Speech Enhancement Techniques

Major Project Report

Submitted in partial fulfillment of the requirements

for the degree of

Master of Technology

in

Electronics And Communication Engineering

(Communication Engineering)

By

Patel Priyankaben kirtilal

(08MECC20)

Department of Electronics & Communication Engineering

Institute of Technology

Nirma University

Ahmedabad-382 481

May 2010

Speech Enhancement Techniques

Major Project Reprot

Submitted in partial fulfillment of the requirements

for the degree of

Master of Technology

in

Electronics And Communication Engineering

(Communication Engineering)

By

Patel Priyankaben Kirtilal

(08MECC20)

Under the Guidance of

Prof. A. S. Ranade

Department of Electronics & Communication Engineering

Institute of Technology

Nirma University

Ahmedabad-382 481

May 2010

Declaration

This is to certify that

i) The thesis comprises my original work towards the degree of Master of Technology in Communication Engineering at Nirma University and has not been submitted elsewhere for a degree.

ii) Due acknowledgement has been made in the text to all other material used.

Patel Priyanka Kirtilal

Certificate

This is to certify that the Major Project entitled **"Speech Enhancement Techniques"** submitted by **Patel Priyankaben Kirtilal (08MECC20)**, towards the partial fulfillment of the requirements for the degree of Master of Technology in Electronics & Communication (Communication) engineering of Nirma University, Ahmedabad is the record of work carried out by her under my supervision and guidance. In my opinion, the submitted work has reached a level required for being accepted for examination. The results embodied in this major project, to the best of our knowledge, haven't been submitted to any other university or institution for award of any degree or diploma.

Date: Place: Ahmedabad

HOD & Guide

(Prof. A. S. Ranade)
Professor, EC

Director

(Dr. K Kotecha)
Director, IT, NU

Acknowledgements

I would like to express my gratitude and sincere thanks to Prof. A. S. Ranade Head of Electrical Engineering Department and Dr. D. K. Kothari Coordinator M.Tech Communication Engineering program for allowing me to undertake this thesis work and for his guidelines during the review process.

I am deeply indebted to my thesis supervisor Prof. A. S. Ranade for his constant guidance and motivation. He has devoted significant amount of his valuable time to plan and discuss the thesis work. Without his experience and insights, it would have been very difficult to do quality work.

My special thanks to Prof. T. H. Zaveri who was the first person I used to approach whenever I got stuck. I wish to thank my friends of my class for their delightful company which kept me in good humor throughout the year.

Last, but not the least, no words are enough to acknowledge constant support and sacrifices of my family members because of whom I am able to complete the degree program successfully.

- Priyanka K. Patel

08MECC20

Abstract

Speech enhancement is concerned with improving some perceptual aspect of speech that has been degraded by additive noise. In most application, the aim of speech enhancement is to improve the quality and intelligibility of degraded speech. The NOIZEUS dataset is used to apply different speech enhancement algorithm. There are various speech enhancement methods available in literatures which are application specific. Speech enhancement algorithms are divided manly into three categories such as statistical model based, spectral subtractive and Subspace algorithm based methods.

The objective of the thesis is to identify limitations of these algorithms and compare results of standard speech enhancement techniques available in literature which includes wiener a priori SNR method, wavelet thresholding method using multitapper spectrum, log MMSE estimator method, spectral subtraction method and multiband spectral subtraction method for speech enhancement. It has been observed from the simulation results that log MMSE and multiband spectral subtraction algorithm outperforms than other compared algorithm. The comparative results of these algorithms in terms of different objective and subjective quality evaluation parameter is presented in this thesis. And from all these result, we see that log MMSE and multiband spectral subtraction algorithms perform best in almost all noise environments at different input SNR.

Contents

List of Tables

List of Figures

Chapter 1

Introduction and Problem Definition

1.1 Introduction

Speech enhancement is concerned with improving some perceptual aspect of speech that has been degraded by additive noise. In most applications, the aim of speech enhancement is to improve the quality and intelligibility of degraded speech. The improvement in quality is highly desirable as it can reduce listener fatigue. Speech enhancement algorithms reduce or suppress the background noise to some degree and are sometimes referred to as noise suppression algorithms as described in[1].

The need to enhance speech signals arises in many situations in which the speech signal originates from a noisy location or is affected by noisy communication channel. There are a wide variety of scenarios in which it is desired to enhance speech. Voice communication, for instance, over cellular telephone systems typically suffers from background noise present in the car, restaurant etc., at the transmitting end. Speech enhancement algorithms therefore be used to improve the quality of speech at the receiving end; that is, they can be used as a pre-processor in speech coding systems employed in cellular phone standards. If the cellular phone is equipped with

a speech reorganization system for voice dialling, then recognition accuracy will suffer in the presence of noise. In this case, the noisy speech signal can be pre-processed by a speech enhancement algorithm before being fed in to the speech recognizer. In an air ground communication scenario, speech enhancement techniques are needed to improve quality, and preferably intelligibility, of the pilot's speech that has been corrupted by extremely high levels of cockpit noise. In this, as well as in similar communication system used by military, it is more desirable to enhance the intelligibility rather than the quality of the speech. In a teleconferencing system, noise sources present in one location will be broadcast to all other locations. The situation is further worsened if the room is reverberant. Enhancing the noisy signal prior to broadcasting it will improve the performance of the teleconferencing system [1].

The forgoing examples illustrate that the goal of speech enhancement varies depending on the application at hand. Ideally, we would like speech enhancement algorithms to improve both quality and intelligibility. In practice, however, most speech enhancement algorithms improve only the quality of speech. It is possible to reduce the background noise, but at the expense of introducing speech distortion, which in turn may impair speech intelligibility. Hence, the main challenge in designing effective speech enhancement algorithms is to suppress the noise without introducing any perceptible distortion in the signal [1].

The solution to the general problem of speech enhancement depends largely on the application at hand, the characteristic of the noise source or interference, the relationship of the noise to the clean signal, and the number of microphones or sensors available [2]. The interference could be noise like (fan noise) or speech like, such as an environment (restaurant) with competing speakers. Acoustic noise can be additive to the clean signal, or convolutive, if it originates from highly reverberant room. Furthermore, the noise may be statistically correlated or uncorrelated with the clean speech signal. The number of microphones available can influence the performance of speech enhancement algorithms. Typically the larger number of microphones, the easier speech enhancement task becomes. Adaptive cancelation techniques can be

used when at least one microphone is placed near the noise source [1].

1.2 Problem Definition

In previous section, we have seen that speech enhancement is concerned with improving some perceptual aspect of speech that has been degraded by additive noise. The different kind of noise affect on the quality of the speech. To solve this problem by different speech enhancement techniques, which is used to improve the quality of speech and reduce the specific noise coming from different sources at different SNRs. *The aim of project is to improve the quality and intelligibility of degraded speech.for some application like hearing aid, cockpit communication, video conferencing, etc..*

1.3 Noise

It is crucial to understand the behavior various types of noise, the differences between the noise sources in terms of temporal and spectral characteristics, and the range of noise levels that may be encountered in real life.

Based on the nature and properties of the noise sources, noise can be classified in the following ways [2]:

1. Background noise: additive noise, which is usually uncorrelated with the signal and present in various environment scenarios like cars, offices, city streets, fans,machines, climatic conditions, factory environment, cockpits, helicopters etc.From these types of noise, Hoth noise (white noise filtered to model long-term average of room noise) is stationary, noises in streets and factories etc. have more dynamic characteristics. Factory and helicopter noise having strong periodic components and noise from fans, and car noise in a hands free environment etc. are real noise and are examples of non-stationary noise having time varying characteristics.

2. Interfering speakers (speech like noise): additive noise, composed of single or multiple competing speakers. The multi-talker babble which also attributes to the phenomenon called cocktail party effect(many voices talking simultaneously, e.g. in a cafeteria, a noisy classroom) is noise, which has characteristics and frequency range very similar to the speech signal of interest.

3. Impulse noise: slamming of doors, noise present in archived gramophone recordings.

4. Non-additive noise: noise due to non-linearities of microphones, speakers and channel distortion (speech on transmission lines).

5. Non-additive noise due to speaker stress: e.g. Lombard effect i.e. the effect induced in presence of noise, wherein the speaker has a tendency to increase his vocal effort . This results in speech having different spectral properties as compared to clean speech. Speech produced under situational and emotional stress also fall in this category.

6. Noise correlated with the signal: reverberations and echos.

7. Convolutive noise: corresponds to convolution in time domain. For instance, changes in speech signal due to changes in room acoustics or changes in microphones etc. These are usually harder to deal with, as compared to additive noise.

8. Multiplicative noise: signal distortion due to fading in cellular channels.

In general, it is more difficult to deal with non-stationary noise, where there is no priori knowledge available about the characteristics of noise. Since non-stationary noise is time varying, the conventional method of estimating the noise from initial

intervals assuming no speech signal is not suitable for estimation. Noise types, which are similar in temporal, frequency or spatial characteristics to speech, are also difficult to remove or attenuate. Multitalker babble, for instance, retains some characteristics of speech and poses a particularly difficult problem for an algorithm intended to isolate speech signal from the background noise.

1.4 Noise and Speech Levels in Various Environments

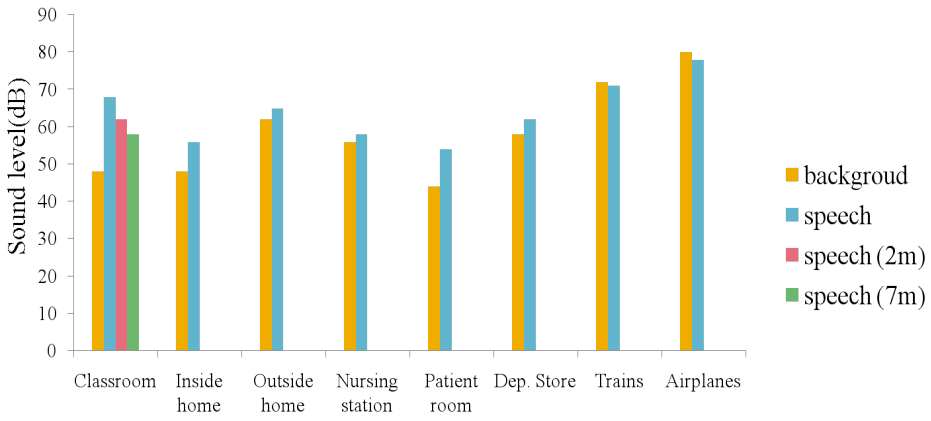

Figure 1.1: Average Noise and Speech Levels in Various Environments

A comprehensive analysis and measurement of speech and noise levels in real world environments was done by Pearson. Figure 1.1 summaries the average speech and noise levels in various environments. Noise levels are lowest in the classroom, hospital, inside house, and in department stores. In these environments, noise levels range between 50 to 55 dB SPL(Sound Pressure Level). The corresponding speech levels range between 60 to 70 dB SPL. This suggests that the effective SNR levels in

these environments range between 5 and 15 dB [2]. Noise levels are particularly high in trains and airplanes, averaging about 70-75 dB SPL. The corresponding speech levels are roughly the same, suggesting that effective SNR levels in these two environments are near 0dB. For the speech enhancement algorithm to be employed in a practical application, it needs to operate at SNRs in the range of 0-15 dB.

1.5 Thesis Organization

The rest of the thesis is organized as follows.

Chapter 2, *literature Survey* , This chapter introduces different speech enhancement algorithms like Spectral Subtractive Algorithms, Statistical Model Based Algorithms, Subspace Algorithm, Wiener-type Algorithm.

Chapter 3, *Statistical model based Algorithms*, describes statestical model based algorithm like wiener a-priori SNR algorithm, Wavelet thresholding algorithm and log MMSE(Minimum Mean Square Error) algorithm for speech enhancement.

chapter 4, *Spectral Subtractive Algorithms*, describes the spectral subtractive algorithms for speech enhancement. Two main algorithm for speech enhancement are described, one is spectral subtraction algorithm and another is Multi-Band spectral subtraction algorithm.

Chapter 5, different quality evaluation parameters which are used to compare these algorithms are described in *Quality Evaluation Parameter*,

Chapter 6, *Simulation Results and Comparision*, shows out come of different speech enhancement algorithm which indicates reduction in noise. Also shows spectrogram of clean, noisy and enhanced speech signals.Also describes comparison of different algorithm in terms of different parameter and represent graphical comparison in terms of overall performance of different algorithm.

Finally, in **chapter 7** conclusion and future scope is presented in this chapter.

Chapter 2

Literature Survey

The problem of improving performance of speech communication systems in noisy environments has been a challenging area for research for more then three decades now. Important applications of noise suppression and speech enhancement systems include improving the performance of [3]

1) digital mobile radio telephony systems, which suffer both from background noise in the environment as well as from channel noise

2) hands free telephone systems suffering from car noise etc.

3) pay phones in a noisy environment (e.g. restaurants, factories, airports)

4) air-ground communication systems in which pilots speech is corrupted by cockpit and engine noise

5) ground-air communication where the cockpit/engine noise corrupts the received signal

6) long distance communication over noisy radio channels

7) teleconferencing systems where a noise source from one location maybe broadcasted to all other locations and

8) hearing aids and cochlear implants in a noisy environment (e.g. classrooms, cafeteria etc).

Efforts to achieve higher quality and/or intelligibility of noisy speech may effectively end up improving performance of other speech applications such as speech

coding/compression and speech recognition etc. Speech enhancement has three major goals [4]:

1. to improve the quality and intelligibility of speech corrupted by background noise, reduce the perceptual fatigue.
2. to make speech coders robust when to input noise.
3. to make speech recognition systems more robust to input noise.

This chapter presents an overview of different speech enhancement methods.

2.1 Classes of Speech Enhancement Algorithms

A number of algorithms have been proposed in the literature of speech enhancement. These algorithms can be divided in to three main classes shown below as per [1]:

a. Spectral subtractive algorithms [5]

b. Statistical model based algorithms [6]

c. Subspace algorithms [1]

2.2 Spectral Subtractive Algorithms

Figure 4.1 shows simple spectral subtractive algorithms [1]. Where p is power exponent, with p = 1 yielding the original magnitude spectral subtraction and p = 2 yielding the power spectral subtraction algorithm. These are by far the simplest enhancement algorithms to implement. They are based on the basic principle that as noise is additive, one can estimate/update the noise spectrum [7] when speech is not present and subtract it from the noisy signal. Spectral subtractive algorithms were initially proposed by Weiss. in correlation domain and later by Boll in Fourier transform domain. The algorithm is computationally simple as it only involves a

forward and inverse Fourier transform. Most of the speech distortion introduced by the Spectral subtraction process mention in [8].

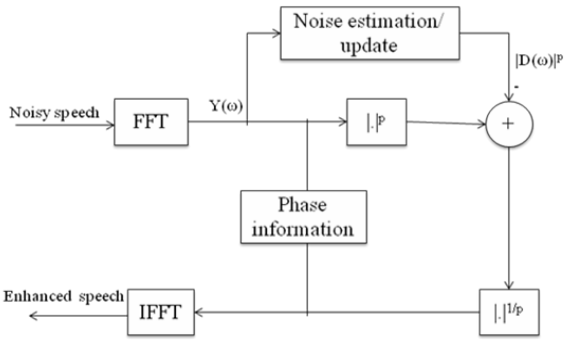

Figure 2.1: General Form of The Spectral Subtractive Algorithms

2.3 Statistical Model Based Algorithms

The majority of statistical model based algorithms perform equally well in terms of overall quality by Minimum mean square error (MMSE) algorithm [9] and wiener a-priori SNR algorithm(wiener as).

The minimum mean square error (MMSE) filter by Ephraim and Malah in 1984 is an important milestone. In these estimation type approaches, the transform coefficients are filtered in each short-time frame and attenuated independently of their intra-frame neighboring coefficients as well as inter-frame neighboring coefficients [**?**].

In wiener type algorithm, There are mainly three types of wiener filter algorithm. Wiener as algorithm, Wavelet Thresholding (WT) algorithm, Audio Suppression algorithm. The functional block diagram for a speech enhancement system with VAD [10] and wiener filter shown in Figure 2.2 [11]. In this $P_y(\omega)$ is noisy speech spectrum and $P_n(\omega)$ is estimated noise spectrum [6].

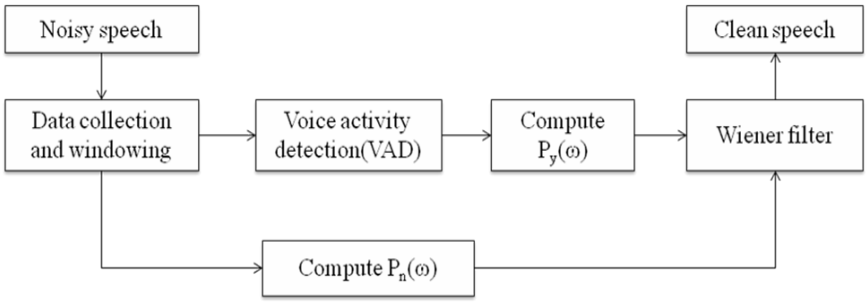

Figure 2.2: Functional Block Diagram for a Speech Enhancement System With a VAD and Wiener Filter

2.4 Subspace Algorithm

The signal subspace algorithm was originally developed by Ephraim and Van Trees in 1995 for white input noise and was later extended to handle colored noise [9] (e.g., speech-shaped noise) by Hu and Loizou in 2002.The underlying principle of the subspace algorithm is based on the projection of the noisy speech vector (consisting of, say, a segment of speech) onto two subspaces: the "signal" subspace and the "noise" subspace. The noise subspace contains only signal components due to the noise, and the signal subspace contains primarily the clean signal [11]. Therefore, an estimate of the clean signal can be made by removing the components of the signal in the noise subspace and retaining only the components of the signal in the signal subspace.

The subspace methods are based on the decomposition of noisy speech signal in to two subspaces: the speech plus noise and the noise only subspace. Once the decomposition is achieved, the noise subspace is discarded, while the clean speech is estimated from the remaining speech plus noise subspace.

Speech enhancement can be either single-channel or multi-channel as per [7]. In single-channel enhancement, speech is available from only a single microphone, whereas multi-channel systems make use of more than one microphone to better characterize and attenuate the noise. In this study we will concentrate on single-

channel systems. Single-channel enhancement methods can be divided into mainly two groups:

a. spectral subtraction-based methods

b. Wiener filtering- based methods

Most of the Wiener filtering-based algorithms are iterative since an estimate of clean speech power spectrum is required in the formulation, whereas spectral subtraction methods are non-iterative. Therefore spectral subtraction methods are computationally more attractive in practical applications.

comparison of speech enhancement algorithms is shown in Table 2.1.

Table 2.1: Comparison of The Different Speech Enhancement Algorithms According to Their Overall Performance.

Speech Enhancement Algorithms	Spectral Subtractive		statistical		
Sub Type	SS	Mband SS	WT	Wiener as	log MMSE
Overall Performance	Average	Best	Poor	Average	Best

2.5 Summary

In this chapter, Some speech enhancement algorithms like Spectral Subtractive Algorithms, Statistical Model Based Algorithms, Subspace Algorithm are explained And comparison of these speech enhancement algorithm in terms of overall performance is also given.

Chapter 3

Statistical Model Based Algorithms

In this chapter several statistical model based speech enhancement algorithms like wiener a-priori SNR algorithm, wavelet thresholding algorithm and log MMSE algorithm are described.

3.1 Wiener Algorithm

This section describes wiener a priori based algorithm for speech enhancement.

3.1.1 Basic Wiener Theory

Consider the statistical filtering problem given in Figure 3.1. The input signal goes through a linear and time invariant (LTI) system to produce an output signal [1]. We are to design the system in such a way that the output signal, $\hat{d}(n)$, is equal(in some sense) to the desired signal d(n), as possible. This can compute the estimation error e(n), and making it small possible optimal filter that minimizes the estimation error is called the wiener filter. Assuming the filter is FIR filter, so we have

$$\hat{d}(n) = \sum_{k=0}^{M-1} h_k y(n-k) \qquad\qquad n = 0, 1, 2...$$

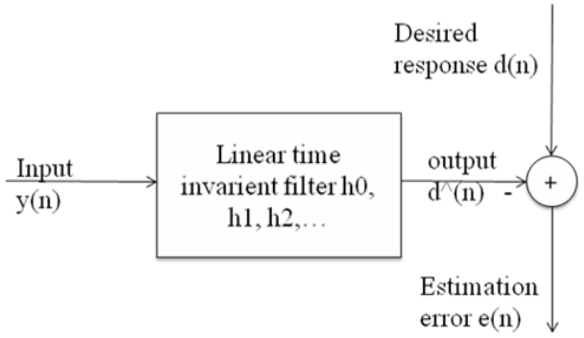

Figure 3.1: Block Diagram of The Statistical Filtering Problem

Where h_k are the filter coefficients, and M is the number of coefficient. Next, we need to compute the filter coefficient so that the estimation error i.e., d(n) - $\hat{d}(n)$ is minimized. The mean square of estimation error commonly used as a criterion for minimization.

3.1.2 Wiener Filter For Noise Reduction

Time Domain

- Basic Principle: The Wiener filter is to obtain an estimate of the clean signal from that corrupted by additive noise.
- This estimate is obtained by minimizing the Mean Square Error (MSE) between the desired signal d(n) and the estimated signal $\hat{d}(n)$.

In speech enhancement applications, the input signal y(n) in Figure 3.2. Is the noisy speech signal:

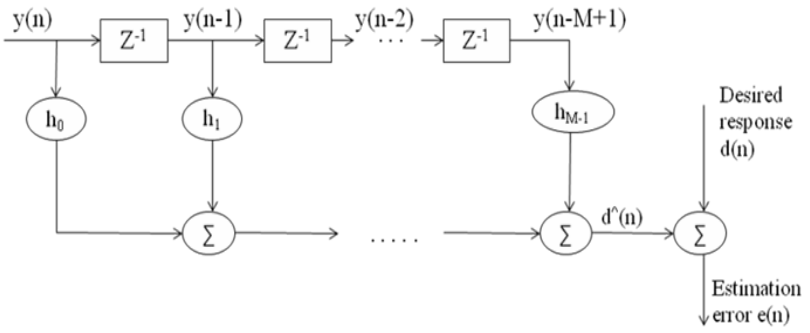

Figure 3.2: Block Diagram of The Statistical Filtering Problem

$$y(n) = x(n) + n(n) \tag{3.1}$$

where x(n) is the clean speech signal, and n(n) is the noise signal. The desired signal d(n) in Figure 3.2 is the clean (noise free) signal x(n), i.e., d(n) = x(n). Optimal filter coefficient

$$h^* = R_{yy}^{-1} r_{yd}^{-} \tag{3.2}$$

To evaluate wiener filter,

$$\begin{aligned} R_{yy} &= E[yy^T] \\ &= E[(x+n)(x+n)^T] \\ &= E[xx^T] + E[nn^T] + E[xn^T] + E[nx^T] \\ &= R_{xx} + R_{nn} \end{aligned} \tag{3.3}$$

Last two terms are zero, because signal and noise are assumed to be uncorrelated and zero mean. The crosscorelation vector r_{yd}^{-} in equation 3.2 is equal to r_{xx} because the signal and noise signals are assumed to be uncorrelated. Therefore the resulting wiener filter in time domain is

$$h^* = (R_{xx} + R_{nn})^{-1} r_{xx} \tag{3.4}$$

h* is a function of autocorrelation of x(n), and therefore is not realizable. Increasing asymptotic relationships about the values of optimal wiener filter h* can be derived as follow,

$$h^* = [(\frac{1}{SNR}) + \hat{R}_{nn}^{-1}\hat{R}_{xx}]^{-1}\hat{R}_{nn}^{-1}\hat{R}_{xx}u_1 \tag{3.5}$$

where

$$SNR = \frac{E[x2(n)]}{E[n2(n)]} = \frac{\sigma_x^2}{\sigma_n^2} \tag{3.6}$$

is the signal to noise ratio(SNR), I is the identity matrix (M x M), u_1^T = [1, 0, 0,,0] (1 x M), and $\hat{R}_{xx} \triangleq R_{xx}/\sigma_x^2$, $\hat{R}_{nn} \triangleq R_{nn}/\sigma_n^2$. From equation 3.5, we can write following asymptotic relationship about the wiener filter for large and small SNR values:

$$\begin{aligned} \lim_{SNR\to\infty} h^* &= u_1 \\ \lim_{SNR\to 0} h^* &= 0 \end{aligned} \tag{3.7}$$

The first relationship shows that when the SNR is extremely large, the wiener filter does not provide noise reduction because u_1^T = y(n), i.e., the observed noisy signal passes unaltered. Consequently, no speech distortion is imparted to the speech signal by the wiener filter when the SNR is large [11]. In contrast, the second relationship suggests that when the SNR is extremely low, the output of wiener filter is heavily attenuated. This attenuation produces undesirable distortion in the speech signal.[12]

Frequency Domain

General form of wiener filter in frequency domain

$$H(\omega_k) = \frac{P_{dy}(\omega_k)}{P_{yy}(\omega_k)} \tag{3.8}$$

$P_{yy}(\omega_k) = E|Y(\omega_k)|^2$ is the power spectrum of y(n), and $P_{dy}(\omega_k) = E|Y(\omega_k)D^*(\omega_k)|$ is the cross power spectrum of y(n) and d(n) taking the Fourier transform of equation 3.1, we get:

$$Y(\omega_k) = X(\omega_k) + N(\omega_k) \tag{3.9}$$

According to equation 3.8, we need to compute $P_{dy}(\omega_k)$ and $P_{yy}(\omega_k)$. Given that $D(\omega_k) = X(\omega_k)$, and using equation 3.9, we get

$$\begin{aligned} P_{dy}(\omega_k) &= E[X(\omega_k)\{X(\omega_k) + N(\omega_k)\}^*] \\ &= E[X(\omega_k)X^*(\omega_k)] + E[X(\omega_k)N^*(\omega_k)] \\ &= P_{xx}(\omega_k) \end{aligned} \tag{3.10}$$

Similarly,

$$\begin{aligned} P_{yy}(\omega_k) &= E[\{x(\omega_k) + N(\omega_k)\}\{X(\omega_k) + N(\omega_k)\}^*] \\ &= E[X(\omega_k)X^*(\omega_k)] + E[N(\omega_k)N^*(\omega_k)] + \\ &\quad E[X(\omega_k)N^*(\omega_k)] + E[X^*(\omega_k)N(\omega_k)] \\ &= P_{xx}(\omega_k) + P_{nn}(\omega_k) \end{aligned} \tag{3.11}$$

Finally, after substituting equation 3.10 and 3.11 in equation 3.8, we get the wiener filter in frequency domain

$$H(\omega_k) = \frac{P_{xx}(\omega_k)}{[P_{xx}(\omega_k) + P_{nn}(\omega_k)]} \tag{3.12}$$

The fact that $H(\omega_k)$ is even and real suggests that the impulse response, h_k, must be even as well. By defining ξ_k

$$\xi_k \triangleq \frac{P_{xx}(\omega_k)}{P_{nn}(\omega_k)} \tag{3.13}$$

as the a priori SNR at frequency ω_k, we can also express the wiener filter in equation 3.12 as

$$H(\omega_k) = \frac{\xi_k}{[\xi_k + 1]} \tag{3.14}$$

Note that $0 \leq H(\omega_k) \leq 1$, and $H(\omega_k) \approx 0$ when $\xi_k \longrightarrow 0$ (i.e., at extremely low SNR regions) and $H(\omega_k) \approx 1$ when $\xi_k \longrightarrow \infty$ (i.e., at extremely high SNR regions). These asymptotic relationships in the frequency domain are in line with the asymptotic relationships in the time domain.

3.1.3 Algorithm for Speech Enhancement System

Figure 3.3 shows the algorithm for speech enhancement by wiener a priori SNR method.In this, a non-iterative Wiener filtering technique is described [6]. The main advantage of the described method is that it makes use of a time varying signal-to-noise ratio (SNR) dependent noise suppression factor. This property gives us the ability to suppress those parts of the degraded signal, where speech is not likely to be present and not to suppress, and hence not to distort the speech segments much. However, in this we use a different SNR estimation method as explained in the algorithm description section. The non-iterative Wiener filtering technique described here produces enhanced speech significantly better than enhanced speech from standard spectral subtraction [13].

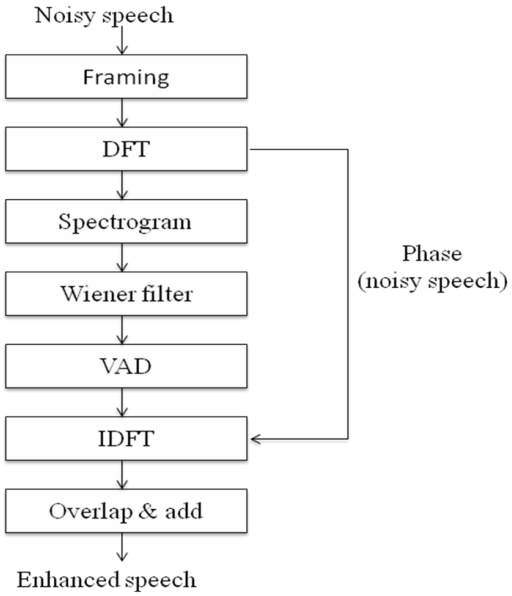

Figure 3.3: Algorithm for speech enhancement using wiener a priori SNR method

Wiener Filter

Wiener filter in equation 3.15 is widely used and its construction requires the expectation of the power of the clean speech and the noise to be known

$$W(u,v) = \frac{\xi(u,v)}{[\xi(u,v)+1]} \tag{3.15}$$

where, a-priori SNR, $\xi(u,v) = E[X(u,v)^2]/\lambda_N E[.]$ the expectation operator and $\lambda_N = E[N(u,v)^2]$. In practical implementations of speech enhancement systems, only the noisy speech spectrum is available. Therefore, $E[X(u,v)^2]$] is unknown and should be estimated. λ_N is assumed to be known in this paper since the background noise is stationary and can be easily estimated during speech pauses [10]. The a-priori SNR can be estimated from the combination of the a-posteriori SNR and the enhanced speech spectrum derived of the previous frame using the popular decision-directed approach [9]:

$$\hat{\xi}(u,v) = \alpha \frac{\hat{X}(u-1,v)^2}{\lambda_N} + (1-\alpha)F[\gamma(u,v)-1] \tag{3.16}$$

where $\gamma(u,v) = Y(u,v)^2/\lambda_N$ is the a-posteriori SNR, $F[.]$ denotes the half-wave rectification function and $\hat{X}(u-1,v)$ is the estimated speech spectrum value of X(u, v) in the previous frame. The value of α controls the behavior of the SNR estimator and is normally set to 0.98.

Estimating the A Priori SNR using Decision-Direct Approch Method

Decision-direct approach was based on the definition of ξ_k and its relationship with the a posteriori SNR γ_k. We know that ξ_k is givem by

$$\xi_k = \frac{E\{X_k^2(m)\}}{\lambda_d(k,m)} \tag{3.17}$$

We also know that ξ_k is related to γ_k by

$$\begin{aligned} \xi_k(m) &= \frac{E\{Y_k^2(m) - D_k^2(m)\}}{\lambda_d(k,m)} \\ &= \frac{E\{Y_k^2(m)\}}{\lambda_d(k,m)} - \frac{E\{D_k^2(m)\}}{\lambda_d(k,m)} \\ &= E\{\gamma_k(m)\} - 1 \end{aligned} \tag{3.18}$$

Combining the two expression for, ξ_k i.e. Equation 3.17 and 3.18 we get:

$$\xi_k(m) = E\{\frac{1}{2}\frac{X_k^2(m)}{\lambda_d(k,m)} + \frac{1}{2}[\gamma_k(m)-1]\} \tag{3.19}$$

The final estimator for ξ_k is derived by making the preceding equation recursive:

$$\hat{\xi}_k(m) = a\frac{X_k^2(m-1)}{\lambda_d(k,m-1)} + (1-a)max[\gamma_k(m)-1, 0] \tag{3.20}$$

where $0 < a < 1$ is the weighting factor replacing the $1/2$ in the equation 3.19

and $\hat{X}_k^2(m-1)$ is the amplitude estimator obtained in the past analysis frame. The max(.) operator is used to ensure the positiveness of the estimator, as $\hat{\xi}_k(m)$ needs to be nonnegative.

This new estimator of ξ_k is a weighted average of the past a priori SNR (Given by the first term) and the present a priori SNR estimate (given by the second term). Note that the present a priori SNR estimate is also the maximum-likelihood estimate of the SNR. Equation 3.20 was called the decision-direct estimator because $\hat{\xi}_k(m)$ is updated using information from the previous amplitude estimate. The decision-direct approach for estimating the a priori SNR was found not only important for MMSE-type algorithms but also in other algorithms. Equation 3.20 needs initial conditions for the first time, i.e. for $m = 0$. The following initial conditions were recommended for $\hat{\xi}_k(n)$: Good results were obtained with $a = 0.98$

$$\hat{\xi}_k(0) = a + (1-a)max[\gamma_k(0) - 1, 0] \tag{3.21}$$

VAD : Voice Activity Detection

A voice activity detector (VAD) [10]was used in most of the speech enhancement methods to update the noise spectrum. More precisely, a statistical-model based voice activity detector (VAD) was used to update the noise spectrum during speech-absent periods. The following VAD decision rule was used as per [6, 12]:

$$1/N \sum_{k=1}^{n-1} \log \wedge_k \lessgtr \delta \tag{3.22}$$

where

$$\wedge_k = \frac{1}{1+\xi_k} \exp\{\frac{\gamma_k \xi_k}{1+\xi_k}\}$$

Where γ_k = posterior SNR and ξ_k = priori SNR and ξ_k is estimated using the decision directed approach ($\alpha = 0.98$). N is the size of the FFT, H1 denotes the hypothesis of speech presence, H0 denotes the hypothesis of speech absence, and η is a preset threshold. In our implementation, $\eta = 0.15$ for all conditions. During the

speech-absent periods, i.e., when the left side of equation 3.22 was smaller than η, the noise power spectrum was updated according to:

$$N_j(k) = (1 - \beta)|Y_j(k)|^2 + \beta N_{j-1}(k) \tag{3.23}$$

where $N_j(k)$ is the estimate of the noise power spectrum at frame j for frequency bin k, $\beta = 0.98$ is a preset smoothing factor, and $|Y_j(k)|$ is the noisy speech magnitude spectrum. The initial estimate of $N_j(k)$ was obtained from the first (speech-absent) 120-ms segment of each sentence.

Discrete Fourier Transform (DFT) & Inverse DFT (IDFT)

DFT of the speech segment can provide a reasonable representation of the frequency-domain characteristic of the speech in this time interval. The DTFT of discrete-time sequence is a continues function of frequency ω. In practice, the sequence x(n) is finite in duration (e.g. consisting of N samples), and we can sample the DTFT at N uniformly spaced frequencies i.e. at $\omega_k = 2\Pi k/N$, k = 0,1,...,N-1. This sampling yields a new transform referred to as the DFT. The DFT of x(n) is given by:

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j\frac{2\Pi kn}{N}} \qquad 0 \leq k \leq N-1 \tag{3.24}$$

Given X(k),we can recover x(n) from its DFT, using inverse DFT (IDFT):

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) e^{j\frac{2\Pi kn}{N}} \qquad 0 \leq n \leq N-1 \tag{3.25}$$

Overlap and Add

After the inverse fourier transform, the enhanced speech is constructed by inverse windowing followed by the overlap and add operation.Overlap and add is method for reconstructing x(n) from its Fourier Transform (FT), which is widely used in speech enhancement.

Spectrogram

The spectrogram is a graphical display of the power spectrum of speech as a function of time and is given by

$$S(n,\omega) = |X(n,\omega)|^2 \tag{3.26}$$

where, $X(n,\omega)$ denotes the Short Time Fourier Transform (STFT) of the speech signal x(n). The quantity $S(n,\omega)$ can be viewed as a two dimensional Power Spectral Density (PSD), the second dimension being time. The spectrogram describes the speech signal's relative energy concentration in frequency as a function of time and, as such, it reflects time varying properties of speech waveform.

3.2 Wavelet Thresholding Algorithm

In this section, Wavelet thresholding which uses multitaper multitapper method for speech enhancement is described.

3.2.1 Wavelet Thresholding Basics

A technique based on wavelet transform is given for noise reduction. It reduces noise by thresholding the wavelet coefficients so that only the coefficient with values above the threshold are retained. Since, signal energy is concentrated on a small number of wavelet coefficients in many signals while wavelet coefficients of noise is spread over a wide number of coefficients appropriate thresholding can lead to high noise reduction with low signal distortion [14].

Traditional wavelet-based speech enhancement algorithm can be summarized by the following three steps [15],

- Wavelet transform of noisy signal
- Thresholding the resulting wavelet coefficients
- Inverse transform to obtain the denoised signal

There are wide varieties of basic thresholding approaches [16].

- Hard thresholding, where all coefficients below predefined threshold value are set to zero.
- Soft thresholding, where in addition the remaining coefficients are linearly reduced in value.
- Nonlinear thresholding, where a smooth function is used to map the original coefficients to a new set, avoiding abrupt value changes.

Illustration of Hard, Soft and Nonlinear thresholding operations are shown in Figure 3.4.

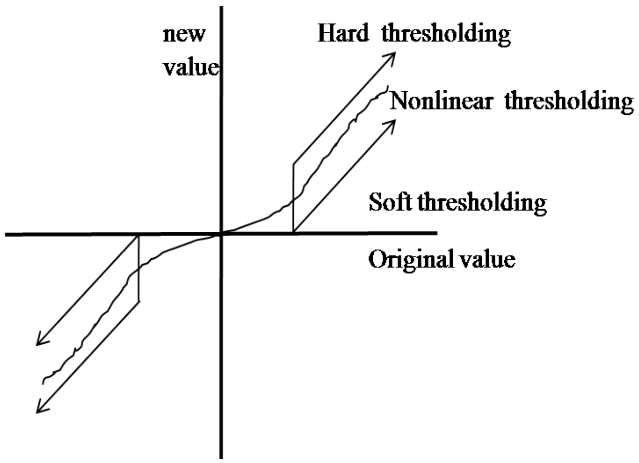

Figure 3.4: Threshold mapping functions

Two of the most common methods are universal thresholding and stein's unbiased risk estimator(SURE), typically implemented with soft thresholding function.

3.2.2 Multitaper Spectrum Estimator

Direct spectrum estimation based on Hamming windowing is the most often used power spectrum estimator for speech enhancement. Although windowing reduces the bias, it does not reduce the variance of the spectral estimate [16]. The idea behind the multitaper spectrum estimator is to reduce this variance by computing a small number (L) of direct spectrum estimators each with a different taper (window), and then average the L spectral estimates. The underlying philosophy is similar to Welch's method of modied periodogram. If the L tapers are chosen to be pairwise orthogonal and properly designed to prevent leakage, then the resulting multitaper spectral estimator will be superior to the periodogram in terms of reduced bias and variance. At best, the variance of the multitaper estimate will be smaller than the variance of each spectral estimate by a factor of 1/L [17] [15] [18].

Multitaper method used to estimate power spectrum of signal [19].

Multitaper spectrum estimator of signal x(n),

$$P_{xx}^{mt}(\omega) = \frac{1}{L}\sum_{k=0}^{L-1} P_k(\omega) \quad L = Number\ of\ Tapers \tag{3.27}$$

$$P_k(\omega) = \mid \sum_{n=0}^{N-1} t_k(n)x(n)e^{-j\omega n} \mid^2 \tag{3.28}$$

$t_k(n)$ is k^{th} data taper [20].

sine taper for $t_k(n)$

$$t_k(n) = \sqrt{\frac{2}{N+1}}\sin(\frac{\Pi k(n+1)}{N+1}) \quad n = 0, 1, 2, .., N-1 \tag{3.29}$$

where N is data length.

Wavelet thresholding techniques can be used to further refine the spectral estimate and produce a smooth estimate of logarithm of spectrum. [15]

Wavelet thresholding used to compute $\xi_k(a\ priori\ SNR)$

$$\hat{\xi_k} = \frac{\hat{P_{xx}}(\omega_k)}{\hat{P_{dd}}(\omega_k)} \tag{3.30}$$

where, $\hat{P_{xx}}(\omega) = P_{yy}(\omega) - \hat{P_{dd}}(\omega)$

3.2.3 Algorithm for Speech Enhancement System

Figure 3.5 shows the algorithm for speech enhancement by wavelet thresholding.

1. Compute multitaper power spectrum $P_{yy}^{mt}(\omega)$ of noisy speech y using equation and estimate multitaper power spectrum $P_{xx}^{mt}(\omega)$ of clean speech signal by

$$P_{xx}^{mt}(\omega) = P_{yy}^{mt}(\omega) - P_{dd}^{mt}(\omega) \tag{3.31}$$

$P_{dd}^{mt}(\omega)$ = multitaper power spectrum of noise obtained using noise sample collected during speech absent frames negative elements of $P_{xx}^{mt}(\omega)$ are floored as follows

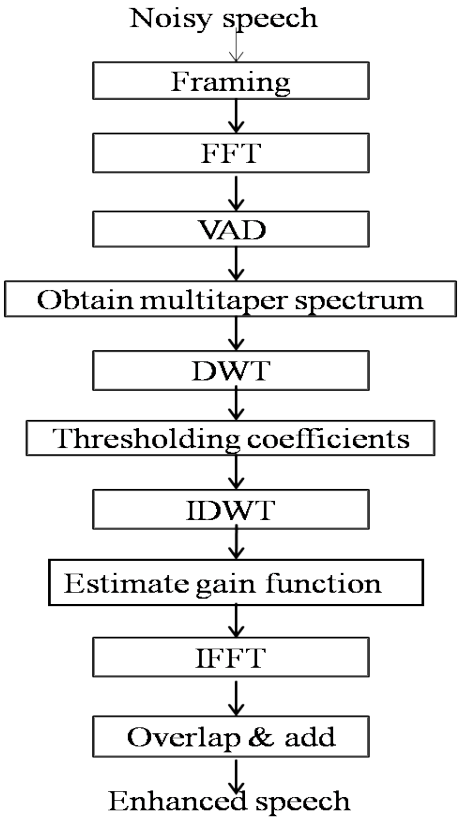

Figure 3.5: Algorithm for speech enhancement using wavelet thresholding

$$\begin{aligned} P_{xx}^{mt}(\omega) &= P_{yy}^{mt}(\omega) - P_{dd}^{mt}(\omega) \quad if \;\; P_{yy}^{mt}(\omega) > (\beta+1)P_{dd}^{mt}(\omega) \\ &= \beta P_{dd}^{mt}(\omega) \qquad\qquad else \end{aligned} \tag{3.32}$$

Where, $\beta = 0.002$

2. Compute $z(\omega) = \log P_{yy}^{mt}(\omega) + c$ where $c = -\phi(L) + \log L$. where, $\phi(\bullet)$ is digamma function.

Apply DWT to $z(\omega)$ to obtain DWT coefficient $z_{j,k}$.

Threshold the wavelet coefficient $z_{j,k}$ and apply inverse DWT to threshold wavelet coefficients to obtain the refined log spectrum $\log P_{yy}^{wmt}(\omega)$ of noisy speech signal [21].

Repeat this procedure to obtain the refined log spectrum, $\log P_{dd}^{wmt}(\omega)$ of noise signal. then,

$$\log P_{xx}^{wmt}(\omega) = logP_{yy}^{wmt}(\omega) - logP_{dd}^{wmt}(\omega) \tag{3.33}$$

So,

$$\hat{\xi}_k = \frac{\hat{P_{xx}^{wmt}}(\omega_k)}{\hat{P_{dd}^{wmt}}(\omega_k)} \tag{3.34}$$

3. Find μ which used in gain function.

$$g(k) = \frac{\xi_k}{\xi_k + \mu} \tag{3.35}$$

$$\begin{aligned} \mu &= \mu_0 - (\frac{SNR_{dB}}{S}) \quad -5 < SNR_{dB} < 20 \\ &= 1 \qquad\qquad\qquad SNR_{dB} \geq 20 \\ &= \mu_{max} \qquad\qquad SNR_{dB} \leq -5 \end{aligned} \tag{3.36}$$

$SNR_{dB} = 10 \log_{10} SNR$

$$SNR = \frac{\sum_{k=0}^{N-1} P_{xx}^{wmt}(\omega_k)}{\sum_{k=0}^{N-1} P_{dd}^{wmt}(\omega_k)} \tag{3.37}$$

4. Estimate g(k) for frequency component ω_k using equation (9).

Enhanced spectrum $\hat{x}(\omega_k) = g(k).y(\omega_k)$.

apply IFFT of $\hat{x}(\omega_k)$ to obtain enhanced speech signal [14].

3.3 Log MMSE algorithm

In this section, log MMSE (Minimum mean square error) algorithm is described.

3.3.1 Log MMSE Basic and Algorithm

The optimal MMSE spectral amplitude estimator minimizes the error of the spectral magnitude spectra. The matric based on the squared error of the log magnitude spectra is more suitable for speech processing. Below, an estimator that minimize the mean-square error of the log-magnitude spectra is described [5]. Figure 3.6 shows the algorithm for speech enhancement by log MMSE estimator.

$$E(logX_k - log\hat{X}_k)^2 \tag{3.38}$$

The optimal log-MMSE estimator can be obtained by evaluating the conditional mean of $\log X_k$, i.e.,

$$log\hat{X}_k = E(logX_k|Y(\omega_k) \tag{3.39}$$

From which we can solve for $\hat{X}_k$:

$$\hat{X}_k = exp(E(logX_k|Y(\omega_k)) \tag{3.40}$$

Here,

$$E(logX_k|Y(\omega_k) = \frac{1}{2}log\lambda_k + \frac{1}{2}log\nu_k + \frac{1}{2}\int_{\nu_k}^{\inf}\frac{e^{-t}}{t}dt \tag{3.41}$$

by putting this equation in equation 3.40, we get the optimal log-MMSE estimator [1]:

$$\hat{X}_k = \frac{\xi_k}{\xi_k + 1}exp\frac{1}{2}\int_{\nu_k}^{\inf}\frac{e^{-t}}{t}dtY_k \tag{3.42}$$

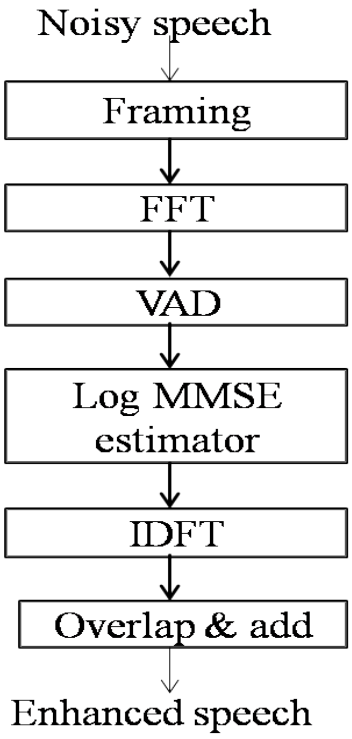

Figure 3.6: Algorithm for speech enhancement using Log MMSE estimator

$$\hat{X}_k =\triangleq G_{LSA}(\xi_k, \nu_k)Y_k \tag{3.43}$$

where ξ_k is the a priori SNR, and $G_{LSA}(\xi_k, \nu_k)$ is the gain function of the log MMSE estimator. and

$$\nu_k = \frac{\xi_k}{\xi_k + 1}\gamma_k \tag{3.44}$$

$$\gamma_k = \frac{(Y_k)^2}{\lambda_d(k)} \tag{3.45}$$

$$\xi_k = \frac{\lambda_x(k)}{\lambda_d(k)} \tag{3.46}$$

γ_k is a posteriori SNR, can be considered the observed or measured SNR of the k th spectral component after noise is added [1].

3.4 Summary

This chapter described several statistical model based speech enhancement algorithms like wiener a-priori SNR algorithm, wavelet thresholding algorithm and log MMSE algorithm.

Chapter 4

Spectral Subtractive Algorithms

In this chapter we will see various spectral subtractive algorithms for speech enhancement like spectral subtraction algorithm and multi-band spectral subtraction algorithm.

4.1 Spectral Subtraction Algorithm

4.1.1 Spectral Subtraction Basics

One of the most popular methods of reducing the effect of background (additive) noise. Assume that y(n), noise corrupted input signal is composed of the clean speech signal x(n) and additive noise signal d(n) [22];

$$y(n) = x(n) + d(n) \tag{4.1}$$

Taking discrete time Fourier transform of both sides

$$Y(\omega) = X(\omega) + D(\omega) \tag{4.2}$$

Y(w) in polar form as:

$$Y(\omega) =\mid Y(\omega) \mid \exp(j\phi_y(\omega)) \tag{4.3}$$

$\mid Y(\omega) \mid$ is magnitude and $\exp(j\phi_y(\omega))$ is phase spectrum of corrupted noisy signal.

Noise spectrum $D(\omega)$can also be expressed in terms of magnitude and phase spectra as $D(\omega) =\mid D(\omega) \mid \exp(j\phi_d(\omega))$.

The magnitude noise spectrum $\mid D(\omega) \mid$ is unknown, but can be replaced by its average value computed during non speech activity . Similarly, noise phase $\phi_d(\omega)$ can be replaced by the noise speech phase $\phi_y(\omega))$, because of the fact that phase that does not effect noise intelligibility [23].

By substituting to equation 4.2, we can obtain an estimate of clean signal spectrum.

$$\hat{X}(\omega) = [\mid y(\omega) - \hat{D}(\omega) \mid] \exp(j\phi_y(\omega)) \tag{4.4}$$

$\mid \hat{D}(\omega) \mid$ is the estimate of the magnitude noise spectrum made during non-speech activity. Enhanced speech signal can be obtained by inverse Fourier transform of $\hat{X}(\omega)$

Block Diagram

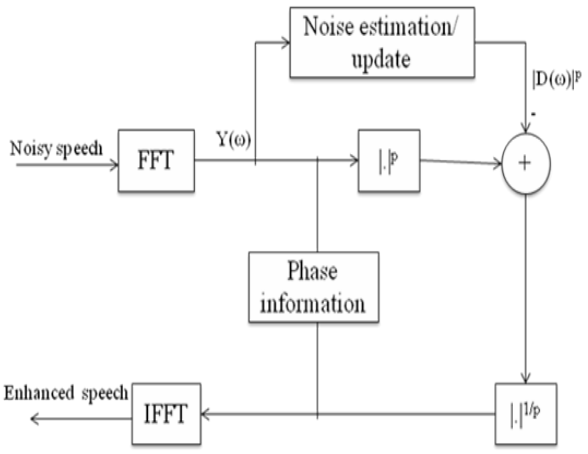

Figure 4.1: General Form of The Spectral Subtractive Algorithms

If the enhanced noise spectrum is negative, which can not be negative; hence; caution needs to be exercised when subtracting the two spectra to ensure that $\mid \hat{X}(\omega) \mid$ is always non negative. One solution to this is two half wave rectify the difference spectra, i.e., set the negative spectral components to zero as follow:

$$\begin{aligned} \mid \hat{X}(\omega) \mid &= \mid Y(\omega) - \hat{D}(\omega) \mid \quad if \mid Y(\omega) \mid > \mid \hat{D}(\omega) \mid \\ &= 0 \qquad\qquad\qquad\qquad else \end{aligned} \tag{4.5}$$

In power spectral domain; to obtain short term power spectrum of the noisy speech we multiply $Y(\omega)$ in equation 4.2 by its conjugate $Y^*(\omega)$. So equation 4.2 becomes

$$\begin{aligned} \mid Y(\omega) \mid^2 &= \mid X(\omega) \mid^2 + \mid D(\omega) \mid^2 + X(\omega)D^*(\omega) + X^*(\omega)D(\omega) \\ &= \mid X(\omega) \mid^2 + \mid D(\omega) \mid^2 + 2Re\{X(\omega)D^*(\omega)\} \end{aligned} \tag{4.6}$$

The term $\mid D(\omega) \mid^2, X(\omega)D^*(\omega), X^*(\omega)D(\omega)$ can not be obtained directly and are approximated as $E\{\mid D(\omega) \mid^2\}, E\{X(\omega)D^*(\omega)\}, E\{X^*(\omega)D(\omega)\}$

where E[.] demotes the expectation power. Typically $E\{\mid D(\omega) \mid^2\}$ is estimated during non-speech activity and is denoted by $\mid \hat{D}(\omega) \mid^2$. Here, if we assume that d(n) is zero mean and uncorrelated with the clean signal x(n), then the terms $E\{X(\omega)D^*(\omega)\}$ and $E\{X^*(\omega)D(\omega)\}$ reduce to zero. So the estimate of the clean speech power spectrum can be obtained as follow [8]:

$$\mid \hat{X}(\omega) \mid^2 = \mid Y(\omega) \mid^2 - \mid \hat{D}(\omega) \mid^2 \tag{4.7}$$

This equation describes the power spectrum subtraction algorithm.

A more generalized version of the spectrum subtraction algorithm [24] is given by

$$\mid \hat{X}(\omega) \mid^p = \mid Y(\omega) \mid^p - \mid \hat{D}(\omega) \mid^p \tag{4.8}$$

Where p is the power exponent, with p = 1 ,the original magnitude s spectrum subtraction, and p = 2, the power spectrum subtraction algorithm as shown in fig. 1

4.1.2 Algorithm for Speech Enhancement System

Figure 4.2 shows the algorithm for speech enhancement by spectral subtraction method.

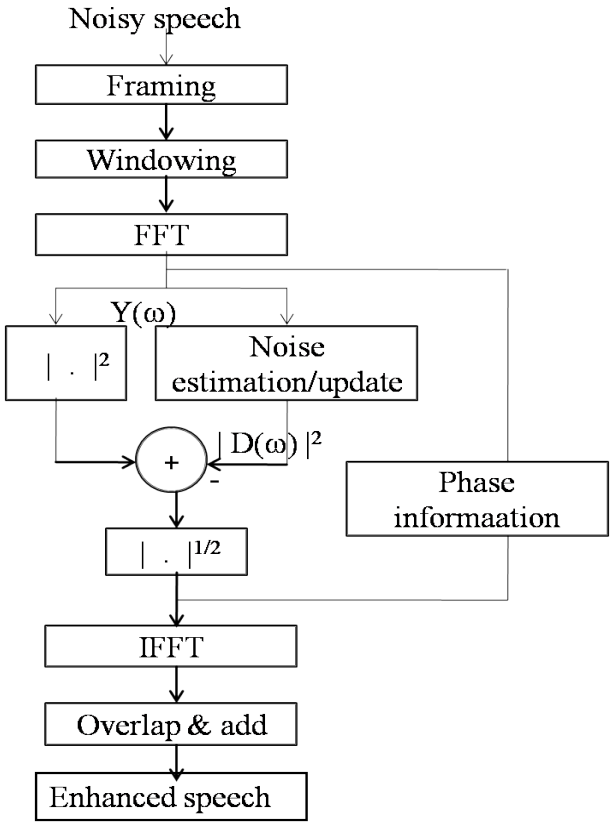

Figure 4.2: Algorithm for speech enhancement using spectral subtraction method

The decision as to weather the noise spectrum should be updated or not is based on comparison of estimated a posteriori SNR to a threshold.

$$SNR(dB) = 10\log_{10}(\frac{\sum \mid \hat{Y}(k) \mid^2}{\sum \mid \hat{D}(k) \mid^2}) \tag{4.9}$$

If SNR < threshold, speech absence than noise spectrum is updated.

And if SNR > threshold, speech is presence than noise spectrum update is stopped.

The estimate of the clean speech spectrum $\mid \hat{X}(k) \mid$

$$\mid \hat{X}(k) \mid^2 = \mid \hat{y}(k) \mid^2 + \alpha \mid \hat{D}(k) \mid^2$$

Where $\mid \hat{X}(k) \mid$ is the preprocessed noisy speech spectrum $\mid \hat{D}(k) \mid$ is the noise spectrum estimate and a overall subtraction factor.

$$\begin{aligned} \alpha &= 4 && SNR < 5 \\ &= 5 - \frac{3}{20}(SNR) && -5 \leq SNR \leq 20 \\ &= 1 && SNR > 20 \end{aligned} \tag{4.10}$$

M = samples of y(n) , M-point FFT

λ is frame index

k = 0,1,2 M-1.

$p(\lambda, k) = \delta p(\lambda - 1, k)^2 + (1 - \delta) \mid Y(\lambda, k) \mid^2$

δ = smoothing constant, range 0.6 to 0.98

4.2 Multiband Spectral Subtraction(MBSS) Algorithm

4.2.1 Multiband Spectral Subtraction(MBSS) basic

noise signal does not affect the speech signal uniformly over the whole spectrum. Some frequencies are affected more adversely than the others depending on the spectral characteristics of the noise [1].
So in MBSS Speech is processed into N ($1 \leqslant N \leqslant 8$) overlapping frequency bands and spectral subtraction is performed independently on each band using band-specific over-subtraction factors.This method provides a greater degree of flexibility and control on the noise subtraction levels that reduces artifacts in the enhanced speech, resulting in improved speech quality [25].

Multiband spectral subtraction-Blockdiagram

Steps of MBSS method

A block diagram of the MBSS method consists of 4 stages [25].

- The signal is windowed and the magnitude spectrum is estimated using the FFT.
- Split the noise and speech spectra into different frequency bands and calculate the over-subtraction factor(αi) for each band.
- Process the individual frequency bands by subtracting the corresponding noise spectrum from the noisy speech spectrum.
- Now, Reconstruct the modified frequency bands and obtain the time signal by using the noisy phase information and take the IFFT.

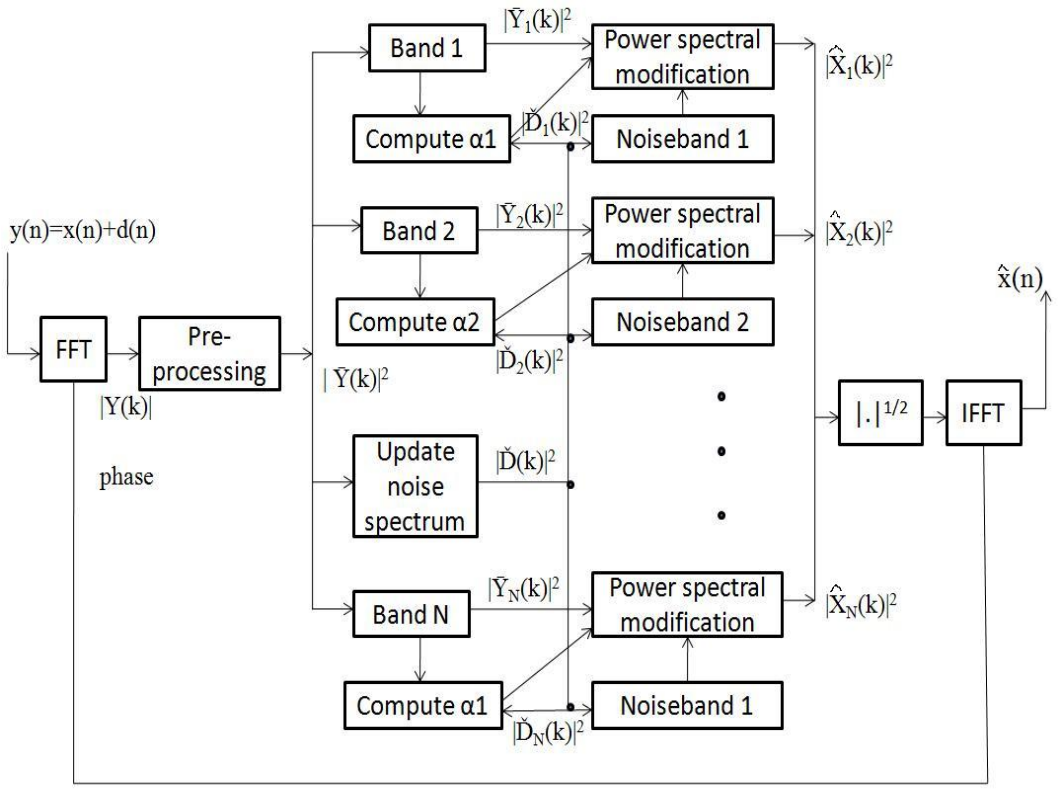

Figure 4.3: Blockdiagram of Multiband spectral subtraction

4.2.2 Algorithm for Speech Enhancement System

Figure 4.4 shows the algorithm for speech enhancement by multiband spectral subtraction method.

Description of algorithm

Assume that y(n), noise corrupted input signal is composed of the clean speech signal x(n) and additive noise signal d(n);

$$y(n) = x(n) + d(n) \tag{4.11}$$

$$| Y(k) |^2 = | X(k) |^2 + | D(k) |^2 \tag{4.12}$$

S(k) and D(k) are the magnitude spectra of clean speech and noise.

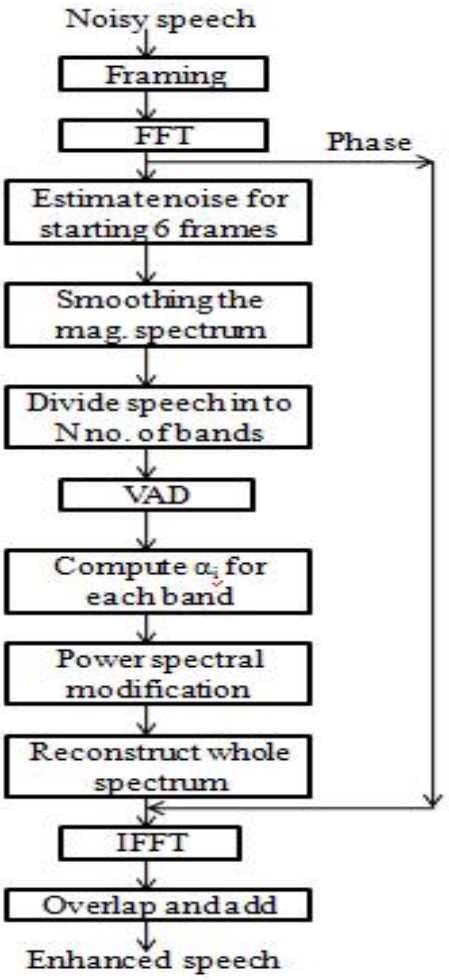

Figure 4.4: Algorithm for speech enhancement using Multiband spectral subtraction method

Estimated $\hat{D}(k)$ is calculated during periods of silence or non-speech activity.

estimated speech(clean) spectrum is obtained by

$$| \hat{X}(k) |^2 = | Y(k) |^2 - \alpha | \hat{D}(k) |^2 \tag{4.13}$$

$$\begin{aligned} | \hat{X}(k) |^2 &= | \hat{X}(k) |^2 \quad if \; | \hat{X}(k) |^2 > \beta | \hat{D}(k) |^2 | \\ &= \beta | \hat{D}(k) |^2 \quad else \end{aligned} \tag{4.14}$$

$\bar{Y}_i(\omega_k)$ is the smoothed noisy speech spectrum of the i th frequency band estimated in preprocessing stage.
Negative values are floored as follow,

$$\begin{aligned} \mid \hat{X}_i(\omega_k) \mid^2 &= \mid \hat{X}_i(\omega_k) \mid^2 \quad if \mid \hat{X}_i(\omega_k) \mid^2 > \beta \mid \bar{Y}_i(\omega_k) \mid^2 \\ &= \beta \mid \bar{Y}(\omega_k) \mid^2 \quad else \end{aligned} \tag{4.15}$$

where the spectral floor parameter β is set to 0.002. α is over subtraction factor, which is function of segmental SNR, in this noise affects speech uniformly.

But in fact, colored noise affects speech spectrum differently at various frequency. So multiband the spectral subtraction. Speech signal is divided in to N non-overlapping bands, and spectral subtraction performed independently on each band.

The estimate of clean speech spectrum in the i th band is obtained by;

$$\mid \hat{X}_i(\omega_k) \mid^2 = \mid \bar{Y}_i(\omega_k) \mid^2 - \alpha_i \delta_i \mid \hat{D}(\omega_k) \mid^2 \tag{4.16}$$

where $\omega_k = 2\pi$k/N (k = 0,1,....,N-1) are the discrete frequencies, $\mid \hat{D}(\omega_k) \mid^2$ is the estimated noise power spectrum,

$b_i \leqslant$ k $\leqslant e_i$

b_i = begining frequency and
e_i = ending frequency bins of the i th frequency band.
α_i is the over subtraction factor of the i th band.
δ_i is the band subtraction factor, that can be individually set for each frequency band to customize the noise removal property.

4.3 Summary

This chapter describes various spectral subtractive algorithms for speech enhancement like spectral subtraction algorithm and multi-band spectral subtraction algorithm. These algorithms are computationally simple as they only involve a forward and an inverse forward Fourier transform.

Chapter 5

Quality Evaluation Parameter

Subjective quality parameters (SIG, BACK, OVRL) and Objective quality parameters (LLR, SNRseg, WSS, PESQ) are computed for the enhanced speech signals produced by different algorithms in various noise environments like Babble, Car, Train, Restaurant having 0 dB, 5 dB, 10 dB, 15 dB of input(noisy) speech signal. These parameters are described as follow as given in [1].

5.1 Objective Quality Parameters

Objective measure of speech quality are implemented by first segmenting the speech signal in to 10-30 msec frames and then computing a distortion measure between the original and proposed signal.

5.1.1 Segmental SNR(Signal to Noise Ratio)

for this measure, it is important that the original and processed speech signals be aligned in time and in same phase [26].

Segmental SNR is defined as:

$$SNR_{seg} = \frac{10}{M}\sum_{m=0}^{M-1} log_{10}\frac{\sum_{n=Nm}^{Nm+N-1} x^2(n)}{\sum_{n=Nm}^{Nm+N-1}(x(n)-\hat{x}(n))^2} \tag{5.1}$$

x(n) is the original (clean) speech signal and $\hat{x}(n)$ is enhanced speech signal. N is frame length and M is number of frames in the signal.

SNR_{seg} measure is based on the geometric mean of the SNRs across all frames of the speech signal. values were limited in the range of (-10dB,35dB).

5.1.2 LLR

LLR(Log-Likelihood Ratio) is the dissimilarity between all-pole models of the clean and enhanced speech signals. [26]

$$d_{LLR}(a_x, \bar{a}_x) = log\frac{\bar{a}_{\hat{x}}^T R_x \bar{a}_{\hat{x}}}{\bar{a}_x^T R_x \bar{a}_x} \tag{5.2}$$

where $\bar{a}_x^T = [1, -\alpha_x(1), -\alpha_x(2)... - \alpha_x(P)$are the LPC co-efficient of the clean signal. $\bar{a}_{\hat{x}}^T = [1, -\alpha_{\hat{x}}(1), -\alpha_{\hat{x}}(2)... - \alpha_{\hat{x}}(P)$ are the LPC co-efficient of the enhanced signal.

R_x is the $(P+1)X(P+1)$ autocorrelation matrix of clean signal.

5.1.3 PESQ

PESQ(Perceptual Evaluation of Speech Quality) is computed as a linear combination of the average disturbance value and the average asymmetrical disturbance value [27].

$$PESQ = 4.5 - 0.1 \cdot d_sym - 0.0309 \cdot d_asym \tag{5.3}$$

the range of PESQ score is 0.5 to 4.5

5.1.4 WSS

WSS(Weighted Spectral Slope) computes the weighted difference between the spectral slops in each frequency band. this spectral slop is obtained as the difference between adjacent spectral magnitudes in decibels [28]. The WSS measure is computed for each frame of speech as:

$$d_{WSS}(C_x, \bar{C}_x) = \sum_{k=1}^{L} W(k)(S_x(k) - \bar{S}_{\hat{x}}(k))^2 \tag{5.4}$$

L is the number of critical bands used. W(k) is the weight for bank k. $C_x and \bar{C}_x$ are original (clean) and enhanced critical band spectra, respectively.

$$S_x(k) = C_x(k+1) - \bar{C}_x(k) \tag{5.5}$$

$$\bar{S}_{\hat{x}}(k) = C_{\hat{x}}(k+1) - \bar{C}_{\hat{x}}(k) \tag{5.6}$$

where $S_x(k) and \bar{S}_{\hat{x}}(k)$ denote the spectral slops of the clean and enhanced signals, respectively.

5.2 Subjective Quality Parameters

5.2.1 SIG, BAK, OVRL

[28]

- The speech signal alone using five-point scale of signal distortion.(SIG) as shown in table 5.2.1
- The background noise alone using five-point scale of background intrusiveness.(BAK) as shown in table 5.2.1
- The overall effect using the scale of the mean opinion score.(OVRL)
  (1 = bad, 2 = poor, 3 = fair, 4 = good, 5 = excellent)

rating	**description**
5	Very natural, no degradation
4	Fairly natural, little degradation
3	Somewhat natural, somewhat degradation
2	Fairly unnatural, fairly degraded
1	Very unnatural, vary degraded

Table 5.1: Scale of Signal Distortion

rating	**description**
5	Not noticeable
4	Somewhat noticeable
3	Noticeable but not intrusive
2	Fairly conspicuous, somewhat intrusive
1	Very conspicuous, very intrusive

Table 5.2: Scale of background intrusiveness

5.3 Summary

Different quality evaluation parameters which are used to compare these algorithms are described in this chapter.

Chapter 6

Simulation Results and Comparison

Database for noisy and clean speech is taken from noisy speech corpus(NOIZEUS). Quality evaluations are also done using NOIZEUS. All algorithm described in this thesis are performed on the speech "read verse out load for pleasure" at different noise environment and different input SNRs.

Given Figure 6.1 shows the time domain of the clean speech signal and spectrogram of it, and Figure 6.2 shows the time domain of noisy speech signal and spectrogram of it. Figure 6.3, Figure 6.4, Figure 6.5, Figure 6.6 and Figure 6.7 shows the time domain representation and spectrogram of Enhanced Speech by Wiener as Method, Enhanced Speech by Wavelet Thresholding Method, Enhanced Speech by log MMSE Method, Enhanced Speech by Spectral Subtractive Method and Enhanced Speech by Multi-band Spectral Subtractive Method respectively.

After passing the noisy speech signal through described algorithms, we get the enhanced speech. Enhanced speech through log MMSE method and Multi-band SS method we can get good quality of speech rather than other algorithms described here. The wavelet thresholding algorithm gives poor performance as observing these enhanced speech signals. As shown in spectrogram, the noise level is reduced after

passing the noisy speech through different speech enhancement algorithm described above.

Enhanced speech quality also depends on the noise environment of the input noisy speech and input SNR of the speech.

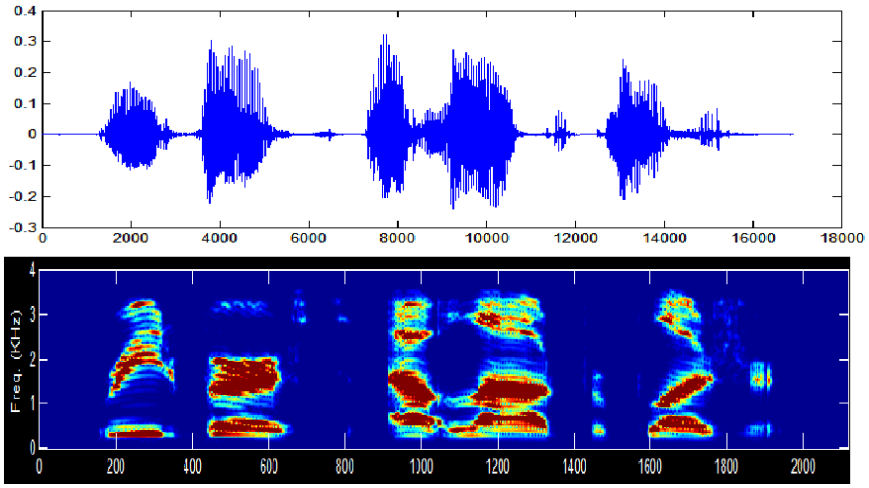

Figure 6.1: Clean Speech

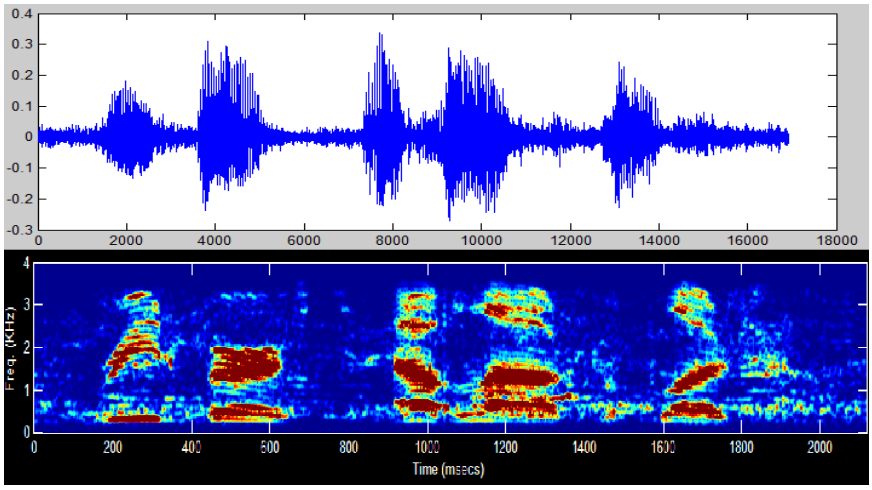

Figure 6.2: Noisy Speech

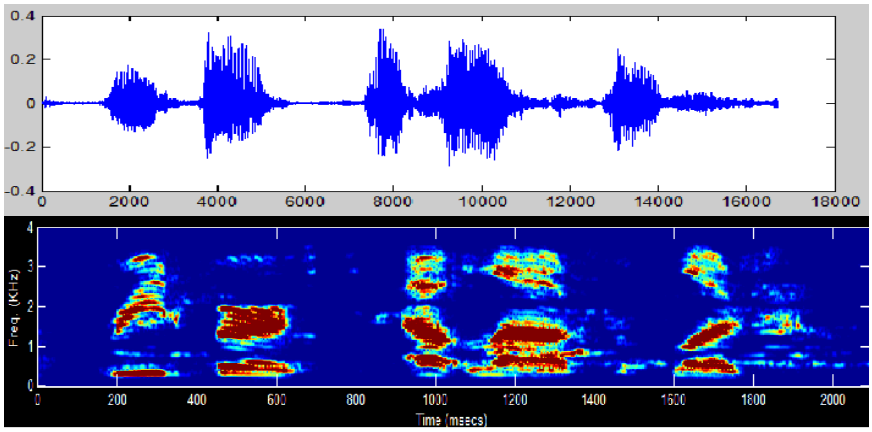

Figure 6.3: Enhanced Speech by Wiener as Method

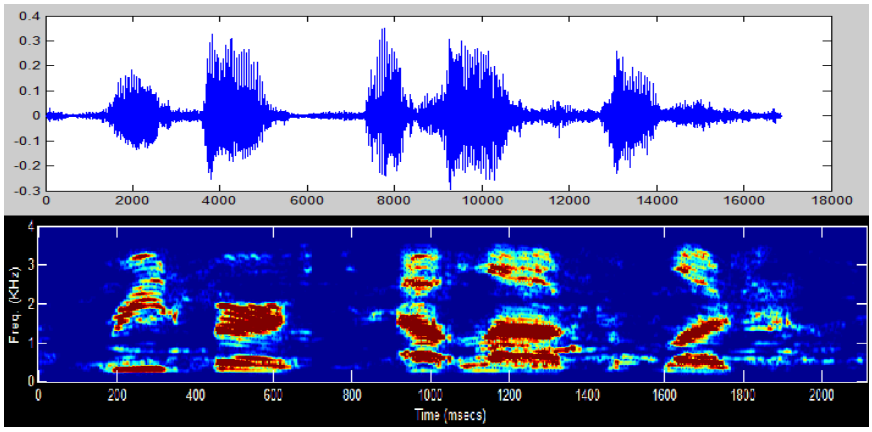

Figure 6.4: Enhanced Speech by Wavelet Thresholding Method

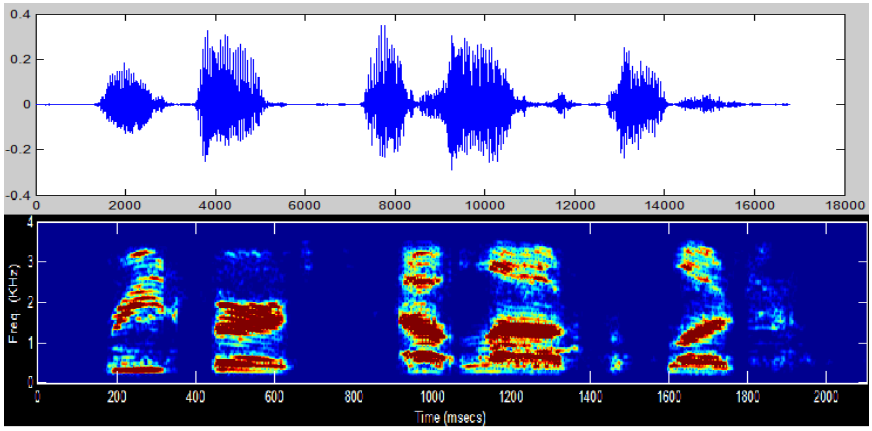

Figure 6.5: Enhanced Speech by log MMSE Method

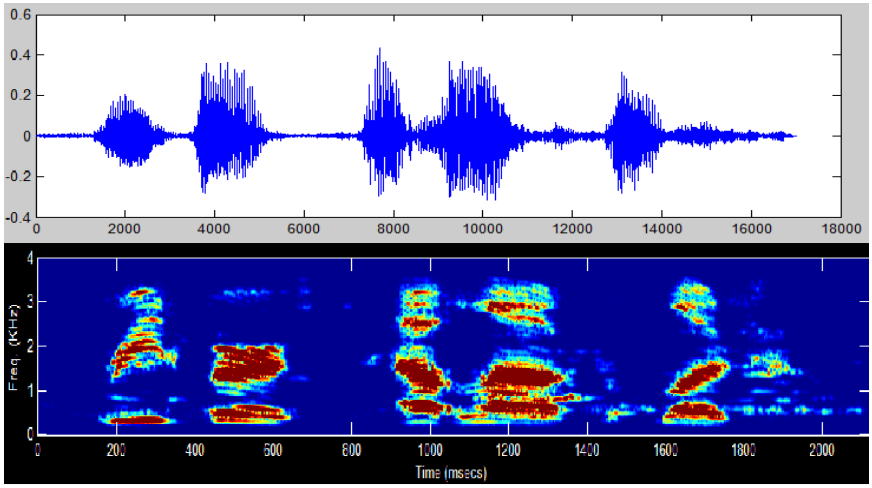

Figure 6.6: Enhanced Speech by Spectral Subtractive Method

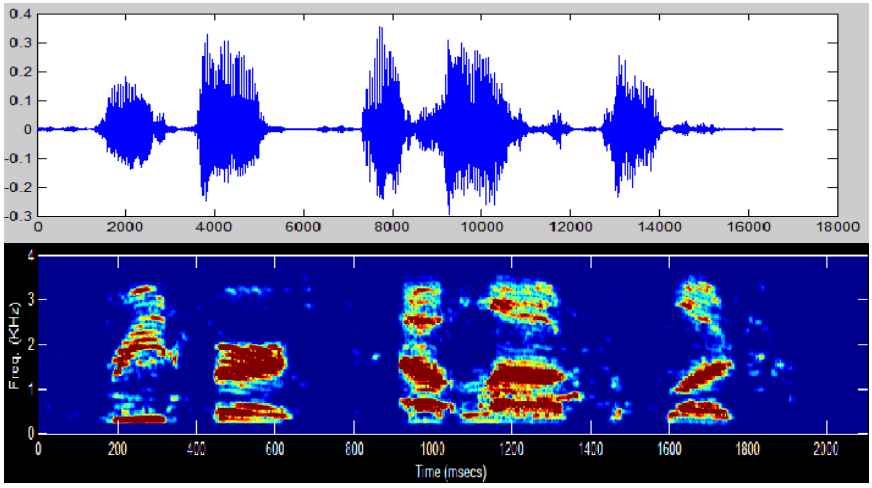

Figure 6.7: Enhanced Speech by Multiband SS Method

Summarization in terms of different quality evaluation parameters, obtained by different algorithms are as in tables given below.

Table 6.1, Table 6.2,Table 6.3and Table 6.3 describes overall quality, and other objective quality parameters like LLR, SNRseg, WSS, PESQ are measured in the babble noise environment at 0 dB, 5 dB, 10 dB, 15 dB of input(noisy) speech signal respectively. As we can see, In terms of overall quality log MMSE and M-band SS algorithms examined performed equally well for most SNR conditions and four types of noise environment.

Table 6.1: Comparison of different quality evaluation parameters in noise environment: Babble (0 dB input SNR)

	Wiener as	**WT**	**SS**	**log MMSE**	**M-band SS**
SIG	1.8476	1	1.8518	2.1216	2.7489
BAK	1.5302	1.5178	1.6603	1.7565	2.062
OVRL	1.6171	1	1.6296	1.875	2.3419
LLR	1.3039	2.1142	1.3585	1.2768	1.0354
SNRseg	-3.0997	-0.3817	-1.7762	-2.6229	-2.4174
WSS	112.0710	84.0341	104.0867	94.0102	62.8168
PESQ	1.8325	1.0377	1.8134	1.9786	2.1338

Table 6.2: Comparison of different quality evaluation parameters in noise environment: Babble (5 dB input SNR)

	Wiener as	**WT**	**SS**	**log MMSE**	**M-band SS**
SIG	2.8286	2.3178	2.6915	2.9137	3.08
BAK	2.1314	2.2758	2.1615	2.2287	2.2887
OVRL	2.3965	2.0223	2.2832	2.4769	2.6289
LLR	0.8509	1.38233	0.9454	0.8807	0.8838
SNRseg	0.0242	1.9280	1.0086	0.1378	-1.0112
WSS	84.181003	54.2374	81.8484	72.4484	57.9538
PESQ	2.2701	1.8828	2.1691	2.2869	2.351689

Table 6.3: Comparison of different quality evaluation parameters in noise environment: Babble (10 dB input SNR)

	Wiener as	**WT**	**SS**	**log MMSE**	**M-band SS**
SIG	3.4269	3.5004	3.3405	3.4702	3.556
BAK	2.5209	2.8672	2.639	2.5922	2.5222
OVRL	2.9011	3.0325	2.8326	2.9322	3.027
LLR	0.6431	0.8046	0.7072	0.6384	0.6530
SNRseg	1.4189	3.9444	3.5657	2.0761	-0.2900
WSS	61.2423	40.1401	59.5918	56.7400	50.2992
PESQ	2.5653	2.6478	2.5066	2.5618	2.6328

Table 6.4: Comparison of different quality evaluation parameters in noise environment: Babble (15 dB input SNR)

	Wiener as	**WT**	**SS**	**log MMSE**	**M-band SS**
SIG	3.8957	3.8078	3.8096	3.9635	4.0169
BAK	2.9495	3.2003	3.0476	3.0113	2.863
OVRL	3.3201	3.3137	3.2404	3.3952	3.4278
LLR	0.5096	0.7097	0.5293	0.4973	0.4901
SNRseg	4.0894	6.6124	6.4660	4.3796	1.4613
WSS	42.9195	30.6794	46.6047	41.9675	35.5065
PESQ	2.8415	2.8545	2.78717	2.9187	2.8985

Table 6.5, Table 6.6,Table 6.7and Table 6.8 describes overall quality, and other objective quality parameters like LLR, SNRseg, WSS, PESQ are measured in the car noise environment at 0 dB, 5 dB, 10 dB, 15 dB of input(noisy) speech signal respectively.

Table 6.5: Comparison of different quality evaluation parameters in noise environment: Car (0 dB input SNR)

	Wiener as	**WT**	**SS**	**log MMSE**	**M-band SS**
SIG	2.4211	1	2.1973	2.694	2.5811
BAK	1.8825	1.7613	1.8636	2.1033	2.003
OVRL	2.0661	1.0874	1.902	2.2878	2.1723
LLR	1.08925	2.2538	1.224	1.0494	1.131
SNRseg	-1.8878	-0.3594	-1.1237	-1.2543	-2.4671
WSS	85.8432	82.466	89.5393	64.2318	57.2494
PESQ	2.025679	1.521344	1.9397	2.0878	1.9356

Table 6.6: Comparison of different quality evaluation parameters in noise environment: Car (5 dB input SNR)

	Wiener as	**WT**	**SS**	**log MMSE**	**M-band SS**
SIG	3.1709	2.6926	2.8348	3.2605	3.0041
BAK	2.3127	2.4743	2.2395	2.4384	2.3184
OVRL	2.6484	2.8371	2.3556	2.7429	2.5432
LLR	0.7403	1.246	0.8694	0.7573	0.9588
SNRseg	0.0732	2.1302	1.4509	0.7554	-0.5351
WSS	62.2963	52.6168	71.2118	54.3564	48.1793
PESQ	2.3224	2.2476	2.1182	2.3814	2.2079

Table 6.7: Comparison of different quality evaluation parameters in noise environment: Car (10 dB input SNR)

	Wiener as	**WT**	**SS**	**log MMSE**	**M-band SS**
SIG	3.1909	2.3496	3.0254	3.4676	3.3701
BAK	2.5498	2.5263	2.5417	2.7288	2.5631
OVRL	2.7084	2.0996	2.4915	2.9298	2.8488
LLR	0.7932	1.47	0.8369	0.7154	0.7966
SNRseg	2.9057	4.3856	4.3089	3.3062	0.8615
WSS	59.8476	46.1187	53.1362	43.6837	39.5642
PESQ	2.4092	1.964	2.1092	2.4942	2.4095

Table 6.8: Comparison of different quality evaluation parameters in noise environment: Car (15 dB input SNR)

	Wiener as	**WT**	**SS**	**log MMSE**	**M-band SS**
SIG	4.0713	3.3904	4.0087	4.2023	3.9072
BAK	3.1354	3.1271	3.2369	3.2429	2.9273
OVRL	3.4886	2.9907	3.4526	3.5906	3.3755
LLR	0.4386	0.9579	0.5001	0.4222	0.6249
SNRseg	5.7525	7.458	7.3583	6.0488	2.1258
WSS	41.9768	34.4122	40.9152	30.2984	31.1042
PESQ	2.9974	2.6447	2.9827	3.0122	2.8809

Table 6.9, Table 6.10,Table 6.11and Table 6.12 describes overall quality, and other objective quality parameters like LLR, SNRseg, WSS, PESQ are measured in the train noise environment at 0 dB, 5 dB, 10 dB, 15 dB of input(noisy) speech signal respectively.

Table 6.9: Comparison of different quality evaluation parameters in noise environment: Train (0 dB input SNR)

	Wiener as	**WT**	**SS**	**log MMSE**	**M-band SS**
SIG	2.027	1.0098	2.1239	2.5094	2.4636
BAK	1.8458	1.706	1.8252	2.0608	2.0155
OVRL	1.8371	1	1.8325	2.1548	2.1375
LLR	1.3989	2.1092	1.2927	1.1856	1.2786
SNRseg	-1.5348	-0.1059	-1.6848	-1.3675	-2.6998
WSS	91.8601	70.2854	83.4481	63.4637	56.2202
PESQ	1.9905	1.1937	1.8441	2.0025	1.9773

Table 6.10: Comparison of different quality evaluation parameters in noise environment: Train (5 dB input SNR)

	Wiener as	**WT**	**SS**	**log MMSE**	**M-band SS**
SIG	2.6204	1.7923	2.4832	2.8457	2.8239
BAK	2.1648	2.2536	2.1372	2.337	2.2424
OVRL	2.2416	1.7558	2.0989	2.4373	2.3811
LLR	1.0855	1.8669	1.1311	1.0451	1.0494
SNRseg	0.1716	1.9061	0.8686	0.6218	-0.6472
WSS	62.4192	57.7351	68.6495	54.3959	48.2593
PESQ	2.0898	1.8904	1.9436	2.1852	2.0648

Table 6.11: Comparison of different quality evaluation parameters in noise environment: Train (10 dB input SNR)

	Wiener as	**WT**	**SS**	**log MMSE**	**M-band SS**
SIG	2.9669	2.3935	3.0439	3.2026	3.2387
BAK	2.5048	2.7022	2.66	2.7006	2.589
OVRL	2.5515	2.2601	2.6516	2.7464	2.8033
LLR	0.9834	1.5992	0.9758	0.9269	0.9501
SNRseg	2.5577	4.9566	4.1481	3.4615	0.9404
WSS	55.5305	44.2343	56.2882	41.3531	39.085
PESQ	2.2977	2.2292	2.4239	2.3807	2.4464

Table 6.12: Comparison of different quality evaluation parameters in noise environment: Train (15 dB input SNR)

	Wiener as	**WT**	**SS**	**log MMSE**	**M-band SS**
SIG	3.644	3.115	3.6336	3.8588	3.6804
BAK	3.0292	3.1635	3.142	3.1476	2.8637
OVRL	3.1767	2.887	3.183	3.3469	3.2052
LLR	0.7508	1.2761	0.7693	0.6616	0.7776
SNRseg	5.4062	7.4071	7.0876	5.7546	1.9935
WSS	39.9005	32.6734	40.7571	31.9835	31.3948
PESQ	2.7905	2.7096	2.8176	2.8764	2.7682

Table 6.13, Table 6.14,Table 6.15and Table 6.16 describes overall quality, and other objective quality parameters like LLR, SNRseg, WSS, PESQ are measured in the restaurant noise environment at 0 dB, 5 dB, 10 dB, 15 dB of input(noisy) speech signal respectively.

Table 6.13: Comparison of different quality evaluation parameters in noise environment: Restaurant (0 dB input SNR)

	Wiener as	**WT**	**SS**	**log MMSE**	**M-band SS**
SIG	2.1918	1.2569	1.9913	2.3782	2.3354
BAK	1.5583	1.8278	1.6656	1.807	1.7613
OVRL	1.7133	1.2824	1.7	2.0162	1.8396
LLR	0.9758	2.0877	1.2502	1.0914	1.0745
SNRseg	-2.7028	-0.997	-2.0356	-2.5767	-2.4846
WSS	96.0101	67.901	99.5929	86.9288	66.3712
PESQ	1.6038	1.5312	1.7929	1.9745	1.5657

Table 6.14: Comparison of different quality evaluation parameters in noise environment: Restaurant (5 dB input SNR)

	Wiener as	**WT**	**SS**	**log MMSE**	**M-band SS**
SIG	2.7515	2.2167	2.6154	2.751	2.9943
BAK	1.9951	2.2185	2.082	2.1882	2.2371
OVRL	2.2906	1.9579	2.1905	2.327	2.5281
LLR	0.8469	1.4348	0.9762	0.9631	0.9268
SNRseg	-1.117	1.6003	0.3122	-0.4762	-1.2984
WSS	84.2808	59.3029	78.325	71.7112	53.2843
PESQ	2.1369	1.8803	2.043	2.1467	2.2132

Table 6.15: Comparison of different quality evaluation parameters in noise environment: Restaurant (10 dB input SNR)

	Wiener as	**WT**	**SS**	**log MMSE**	**M-band SS**
SIG	3.3468	2.9677	3.663	3.6638	3.7196
BAK	2.6286	2.929	2.8312	2.7679	2.7512
OVRL	2.8959	2.7411	3.1134	3.124	3.2249
LLR	0.7307	1.2815	0.5613	0.6035	0.6663
SNRseg	2.99	5.4503	4.4058	2.8972	1.0067
WSS	66.4655	41.877	53.9043	49.4399	41.2314
PESQ	2.66	2.6041	2.7132	2.7144	2.8083

Table 6.16: Comparison of different quality evaluation parameters in noise environment: Restaurant (15 dB input SNR)

	Wiener as	**WT**	**SS**	**log MMSE**	**M-band SS**
SIG	3.6717	3.3624	3.6377	3.785	3.8686
BAK	2.848	3.1282	3.0185	2.9356	2.8098
OVRL	3.1153	3.0024	3.1149	3.2052	3.2958
LLR	0.5282	1.0132	0.6056	0.5137	0.5413
SNRseg	5.0253	7.1341	7.1653	5.1978	1.8108
WSS	58.5581	35.0217	54.1965	48.8109	39.8643
PESQ	2.735	2.6985	2.7458	2.7527	2.8049

Comparison of Different Algorithm

1= bad, 2 = poor, 3 = fair, 4 = good, 5 = excellent

Figure 6.8, Figure 6.9, Figure 6.10, Figure 6.11 shows the mean scores for OVRL scales for speech processed by five different speech enhancement algorithms evaluated in 4 type of background noise (babble, car, train, restaurant) and at four different SNR levels (0,5,10and 15 dB) respectively.

In terms of overall quality, log MMSE and Multi-band spectral subtraction algorithm performs best and equally well. Wavelet thresholding algorithm performs extremely poor among all these algorithm at 0dB SNR as shown in Figure 6.8. At 15 dB SNR almost all algorithm performs well and good as shown in Figure 6.11. Figure 6.12 shows comparison of different speech enhancement algorithm in terms of objective parameter at 0 dB as in [1]

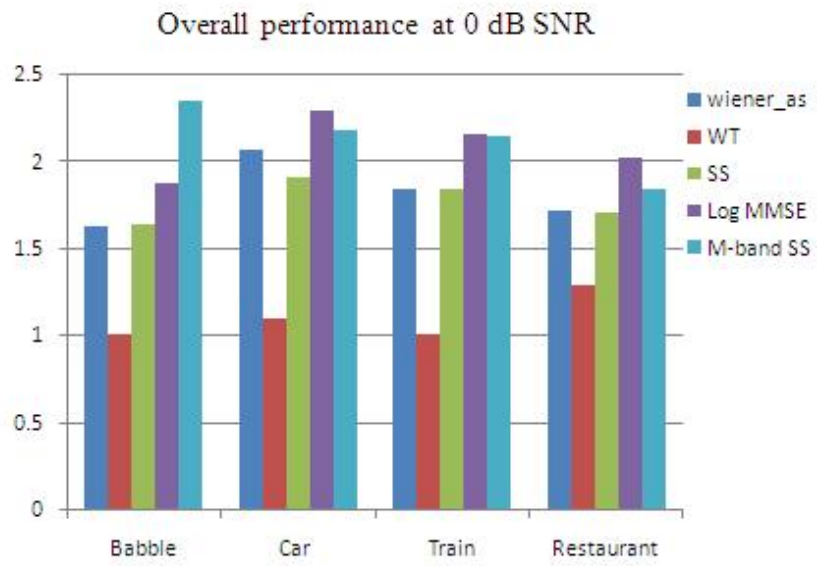

Figure 6.8: Overall performance of different speech enhancement algorithm at 0dB input SNR in different Noise environment

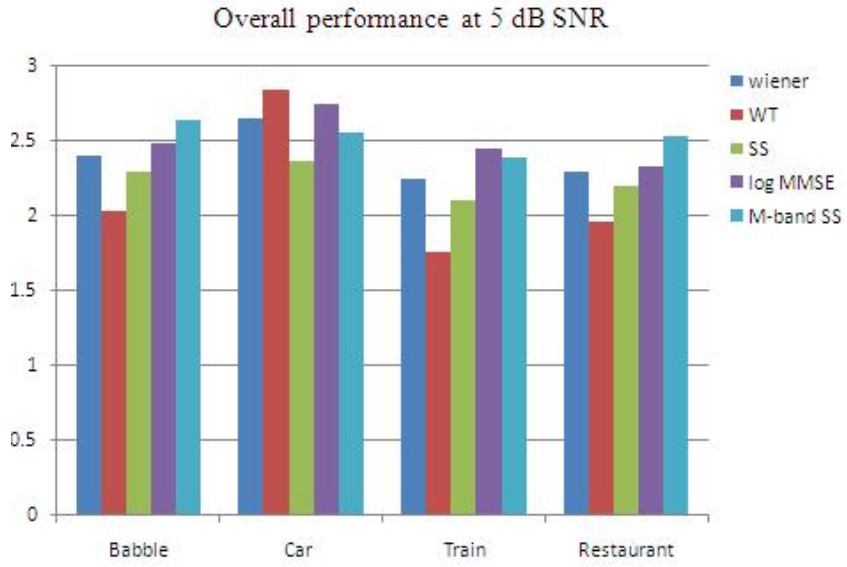

Figure 6.9: Overall performance of different speech enhancement algorithm at 5dB input SNR in different Noise environment

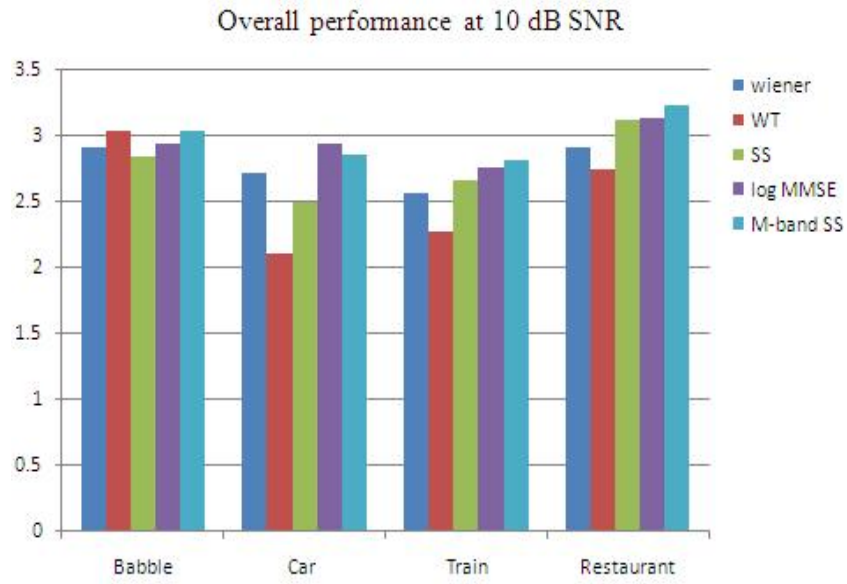

Figure 6.10: Overall performance of different speech enhancement algorithm at 10dB input SNR in different Noise environment

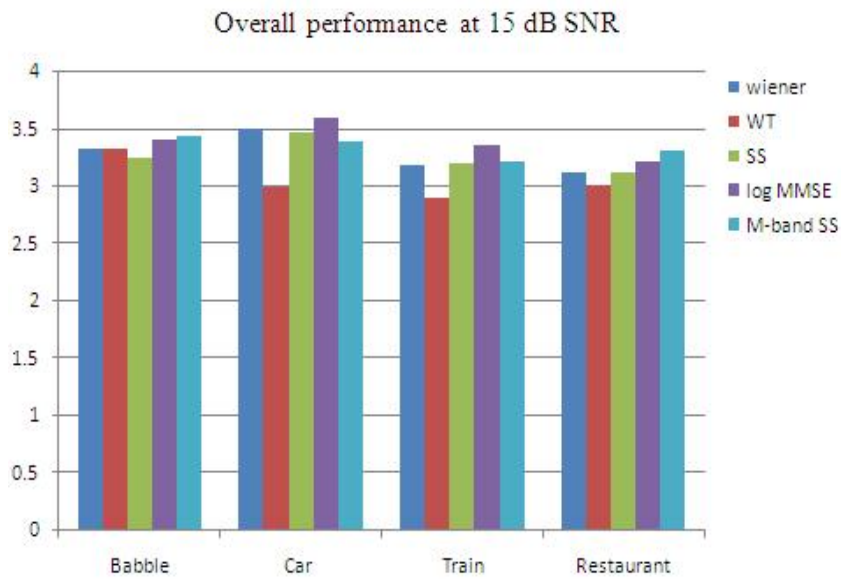

Figure 6.11: Overall performance of different speech enhancement algorithm at 15dB input SNR in different Noise environment

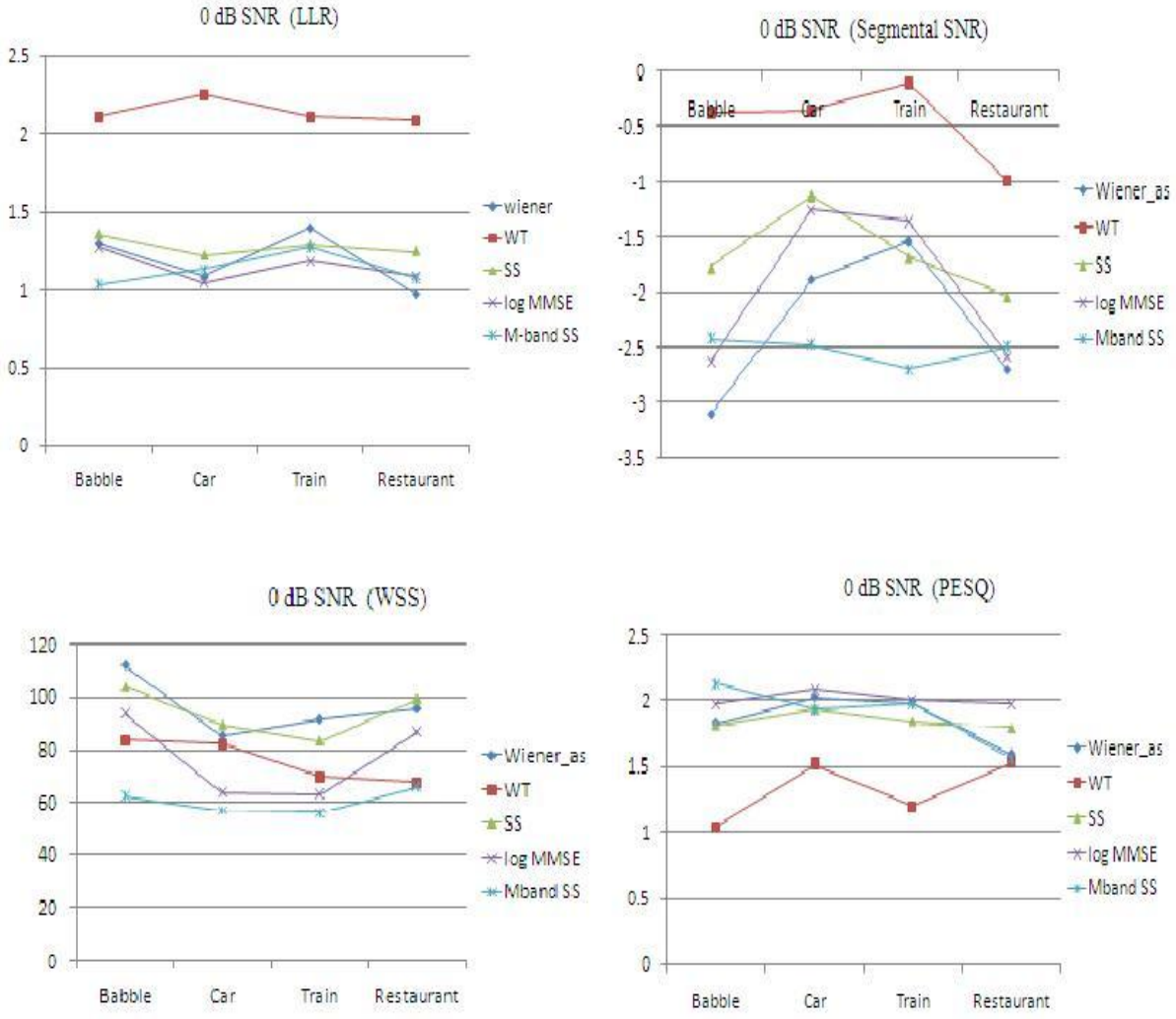

Figure 6.12: Comparison of different speech enhancement algorithm in terms of objective parameter at 0 dB

Chapter 7

Conclusion

7.1 Conclusion

In this thesis, we discussed five different speech enhancement algorithm and applied them on the noisy speech having different background noise and at different SNR. The Wiener-as method which relies on the decision-directed approach to estimate the a priori SNR, while the multi-band spectral subtraction algorithm does not make use of a priori SNR information.

From the results we can see that,In terms of overall quality, the following algorithm performed the best: log MMSE and multi-band spectral subtraction. These algorithms also yielded the lowest speech distortion. The wiener-as algorithm also performed well in some conditions. The VAD algorithms for updating the noise spectrum produce significant improvement in performance.

In terms of low computational complexity and good performance, the two winners are the wiener as algorithm and multi-band spectral subtraction algorithm. The multi-band spectral subtraction algorithm performed as well as the log MMSE algorithm of statistical model based algorithm in nearly all conditions.

7.2 Future Scope

Future work will include the development of these algorithm by changing the criteria and parameters used in these algorithm. Also by changing the transform applied in these algorithms we can get results and can compare it with the present results.

References

[1] Yi Hu and P. C. loizou. *Speech Enhancement Techniques Theroy and Practice*. July 2007.

[2] Kotta Manohar and Preeti Rao. Speech enhancement in non-stationary noise environements using noise properties. *Science Direct Speech Communication*, August 2005.

[3] Dr. P. loizou, Dr. mohammad Saquib, and Dr. john. Modified spectral subtraction method combined with perceptual weighting for speech enhancement. *University of texas at dallas*, August 2002.

[4] Ephraim Y. and Van Trees. A signal subspace approach for speech enhancement. *IEEE Transactions on Speech and audio Processing*, july 1995.

[5] P. C. Loizou, Arthur Lobo, and Yi Hu. Subspace algorithm for noise reduction in cochlear implants. *Journal Acoust Soc. Am NIH Public Access*, November 2005.

[6] Levent M. Arslan. Modified wiener filtering. *Science Direct Signal Processing*, May 2005.

[7] Sundarrajan Rangachari and Philipos C. Loizou. A noise-estimation algorithm for highly non-stationary environments. *Science Direct Speech Communication*, August 2005.

[8] P. C. Loizou and Yi Hu. A generalized subspace approach for enhancing speech corrupted with colored noise. *IEEE Transactions on Speech and Audio Processing*, July 2003.

[9] Pans France The author is with TELECOM Paris, Dtpartement SIGNAL. Elimination of the musical noise phenomenon with the ephraim and malah noise suppressor. *IEEE transactions on speech and audio processing*, April 1994.

[10] Jianjun Lei, Jiachen Yang, Jian Wang, and Zhen Yang. A robust voice activity detection algorithm in nonstationary noise. *International Conference on Industrial and Information Systems*, 2009.

[11] H. S. Kim, Y. M. Cho, and H. J. Kim. Speech enhancement via mel-scale wiener filtering with a frequency-wise voice activity detector. *Journal of Mechanical Science and Tehnology*, March 2007.

[12] Huijun Ding, I. Y. Soon, S. N. Koh, and C. K. Yeo. A spectral filtering method based on hybried wiener filter for speech enhancement. *Science Direct Speech Communication*, September 2008.

[13] Ing Yann Soon and Soo Ngee Koh. Speech enhancement using 2-d fourier transform. *IEEE transactions on speech and audio processing*, November 2003.

[14] Mohammed Bahoura and Jean Rouat. Wavelet speech enhancement based on the teagerenergy operator. *ERMETIS, DSA, Canada*.

[15] V. Balakrishnan, Nash Borges, and Luke Parchment. Wavelet denoising and speech enhancement. *The Johns Hopkins University, Baltimore, MD*, July 2005.

[16] Saeed Ayat, M.T. Manzuri-Shalmani, and Roohollah Dianat. An improved wavelet-based speech enhancement by using speech signal features. *Computers and Electrical Engineering, speech and audio processing*, January 2006.

[17] Yi Hu and Philipos C. Loizou. Speech enhancement based on wavelet thresholding the multitaper spectrum. *IEEE transactions on speech and audio processing*, January 2004.

[18] Jong Kwan Lee and Chang D. Yoo. Wavelet speech enhancement based on voiced/unvoiced decision. *The 32nd International Congress and Exposition on Noise Control Engineering Jeju International Convention Center, Seogwipo, Korea*, August 2003.

[19] Mohammed Bahoura and Jean Rouat. Wavelet speech enhancement based on timescale adaptation. *Speech Communication*, June 2006.

[20] S.MANIKANDAN. Speech enhancement based on wavelet denoising. *acadamic open internet journal*, 2007.

[21] B.Mohan Kumar, M.Arun, and M.G.Sumithra. A novel approach in wavelet based signal enhancement using labview. *Department of ECE, Amrita Vishwa Vidyapeetham, Coimbatore*, December 2006.

[22] Ephraim Y.and Malah D. Speech enhancement using a minimum mean square error log-spectral amplitude estimator. *IEEE Transactions on Acoustics, Speech, Signal Processing*, april 1985.

[23] Md. Kamrul Hasan, Sayeef Salahuddin, and M. Rezwan Khan. A modified a priori snr for speech enhancement using spectral subtraction rules. *IEEE signal processing latters*, April 2004.

[24] P. C. Loizou. Speech enhancement based on perceptually motivated bayesian estimators of the speech magnitude spectrum. *IEEE Transactions on Speech and Audio Processing*, Sept 2005.

[25] Sunil D. Kamath and Philipos C. Loizou. A multi-band spectral subtraction method for enhancing speech corrupted by colored noise. *speech and audio processing*, March 2007.

[26] Yi Hu and P. C. Loyzou. Subjective comparision and evaluation of speech enhancement algorithms. *Speech Communication NIH Public Access*, July 2007.

[27] Shihua Wang, Andrew Sekey, and Allen Gersho. An objective measure for predicting subjective quality of speech coders. *IEEE journal on selected areas in communications*, June 1992.

[28] Stavros Ntalampiras, Todor Ganchev, Ilyas Potamitis, and Nikos Fakotakis. Objective comparison of speech enhancement algorithms under real world conditions. *Speech Communication*, 2008.