Here we have used a hat over some symbols to distinguish them from the values obtained using the steepest descent algorithm. First, P and R are estimated by using *instantaneous estimates* that are based on sample values of the tap $Y(n)$ and $r(n)$

$$\hat{R}(n) = y(n)\, y(n)^T \qquad \textit{[eq 9]}$$

$$\hat{P}(n) = y(n)\, r(n) \qquad \textit{[eq 10]}$$

Correspondingly, the new recursive relation is

$$\hat{w}(n+1) = \hat{w}(n) + \mu\, y(n)\left(r(n) - \hat{w}(n)^T\, y(n)\right) \qquad \textit{[eq 11]}$$

$$= \hat{w}(n) + \mu\, y(n)\, e(n) \qquad \textit{[eq 12]}$$

Comparing with the method of steepest descent, we see that the expectation operator $E[\cdot]$. Is missing in the LMS algorithm. Accordingly, the recursive computation of each tap weight in the LMS algorithm suffers from a *gradient noise*, which causes $\hat{w}(n)$ to move randomly around the minimum point of the error-performance surface rather than terminating on the Wiener solution W_{opt} as before. Since the LMS algorithm involves feedback in its operation, an issue of stability is raised. In this case, a meaningful criterion is to require

$$J(n) \to J(\infty) \text{ as } n \to \infty \qquad \textit{[eq 13]}$$

Where J(n) does the LMS algorithm produce the mean-squared error at time and its final value, $J(\infty)$ is a constant. An algorithm that satisfies this requirement is said to be *convergent in the mean square*. For the LMS algorithm to satisfy this criterion, the step size parameter μ has to satisfy a certain condition related to the eigenstructure of the correlation matrix of the tap inputs, i.e.

$$0 < \mu < \frac{2}{\lambda_{\max}} \qquad \textit{[eq 14]}$$

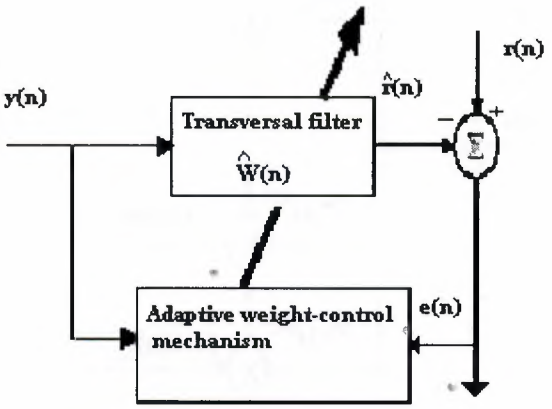
Where λ_{max} is the largest eigenvalue of the input signal correlation matrix R. In practice, the value of λ_{max} is not available. Thus, a conservative estimate for λ_{max} is to use tr.[R] since $\lambda_{max} \leq R$ where tr.[.] Is the trace of a matrix. A more restrictive bound can be obtained by assuming that R is not only positive definite but also Toeplitz with all the elements on its main diagonal equal to $r(0)$. The assumption is true for a stationary input signal. Thus,

$$tr[R] = Mr(0) = \sum_{k=0}^{M-1} E[y(n-k)^2] \qquad [eq\ 15]$$

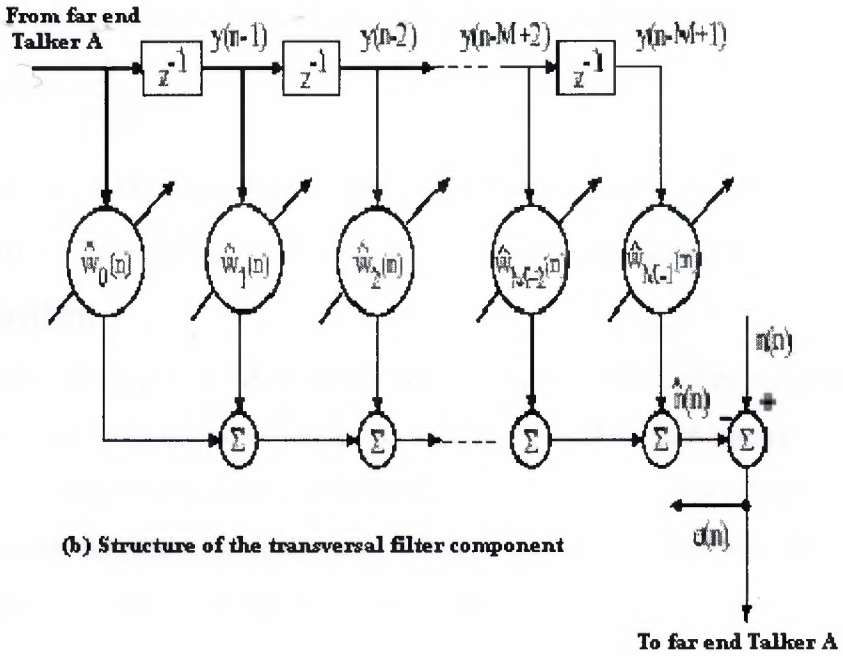
Where M is the filter length; and (14) becomes

$$0 < \mu < \frac{2}{\sum_{k=0}^{M-1} E[y(n-k)^2]} \qquad [eq\ 16]$$

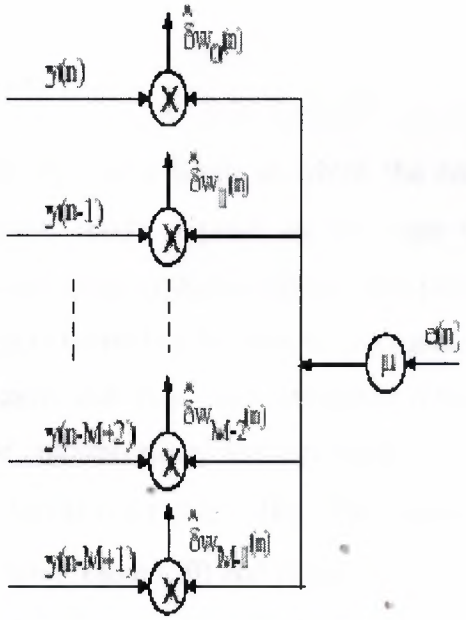
In (12), the correction $\mu y(n)\, e(n)$ is directly proportional to $y(n)$ thus, when $y(n)$ is large, the LMS algorithm experiences a *gradient noise amplification* problem. Another drawback is that the convergence rate fluctuate considerably if the condition number of input signal correlation matrix R is large. One of the sources of such variability in eigenvalues is the change of input signal level. The *normalized LMS* (NLMS) algorithm is



(a) Bolck-diagram of adaptive transversal



(b) Structure of the transversal filter component



(C) Structure of the adaptive weight- control mechanism

Figure 10: LMS adaptive transversal filter

Used to overcome these two problems and (12) can be rewritten as

$$\hat{w}(n+1) = \hat{w}(n) + \frac{\hat{\mu}}{|y(n)|^2} y(n)\, e(n) \qquad [eq\ 17]$$

where $\hat{\mu}$ is a positive real scaling constant. Comparing (12) with (17) we can observe that the normalized LMS algorithm has a time-varying step-size parameter because

$$\mu(n) = \frac{\hat{\mu}}{|y(n)|^2} \qquad \textit{[eq 18]}$$

The normalized LMS algorithm is convergent in the mean square if $\hat{\mu}$ satisfies the condition of $0 < \hat{\mu} < 2$ when $|y(n)^2|$ can be estimated as Mr.(0) as in (15).

LS algorithm

As mentioned earlier, the LMS algorithm estimates the statistical expectation operator E [.] by using the instantaneous values as in (9)-(12). The least-square (LS) algorithm avoids such estimation because it is deterministic in approach. Specifically, it minimizes a cost function that consists of the sum of error squares by choosing optimally the tap weights $(w_0, w_1, \ldots\ldots, w_{n-1})$ of the transversal filter:

$$J(w) = \sum_{i=i_1}^{i_2} e(i)^2 \qquad \textit{[eq 19]}$$

$$= \sum_{i=i_1}^{i_2} \left(r(i) - \hat{r}(i)\right)^2 \qquad \textit{[eq 20]}$$

Where i_2 and i_2 define the index limits at which the error minimization occurs. The values assigned to these limits depend on the type of data windowing methods employed. It can be covariance, autocorrelation, and prewindowing or postwindowing method. The discussion is limited to the prewindowing method which assumes the input data prior to $i=0$ are zero, but make no assumption about the data after $i=N-1$. Thus, $i_1=0$ and $i_2=N-1$. For minimization, the tap weights : $w_0, w_1, \ldots\ldots, w_{m-1}$ are held constant during the interval $i_1 \leq i \leq i_2$. The filter resulting from the minimization is termed as a linear LS filter. From (20) and using $\hat{r}(i) = w^T y(i)$, the linear LS estimator can be found by minimizing

$$J(w) = \sum_{i=0}^{N-1} \left(r(i) - w^T y(i)\right)^2 \qquad \textit{[eq 21]}$$

$$= (r - Hw)^T (r - Hw) \qquad \textit{[eq 22]}$$

The matrix H is a known NxM matrix of full rank M and it is referred as the input sample observation matrix:

$$H = \begin{pmatrix} y(0) & 0 & ....0 \\ y(0) & 0 & ....0) \\ y(0) & 0 & ....0 \\ . & . & . \ . \\ . & . & . \ . \\ y(N-1) & y(N-1) & y(N-1) \end{pmatrix} \qquad [eq\ 23]$$

And the vector $r = (r(0), r(1), r(2),.....,r(N-1))^T$ is the $Nx1$ vector of the echo signal. The minimization is easily accomplished by setting the gradient to zero,

$$\nabla J(w) = -2H^T r + 2H^T Hw = 0 \qquad [eq\ 24]$$

And yields the LS estimator,

$$\hat{w} = (H^T H)^{-1} H^T r \qquad [eq\ 25]$$

The equation $H^T Hw = H^T r$ to be solved for $\hat{w}$ is termed the normal equation. The assumed full rank of H guarantees the inevitability of $H^T H$. The inversion requires $0(M^3)$ computations. However, since the filter structure is transversal FIR, then the matrix H is Toeplitz. This property indicates that the inversion can be performed in $0(M^2)$ operations [10].

The alternative algorithm to LS algorithm is the recursive least-squares (RLS) algorithm, which obtains an optimum estimate of filter tap weight recursively sample-by-sample. It minimizes a weighted sum of squared errors that is usually defined as,

$$J(n) = \sum_{i=0}^{n} \beta(n,i)\, e(i)^2 \qquad [eq\ 26]$$

Two common choices for $\beta(n,i)$ is *exponential weighting factor* or *forgetting factor* and *sliding window factor* which are defined respectively by (27) and (28),

$$\beta(n, i) = \lambda^{n-i}, \quad 0 < \lambda < 1 \qquad [eq\ 27]$$

$$\beta(n,i) = \begin{cases} 1 & \text{for} \quad 0 \le i \le k-1 \\ 0 & \text{otherwise} \end{cases} \qquad [eq\ 28]$$

The factor (27) fades out the effect of past samples exponentially; and the factor (28) uses only the most recent samples to do the estimation and weighs these samples equally. For simplicity, (27) is used and (26) becomes,

$$J(n) = \sum_{i=0}^{n} \lambda^{n-1} \left(r(i) - \sum_{i=0}^{M-1} wn(L) y(i-L) \right)^2 \qquad [eq\ 29]$$

Similar to the LS method, the optimum value of the tap weight vector, $\hat{w}(n)$ is defined by the normal equations written in matrix form:

$$\phi(n)\,\hat{w}(n) = z(n) \qquad [eq\ 30]$$

where $\phi\ (n)$ is the MxM weighted correlation matrix of the tap inputs, $y(I)$ and $z(m)$ is the $Mx1$ weighted crosscorrelation vector between the tap inputs, $y(I)$ and the desired response, $r(i)$. The correlation matrix, $\phi\ (n)$ is defined as

$$\phi(n) = \sum_{i=0}^{n} \lambda^{n-i}\, y(i)\, r(i)^T \qquad [eq\ 31]$$

and the cross-correlation vectors(n) is defined as

$$Z(n) = \sum_{i=0}^{n} \lambda^{n-i}\, y(i)\, r(i) \qquad [eq\ 32]$$

Which arise naturally from (29). By re-arranging (31) and (32), the following recursive equations are obtained:

$$\phi(n)=\lambda\phi(n-1)+y(n)y(n)^T \qquad [eq\ 33]$$

$$Z(n)=\lambda Z(n-1)+y(n)\ r(n) \qquad [eq\ 34]$$

To compute $\hat{w}(n)$ by using (30), (33) and (34) requires the inversion of $\phi(n)$, which involves $0(M^3)$ operations for nonToeplitz matrix. In practice, such a computation should be avoided. This objective can be realized by using the well-known *matrix inversion lemma* that subsequently yields the following RLS algorithm for each time instant

$$K(n) = \frac{\lambda^{-1} p(n-1)y(n)}{1+\lambda^{-1} y(n)^T\, p(n-1)y(n)} \qquad [eq\ 35]$$

$$\hat{w}(n) = \hat{w}(n\text{-}1) + K(n)\ (r(n)\text{-}\hat{w}(n\text{-}1)^T\, y(n) \qquad [eq\ 36]$$

$$p(n)=\lambda^{-1}p(n\text{-}1)+ \lambda^{-1}K(n)y(n)^T\, p(n\text{-}1) \qquad [eq\ 37]$$

Where $p\ (n)=\phi(n)^{-1}$ and $K\ (n)$ is the gain vector. The recursive algorithm needs an initial value for P (n) and $\hat{w}(n)$. As a simple procedure, $p\ (0)=\sigma^{-1}\ I$ and $\hat{w}(0) = 0$.

Where I is MxM and identity matrix and is a small constant. It is of a great interest that no matrix inversion is required and the rate of convergence can be shown to be invariant with respect to the condition number of the ensemble-averaged correlation matrix σ of the input vector $y(n)$ The computation is now in the order of the RLS algorithm can also be deduced in the exact form directly from the covariance Kalman filtering by using the state-space model that matches the RLS problem [11]. Thus, the RLS algorithm can be viewed as a special case of Kalman filter, i.e. the deterministic approach of Kalman filter.

3.5.3 Other Algorithms

Before ending the discussion in this section, let us briefly discuss other methods that may improve the algorithms that are discussed so far. When adaptive filters are implemented in any real application, the input data *y (n)* desired response d *(n)* and filter weights *w (n)* are necessarily represented by a finite number of bits. Similarly, the numerical operations involved are carried out using finite precision arithmetic. Thus, in a digital implementation of an adaptive filter, there are two sources of error, namely analog to digital conversion (quantization) and finite word-length arithmetic (round-off)[2] .

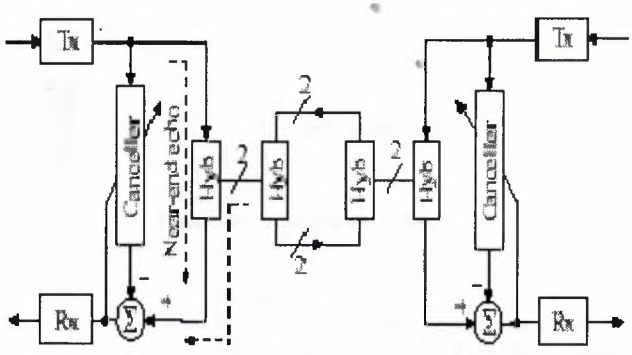
The recursive nature of the LMS algorithm means that the word-length needed for computation can grow unlimited and thus causing some bits to be discarded during the tap weight updating operations. Consequently, its performance deviates from the ideal (i.e. *infinite-precision*) form of the filter. In practice, introducing a leakage factor to the standard LMS algorithm as mentioned in [12] can counteract finite-precision effects. Basically, the leakage prevents the occurrence of overflow in a limited-precision environment by providing a compromise between minimizing the mean-squared error and containing the energy in the impulse response of the adaptive filter. The RLS has faster convergence rate and a better minimum mean squared error performance. Although its requirement of $O(M^2)$ operations per iteration is less than the $O(M^3)$ computations as in the LS algo-2. The types of error are mentioned in the brackets. Rithm, it is still much more than the LMS algorithm, which requires only O (M) operations per iteration. A variant of RLS, which has computation complexity comparable to LMS (i.e. increases linearly with the number of adjustable parameters), is known as *fast RLS*. Similar to the LS algorithm, the computation requirements can be reduced if the filter structure is a transversal FIR. Two important members of such algorithm family which use adaptive transversal filter structures are called the *fast transversal filter* (FTF) algorithm [13] and the *fast QR-decomposition-based recursive least-square* (FQR-RLS) algorithm [14].

3.6 Digital Data Transmission on Subscriber's Loop

The two-wire telephone line of subscriber's loop can be used for transmission of data

through a modem. Using the entire bandwidth of the wire can do this or transmitting the data on a bandwidth above the one that is used to carry speech signal. On an analog subscriber's loop, the speech signal occupies the bandwidth between 300 to 3.4kHz. A higher bit rate of up to 16kbps is transmitted by modulating the data signal appropriately onto a carrier signal at a band above 4kHz. Echo cancellation is needed to enable full-duplex communication within the same bandwidth over the subscriber's loop as shown in Fig. 3.6. The ITU-T Recommendation V.32 requires echo cancellers to be placed at the line interface where the hybrids connect the modem to the two-wire subscriber's loop. Several problems are associated with this type of application. Firstly, it is not practical to freeze the adaptation during double-talking (i.e. full duplex operation) because the echo path's characteristic is likely to change during this lengthy communication session. Secondly, the far-end echo (that is returned from the far-end hybrid) must also be taken into account. Subsequently, the entire echo delay becomes very large which is unique to echo cancellation at the station location. If the circuit includes a satellite communication in its four-wire link, the far-end echo will be delayed for at least 500msec. Therefore, two cancellers are required, each for the near-end and the far-end echo at the station location. Lastly, a considerable higher level of echo cancellation is required. The data signal coming from a farend modem may be attenuated by 40 to 50dB.

End echo (that is returned from the first hybrid at the local station) can be 40 to 50dB higher than the desired signal. For reliable communication, the echo canceller must be able to attenuate the near-end echo by 50 to 60dB so that the signal power is at about 10dB above the echo.



Notation

Hyb	**Hybrid**
Rx	**Receiver**
Tx	**Transmitter**

Figure 3.6

Figure 3.6, Echo cancellers at station locations for full-duplex voice-band modems

In [15], a data-driven echo canceller for full-duplex data transmission with multitude modulation is presented. The proposed method has advantages over the signal-driven echo chancellors, which are inherently affected by the condition number problem. It also makes use of frequency-domain updating, but time-domain implementation of the canceller. This results in performance improvement in the dynamic range of achievable echo cancellation in finite precision and less computational requirements.

3.7 Cellular to Fixed Telephone Call

In digital cellular communication, the combination of speech coding, channel coding and signal processing involves considerable delays. In most cases, the delays are increased further by time division multiple access framing. The total one way delay can be from 30 to 120msec. Fig. 3.7, shows that only one canceller facing the public switched telephony network (PSTN) is needed in a digital cellular application. However, this is only true if the cellular telephone is assumed to behave in a perfect (or near perfect) 4-wire fashion with no significant echo path between the microphone and the earpiece. Hence no echo control device is needed to remove acoustic echo returned from the cellular telephone.

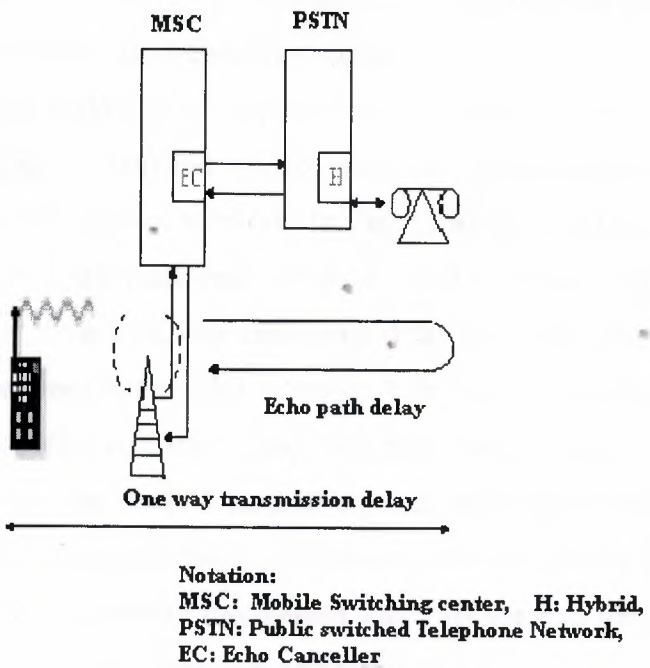


Figure 3.7, Cellular to Fixed Telephone Call

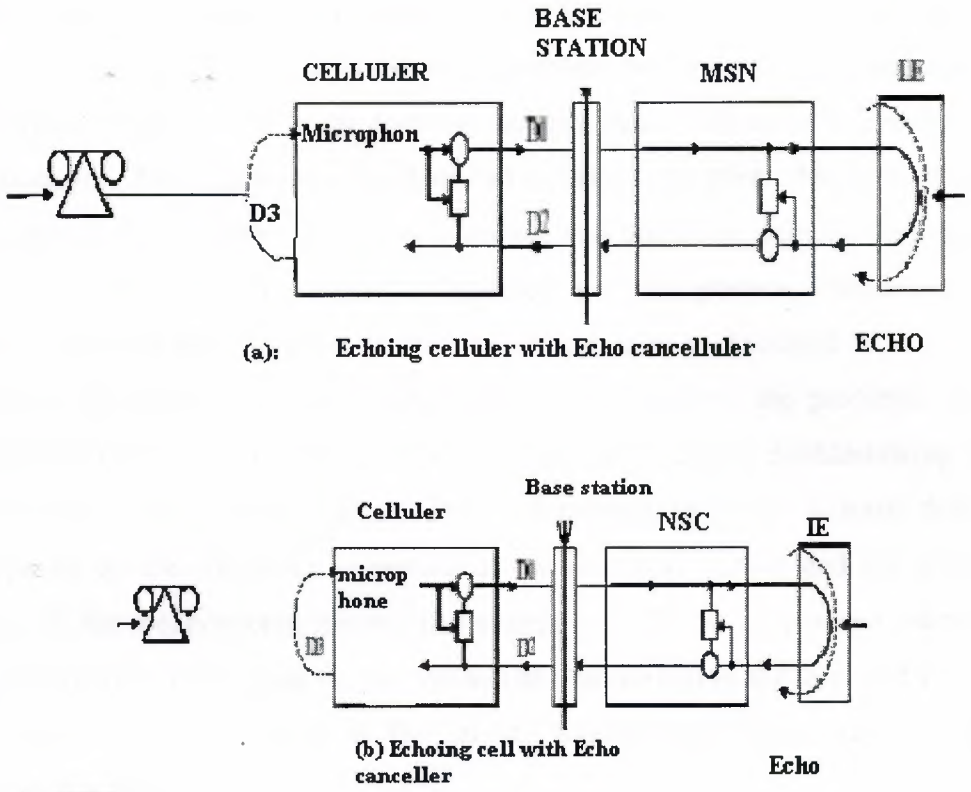


Figure 3.7.1, Cellular to Fixed Telephone Call with Echoing Cellular

Although the acoustic path between the earpiece to the microphone may be negligibly small, the possibility of using the hands-free operation in vehicle-supported mobile is a concern. This will cause additional echo path of 10 to 20msec between the loudspeaker and the microphone. Fig. 3.7.1(a) and 3.7.1(b) show two configurations to remove echo from the mobiles and the national network. In Fig. 13(a) the mobile echo canceller is located in the MSC, with the radio path delays D1 and D2 now enlarged by acoustic echo D3 in the mobile. If the minimum radio path delay is 70msec, the total echo delay will be 80 to 90msec as seen by the echo canceller A. in Fig. 3.7.1(b), the echo canceller is placed in the cellular telephone itself. The main disadvantage is that many more such devices are needed (one per mobile instead of one per radio channel) and it adds complexity to the already compact cellular telephone. However, it has the advantage of being closer to the echo-generating mechanism and thus the end loop delay is much shorter. Methods to remove this acoustic echo will be discussed

3.8 Teleconference/Videoconference Communication Systems

In Sec. 3.7, the acoustic echo problem in hands-free operation of a cellular telephone that is used in an enclosed environment, such as in a vehicle or room is introduced. This is a known problem that exists also in teleconference/videoconference systems, public address systems and hearing aids due to the acoustic feedback coupling of the sound waves between the loudspeakers and microphones. The cancellation of this type of echo differs from the cancellation of line echo as discussed in Sec. 3.2. And 3.3 due to the different nature of the echo paths. However, the problems of designing effective echo cancellers still remain, i.e. divergence during double-talking and no far-end signal; and residual echoes. The total round-gain of the acoustic feedback loop depends on the frequency responses of the electrical section and the acoustic signal path. If the speaker-room-microphone system is excited at a frequency whose loop gain is greater than unity, then the far-end signal is amplified in the loop and a distinguished howling or echoes results at the far-end loudspeaker. There are few methods for removing the echo:

Install a phase or frequency shifter in the electrical section of the feedback loop. A few hertz in the loop will shift the far-end signal before being retransmitted at the far-end

Loudspeaker. This method reduces the howling but not the overall echo. Reduce the total round-gain at those frequencies where the acoustic echo energy is concentrated by using an adaptive notch filter. The disadvantage is that some distortion on the desired signal frequencies also occurs. Use an adaptive echo canceller, which has the same working-mechanism as in a line echo canceller, i.e. attempts to synthesis a replica of the acoustic feedback at its output.

3.8.1 Adaptive Filter Structure and Algorithm

Both the NLMS and RLS adaptation algorithms as described in Sec. 3.5 can be used in an adaptive acoustic echo canceller. However, acoustic echo cancellation is Far more challenging than line echo cancellation for a number of reasons: The duration of the impulse response of the acoustic echo path is usually several times longer (100 to 400msec), which implies that impracticably large transversal FIR filters with thousands of taps are required. Perhaps modeling the echo path as a recursive IIR filter can reduce the number of filter coefficients. The characteristic of the echo path is more non-stationary, e.g. due to opening or closing of a door or a moving person, while the line

echo path is almost stationary once a call connection is established. Acoustic echo is due to reflection of signal from a multitude of different path, e.g. off the walls, the floor, the ceiling, the windows, etc. The echo path is not well approximated by a FIR or an IIR linear filter because it has a mixture of linear and non-linear characteristics. The reflection of acoustic signals inside a room is al-most linearly distorted, but the loudspeaker introduces nonlinearities. The main causes of this non-linearity are suspension non-linearity, which affects distortion at low frequency and in homogeneity of flux density, which produces non-linear distortion at large output signal levels. Due to these reasons, acoustic echo cancellers require more computing power to compensate the length of the impulse response and to obtain faster converging algorithms. The general LMS algorithm performs badly for long impulse responses and with speech as input signal. Although the RLS can increase convergence speed, it is still infeasible for long impulse responses due to complex computation. Therefore, other methods must be explored which are introduced briefly in Sec. 2.4. Some of the variations of algorithms that have already been used are summarized below: The LMS algorithm has a convergence behavior, which is very dependent on the relative strengths of the eigenvalues of the autocorrelation matrix of the input signal. The eigenvalues can be regarded as the strength of the predominant frequency components in the signal. If the eigenvalues are widely spread the LMS algorithm will converge very slowly for the weak signal. If the eigenvalues are nearly the same strength (white noise), the algorithm converges at the same rate for all eigenvectors. Thus, "whitening" the speech signal for LMS-adaptation by using linear predictive coding inverse filter can lead to faster convergence because the eigenvalues are less spread out [16]. In [17], a new adaptive IIR based on gradient instrumental variable (IV) algorithm is presented for echo cancellation application. This method is able to update the filter's coefficients during double-talking and is guaranteed to converge to a unique global minimum. It can also avoid the need to invert nonsymmetrical cross-correlation matrices, as in the traditional IV method. This gives the ad-vantage of having robust numerical stability and a computational complexity that is comparable to the equation error IIR LMS algorithm. As described in Sec 3.5.3, the *fast RLS* algorithm family is a potential solution to the high computation complexity of LS by reducing the computation to an order of O (M). However, its practical use is prevented in the past because of divergence due to numerical error accumulation in its linear prediction parameters. Efficient stabilization

technique to solve the problem of numerical instability is proposed in [18]. However, the proposed stabilized FTF can introduce complexity problem in the implementation of acoustic echo cancellation where long adaptive filter is required. The *fast Newton transversal filter* (FNTF) method in [19] provides an alternative solution with its 12L +2M multiplication' s complexity instead of in [18]. It estimates the covariance matrix of order MxM by extrapolating from a lower order LxL estimate. This is feasible because of the fact that the prediction part of the *fast RLS* is allowed to be lower than the filter size of 8M In general, this new method reduces the complexity and thus allows the practical use of *fast RLS* algorithm on long filters. It also provides possible tradeoffs between complexity and performance by appropriately choosing the prediction order versus the filter length.

Conclusions

The echo generating-mechanisms and the methods to remove echo are examined. The type of echo to be produced depends on the generating-mechanisms. The line echo results from the impedance mismatch at the telephone exchanges' hybrids, the acoustic echo is due to the reflection of a loudspeaker's sound from walls, floor, ceiling, etc. back to the microphone in an enclosed environment.

Echo suppressors can be used for all the mentioned applications, but they can cause damaging interruptions especially during double-talking.

Therefore, their performances are inferior than the adaptive echo cancellers, which are capable to gradually match their impulse responses to the impulse response of the echo path. This characteristic is important because the echo path can be quite non-stationary particularly in acoustic echo problem.

There are several factors that must be taken into consideration when designing effective echo cancellers, i.e. divergence due to double-talking and silent fared signal; and residual echoes. Unfortunately, there is not much research activity done on this subject. The design of echo cancellers also depends on the choice of the filter structure and the adaptation algorithm. Most practical systems use variations of the LMS algorithm on a transversal FIR digital filter. The main reasons for adopting these selections are the proven stability and convergence property of the transversal FIR filter and the low computation cost of LMS

The problems seems to be resolved because the implementation of echo cancellers at the ISC can afford to use high performance (and expensive) equipment to overcome the common problems of slow convergence and complex algorithm. In addition, the echo path is almost stationary once a call connection is established. In contrast, the acoustic echo problem is the main focus of current research works. The acoustic echo cancellers demand more computing power to compensate for the long impulse response and to have faster convergence.

Therefore, more advance methods are required such as the fast Newton transversal filter (FNTF).

References

[1] S. Haykin. (1996). *Adaptive Filter Theory*. Prentice-Hall, Upper Saddle River, NJ.

[2] S. Vaseghi.(1996). *Advanced Signal Processing and Digital Noise Reduction*. John Wiley & Sons, Chich-ester, NY.

[3] M. M. Sondhi and W. Kellermann.(1991). *Advances in Speech Signal Processing*, chapter Adaptive echoes cancellation for speech signals, pp. 327–356. Marcel-Decker.

[4] K. Murano, S. Unagami, and F. Amano. (1990). Echo cancellation and applications, *IEEE Comm. Magazine*, 28(1): 49–55, Jan.

[5] R. H. Moffett, 1987. Echo and delay problems in some digital communication systems, *IEEE Comm. Magazine*, 25(8): 41–47, Aug.

[6] C. W. K. Gritton and D. W. Lin. (April 1984). Echo cancellation algorithms, *IEEE Acoustics, Speech and Signal Processing Magazine*, 1(2): 30–38,

[7] A. Eriksson, G. Eriksson, J. Karlsen, 1996. A. Roxstrom, and T. V. Hulth, Ericsson echo cancellers, *Ericsson Review*, (1): 25–33.

[8] S. B. Weinstein. (1977). Echo cancellation in the telephone network, *IEEE Comm. Society Magazine*, 1(1): 8–15, Jan.

[9] B. Widrow and Jr. M. E. Hoffl (960). Adaptive switching circuits, In *IRE WESCON Conv. Rec.*, pp. 96–104.

[10] S. Marple, Feb. 1981. Efficient least-squares fir system identification, *IEEE Trans. on Acoust. Speech, and Signal Processing*, 29(1):62–73.

[11] A. H. Sayed and T. Kailath, July. (1994).A state-space approach to adaptive rls filtering, *IEEE Signal Processing Magazine*, 11(3):18–60.

[12] K. Mayyas and T.April. (1997). Aboulnasr, Leaky lms algorithm: Mse analysis for gaussian data, *IEEE Trans. on Signal Processing*, 45(4):927–934.

[13] J. M. Cioffi and T. Kailath, April. (1984). Fast, recursive-least-squares transversal filters for adaptive filtering, *IEEE Trans. on Acoust. Speech, and Signal Processing*, 32(2): 304–337.

[14] Z. Liu, , March.(1995).Qr methods of o(n) complexity in adaptive parameter estimation, *IEEE Trans. on Signal Processing*, 43(3):720–729.