ACOUSTIC ECHO CANCELLATION FOR THE ADVANCEDMENT IN TELECOMMUNICATION

BY

S. I. M. M. RATON MONDOL

ID: 072-19-701

A.A.M. MUZAHID

ID: 071-19-599

NASRIN AKTER

ID: 072-19-675

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Electronics and Telecommunication Engineering

Supervised By

Dr. A.K.M Fazlul Haque
Associate Professor
Department of ETE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

DECEMBER 2011

CHAPTER 1

INTRODUCTION

1.1 General Introduction

Echo is the reflected copy of the voice heard some time later and a delayed version of the original sound or electrical signal is reflected back to the source. Echo is a congenital problem which mainly occurs in PSTN (Public Switching Telephone Network). Echo occurs in analogy part of a telecommunication system [1]. Echoes of our speech are heard as they are reflected from the floor, walls and other neighboring objects. If a reflected wave arrives after a very short time of direct sound, it is considered as a spectral distortion or reverberation [2]. However, when the leading edge of the reflected wave arrives a few tens of milliseconds after the direct sound, it is heard as a distinct echo. In data communication, the echo can incur a big data transmit error. In applications like hands-free telecommunications, the echo, with rare exceptions, conversations take place in the presence of echoes [3].

The advent of telephony echoes have been a problem in communication networks. The most important factor in echoes is called end-to-end delay, which is also known as latency. Latency is the time between the generation of the sound at one end of the call and its reception at the other end. Round trip delay, which is the time taken to reflect an echo, is approximately twice the end-to-end delay. Echoes become annoying when the round trip delay exceeds 30ms. Such an echo is typically heard as a hollow sound. Echoes must be loud enough to be heard. Those less than thirty (30) decibels (dB) are unlikely to be noticed. However, when round trip delay exceeds 30 ms and echo strength exceeds 30 dB, echoes become steadily more disruptive. However, not all echoes reduce voice quality. In order for telephone conversations to sound natural, callers must be able to hear themselves speaking. For this reason, a short instantaneous echo, termed side tone, is deliberately inserted. The side tone is coupled with the caller's speech from the telephone mouthpiece to the earpiece so that the line sounds connected [4].

1.2 Need for Echo Cancellation

The rapid growth of technology in present decades has changed the whole dimension of global communications. New safety regulations are leading the field of tele-communications towards hands-free radio/telephones. With such a system, the speaker (operator) can talk freely and still concentrate on his driving task [5]. Now wireless phones are regarded as essential communications tools and have a direct impact on people's day-to-day personal and business communications. Examples of such systems are mobiles, VOIP calls by using, for instance, Skype, the teleconferencing for meetings or remote educations etc. and the hands-free operations have gained more and more popularity in recent years. But echo can degrade the quality of service, and echo cancellation is an important part of telecommunication systems. The development of echo reduction began in the late 1950s, and continues today.

Subscriber demand for enhanced voice quality over wireless networks has driven a new and key technology termed echo cancellation, which can provide near wire line voice quality across a wireless network. Today's subscribers use speech quality as a standard for assessing the overall quality of a network. Regardless of whether or not the subscribers' opinion is subjective, it is the key to maintaining subscriber loyalty. For this reason, the effective removal of hybrid and acoustic echoes, which are inherent within the telecommunications network infrastructure, is the key to maintaining and improving the perceived voice quality of a call [4].

1.3 Literature Study

According to asterisk echo cancellation previously called carbon profile is operated by generating multiple copies of the received signal, each delayed by some small time increment. These delayed copies are then scaled and subtracted from the original received signal. Srinivasaprasath Raghavendran et al [4] has proposed an echo cancellation process using MATLAB but there the far end signal and the near end signal is taken separately and then tested whether there is echo or not by Double talk detector. This process also includes NLMS and subtraction. Jerker Taudien el al [6] suggested Line probing is a method of inserting a known signal at the far-end and recording the near-end signal. The two signals are then analyzed together for various impediments. Three tone sweeps of different power levels are used to probe the line in the non-linear distortion analysis tool. The tone sweeps are recorded in three different power levels to detect clipping. Finally, Patrashiya Magdolina Halder et al [1] has proposed an echo cancellation process using Inverse filtering in MATLAB, which analyse the received signal and remove echo from the acquired signal for the field of VOIP. First voice signal is acquired with a additional speech recorder. Then this acquired voice signal is used to create **.wave** file using the audio signal.

In all these above proposed echo cancellation method require additional things beside MATLAB than this proposed method. This project has suggested the acoustic echo cancellation algorithm for both in Voice Over Internet protocol (VOIP) and telecommunication system using the MATLAB, without any additional software. A simple Frequency Domain Adaptive Filter (FDAF) is used here for cancelling echo without clipping and distorting the main signal.

1.4 Objectives of the Project

Acoustic echo control is challenging due to the complex nature of the echo signal [9]. The computational complexity is a main reason to why simpler echo control methods, often using voice-activity detectors, are used in many applications instead of AEC. Those methods do not allow full-duplex communication. The need for a robust and low-complex echo control method, that allows full-duplex communication [7]. The growing problem of acoustic echo from wireless calls is acknowledged, but many solution vendors have attempted to adapt hybrid echo cancellation methods to acoustic echo control, with poor results. However, since the computation power of regular home personal computers, (PCs), has increased tremendously and powerful software has evolved, it is now possible to perform real-time signal processing in the PC environment as well. The advent of this growing capability was the motivation for this research. The objective of the research was the implementation of a software acoustic echo canceller by design of a Frequency Domain Adaptive Filter (FDAF) running natively on a PC with the help of the MATLAB software.

1.5 Achievement

The goal of the project was to eliminate acoustic echo from the microphone signal as like the near end signal. It is not possible to remove echo 100% from echoed signal because if echo is tried to be eliminated completely then the attempt may distort the main signal. That is why echo we cannot be eliminated echo perfectly but the echo to a tolerable range. Design of a Frequency Domain Adaptive Filter (FDAF) for acoustic echo cancellation running natively on a PC with the help of the MATLAB software the echo is minimized to a level so that the received signal seems echo free.

1.6 Outline of the Project

This thesis provides an overview of an improved acoustic echo cancellation technique using a Frequency Domain Adaptive Filter (FDAF) with the help of the MATLAB software. This section is intended to give a short overview of the thesis, by describing the outline of each chapter.

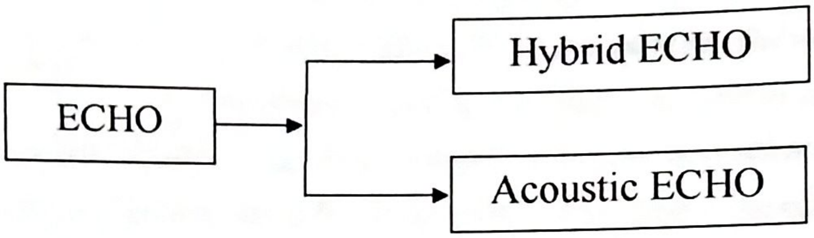
- **Chapter 1** discusses the definition of echo, the necessity of echo cancellers in telecommunications network and the study of different proposed method.
- **Chapter 2** gives an overview of the types of echo and their sources. It also discusses, in great detail the echo phenomena in telecommunication systems.
- **Chapter 3** presents all the theory backgrounds about the acoustic echo cancellation process and some other issues.
- **Chapter 4** discusses the simulation of the proposed algorithm, details of the simulation environment and the results obtained.
- Finally **Chapter 5** provides the conclusions of the work and some ideas concerning further work in this field.

CHAPTER 2

ECHOES IN TELECOMMUNICATION NETWORK

2.1 Echo Types

It is difficult for a listener to differentiate echo. There are two types of echo existing in telecommunication networks, namely *Hybrid echo* and *Acoustic echo*.



Hybrid echo (also known as "electrical echo") is caused by an impedance mismatch on the 4-wire to 2-wire conversion in wire line networks. It is the primary network-induced echo in today's networks. Acoustic echo is created as a result of insufficient acoustic isolation between the earpiece and the microphone in small handsets, or when acoustic waves are reflected against a wall or enclosure, typically when using a hands-free unit [2].

2.2 Hybrid Echo

Hybrid echoes have been inherent within the telecommunications networks since the advent of the telephone. This echo is the result of impedance mismatches in the analog local loop. For example, this happens when mixed gauges of wires are used, or where there are unused taps and loading coils.

In a wire line PSTN network, the subscriber is linked to the local exchange (central office) by a 2-wire analog connection know as the "local loop." From the local exchange, a 4-wire digital link is used to carry the signal longer distances. For this link, the send and receive paths use separate wire pairs. Between the two link methods is the *hybrid*, which converts the 4-wire interface to the 2-wire interface. The hybrid is a 4-port device where the fourth port is terminated with balancing impedance.

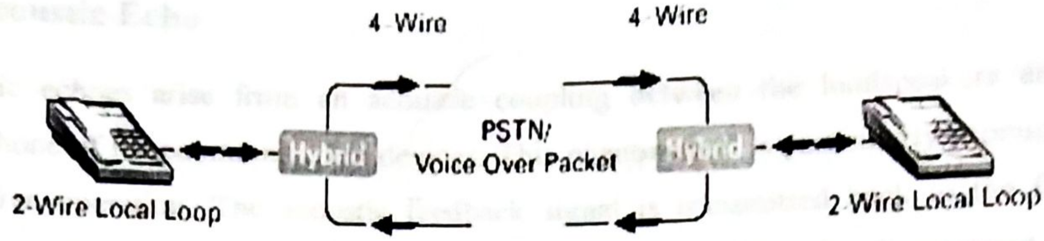


Figure 2.1 Hybrids in a PSTN

To avoid signal reflections in the hybrid, the balancing impedance of the hybrid must match the impedance of the 2-wire line terminated by the telephone. The impedance of the 2-wire line depends on many parameters, such as the length and type of cable, as well as the impedance of the telephone sets at the customer premises. In practice, the balance of the hybrid is only nominally achieved because the 2-wire loop's impedance cannot be determined in advance. Therefore, a fraction of the signal is reflected back to the sender, which is heard as echo.

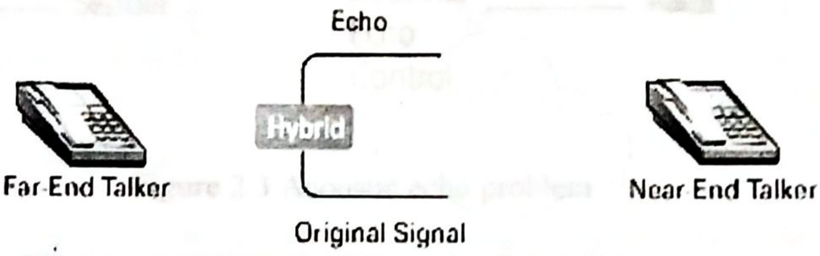


Figure 2.2 Echo from the Hybrid

In the early years, when the public network was entirely circuit switched, the hybrid echo was the only significant source of echo. Since the locations of hybrids and most other causes of impedance differences in circuit switched networks were known, adequate echo control could be planned and provisioned. However, in today's digital networks the points where two wires split into four wires is typically also the point where analog to digital conversion takes place. Regardless of whether the hybrid and analog to digital conversion is implemented in the same device or in two devices, the two to four wire conversions constitute an impedance mismatch and echoes are produced [4].

The degree of imbalance of the hybrid determines the strength of the echo reflection. This strength of the echo reflection is expressed in terms of Echo Return Loss (ERL). Echo Return Loss and additional echo metrics are explained in the section titled "Measuring Echo's Effect on Quality of Service."

2.3 Acoustic Echo

Acoustic echoes arise from an acoustic coupling between the loudspeakers and the microphone of telecommunication devices. This phenomenon is particularly prominent in hands-free operation. The acoustic feedback signal is transmitted back to the far-end subscriber, who notices a delayed version of his own speech. Echo signals represent a very distracting disturbance and can even inhibit interactive, full-duplex communication [8].

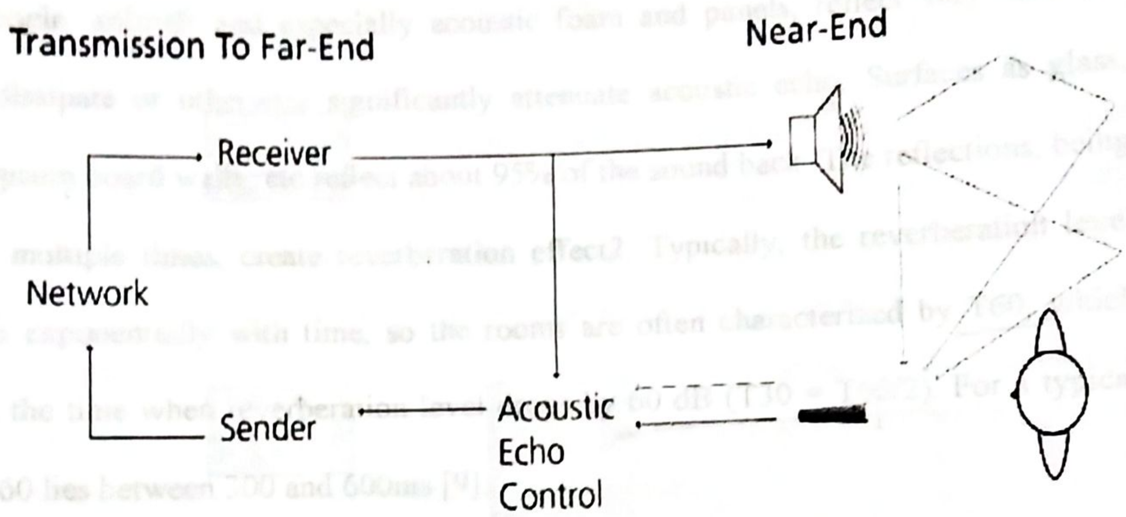


Figure 2.3 Acoustic echo problem

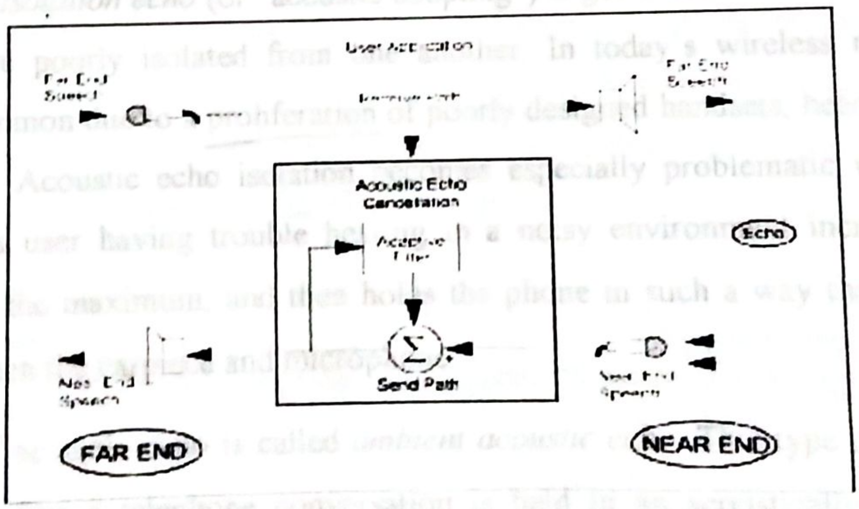


Figure 2.4 Acoustic Echo canceller block diagram

Simply Acoustic echo occurs when some of the sound from the speaker part of the telephone gets picked up and transmitted back by the microphone. Ambient acoustic echo is most likely to occur with "hands free" kits and speakerphones. About 10% of calls in today's wireless networks have acoustic echo (5% each for off-net and on-net callers) [5].

2.4 Sources of Acoustic Echo

Acoustic echo is formed when the sound emitted by a speakerphone's loudspeaker gets reflected from the walls, ceilings, floor, furniture, people, etc. back to the speakerphone's microphone. Sound pressure level decreases with each reflection. Some surfaces, as heavy carpet, soft furniture, open half-full bookshelves with varying format books in random order, people, animals and especially acoustic foam and panels, reflect very little but absorb, dissipate or otherwise significantly attenuate acoustic echo. Surfaces as glass, brick, gypsum board walls, etc reflect about 95% of the sound back. The reflections, being repeated multiple times, create reverberation effect2. Typically, the reverberation level decreases exponentially with time, so the rooms are often characterized by T60, which specifies the time when reverberation level drops by 60 dB (T30 = T60/2). For a typical office, T60 lies between 300 and 600ms [9].

There are two typical sources of acoustic echo:

First One, *Acoustic isolation echo* (or "acoustic coupling") is generated when the earpiece and microphone are poorly isolated from one another. In today's wireless networks, acoustic echo is common due to a proliferation of poorly designed handsets, headsets, and Bluetooth headsets. Acoustic echo isolation becomes especially problematic when, for example, a wireless user having trouble hearing in a noisy environment increases the earpiece volume to the maximum, and then holds the phone in such a way that there is poor isolation between the earpiece and microphone.

The second form of acoustic echo is called *ambient acoustic echo*. This type of acoustic echo is generated when a telephone conversation is held in an acoustically reflective environment. In this situation, the handset microphone first picks up the original audio stream, followed by the speech that is reflected from the walls.

2.5 Hybrid and Acoustic Echo Differences

2.5.1 Stationarity

The hybrid echo path is stationary, which means that it is invariant over time. Once the call path is established, the echo delay does not change during the course of the call. Acoustic echo, on the other hand, varies based on a multitude of external factors like the position of the talker in the room, or even head movements relative to the handset, which makes the acoustic echo a highly nonstationary signal.

2.5.2 Linearity

Linearity is how well the waveform of echo signal matches the original signal. Hybrid echo is a linear signal, which means that a linear mathematical model constructed inside the echo canceller can accurately predict the hybrid echo signal. Acoustic echo is not a linear signal. First, nonlinearities might be created by the analog circuitry. In the case of the handset/headset it includes the microphone, the microphone amplifier, the loudspeaker amplifier, and the loudspeaker. More significantly, for wireless and many VoIP calls, the voice codec processing introduces additional nonlinearities.

2.5.3 Dispersion

An echo signal is not a single reflection of the original signal, but is a consecutive reflection over a period of time. Echoes have a certain duration, or dispersion time, which is the period of time during which the echo reflection occurs. A hybrid echo has a typical dispersion of less than 10 ms. However, since acoustic echo can be generated by reflections from the environment, acoustic echo is more dispersive, with dispersion times of up to 100ms [10].

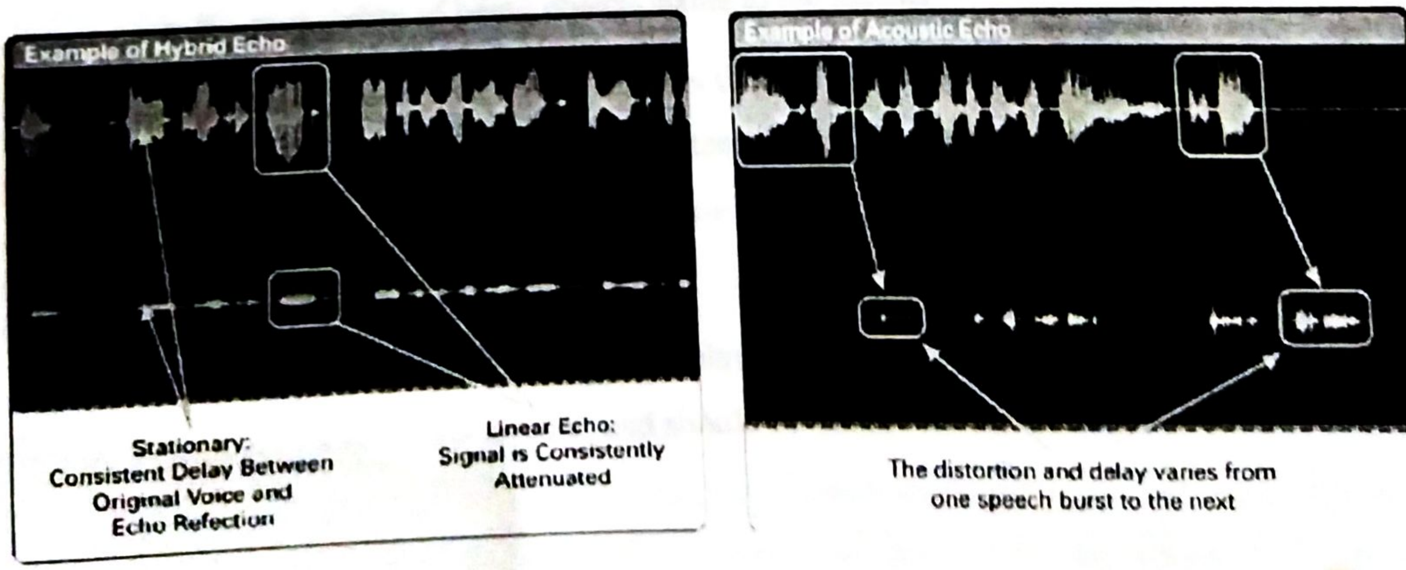


Figure 2.5 Hybrid and Acoustic Echo Waveform

2.6 Echo Objection Rate

Recommendation G.131 (Talker Echo and its Control) from the Telecommunication Standardization Sector of the International Telecommunications Union (ITU-T) presents results on the degree of annoyance of echo as a function of the amount of delay and Talker Echo Loudness Rating (TELR). TELR the echo loss as perceived by the listener, which is the loss between the talker's mouth and ear via the phone and echo path. If we account for 10 dB of loss introduced by the typical phone (per ITU-T G.121), the echo tolerance curve from G.131 can be shown as a function of ERL (Figure 2.5) [10].

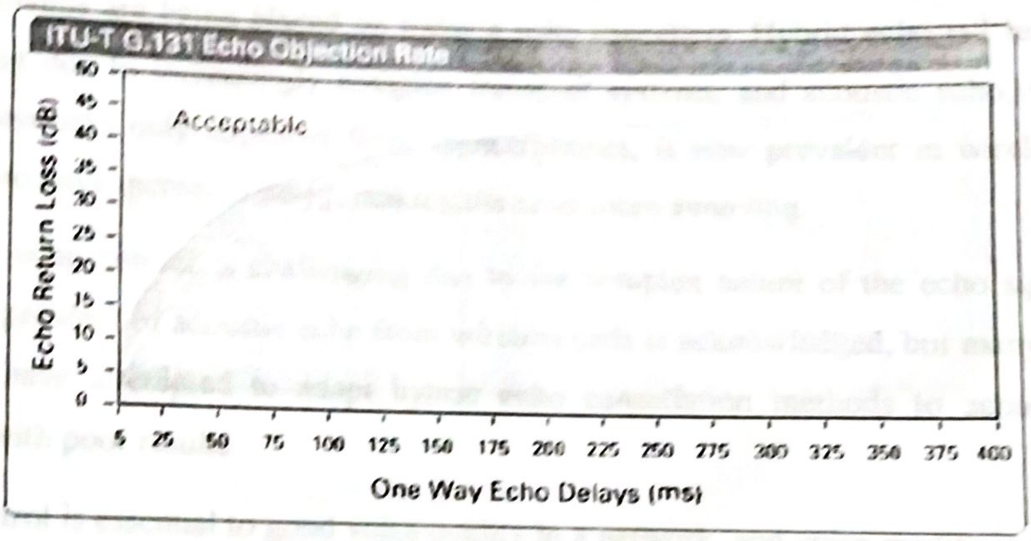


Figure 2.6 Echo Objection Rate as a Function of ERL and Delay

The areas "Acceptable," "Limiting Case," and "Unacceptable" shown in Figure 2.5 correspond to the probability of encountering objectionable echo as perceived by listeners.

- **Acceptable** – Echoes with delay and ERL in the Acceptable section have less than a 1% probability of being objectionable to the listener.
- **Limiting Case** – The ITU-T advises that the Limiting Case threshold should only be allowed in exceptional circumstances and should be avoided. An echo with delay and ERL in the Limiting Case section has a probability of up to 10% of being objectionable to the listener.
- **Unacceptable** – An echo with delay and ERL in the Unacceptable section is objectionable to the listener, and should be cancelled or prevented.

CHAPTER 3

ACOUSTIC ECHO CANCELLETION

The acoustic echo problem is introduced and its traditional solution, the acoustic echo canceller (AEC) is studied. Two basic adaptive algorithms are reviewed. The chapter concludes with an analysis of the computational complexity of AEC.

3.1 Echo Problem in Today's Networks

New demands are being placed on today's echo cancellers. Hybrid echo tail lengths are increasing due to increasingly complex transport systems, and acoustic echo, which on PSTN networks only appeared from speakerphones, is now prevalent in wireless calls, which also have increased delay, making the echo more annoying.

Acoustic echo control is challenging due to the complex nature of the echo signal. The growing problem of acoustic echo from wireless calls is acknowledged, but many solution vendors have attempted to adapt hybrid echo cancellation methods to acoustic echo control, with poor results.

Echo control is essential to good voice quality in a network, and voice quality is becoming increasingly important as wireless competition increases and voice impairment issues remain a barrier to VoIP migration. [5]

3.2 The Process of Acoustic Echo Cancellation

An echo canceller consists of three main functional components, that combine to form an echo canceller are:

1. Adaptive Filter
2. Doubletalk Detector
3. Nonlinear Processor

3.2.1 Adaptive Filtering

As previously demonstrated, the best solution for reducing the echo is to use some form of adaptive algorithm. Basically filtering is a signal processing technique whose objective is to process a signal in order to manipulate the information contained in the signal. In other words, a filter is a device that maps its input signal into another output signal by extracting only the desired information contained in the input signal. An adaptive filter is necessary when either the fixed specifications are unknown or time-invariant filters cannot satisfy the specifications. Strictly speaking an adaptive filter is a nonlinear filter since its characteristics are dependent on the input signal and consequently the homogeneity and additivity conditions are not satisfied. Additionally, adaptive filters are time varying since their parameters are continually changing in order to meet a performance requirement. In a sense, an adaptive filter is a filter that performs the approximation step on line.

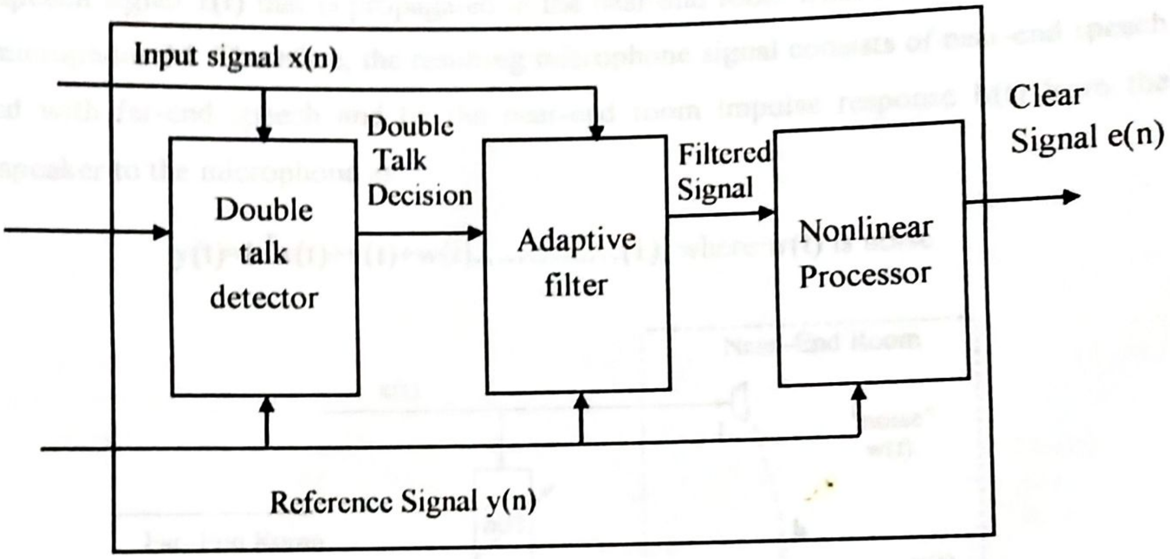


Figure 3.1 Block Diagram of a Generic Echo Canceller

3.2.2 Doubletalk Detector

A doubletalk detector is used with an echo canceller to sense when far-end speech is corrupted by near-end speech. The role of this important function is to freeze adaptation of the model filter when near-end speech is present. This action prevents divergence of the adaptive algorithm [2].

3.2.3 Nonlinear Processor

The non-linear processor evaluates the residual echo, which is nothing but the amount of echo left over after the signal has passed through the adaptive filter. The nonlinear processor removes all signals below a certain threshold and replaces them with simulated background noise which sounds like the original background noise without the echo.

3.3 The Acoustic Echo Cancellation

Acoustic Echo Cancellers are needed for removing the acoustic echoes resulting from the acoustic coupling between the loudspeaker(s) and the microphone(s) in communication systems. In Fig. 3.2, a typical setup for AEC is shown. The main purpose of the setup is that the near-end speech signal **v(t)** is to be picked up by the microphone **M** and propagated to the far-end room while far-end speech is to be emitted by the loudspeaker **L** into the near-end room. But the microphone signal **y(t)** is corrupted by the echo of the far-end speech signal **x(t)** that is propagated in the near-end room from the loudspeaker **L** to the microphone **M**. Therefore, the resulting microphone signal consists of near-end speech mixed with far-end speech and by the near-end room impulse response **h(t)** from the loudspeaker to the microphone as:

$$y(t)=h^{T}x(t)+v(t)+w(t)\ldots\ldots\ldots\ldots(1),\ \text{where } w(t) \text{ is noise}$$

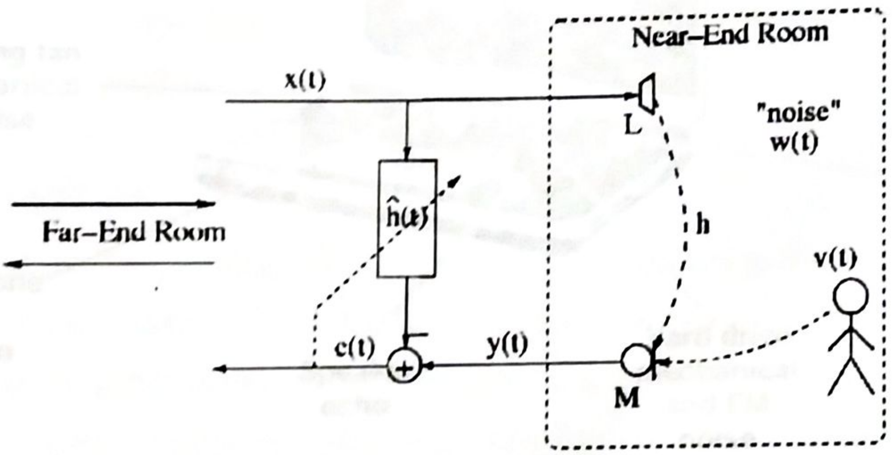


Figure 3.2 Typical AEC setup

The room impulse response is varying with time since movements (e.g., people moving around) may occur in the room. Thus, usually in order to remove the undesired echo an frequency domain adaptive filter estimate **h(t)** of **h** of is used to predict the far-end speech contribution $h^{T}x(t)$ and subtract it from the microphone signal **y(t)**. Thereby, we get the error free signal

$$e(t)= y(t)- h^{T}(t)x(t)= v(t)+ h^{T}x(t)- h^{T}(t)x(t)+w(t)\ldots\ldots\ldots\ldots(2)$$

that ideally should be equal to the near-end speech signal **v(t)**. Note that in (2), for simplicity, we have assumed that **h(t)** and **h** are of the same length. If that is not the case, then (2) has to be modified accordingly [11].

3.4 Some other Noise issues

During hands-free communication for PC applications, a lot of noise may exist and disturb the speech going to the microphone. The noise problem is especially worth concern in the situation of using the internal microphone of a laptop. The amplification of the internal microphone equipped in the laptop is usually high to be able to pick up the near-end speech. Hence, a variety of noises such as the hard-disk, fan of the laptop, the typing on the keyboard, mouse clicking as well as various ambient noise are likely to be picked up The possible noise sources when using internal microphone of the laptop is illustrated in Figure 3.2.

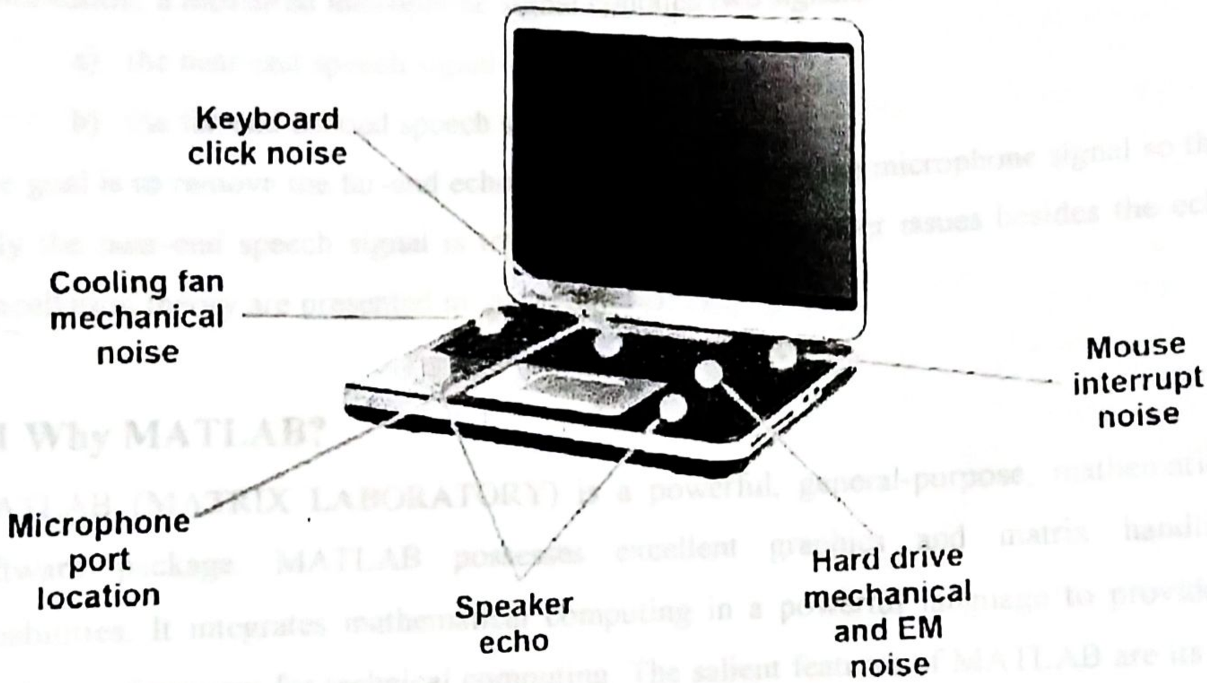


Figure 3.3 Common PC/Laptop Noise Sources

The hard drives and cooling fan are close the microphone, so as to together with other mechanical sounds, be transmitted to the internal mike through vibrations. The clicking sounds of the keyboard and the mouse are also major noise sources in this case. Since the keyboard is usually close the position of the microphone, the typing noise can be quite loud which makes it the most annoying noise source in this case. The situation can be improved much more when a good-quality external microphone is adopted [2].

MATLAB SIMULATION AND RESULTS

Most AEC products are based on the adaptive LMS or NLMS digital filter, which is a well-defined algorithm that has been used for years. To achieve larger echo attenuation without the help from other devices as Nonlinear Processor, the Acoustic Echo Suppressor based on a Frequency Domain Adaptive filter (FDAF) is a good option.

Acoustic echo cancellation is important for audio teleconferencing when simultaneous communication (or full-duplex transmission) of speech is necessary. In acoustic echo cancellation, a measured microphone signal contains two signals:

a) the near-end speech signal

b) the far-end echoed speech signal

The goal is to remove the far-end echoed speech signal from the microphone signal so that only the near-end speech signal is transmitted. Also some other issues besides the echo cancellation theory are presented in this chapter.

4.1 Why MATLAB?

MATLAB (**MATRIX LAB**ORATORY) is a powerful, general-purpose, mathematical software package. MATLAB possesses excellent graphics and matrix handling capabilities. It integrates mathematical computing in a powerful language to provide a flexible environment for technical computing. The salient features of MATLAB are its in-built mathematical toolboxes and graphic functions. Additionally, external routines that are written in other languages such as C, C++, Fortran and Java, can be integrated with MATLAB applications. MATLAB also supports importing data from files and other external devices. Most of the functions in MATLAB are matrix-oriented and can act on arrays of any appropriate dimension [4]. MATLAB also has a separate toolbox for signal processing applications, which provided simpler solutions for many of the problems encountered in this project.

The MATLAB software environment suited the needs of this project for the following reasons:

- The input voice signals (far-end and near-end talker signals) were recorded and stored as *.wav* files and the *.wav* files were easily imported into the code.

- The intermediate signals (echo signals) and output signals (error signal and signals obtained after echo cancellation) could be literally be heard, which aided immensely judgments with respect to the results obtained.
- The signal processing toolbox has in-built functions for almost all signal processing applications.
- Since MATLAB supports graphics, the results of a simulation could be presented in a graphical format with ease.

4.2 Simulation Flowchart

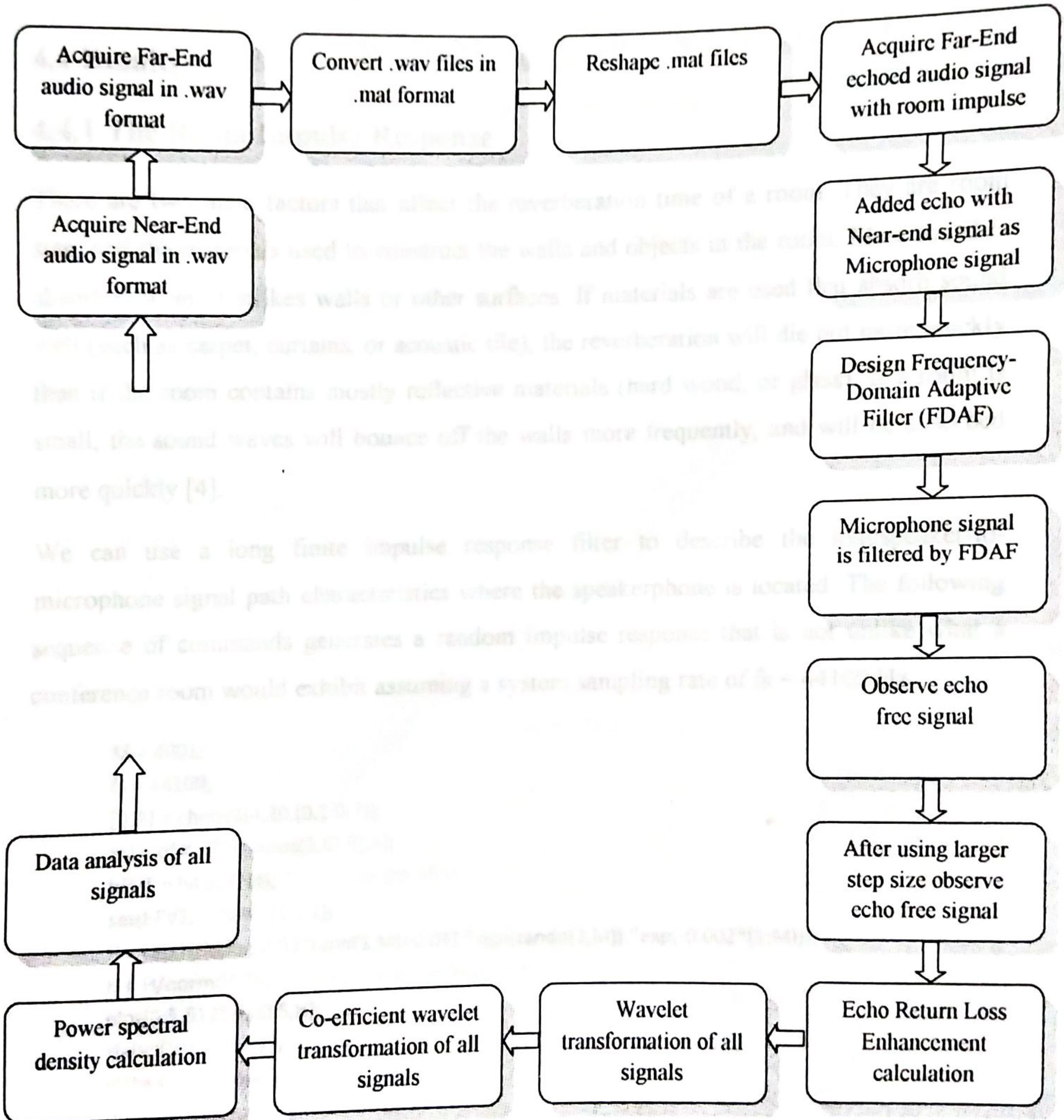


Figure 4.1 Total work done process

4.3 Description of the Simulation Process

The simulation process can be described as:

- The input signals, both far-end and near-end signals, were simulated and given to the AEC, which executed on a PC with the MATLAB environment.
- The input signals 30 seconds in duration.
- A sampling rate of 44100 Hz was used for all the signals in the simulation.
- The graphs plotted have x-axes denoting the time and y-axes denoting the amplitude or magnitude of the signal.

4.4 Results

4.4.1 The Room Impulse Response

There are two main factors that affect the reverberation time of a room. They are room size, and the materials used to construct the walls and objects in the room. Most sound is absorbed when it strikes walls or other surfaces. If materials are used that absorb sound well (such as carpet, curtains, or acoustic tile), the reverberation will die out more quickly than if the room contains mostly reflective materials (hard wood, or glass). If a room is small, the sound waves will bounce off the walls more frequently, and will be absorbed more quickly [4].

We can use a long finite impulse response filter to describe the loudspeaker-to-microphone signal path characteristics where the speakerphone is located. The following sequence of commands generates a random impulse response that is not unlike what a conference room would exhibit assuming a system sampling rate of fs = 44100 Hz

```
M = 4001;
fs = 44100;
[B,A] = cheby2(4,20,[0.1 0.7]);
Hd = dfilt.df2t([zeros(1,6) B],A);
hFVT = fvtool(Hd); % Analyze the filter
set(hFVT, 'Color', [1 1 1])
H = filter(Hd,log(0.99*rand(1,M)+0.01).*sign(randn(1,M)).*exp(-0.002*(1:M)));
H = H/norm(H)*4; % Room Impulse Response
plot(0:5.5125/fs:0.5,H);
xlabel('Time [sec]');
ylabel('Amplitude');
title('Room Impulse Response');
set(gcf, 'Color', [1 1 1])
```

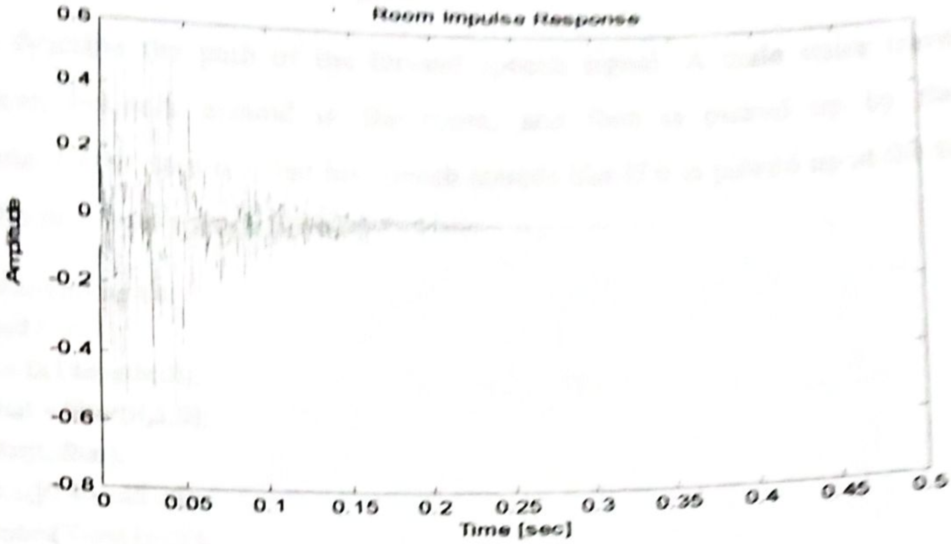


Figure 4.2 Room Impulse Response

4.4.2 The Near-End Speech Signal

The teleconferencing system's user is typically located near the system's microphone. Here is what a male speech sounds like at the microphone.

```
%Near-End signal
load E.mat
n = 1:length(C);
t = n/fs;
plot(t,C);
axis([0 30 -.2 .25]);
xlabel('Time [sec]');
ylabel('Amplitude');
title('Near-End Speech Signal');
set(gcf, 'Color', [1 1 1])
p8 = audioplayer(C,fs);
playblocking(p8);
```

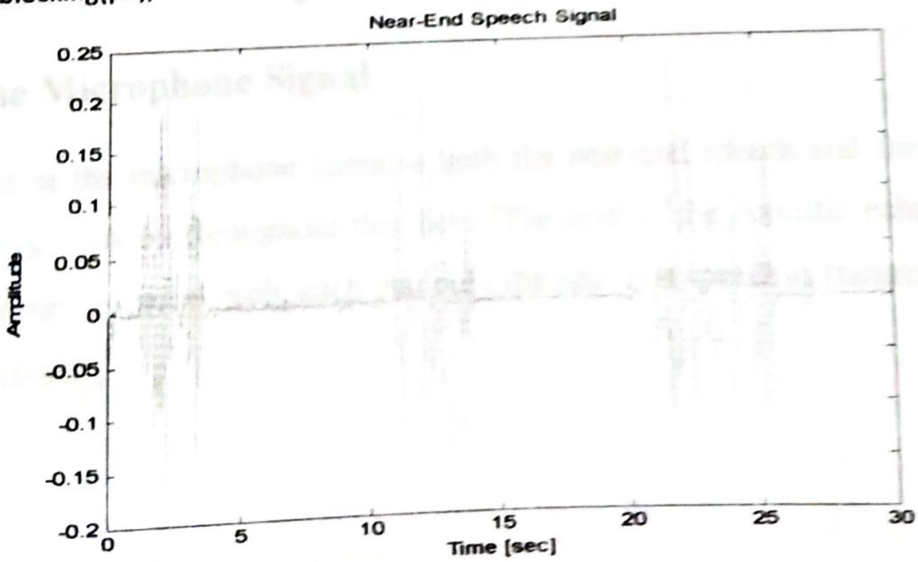


Figure 4.3 The Near-End Speech Signal

4.4.3 The Far-End Speech Signal

Now we describe the path of the far-end speech signal. A male voice travels out the loudspeaker, bounces around in the room, and then is picked up by the system's microphone. Let's listen to what his speech sounds like if it is picked up at the microphone without the near-end speech present.

```
%Far-End signal
load F.mat
D = D(1:length(D));
dhat = filter(H,1,D);
plot(t,dhat);
axis([0 30 -.35 .3]);
xlabel('Time [sec]');
ylabel('Amplitude');
title('Far-End Echoed Speech Signal');
set(gcf, 'Color', [1 1 1])
p8 = audioplayer(dhat,fs);
playblocking(p8);
```

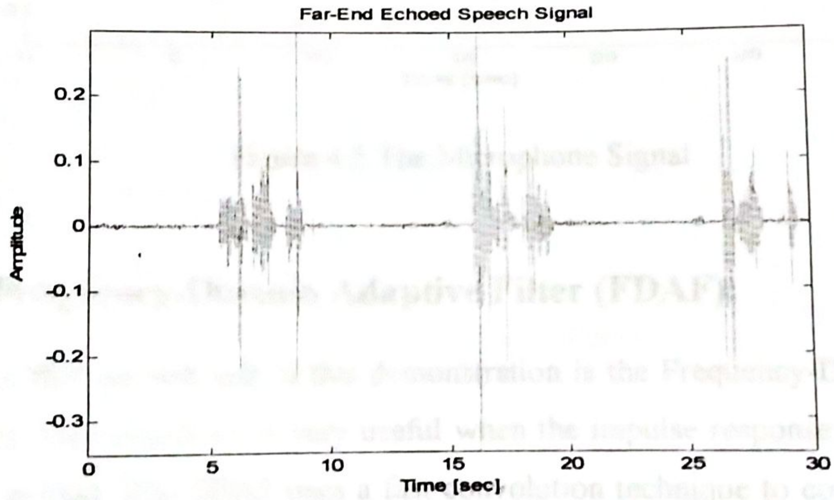


Figure 4.4 The Far-End Speech Signal

4.4.4 The Microphone Signal

The signal at the microphone contains both the near-end speech and the far-end speech that has been echoed throughout the room. (The goal of the acoustic echo canceller is to cancel out the far-end speech, such that only the near-end speech is transmitted back to the far-end listener.)

```
%Microphone signal (Near+far)
d = dhat + C+0.001*randn(length(C),1);
plot(t,d);
axis([0 30 -.35 .3]);
xlabel('Time [sec]');
ylabel('Amplitude');
title('Microphone Signal');
set(gcf, 'Color', [1 1 1])
p8 = audioplayer(d,fs);
playblocking(p8);
```

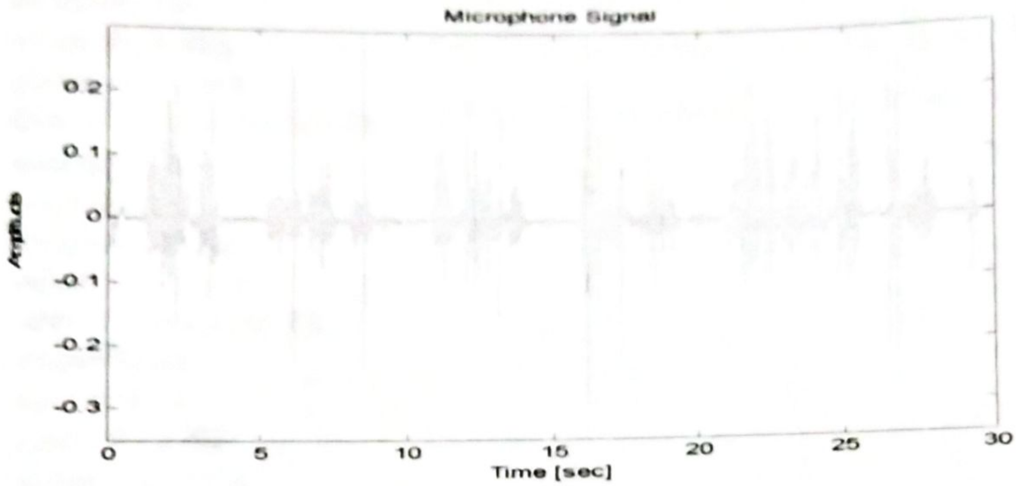


Figure 4.5 The Microphone Signal

4.4.5 The Frequency-Domain Adaptive Filter (FDAF):

The algorithm that we will use in this demonstration is the Frequency-Domain Adaptive Filter (FDAF). This algorithm is very useful when the impulse response of the system to be identified is long. The FDAF uses a fast convolution technique to compute the output signal and filter updates. This computation executes quickly in MATLAB®. It also has improved convergence performance through frequency-bin step size normalization. We'll pick some initial parameters for the filter and see how well the far-end speech is cancelled in the error signal.

```
mu = 0.025;
W0 = zeros(1,2048);
del = 0.01;
lam = 0.98;
D = D(1:length(W0)*floor(length(D)/length(W0)));
d = d(1:length(W0)*floor(length(d)/length(W0)));
% Construct the Frequency-Domain Adaptive Filter
hFDAF = adaptfilt.fdaf(2048,mu,1,del,lam);
[y,e] = filter(hFDAF,D,d);
```

```
n = 1:length(e);
t = n/fs;
pos = get(gcf,'Position');
set(gcf,'Position',[pos(1), pos(2)-100,pos(3),(pos(4)+85)])
subplot(3,1,1);
plot(t,C(n),'g');
axis([0 30 -.2 .25]);
ylabel('Amplitude');
title('Near-End Speech Signal');
subplot(3,1,2);
plot(t,d(n),'b');
axis([0 30 -.35 .3]);
ylabel('Amplitude');
title('Microphone Signal');
subplot(3,1,3);
plot(t,e(n),'r');
axis([0 30 -.2 .3]);
xlabel('Time [sec]');
ylabel('Amplitude');
title('Output of Acoustic Echo Canceller');
set(gcf, 'Color', [1 1 1])
p8 = audioplayer(e/max(abs(e)),fs);
playblocking(p8);
```

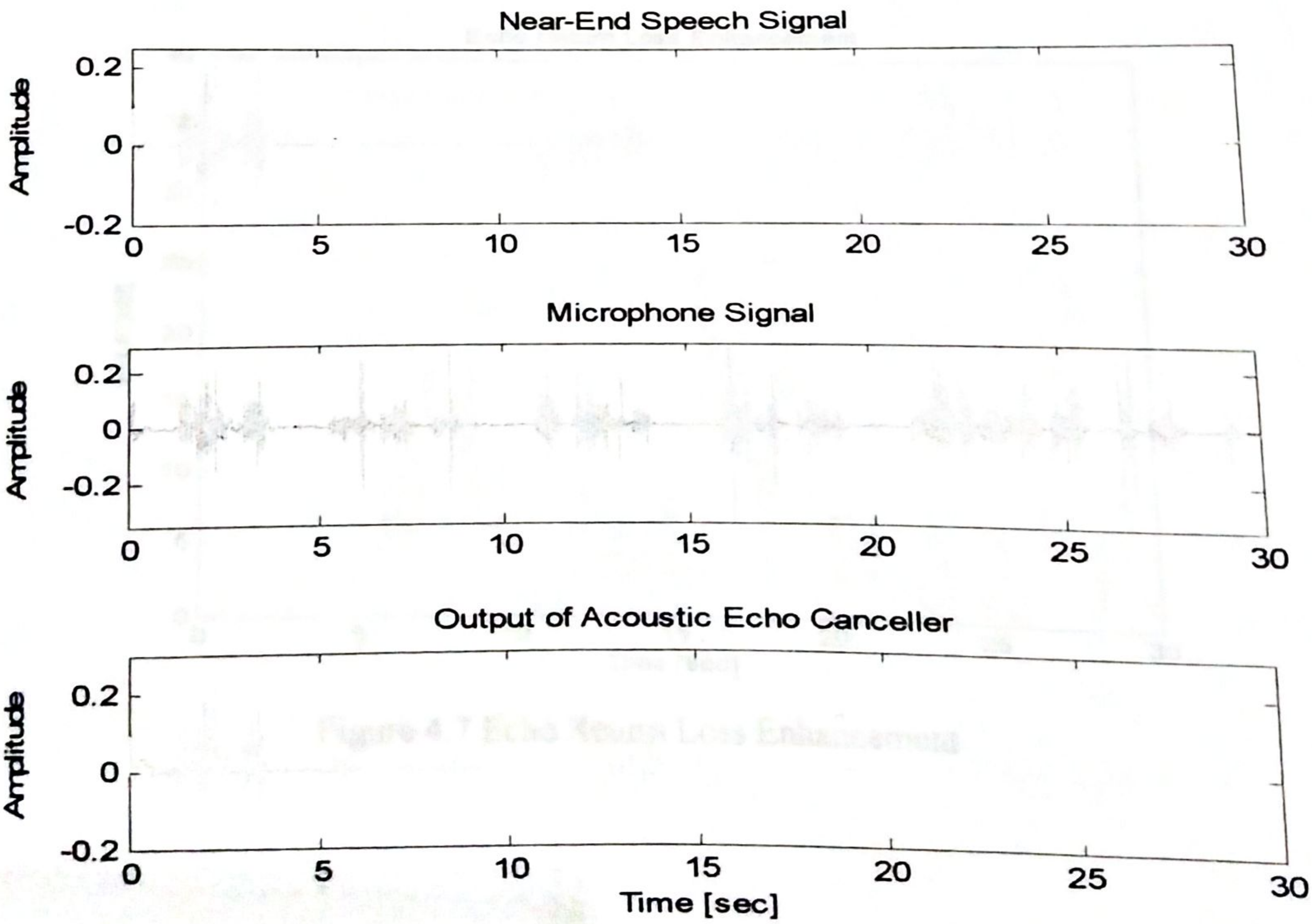


Figure 4.6 Frequency-Domain Adaptive Filter Output

4.4.6 Echo Return Loss Enhancement (ERLE)

Echo Return Loss Enhancement (ERLE) is the most important measure of how much in dB the echo is suppressed by the acoustic echo cancellation. It is defined as the power of the original echo over the power of the residual echo signal after cancellation in dB unit:

ERLE=log10(power of the echoed microphone signal)/(power of the residual signal)

A precise measure of ERLE should be performed in the portion where there is no near-end signal but only the echo. The higher the ERLE is, the better the AEC works. Since we have access to both the near-end and far-end speech signals, we can compute the echo return loss enhancement (ERLE), which is a smoothed measure of the amount (in dB) that the echo has been attenuated. From the plot, we see that we have achieved about a **30**dB ERLE.

```
Hd2 = dfilt.dffir(ones(1,1000));
erle = filter(Hd2,(e-C(1:length(e))).^2)./ ...
   (filter(Hd2,dhat(1:length(e)).^2));
erledB = -10*log10(erle);
plot(t,erledB);
axis([0 30 0 40]);
xlabel('Time [sec]');
ylabel('ERLE [dB]');
title('Echo Return Loss Enhancement');
set(gcf, 'Color', [1 1 1])
```

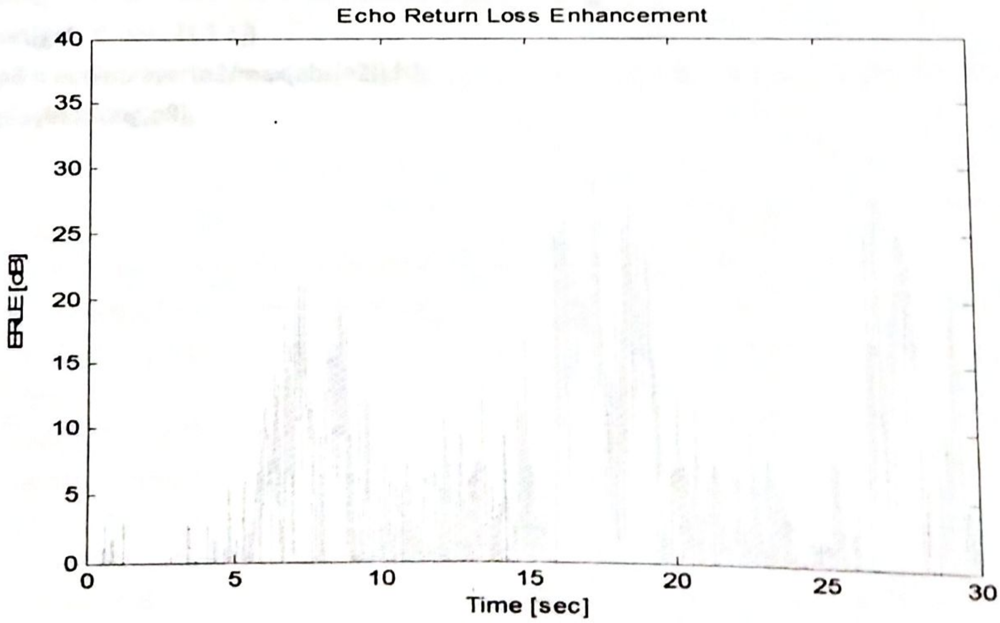


Figure 4.7 Echo Return Loss Enhancement

4.4.7 Effects of Different Step Size Values

To get faster convergence, we can try using a larger step size value. However, this increase causes another effect, that is, the adaptive filter is "mis-adjusted" while the near-end speaker is talking. Listen to what happens when we choose a step size that is 60\% larger than before

```
newmu = 0.04;
set(hFDAF,'StepSize',newmu);
[y,e2] = filter(hFDAF,D,d);
pos = get(gcf,'Position');
set(gcf,'Position',[pos(1), pos(2)-100,pos(3),(pos(4)+85)])
subplot(3,1,1);
plot(t,C(n),'g');
axis([0 30 -.2 .25]);
ylabel('Amplitude');
title('Near-End Speech Signal');
subplot(3,1,2);
plot(t,e(n),'r');
axis([0 30 -.2 .3]);
ylabel('Amplitude');
title('Output of Acoustic Echo Canceller, \mu = 0.025');
subplot(3,1,3);
plot(t,e2(n),'r');
axis([0 30 -.2 .3]);
xlabel('Time [sec]');
ylabel('Amplitude');
title('Output of Acoustic Echo Canceller, \mu = 0.04');
set(gcf, 'Color', [1 1 1])
p8 = audioplayer(e2/max(abs(e2)),fs);
playblocking(p8);
```

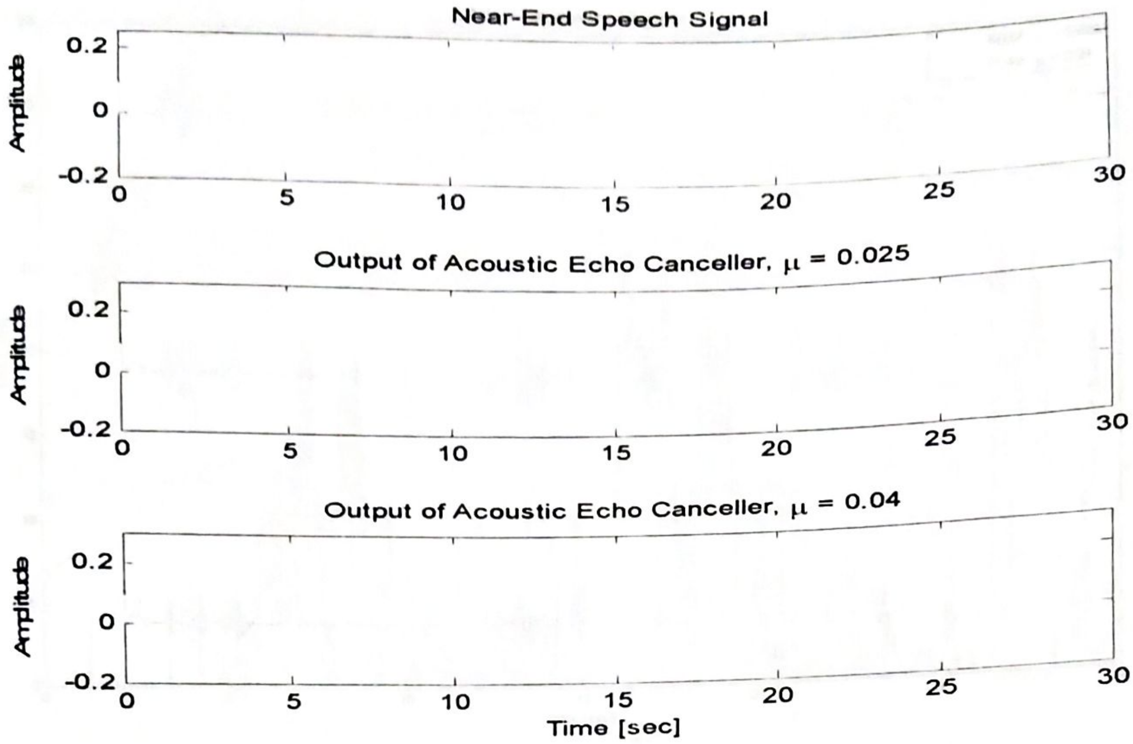


Figure 4.8 output signal effects for Different Step Size Values

4.4.8 Echo Return Loss Enhancement Comparison

With a larger step size, the ERLE performance is not as good due to the misadjustment introduced by the near-end speech. To deal with this performance difficulty, acoustic echo cancellers include a detection scheme to tell when near-end speech is present and lower the step size value over these periods. Without such detection schemes, the performance of the system with the larger step size is not as good as the former, as can be seen from the ERLE plots.

```
close;
erle2 = filter(Hd2,(e2-C(1:length(e2))).^2)./...
   (filter(Hd2,dhat(1:length(e2)).^2));
erle2dB = -10*log10(erle2);
plot(t,[erledB erle2dB]);
axis([0 30 0 40]);
xlabel('Time [sec]');
ylabel('ERLE [dB]');
title('Echo Return Loss Enhancements');
legend('FDAF, \mu = 0.025','FDAF, \mu = 0.04');
set(gcf, 'Color', [1 1 1])
```

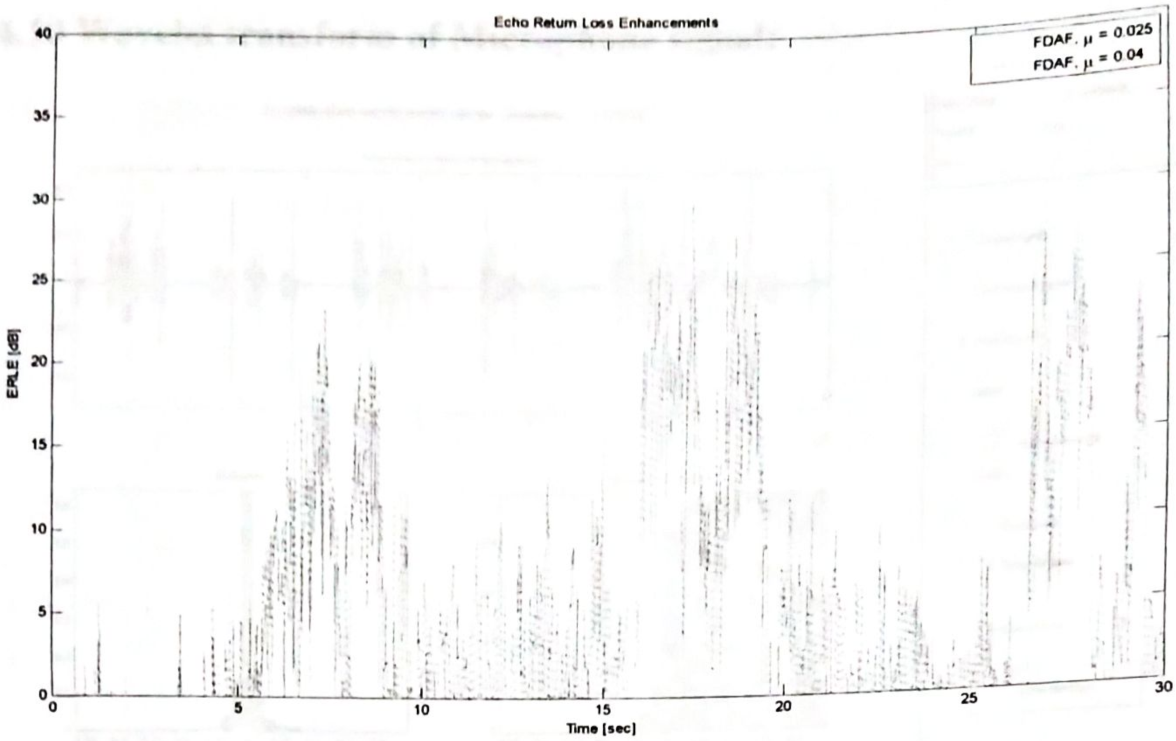


Figure 4.9 Echo Return Loss Enhancement Comparison

4.4.9 Wavelet transform of Far-End signal

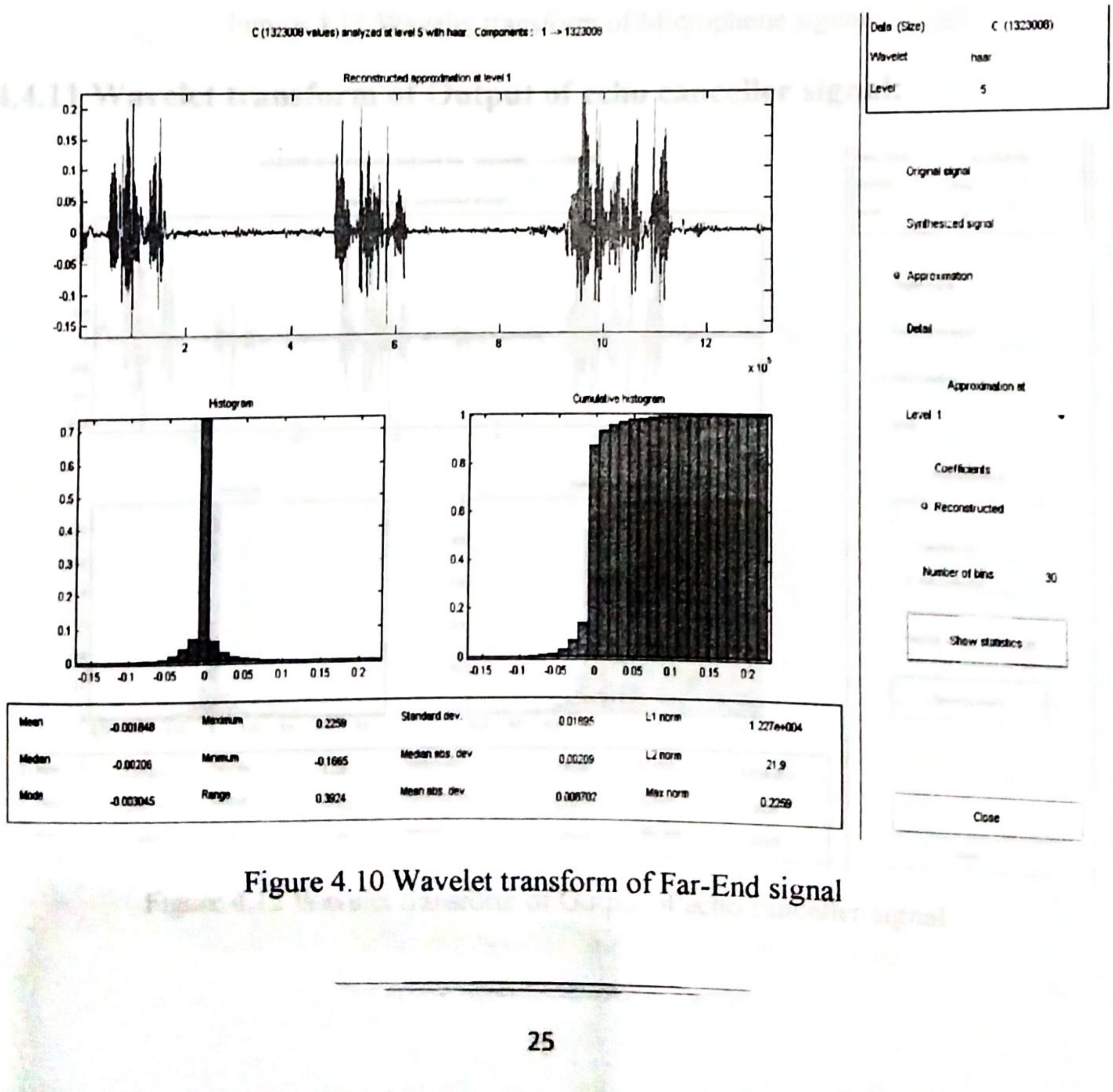


Figure 4.10 Wavelet transform of Far-End signal

4.4.10 Wavelet transform of Microphone signal:

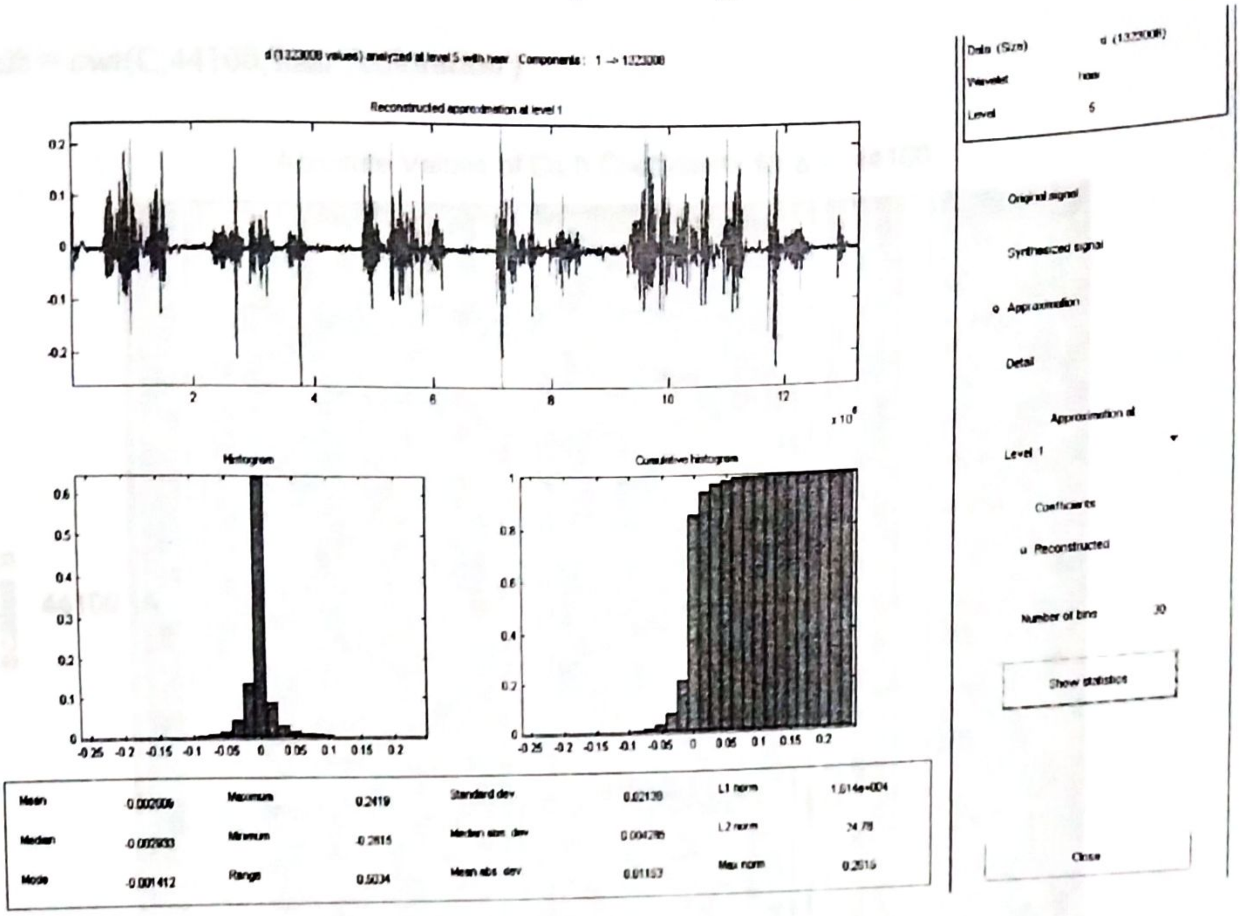


Figure 4.11 Wavelet transform of Microphone signal

4.4.11 Wavelet transform of Output of echo canceller signal:

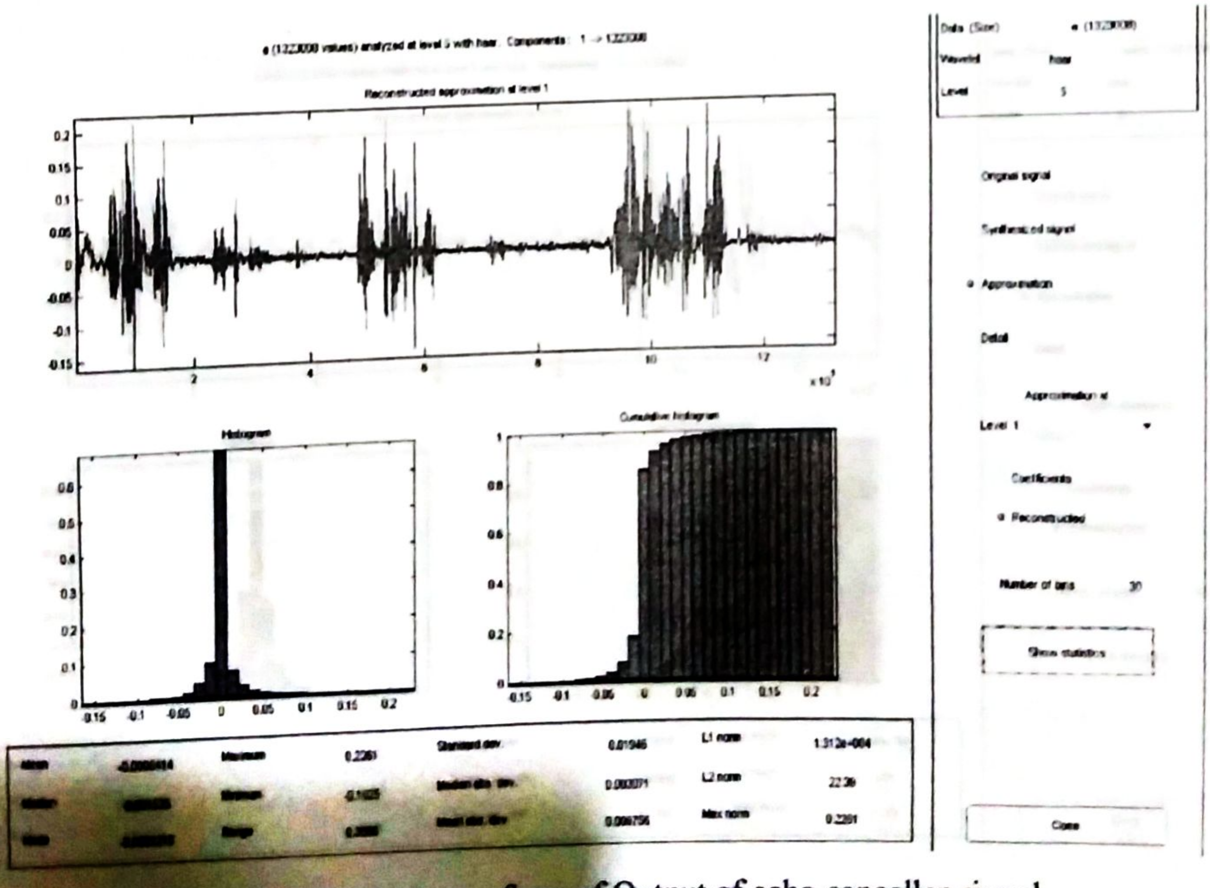


Figure 4.12 Wavelet transform of Output of echo canceller signal

4.4.12 Coefficient wavelet transform of Far-End signal

coefs = cwt(C,44100,'haar','coloration')

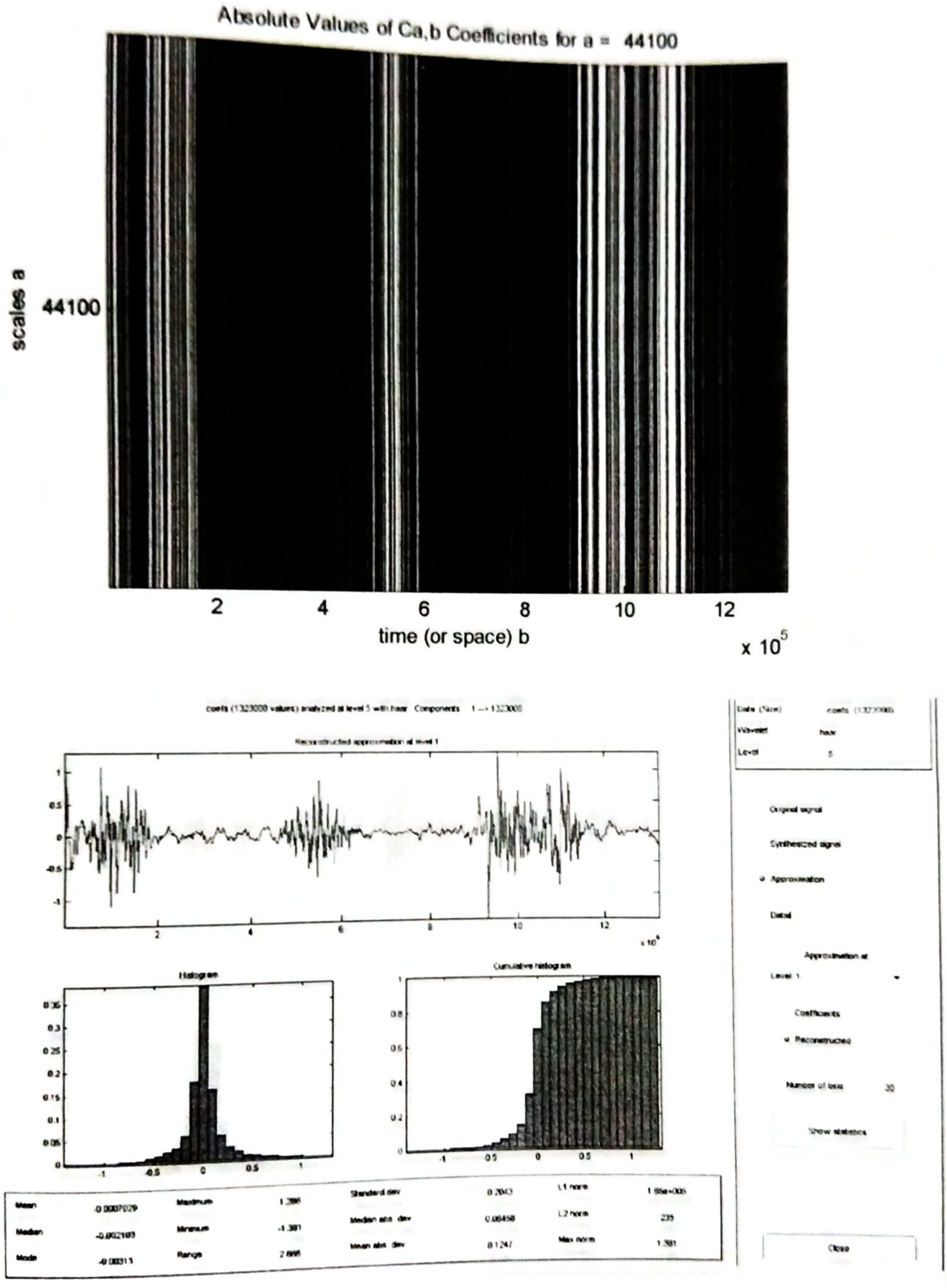


Figure 4.13 Coefficient wavelet transform of Far-End signal

4.4.13 Coefficient wavelet transform of Microphone signal

dcoefs= cwt(d,44100,'haar','coloration')

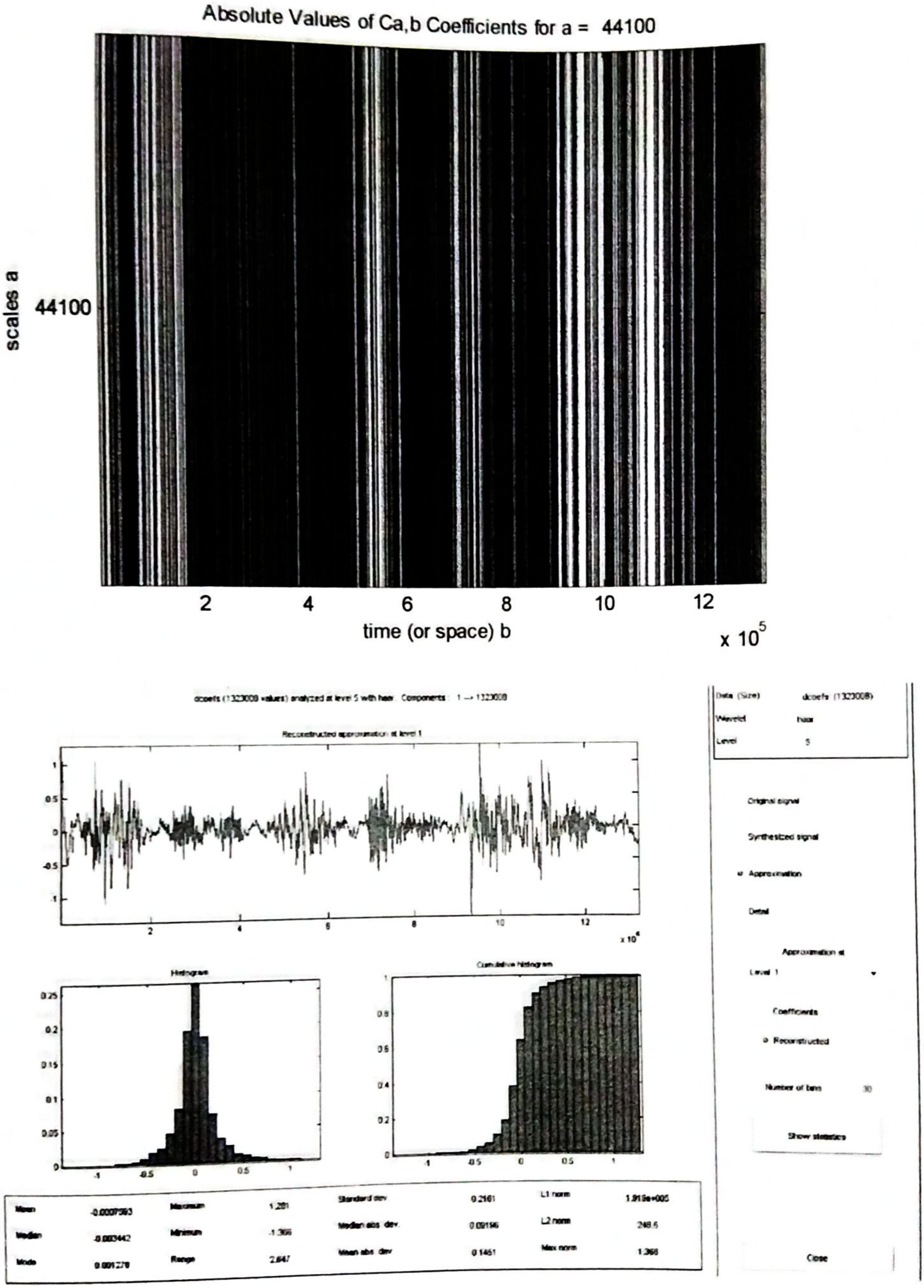


Figure 4.14 Coefficient wavelet transform of Microphone signal

4.4.14 Coefficient wavelet transform of Output of acoustic Echo canceller

ecoefs= cwt(e,44100,'haar','coloration')

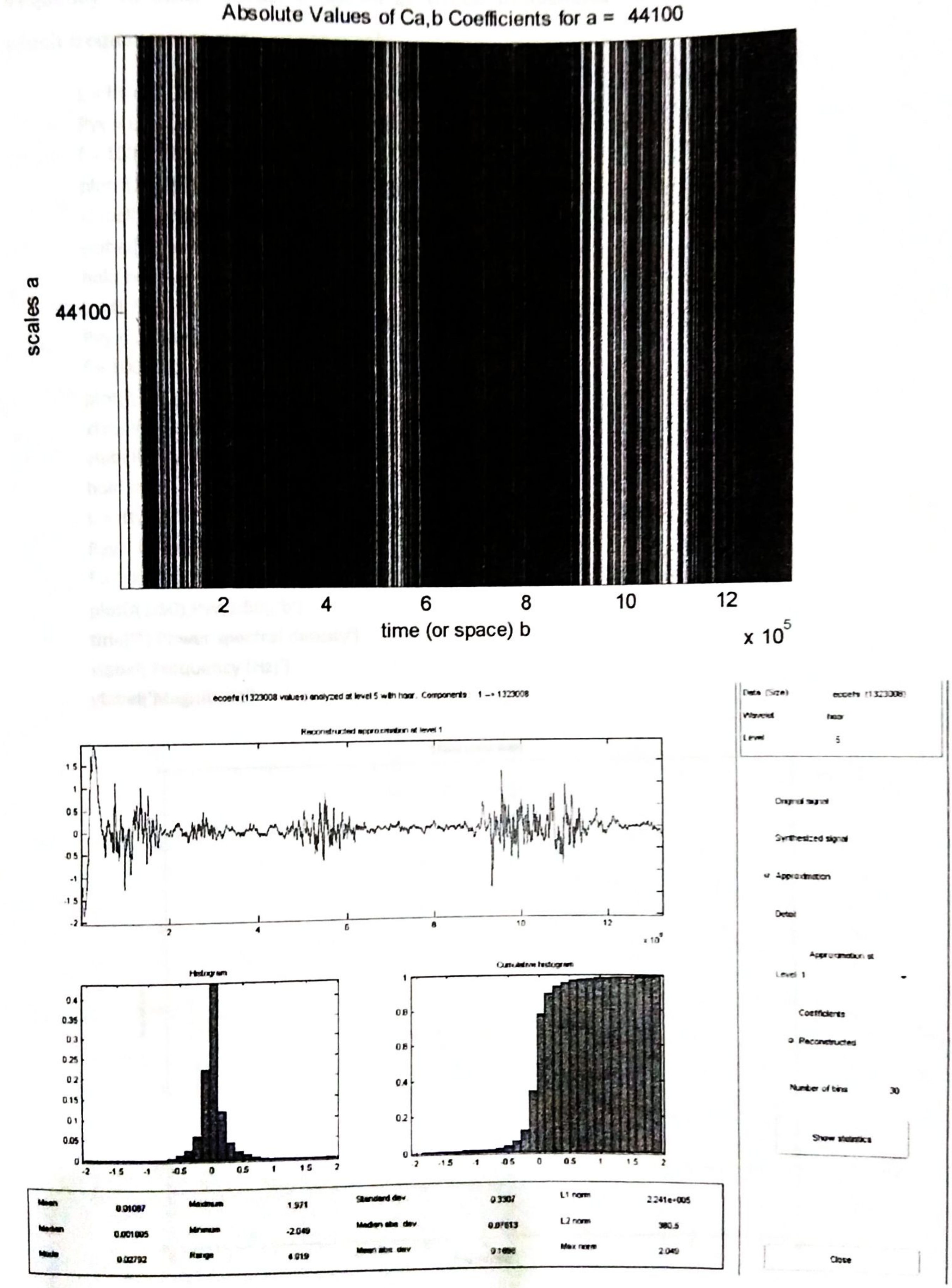


Figure 4.15 Coefficient wavelet transform of Output of acoustic Echo canceller

4.4.15 Power spectral density calculation

Power spectral density shows the strength of the variations (energy) as a function of frequency. In other words, it shows at which frequencies variations are strong and at which frequencies variations are weak.

```
L = fft(C,8192);
Pyy = L.* conj(L) / 8192;
f = 1000*(0:4096)/8192;
plot(f(1:50),Pyy(1:50),'r')
xlabel('Frequency (Hz)')
ylabel('Magnitude')
hold on;
L = fft(d,8192);
Pyy = L.* conj(L) / 8192;
f = 1000*(0:4096)/8192;
plot(f(1:50),Pyy(1:50),'g')
xlabel('Frequency (Hz)')
ylabel('Magnitude')
hold on
L = fft(e,8192);
Pyy = L.* conj(L) / 8192;
f = 1000*(0:4096)/8192;
plot(f(1:50),Pyy(1:50),'b')
title('*) Power spectral density')
xlabel('Frequency (Hz)')
ylabel('Magnitude')
```

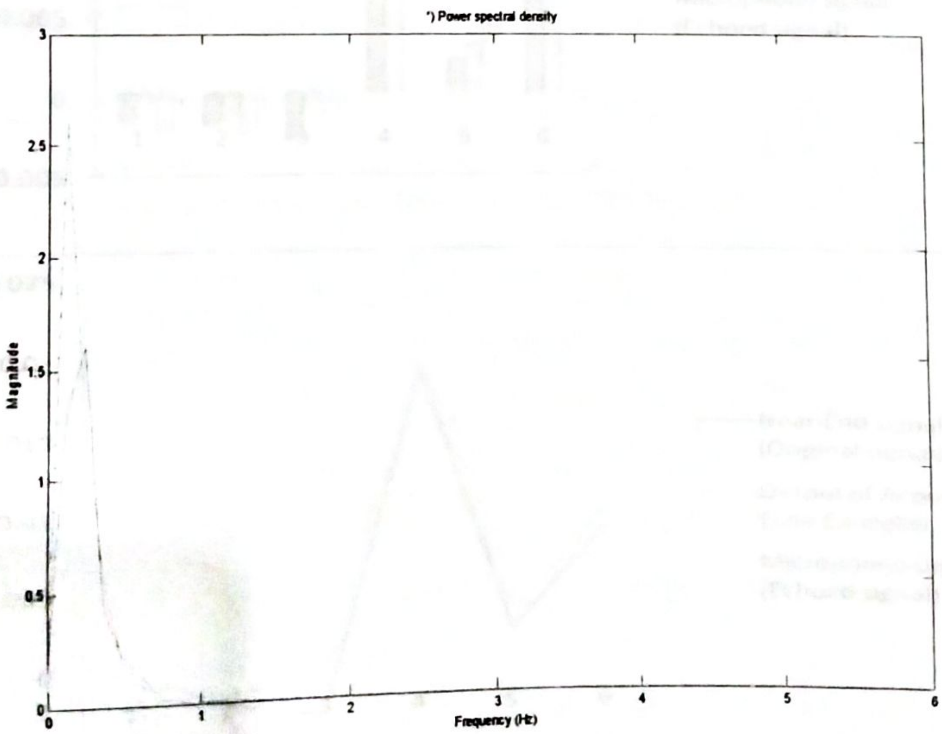
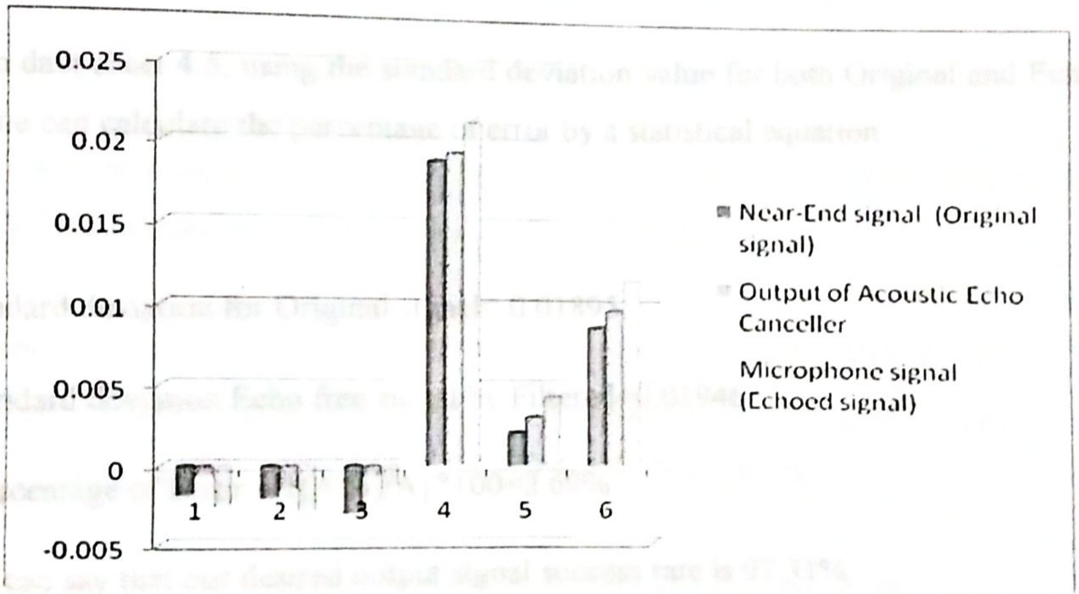


Figure 4.16 Power spectral density of three signals

4.5 Data Analysis of different signals

	Near-End signal (Original signal)	Microphone signal (Echoed signal)	Output of Acoustic Echo Canceller
Mean	-0.001848	-0.002609	-0.0006414
Median	-0.00206	-0.002933	-0.001535
Mode	-0.003045	-0.001412	-0.0006215
Maximum	0.2259	0.2419	0.2261
Minimum	-0.1665	-0.2615	-0.1625
Range	0.3924	0.5034	0.3886
Standard deviation	0.01895	0.02138	0.01946
Median abs. dev.	0.00209	0.004285	0.003071
Mean abs. dev.	0.008702	0.01153	0.009756



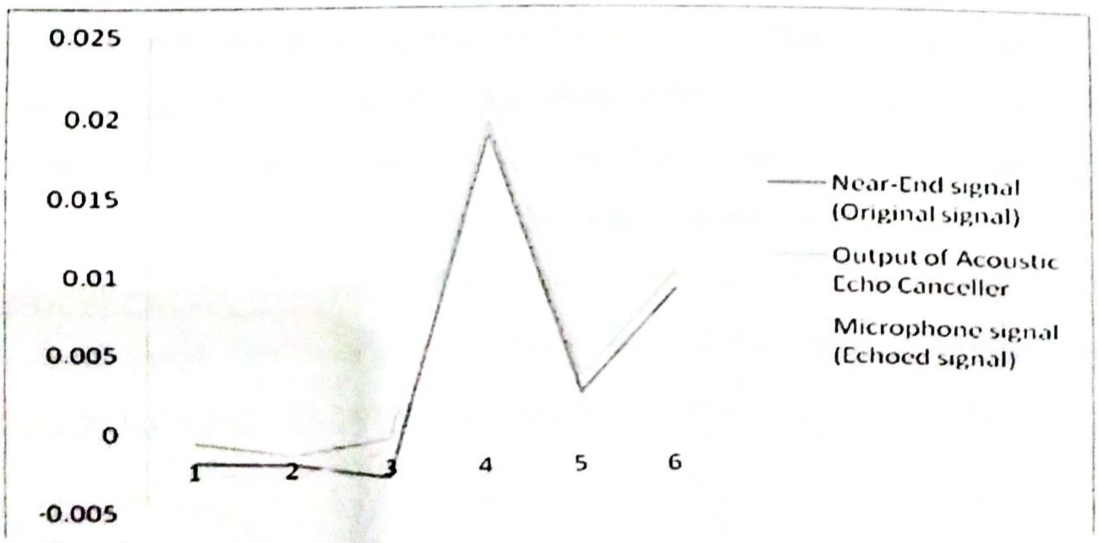


Figure 4.17 Comparing original signal, echoed signal and echo canceller by column & line graph

By the wavelet transform, we can analyze the entire echoed and echo free signal. After compressing the signal and from the statistics we can see that the output of acoustic echo canceller is similar as the original signal. Histogram is also used to analyze the speech signal.

We observe the difference between signal with echo and signal without echo, generally we can say that echo is removed by seeing the waveforms. Both the wavelet and coefficient wavelet transformation are used for better analysis. For further analyzing we extract values from our experiment and taking the values we plot a table and two graphs also, which will help us to prove that echo is removes from the signal. The mean, median, mode and standard deviation are analyzed more over the compressed signal is totally matched with the signal without echo, which is our goal.

4.6 Error Analysis

From to data sheet 4.5, using the standard deviation value for both Original and Echo free signal we can calculate the percentage of error by a statistical equation.

Let,

A=Standard deviation for Original signal= 0.01895

B= Standard deviation Echo free signal by Filtered=0.01946

The Percentage of Error = {(A-B)/A}*100=2.69%

So, we can say that our desired output signal success rate is 97.31%

With the world shrinking into a global village because of superior communications, telephones, both conventional and hands-free sets, occupy a prominent position in solving people's communication needs. One of the major problems in a telecommunication application over a telephone system is echo. The Echo cancellation algorithm presented in this project successfully attempted to find a software solution for the problem of echoes in the telecommunications environment. The proposed algorithm was completely a software approach without utilizing any DSP hardware components. The algorithm was capable of running in any PC with MATLAB software installed. This technique is faster and provides almost perfect results for canceling acoustic echoes without clipping of the reference speech signals. In addition, the results obtained were convincing. The audio of the output speech signals were satisfactory and validated the goals of this research.

Further Work

A number of additional studies would be interesting to perform on AEC. We conclude the thesis by considering some possibilities for further work:

Real-time simulations. In this study the simulations were run "off-line". To be able to fully simulate AEC, the system should be implemented to run in real-time. That would make it possible to test AEC under more realistic circumstances.

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