Master Thesis
Signal Processing
Thesis no.
December 2011

Single Channel Speech Enhancement Using Spectral Subtraction Based on Minimum Statistics

Md. Zameari Islam
G.M. Sabil Sajjad

This thesis is presented as part of Degree of Master of Science in Electrical Engineering with emphasis on Signal Processing

Blekinge Institute of Technology
December 2011

School of Engineering
Department of Electrical Engineering
Blekinge Institute of Technology, Sweden
Supervisor: Dr. Benny Sällberg
Co-supervisor: Dr. Nedelko Grbic, Associate Professor
Examiner: Dr. Nedelko Grbic, Associate Professor

Contact Information:

Authors:

Md. Zameari Islam

email: zameariislam@yahoo.com

G.M. Sabil Sajjad

email: gmsabilsajjad@gmail.com

Supervisor:

Dr. Benny Sällberg

Department of Signal Processing

Blekinge Institute of Technology

SE-371 79 Karlskrona, Sweden

tel +46-455-385000

fax +46-708-178744

email: benny.sallberg @bth.se

Co-supervisor & Examiner:

Dr. Nedelko Grbic, Associate Professor

Department of Signal Processing

Blekinge Institute of Technology

SE-371 79 Karlskrona, Sweden

tel +46-455-385000

fax +46-708-178744

email: nedelko.grbic@bth.se

Abstract

Speech is an elementary source of human interaction. The quality and intelligibility of speech signals during communication are generally degraded by the surrounding noise. Corrupted speech signals need therefore to be enhanced to improve quality and intelligibility. In the field of speech processing, much effort has been devoted to develop speech enhancement techniques in order to restore the speech signal by reducing the amount of disturbing noise. This thesis focuses on a single channel speech enhancement technique that performs noise reduction by spectral subtraction based on minimum statistics. Minimum statistics means that the power spectrum of the non-stationary noise signal is estimated by finding the minimum values of a smoothed power spectrum of the noisy speech signal and, thus, circumvents the speech activity detection problem. The performance of the spectral subtraction method is evaluated using single channel speech data and for a wide range of noise types with various noise levels. This evaluation is used in order to find optimum method parameter values, thereby improving this algorithm to make it more appropriate for speech communication purposes.

The system is implemented in MATLAB and validated by considering different performance measure and for different Signal to Noise Ratio Improvement (SNRI) and Spectral Distortion (SD). The SNRI and SD were calculated for different filter bank settings such as different number of subbands and for different decimation and interpolation ratios. The method provides efficient speech enhancement in terms of SNRI and SD performance measures.

To our parents

Acknowledgement

First and the foremost we would like to thank to our thesis supervisor, Dr. Benny Sällberg for giving us such an interesting thesis topic to work with. We are very much grateful to him for his thorough guidance and all out support throughout our thesis work.

We are also thankful to our co-supervisor Dr. Nedelko Grbic for his persistent help during the whole thesis work. He guided us throughout our work in a very nice way.

We would like to thank BTH for providing us with a nice educational environment where we were able to gain valuable knowledge to move forward with our project work.

Finally, we would like to thank our family members for their moral and financial support throughout our educational career. We would also like to thank all of our friends and staff at BTH.

Thank you all.

Zameari

Sabil

Contents

Abstract iii

Acknowledgement v

List of Figures viii

List of Tables ix

1 Introduction 1

1.1 Introduction 1

1.2 Outline 2

2 Background and Related Work 3

2.1 Introduction 3

2.2 Noise Analysis 3

2.3 Speech Enhancement Methods 5

2.3.1 Single Channel Speech Enhancement 5

2.3.2 Multichannel Speech Enhancement 6

2.4 Spectral Subtraction 8

2.4.1 Spectral Subtraction Basic 8

3 Spectral Subtraction Based on Minimum Statistics 11

3.1 Introduction 11

3.2 Description of Algorithms 12

3.2.1 Noise Power Estimation 16

3.2.1.1 Subband Signal Power Estimation 16

3.2.1.2 Subband Noise Power Estimation 17

3.2.2 SNR and Oversubtraction Factor Calculation 18

3.2.3 Subtraction Rule 18

3.2.4 Reconstruction in Time Domain 19

4 Implementation and Results 21

4.1 Introduction 21

4.2 Implementation 21

4.3 Results 28

4.4 Computational Complexity 42

5 Conclusion 44

References 46

List of Figures

2.1 Single Channel Speech Enhancement System 5

2.2 Basic block Diagram of Spectral Subtraction 8

3.1 Spectral subtraction based on Minimum Statistics 12

3.2 Basic Filter Bank Diagram 13

3.3 Framing of the Input Signal 15

3.4 Overlap Add in Time Domain 20

4.1 Experimental Setup 22

4.2 Average SNRI by changing α, γ and $subf$ 24

4.3 Average SD by changing α, γ and $subf$ 24

4.4 SNRI by changing γ value 25

4.5 SD by changing γ value 25

4.6 Power Spectral Density of Car Noise 26

4.7 Power Spectral Density of Factory Noise 26

4.8 Power Spectral Density of Wind Noise 27

4.9 Power Spectral Density of Cafeteria Noise 27

4.10 Average SNRI Using Male Speech Signal with 75% Overlap 39

4.11 Average SNRI Using Male Speech Signal with 50% Overlap 39

4.12 Average SNRI Using Female Speech Signal with 75% Overlap 40

4.13 Average SNRI Using Female Speech Signal with 50% Overlap 40

4.14 Average Spectral Distortion for Male Speech Signal 41

4.15 Average Spectral Distortion for Female Speech Signal 41

List of Tables

Table 4.1: SNRI and SD for Male Speech Signal with Gaussian Noise at 75% Overlap 29

Table 4.2: SNRI and SD for Female Speech Signal with Gaussian Noise at 75% Overlap 29

Table 4.3: SNRI and SD for Male Speech Signal with Gaussian Noise at 50% overlap 30

Table 4.4: SNRI and SD for Female Speech Signal with Gaussian Noise at 50% Overlap 30

Table 4.5: SNRI and SD for Male Speech Signal with Car Noise at 75% Overlap 31

Table 4.6: SNRI and SD for Female Speech Signal with Car Noise at 75% Overlap 31

Table 4.7: SNRI and SD for Male Speech Signal with Car Noise at 50% Overlap 32

Table 4.8: SNRI and SD for Female Speech Signal with Car Noise at 50% Overlap 32

Table 4.9: SNRI and SD for Male Speech Signal with Factory Noise at 75% Overlap 33

Table 4.10: SNRI and SD for Female Speech Signal with Factory Noise at 75% Overlap 33

Table 4.11: SNRI and SD for Male Speech Signal with Factory Noise at 50% Overlap 34

Table 4.12: SNRI and SD for Female Speech Signal with Factory Noise at 50% Overlap 34

Table 4.13: SNRI and SD for Male Speech Signal with Wind Noise at 75% Overlap 35

Table 4.14: SNRI and SD for Female Speech Signal with Wind Noise at 75% Overlap 35

Table 4.15: SNRI and SD for Male Speech Signal with Wind Noise at 50% Overlap 36

Table 4.16: SNRI and SD for Female Speech Signal with Wind Noise at 50% Overlap 36

Table 4.17: SNRI and SD for Male Speech Signal with Cafeteria Noise at 75% Overlap 37

Table 4.18: SNRI and SD for Female Speech Signal with Cafeteria Noise at 75% Overlap 37

Table 4.19: SNRI and SD for Male Speech Signal with Cafeteria Noise at 50% Overlap 38

Table 4.20: SNRI and SD for Female Speech Signal with Cafeteria Noise at 50% Overlap 38

Table 4.21 Computation Complexity of SSBMS Algorithm 43

Chapter 1

Introduction

1.1 Introduction

In today's technological era speech is the most important way of communication that began with fixed land-line telephony systems. In all forms of speech communication systems such as cellular phones, maintaining the speech quality and intelligibility in information exchange is the main challenge for the researchers. The performance of these systems in real-life applications dramatically degraded due to the presence of surrounding noise such as background noise, babble noise, impulse noise, musical noise and car noise causing distorted information exchange. The success of these innovative systems depends on the restoration of desired speech signal from the mixture of speech and noise and remains main goals in speech processing research.

Many algorithms have been introduced to improve the perceptual quality of the speech signals from the corrupted input signals in communication systems [5] [6] [8] [12]. It is generally difficult to restore desired signal without distorting speech signal and the performance is limited by the trade-off between speech distortion and noise reduction. The most common scenario is the single channel system [24] where noise and speech come from the individual sources and a microphone records speech and noise, and it is the difficult situation to handle because, in recorded signal, speech and noise are correlated with each other. The computational complexity and cost of implementation in real-time applications such as mobile communications, hearing aids, intelligent hearing protectors and so forth is an important issue during proposed a speech enhancement algorithm. The spectral subtraction is one of the ways for speech enhancement. The spectral subtraction algorithm estimates the noise power spectrum from the noisy speech power spectrum and then, estimates the clean speech power spectrum by subtracting this noise power spectrum from the noisy speech power spectrum. Since, last few decades many researches have been carried out on the spectral subtraction based methods because of its simplicity and ease of implementation on portable devices such as mobile communications [25].

In this thesis, we see the performance of Spectral Subtraction Based on Minimum Statistics (SSBMS) algorithm in different noisy environment with various noise levels by changing the number of subband values as well as its method parameter values and find out the optimum values for which the algorithm gives the better SNRI and less SD. This method uses minimum statistics that eliminate the problem of the speech activity detector, gives a superior performance as compared to the conventional method of power spectral subtraction and decreases the residual noise [12].

1.2 Outline

The thesis report is divided into five chapters. The remaining paper is organized as follows. Chapter 2 provides information about the speech enhancement techniques in both single and multichannel. It further introduces some noise characteristics and brief discussion on spectral subtraction. The theory behind the SSBMS algorithm is presented in chapter 3. Chapter 4 provides both the implementation of the algorithm and results. Finally, in chapter 5 the thesis is concluded and provides future research direction on SSBMS algorithm.

Chapter 2

Background and Related Work

2.1 Introduction

In speech communication system, noise removed from corrupted speech signal has been a big challenge for the researchers since last few decades [26]. Many algorithms have been proposed that aimed at improvement in intelligibility, clarity and overall perceptual quality of degraded speech signal. Noise suppression and speech enhancement has many applications. Some of the important applications among these are as follows:

- Mobile communication.
- air-ground communication
- ground-air communication
- Emergency equipment like elevator, SOS alarm, vehicular emergency telephones.
- Teleconferencing systems
- Intelligent Hearing Protectors
- Hearing aids
- Speech recognition in noisy environments
- VoIP

2.2 Noise Analysis

The problem of removing the noise poses a difficulty due to the random nature of the noise and the intrinsic complexities of speech [27]. So it is necessary to understand the noise characteristics to get the better performance from various speech enhancement methods. One method may perform well with one type of noise but the same may not perform well with different type of noise, so it is necessary to experiment on the method with different types of noise. Noise characteristics are dependent on the statistical properties of the noise. Based on the nature and properties of the noise we can generally classify the noise into the following categories.

Background Noise: In acoustical engineering, background noise is the random signals and come from all sources that are undesired. Background noise is additive noise that is normally uncorrelated with the speech signal and occurs in the different communication environment like traffic noise, crowded city streets, electrical and mechanical equipment noise, industrial environment, atmosphere conditions, etc.

Babble noise: Babble noise in encountered whenever a crowd or group of people are talking together simultaneously (i.e. in a cafeteria, crowded classroom, party place), which has the characteristics and frequency range very close to the desired speech signal [1]. This phenomenon is also known as ‘’cocktail party effect’’.

Impulse noise: Impulse noise is a high energy noise that generates almost instantaneous sharp sounds like slamming of doors, clicks and pops.

Non-additive noise: It occurs due to non-linear behaviour of microphones and speakers, e.g., Lombard's effect due to speaker stress [2]. This effect is introduced when speech is produced in the presence of noise since the speaker has a trend to increase his vocal effort [3]. Due to this effect, the speech spectral properties are changed continuously compared to clean speech.

Convolutive noise: This type of noise convolves with the signal in a time domain, e.g. changes in speech signal due to changes in environment or changes in microphones, etc. It is usually difficult to work with it as compared to additive noise.

Some of the other types of noises are correlated noise (reverberations and echoes), multiplicative noise (signal distortion due to fading), etc.

2.3 Speech Enhancement Methods

The term speech enhancement refers to methods aiming at recovering speech signal from a noisy observation. There are many ways to categorize speech enhancement algorithms. Each method has several specializations that are based on certain assumptions and constraints that depend on the distinct application and the environment scenarios. Therefore, it is almost impossible for a specific algorithm to perform optimally across all noise types.

The noise reduction systems generally can be classified based on the number of input channels (one/multiple), domain of processing (time/frequency/spatial) and type of algorithm (non-adaptive/adaptive) [4][5][6][7]. The speech enhancement techniques can be divided into two broad classes based on single-microphone speech enhancement and multi-microphone speech enhancement techniques.

2.3.1 Single Channel Speech Enhancement

This algorithm estimates the clean speech signal from the noisy speech signal which is available in a single channel provided by one microphone, shown in Fig.1.

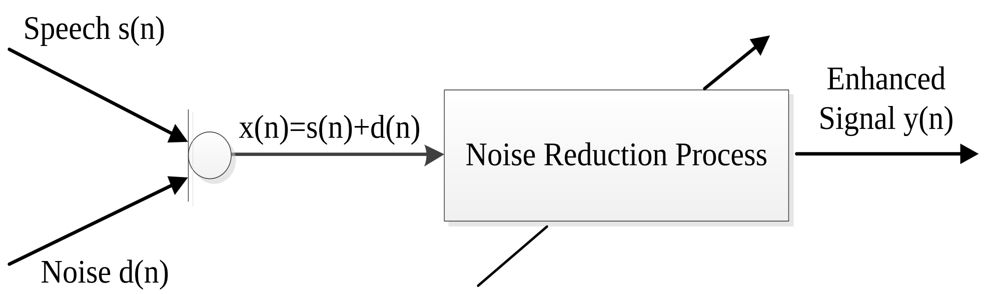

Figure 2.1 Single Channel Speech Enhancement System

Most of the speech enhancement algorithms are based on this technique [28] and mostly applied in real time applications, for example, mobile communication, intelligent hearing protectors, hearing aids and many more. Some proposed algorithms for single channel speech enhancement are [8]:

- Short time spectrum based algorithms
- Speech separation algorithms
- Statistical model based algorithms
- Hearing model based algorithms
- Wavelet algorithm

These methods are easy to build up, since these have less computational complexity and in addition, these have more constraints than multi-channel systems. In general single channel systems constitute by depending on different statistics of speech and unwanted noise that, work in most difficult situations where no prior knowledge of noise is available. The behaviour of these methods depends on Signal to Noise Ratios (SNR) and the features of the noise. Usually the methods assume that the noise is stationary when speech is active. They normally allow non-stationary noise between speech activity periods but in reality when the noise is non-stationary, the performance is dramatically decreased.

2.3.2 Multichannel Speech Enhancement

Multi-microphone method uses multiple signals to enhance the speech quality coming from more than one microphone. These methods usually perform better in very low SNR and non-stationary noise than single channel. However, multi-microphone systems are more complex, since they have fewer constraints than single-microphone systems and are often difficult to carry out due to the equipment size limitation as these need minimum distance among the microphones to set up. These methods use spatiotemporal filtering or beam forming algorithms, which are given below.

- Adaptive Noise Cancellation (ANC)
- Blind Source Separation (BSS)
- Delay and Sum Beam forming (DSB).
- Linear Constraint Minimum Variance (LCMV)
- Generalized Sidelobe Cancellation (GSC)

The Adaptive Noise Cancellation (ANC) is a well known speech enhancement technique that uses a primary channel containing corrupted signal and a reference channel containing noise correlated with primary channel noise to cancel highly correlated noise [9]. The reference input is filtered by an adaptive algorithm and subtracted from primary input signal in order to extract the desired speech signal. This algorithm has some leakage problem; if the primary signal is leaked into the reference signal then some original speech is cancelled and thus the speech quality decreases [10].
The Blind Source Separation (BSS) is used to separate a set of signals from a mixed signal and it is designed in such a way that it only performs in the criteria when speech and noise are independent [11].

The Delay and Sum beam forming (DSB) is the simplest algorithm for beam forming and its efficiency depends on the number of microphone used in a system. The Linear Constrained Minimum Variance (LCMV) algorithm is another kind of beam forming that takes the present signal and delayed samples to enhance the speech quality which may give the better result than DSB algorithm. The Generalized Sidelobe Cancellation (GSC) algorithm uses the microphone array for speech enhancement, and it is very attractive due to its efficient implementation.

In this thesis, we worked on the single channel spectral subtraction based speech enhancement method. Many researches have been carried out for many decades on this method, so the rest of the discussion of this chapter would be on the basics of spectral subtraction.

2.4 Spectral Subtraction

A basic block diagram of the spectral subtraction is given in figure 2.2 [5]. The noisy speech signal $x(n)$ is the input to the system. Initially, the input signal $x(n)$is segmented into many short frames by the window function, and then DFT filter bank is applied to each of the frame for analysis and synthesis. The DFT signals are converted into phase and amplitude. The square magnitude $|X(\omega)|^2$ has been modified by using different noise estimation and the noise subtraction rule. This modified amplitude is added with the phase and then inverse DFT with overlap add is applied to this signal to get the enhanced signal $y(n)$.

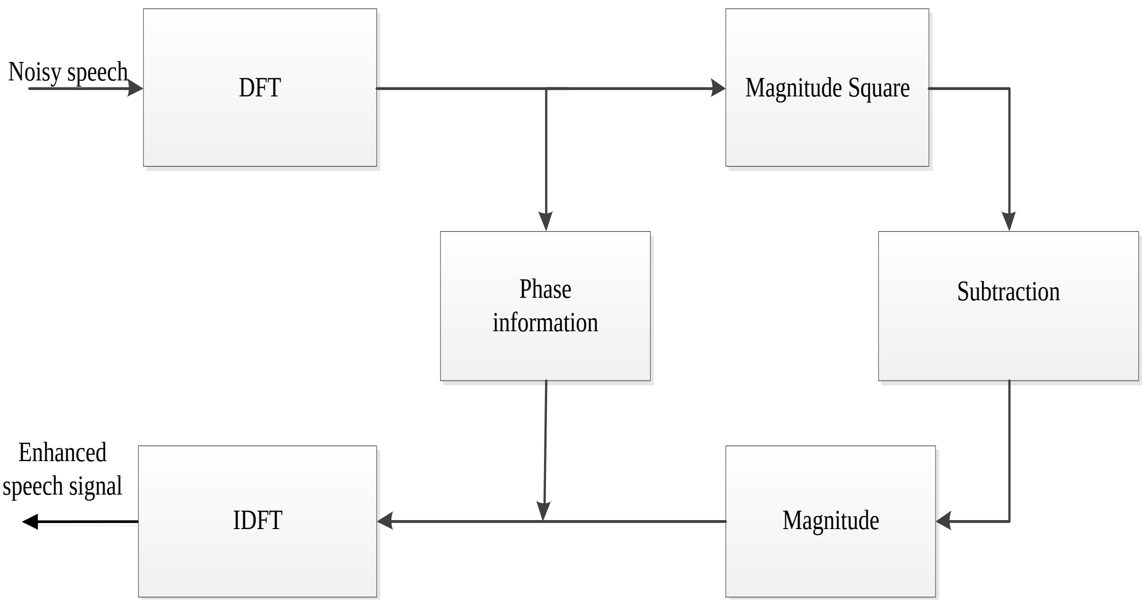

Figure 2.2 Basic Block Diagram of Spectral Subtraction

2.4.1 Spectral Subtraction Basic

The spectral subtraction method generally performs better in additive type of noise, where the power or the magnitude spectrum is recovered through the subtraction of the noisy speech signal spectrum by the estimated noise spectrum, and this is the most common concept for the subtractive type algorithms that have a group of methods based on the subtraction rules [4][21]. These systems assume that the noise is stationary or a less varying process, and operate in the frequency domain. It estimates that the noise spectrum from the noisy speech and updates the spectrum when the speech signal is absent. This updating is possible when the noise signal does not change significantly. In order to transform the frequency domain signal to time domain signal, the phase of the noisy speech signal is combined with modified magnitude spectrum, and then Inverse Discrete Fourier Transform (IDFT) is applied.

Suppose $x(n)$ is a noise corrupted input speech signal which contains clean speech signal $s(n)$ and uncorrelated additive noise signal $d(n)$, so the corrupted signal can be represented as:

$$x(n) = s(n) + d(n) \tag{2.1}$$

The spectral based speech enhancement is carried out frame by frame; therefore, a window $w(n)$ is multiplied with the input signal.

The windowed signal can be expressed as:

$$x_w(n) = s_w(n) + d_w(n) \tag{2.2}$$

The DFT of the windowed signal can be written as:

$$X_w(\omega) = S_w(\omega) + D_w(\omega) \tag{2.3}$$

The DFT of $x_w(n)$ is given by:

$$X_w(\omega) = \sum_{n=0}^{N-1} x_w(n)\, e^{-j\frac{2\pi\omega n}{N}} = |X_w(\omega)|e^{j\phi(\omega)} \tag{2.4}$$

Where $|X_w(\omega)|$, $\phi(\omega)$ is the amplitude and phase of noise corrupted input signal and N is the window length.

To get the power spectrum of the noisy speech signal, the equation (2.3) is multiplied by their complex conjugate and the equation becomes.

$$|X_w(\omega)|^2 = |S_w(\omega)|^2 + |D_w(\omega)|^2 + S_w(\omega).D^*{}_w(\omega) + S^*{}_w(\omega).D_w(\omega) \tag{2.5}$$

By taking the expected value of equation (2.5), we get

$$E\{|X_w(\omega)|^2\} = E\{|S_w(\omega)|^2\} + E\{|D_w(\omega)|^2\} + E\{S_w(\omega).D^*{}_w(\omega)\} + E\{S^*{}_w(\omega).D_w(\omega)\} \tag{2.6}$$

In power spectral subtraction, considering that the noise signal $d(n)$ has zero mean and uncorrelated with clean speech signal $s(n)$, the terms $E\{S_w(\omega).D^*{}_w(\omega)\}$ and $E\{S^*{}_w(\omega).D_w(\omega)\}$ becomes zero. Taking the above assumption into consideration the power spectral subtraction, subtracts the average estimated noise from the spectrum of the corrupted noise and thus the results of estimated clean speech signal are obtained. So the equation (2.6) becomes

$$E\{|S_w(\omega)|^2\} = E\{|X_w(\omega)|^2\} - E\{|D_w(\omega)|^2\} \tag{2.7}$$

Now the phase adds directly with the amplitude of the estimated clean speech and the enhanced speech signal in the time domain is obtained according to:

$$E\{y(n)\} = IDFT[sqrt(\{E|S(\omega)|^2\}).e\hat{}j\phi\omega\ ] \tag{2.8}$$

Chapter 3

Spectral Subtraction based on Minimum Statistics

3.1 Introduction

The Spectral Subtraction Based on Minimum Statistics (SSBMS) is one of the influential method for speech enhancement, which is usually able to track non stationary noise signals [12]. The problem of conventional spectral subtraction method is the requirement of speech activity detector during noise power estimation [13], which increases computational complexity while spectral subtraction based on minimum statistics uses a finite window of sub-band noise power to estimate the noise power [14]. We have selected this algorithm because this needs no additional equipment due to its simplicity.

The block diagram of the spectral subtraction method based on minimum statistics is shown in figure 3.1. This algorithm uses DFT filter bank for the analysis of disturbed speech signal, and modifies the short time spectral magnitude to make the synthesized signal as close to desired speech signal. The SNR of each sub-band is calculated by using estimated noise power to control the oversubtraction factor and this factor reduces the residual noise. The subtraction rule is designed by using estimated noise power with oversubtraction factor for computing the optimal weighting of spectral magnitudes.

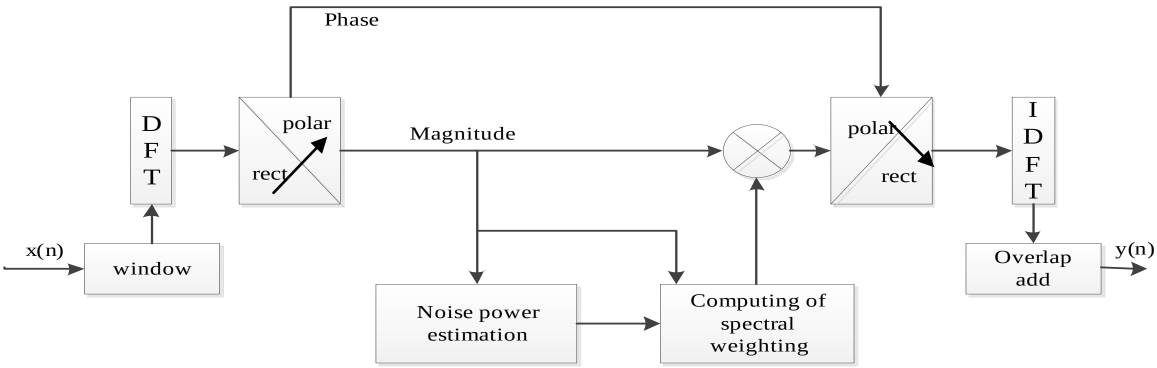

Figure 3.1 Spectral Subtraction Based on Minimum Statistics

3.2 Description of Algorithm

Consider an input signal $x(n)$ contains zero mean speech signal $s(n)$ and zero mean noise signal $d(n)$ and that signals are statistically independent.

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

Where, n denotes the discrete time index.

The spectral processing is based on a DFT filter bank with w_{DFT} sub-bands and with decimation/interpolation ration R [15].The filter bank uses an array of band pass filter in which the signal divides into multiple components, where each component having a certain frequency sub-band of the original signal is shown in figure 3.2.

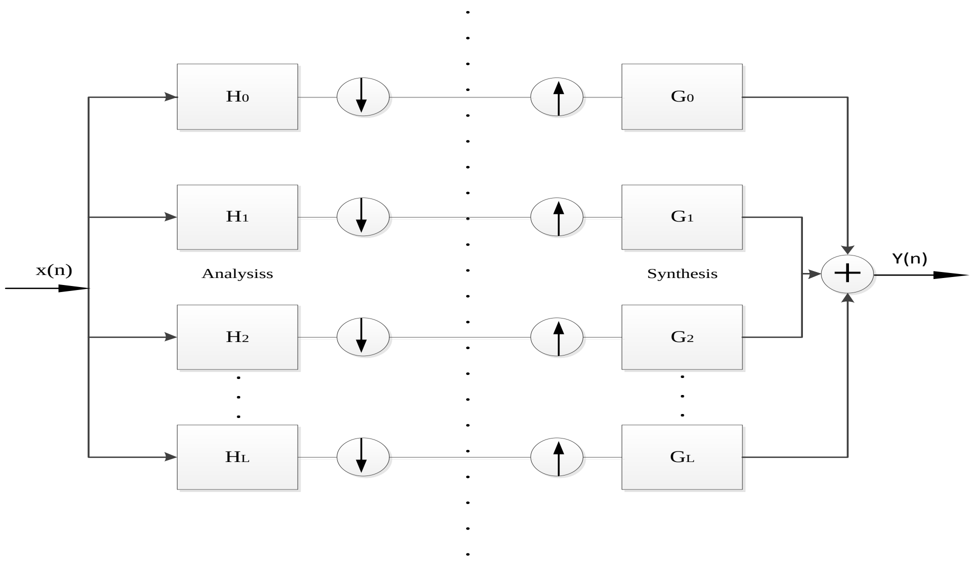

Figure 3.2 Basic Filter Bank Diagram

The DFT of input signal $x(n)$ with window function $h(n)$ is given by [16]

$$X(\lambda, k) = \sum_{\mu=0}^{W_{DFT}-1} x(\lambda R + \mu) h(\mu) e^{-j\frac{2\pi\mu k}{W_{DFT}}} \tag{3.2}$$

Where, λ is the decimated time index and k is the frequency bins, $\lambda\epsilon\ 0,1, \ldots \ldots .., P-1$ and $k\epsilon\ 0,1, \ldots \ldots .., W_{DFT}-1$.

First of all, the long-time input signal is segmented into many short frames (P) by the window function; typically the range of frame duration is in between 1ms to 100ms [17].The amplitude of the short time signal depends on the chosen window function. Many window functions are available with different spectral characteristics and these should be chosen due to the requirements of analysis. The most elementary window is the rectangular window that provides a distorted analysis and its frequency response has high magnitude side lobes [22]. The DFT applies on the windowed short time signals for analysis of the spectrum and represent a variation in the spectrums of signals over time. Overlap is common in time windows since this gives better spectral analysis.

The overlap is usually given as

$$O = \frac{W_{DFT}}{R} \tag{3.3}$$

Typically the filter lengths are 64, 128, 256 and 512, and for these the overlap is 50% or 75%.

The total number of spectral frame can be calculated as:

$$P = \frac{N}{R} \tag{3.4}$$

Here, *N* is the input signal length.

For a sampling rate f_s, the corresponding time index is given by

$$t_\lambda = \lambda \cdot (R/f_s) \tag{3.5}$$

Consider a signal $x(n)$ with length *N* is the input to the system. The system has a window $h(n)$ of length W_{DFT} and decimation/interpolation ratio R, so that time overlapping is occurring. The framing procedure is shown in Figure 3.3.

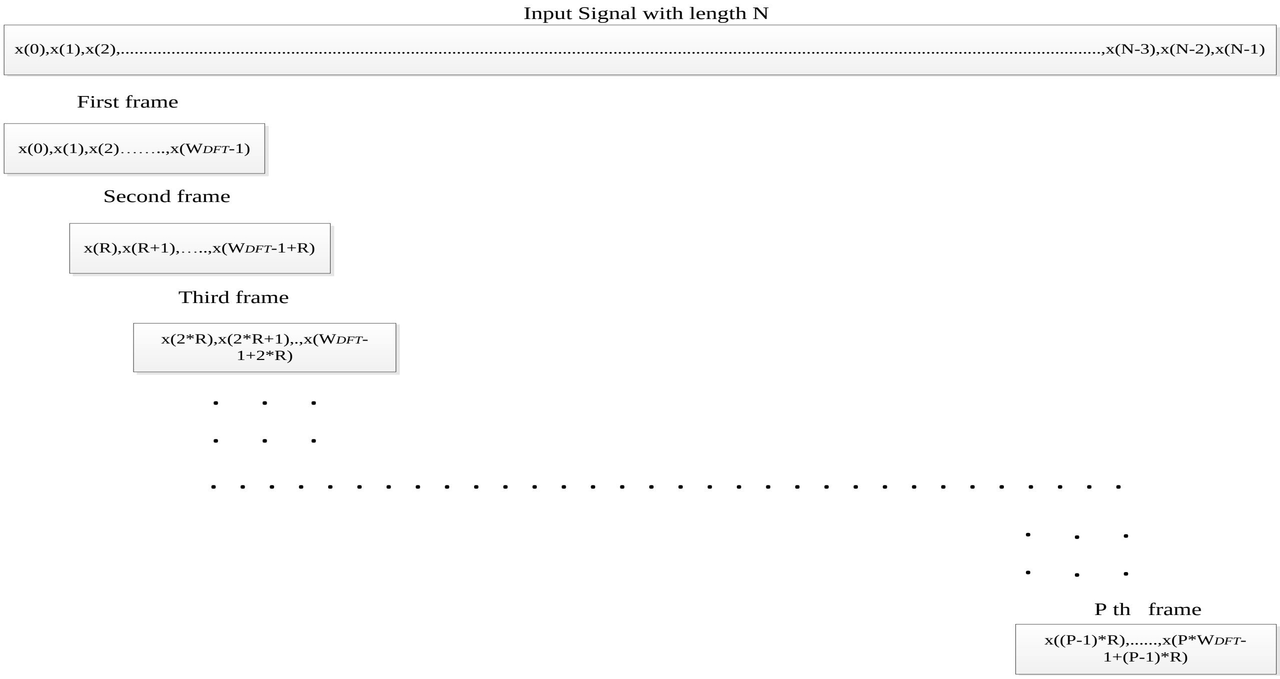

Figure 3.3 Framing of the Input Signal

The DFT filter bank then can be applied on each frame of the input signal.

The DFT filter bank output is then converted to polar form for spectral analysis, i.e. in terms of phase and magnitude by the following equation (3.7).

$$X(\lambda,k) = a(\lambda,k) + ib(\lambda,k) = r(\lambda,k)e^{j\theta(\lambda,k)} \tag{3.7}$$

Where,

$a(\lambda,k)$ corresponds to the real part of $X(\lambda,k)$ and $b(\lambda,k)$ corresponds to the imaginary part of $X(\lambda,k)$

Here,

$$r(\lambda,k)=|X(\lambda,k)| = \sqrt{a(\lambda,k)^2 + b(\lambda,k)^2} \text{ and } \theta(\lambda,k) = tan^{-1}\left({b(\lambda,k)}/{a(\lambda,k)} \right) \quad (3.8)$$

The magnitude spectrum is modified by the noise estimation, and the subtraction rules, and hence this carries significant information. The phase without any modification is combined with the estimated spectrum for time domain restoration, since it is difficult to get an estimation of the phase [4], and from perceptual point of view it is believed that it does not carry any useful information in noise suppression [24].

Magnitude spectrum analysis is a combination of two main procedures

- Noise power estimator
- Subtraction rules

3.2.1 Noise Power Estimation

In our thesis we use minimum statistics algorithm for noise estimation which is proposed by Martin [12]. This algorithm can estimate instantaneous SNR of speech signals by using the combination of estimated minimum values of a smoothed power and instantaneous power spectrum with low computational complexity.

3.2.1.1 Subband Signal Power Estimation

To get the short time subband signal power $P_x(\lambda,k)$, the recursively smoothed periodograms is used [18]. The short time subband signal power is updated on a frame by frame basis which is given by:

$$P_x(\lambda,k) = \alpha \cdot P_x(\lambda-1,k) + (1-\alpha)\cdot |X(\lambda,k)|^2 \quad (3.9)$$

Where, $|X(\lambda, k)|$ is the magnitude of the input signal spectra and α is the smoothing constant that takes the values in between 0.90 to 0.95 [12].

3.2.1.2 Subband Noise Power Estimation

The minimum power is obtained from the short time subband signal power. For calculating the minimum power we have taken a window W, of length D. In order to reduce computational complexity $(D - 1)$ numbers of variables are added at the beginning of the short time subband power $P_x(\lambda = \lambda_1, k)$. Then the minimum power $P_{min}(\lambda, k)$ from the short time subband power $P_x(\lambda, k)$.is found by a sample wise comparison of the values within the window and then the minimum power is stored in the last position of the window. Whenever one minimum value is obtained, the window is updated by taking next short time subband power and the next minimum subband power is found in the same way. The window update for finding minimum noise power is continued until last subband power $P_x(\lambda = \lambda_P, k)$ is reached.

The noise power estimation $P_n(\lambda, k)$ is then calculated by using the minimum power $P_{min}(\lambda, k)$ of the short time sub band signal power within the window of length D [15].

$$P_n(\lambda, k) = omin \cdot P_{min}(\lambda, k) \tag{3.10}$$

Where, $omin$ is the overestimation factor that is used to make the minimum power as noise power, with typically set values in the range 1.3 to 2 [19]. When $omin$ is set at 1.5, it gives better performance [12].

3.2.2 SNR and Oversubtraction Factor Calculation:

In general Signal to noise ratio (SNR) is defined as the ratio of the signal power to the noise power. SNR in each sub band is calculated to adjust the over subtraction factor $osub(\lambda, k)$ as

$$SNR_x(\lambda, k) = 10 \cdot \log_{10}\left(\frac{P_x(\lambda, k) - min(P_n(\lambda, k), P_x(\lambda, k))}{P_n(\lambda, k)}\right) \tag{3.11}$$

Oversubtraction factor can eliminate the residual spectral peaks. The large over subtraction factor not only remove the residual spectral peaks but also suppress some of the low energy components of the speech signal [12]. The speech quality is degraded by this undesirable effect. We calculated the over subtraction factor as a function of $SNR_x(\lambda, k)$ and frequency bin k to maintain the speech quality [20].

$$osub(\lambda, k) = \begin{cases} 5 & SNR_x(\lambda, k) < -5 \\ 4 - \frac{3}{20} SNR_x(\lambda, k) & -5 \leq SNR_x(\lambda, k) \leq 20 \\ 1 & SNR_x(\lambda, k) > 20 \end{cases} \tag{3.12}$$

3.2.3 Subtraction Rule

The short time signal power $\overline{|X(\lambda, k)|}^2$ is calculated by smoothing the squared magnitude of the input spectra with a first order recursive network.

$$\overline{|X(\lambda, k)|}^2 = \gamma \,.\, \overline{|X(\lambda - 1, k)|}^2 + (1 - \gamma) \cdot |X(\lambda, k)|^2 \tag{3.13}$$

Where γ is the smoothing constant and $\gamma \leq 0.9$. We used the Berouti et. al. proposal to subtract the spectral magnitude [21]. According the proposal spectral magnitude is subtracted with an over subtraction factor $osub(\lambda, k)$ and the maximum subtraction is limited by a spectral floor constant $subf$ $(0.01 \leq subf \leq 0.05)$ [12]. The modified magnitude can be obtained by the following way

$$|Y(\lambda,k)| = \begin{cases} \sqrt{subf.P_n(\lambda,k)} & if \quad |X(\lambda k)| \cdot Q(\lambda,k) \leq \sqrt{subf.P_n(\lambda,k)} \\ |X(\lambda k)| \cdot Q(\lambda,k) & else \end{cases} \tag{3.14}$$

Where,

$$Q(\lambda,k) = \left(1 - \sqrt{osub(\lambda,k) \cdot \frac{P_n(\lambda,k)}{|X(\lambda,k)|^2}}\right)$$

3.2.4 Reconstruction in Time Domain

The modified magnitude $|Y(\lambda,k)|$ is directly added to the phase $\theta(\lambda,k)$ by the following equation:

$$M(\lambda,k) = |Y(\lambda,k)| \cdot (\cos\theta(\lambda,k) + isin\,\theta(\lambda,k)) \tag{3.15}$$

Many techniques are available to construct time domain signal from frequency domain signal [21]. The overlap-add IDFT is generally used for the filter bank analysis data to reconstruct the time domain signal. The IDFT is applied in each of the DFT frames to get a series of short time signals. These signals are then added together to reproduce the time domain signal with the same overlap which is used in the DFT filter bank.

The IDFT of the signal $M(\lambda,k)$ with the same window function $h(n)$ is given by [22]

$$y(\lambda R + \mu) = h(\mu) \cdot \frac{1}{W_{DFT}} \sum_{k=0}^{W_{DFT}-1} M(\lambda,k) e^{j\frac{2\pi\mu k}{W_{DFT}}} \tag{3.16}$$

We consider the system with same window $h(n)$ of length W_{DFT} and interpolation ratio R, so that there is an occurrence of time overlapping. The overlap-add procedure is shown in Fig 3.4.

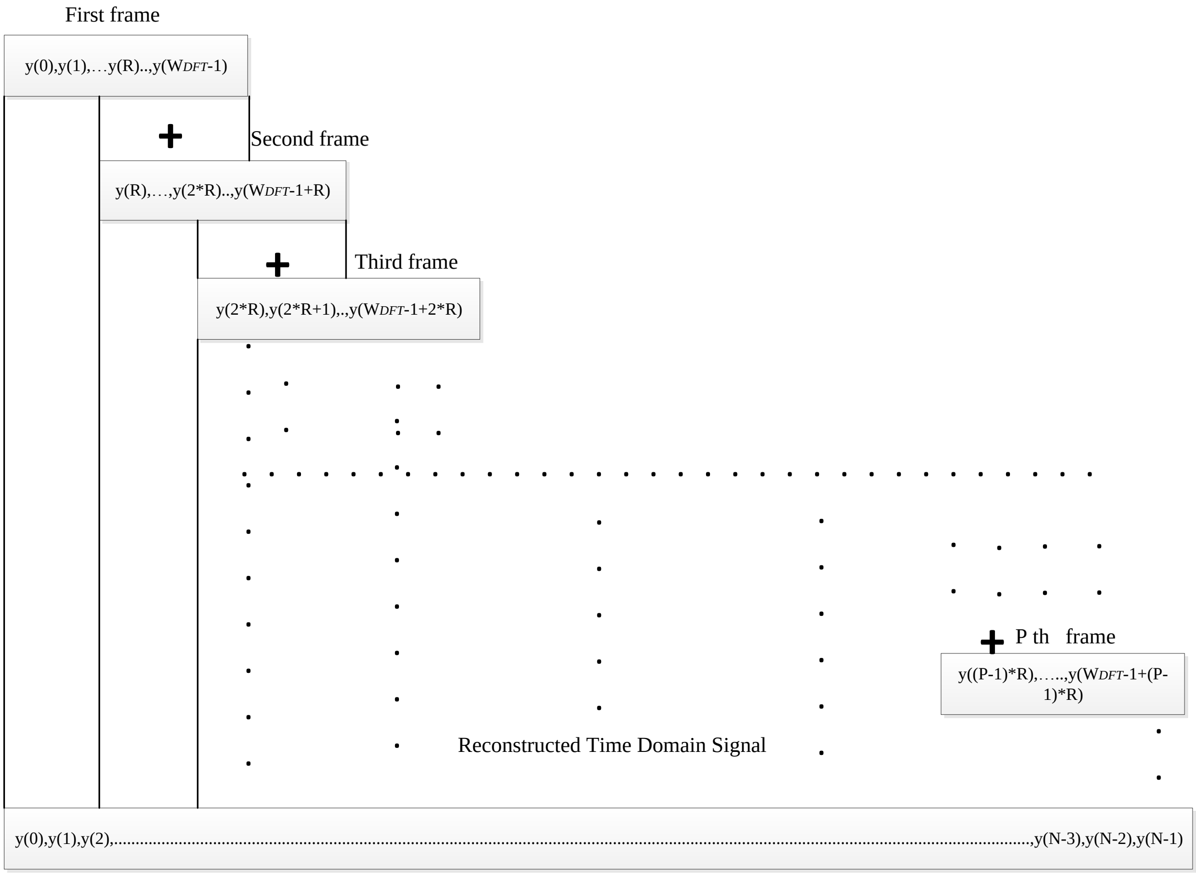

Figure 3.4 Overlap Add in Time Domain

Chapter 4

Implementation and Results

4.1 Introduction

In this chapter we present the implementation and analysis of the Spectral Subtraction Based on Minimum Statistics (SSBMS) algorithm which is discussed in the previous chapter. Section 4.2 describes the details of the implementation and experimental setup of the system. This section also gives the optimum configuration by considering various parameters of the system. In section 4.3, we demonstrate the result from the performance evaluation.

4.2 Implementation

The offline implementation and evaluation of the SSBMS method are carried out in the MATLAB, as the implementation of any algorithm on the real-time system requires preliminary investigation. It is necessary to optimize the MATLAB code in order to reduce the computational load of the algorithm. The use of more 'for' loops degrades the efficiency of the program because of the access of the array elements. The matrix multiplication reduces the time complexity and ensures faster data processing. We have changed the 'for' loops by matrix processing to optimize the program.

The experimental setup for the validation of single channel speech enhancement technique based on SSBMS is shown in figure 4.1. In this figure, $s(n)$ is the clean speech signal, $d(n)$ is the noise signal and $x(n)$ is the system input signal which contains the clean speech signal and the noise signal($s(n) + \beta \cdot d(n)$), and β is scaled by the desired SNR level.

Where,

$$\beta = 10^{\frac{-SNR}{20}}$$

The same filter bank is used for synthesis of the signals. S_M, S_P, D_M, D_P, X_M and X_P are the magnitude and phase of the signals $s(n)$, $d(n)$ and $x(n)$ respectively. The gain function G is calculated when X_M is passed through the system. Each signal after passing through G is added with the corresponding phase and then IDFT is applied to get the output signals $y_s(n)$, $y_d(n)$, $y_x(n)$.

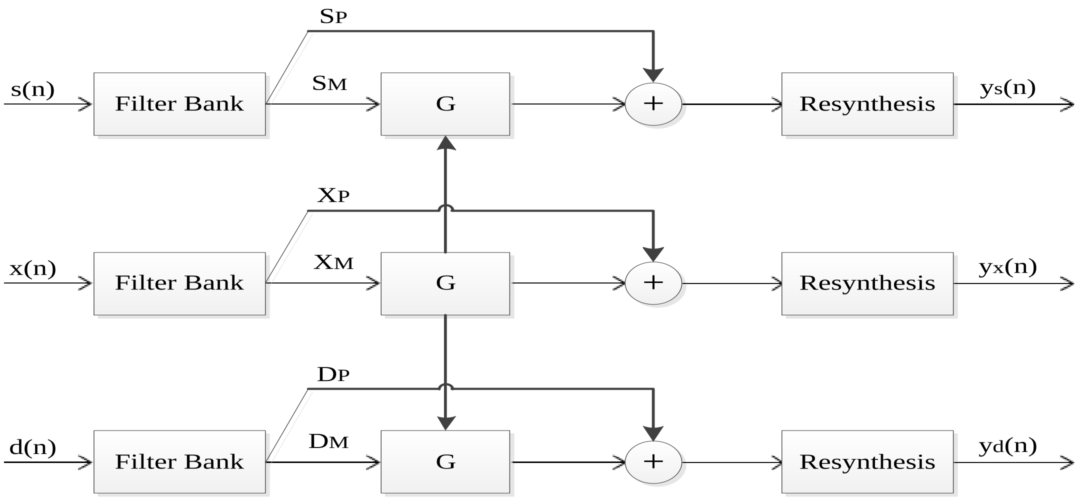

Figure 4.1 Experimental Setup

In this thesis, we use ITU-T P.50 male and ITU-T P.50 female speech signal at the sampling frequency of 16 KHz as clean speech signal. ITU-T P.50 are the artificial voices that are used as test signals in telecommunication systems. The use of recommended artificial voices instead of real speech is the convenient way for the effective validation of the system. These ITU-T P.50 voices include 16 recorded sentences in each of 20 languages and are developed by some ITU members [23]. Both signals are corrupted with the Gaussian Noise (GN), Car Noise (CAN), Factory Noise (FN), Wind Noise (WN) and

Cafeteria Noise (CN) at -5 dB, 0 dB, 5 dB and 10 dB SNR for testing the system. The performance of the system is measured by SNRI and SD. The results are observed by changing the number of subbands and decimation/interpolation ratios. We have used 64,128, 256 and 512 numbers of subbands with 75% and 50% overlapping. During experiment various values of the α, γ and $subf$ are used that created less effect to the algorithm performance in terms of SNRI and SD as shown in figure 4.2 and figure 4.3. In figure 4.4 and figure 4.5 the average SNRI and SD are obtained by using one fixed α and one $subf$ values. Then SNRI and SD are taken by varying γ values from 0.80 to 0.89 and finally one SNRI and SD values are obtained from its average. By varying γ values from 0.50 to 0.89, SNRI and one SD values are taken for further investigation by keeping α = 0.9 and $subf$ = 0.01 as is shown in figure 4.4 and figure 4.5. From the figure 4.2 to figure 4.5 it is clear that the SSBMS algorithm gives comparatively better performance if the values of α, γ, $subf$ and $omin$ are set at 0.90, 0.86, 0.01 and 1.5. The performance of the algorithm is evaluated in different noisy environment from the above setting. The power spectral of the noises is shown in figure 4.6 to figure 4.9.

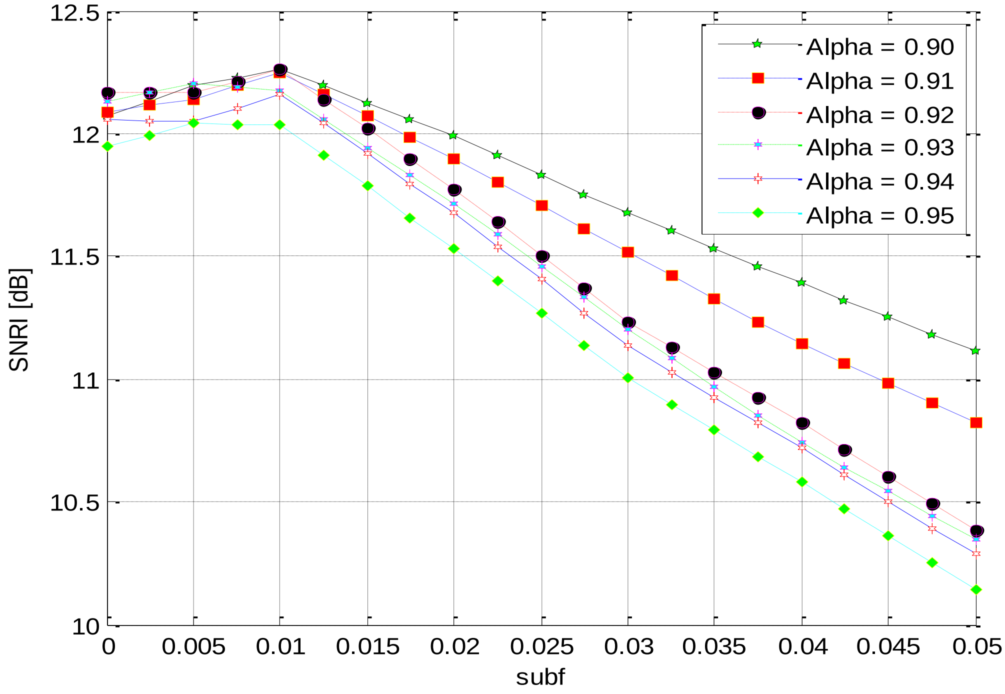

Figure 4.2 Average SNRI by changing α, γ and $subf$

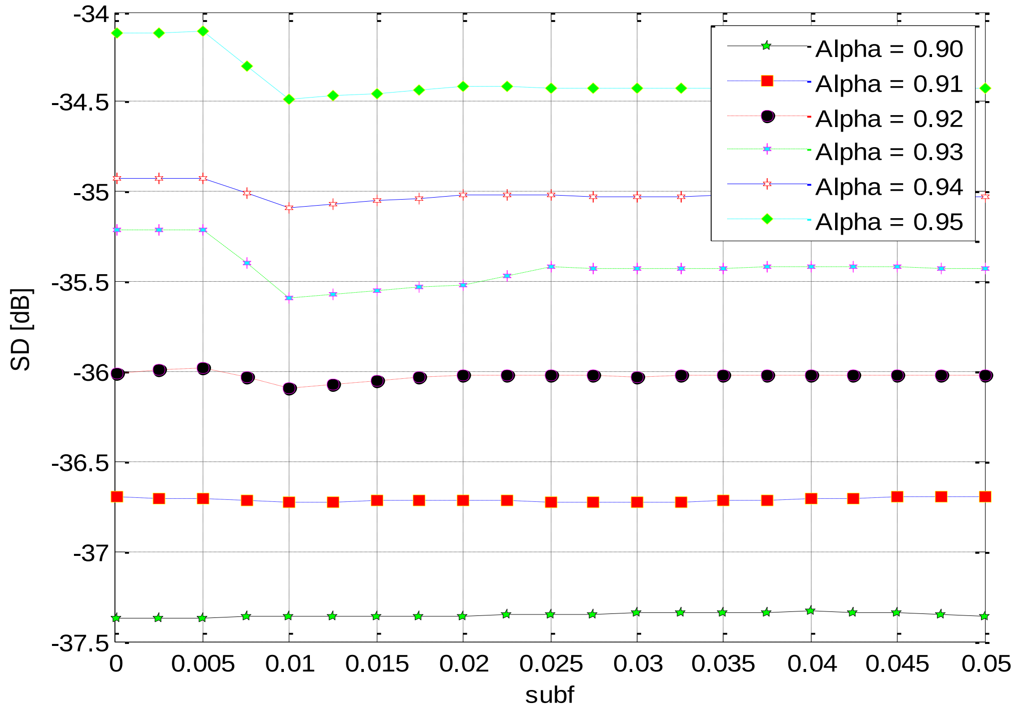

Figure 4.3 Average SD by changing α, γ and $subf$

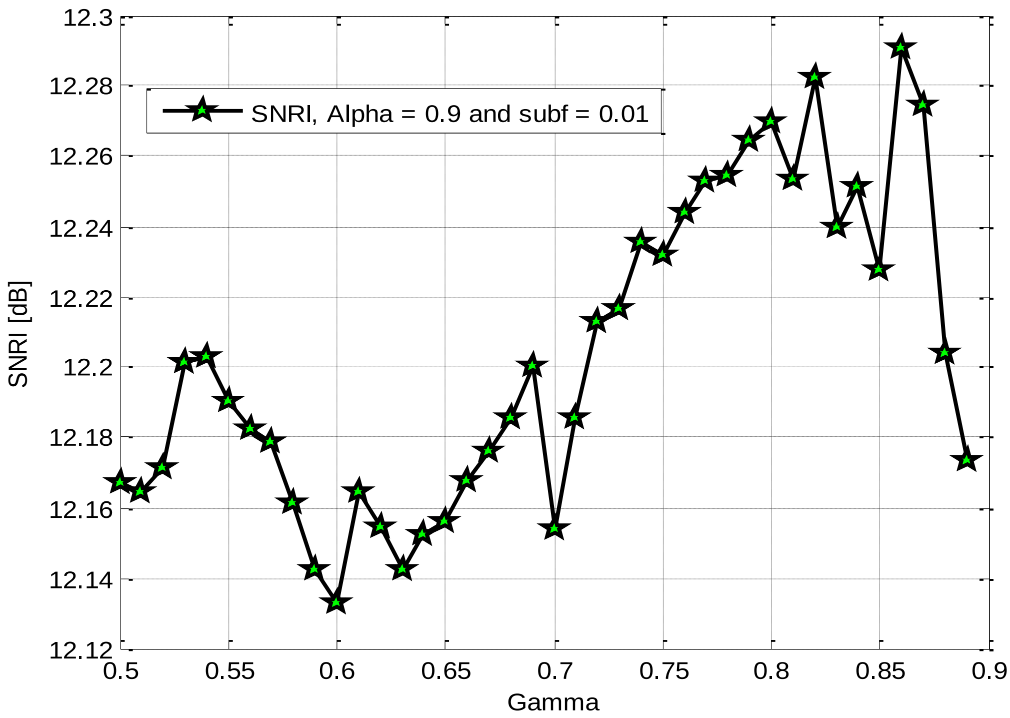

Figure 4.4 SNRI by changing γ value

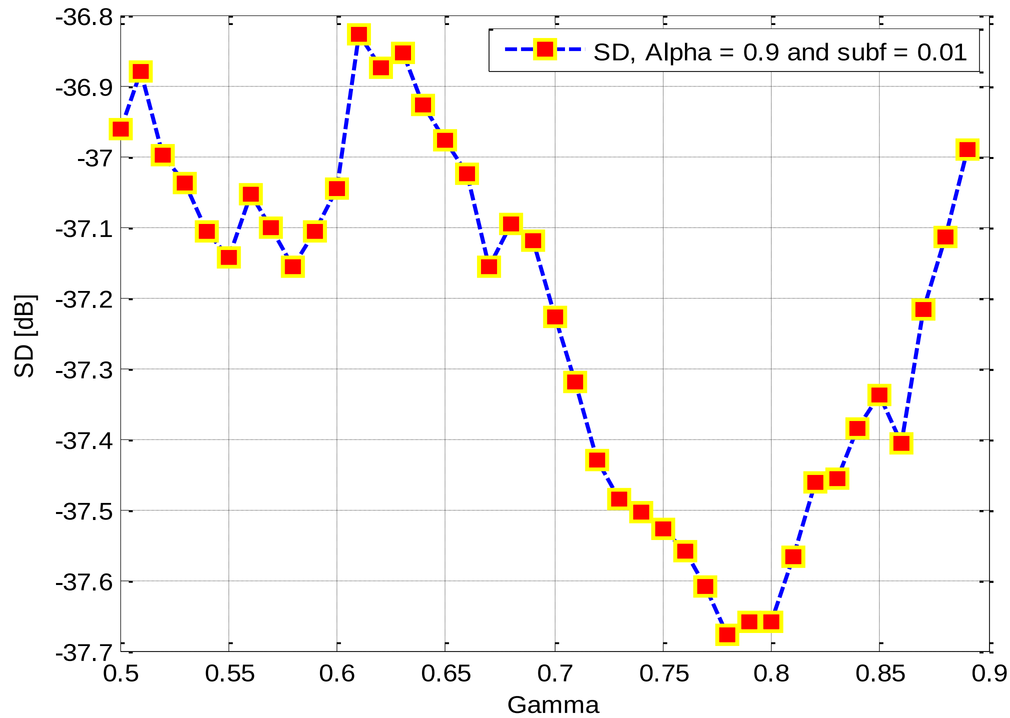

Figure 4.5 SD by changing γ value

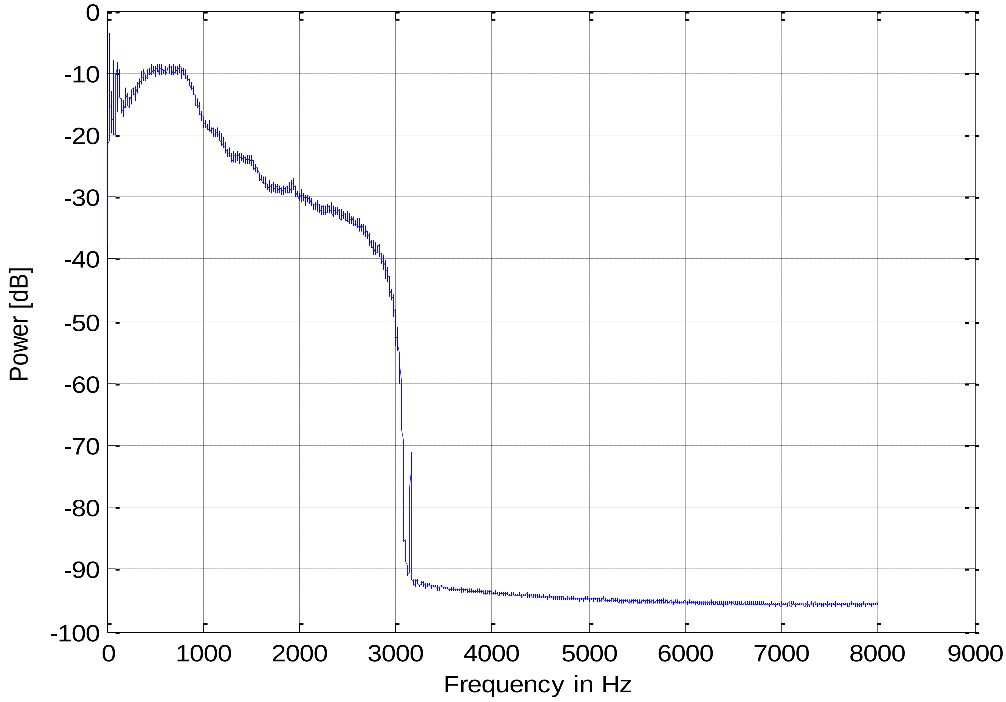

Figure 4.6 Power Spectral Density of Car Noise

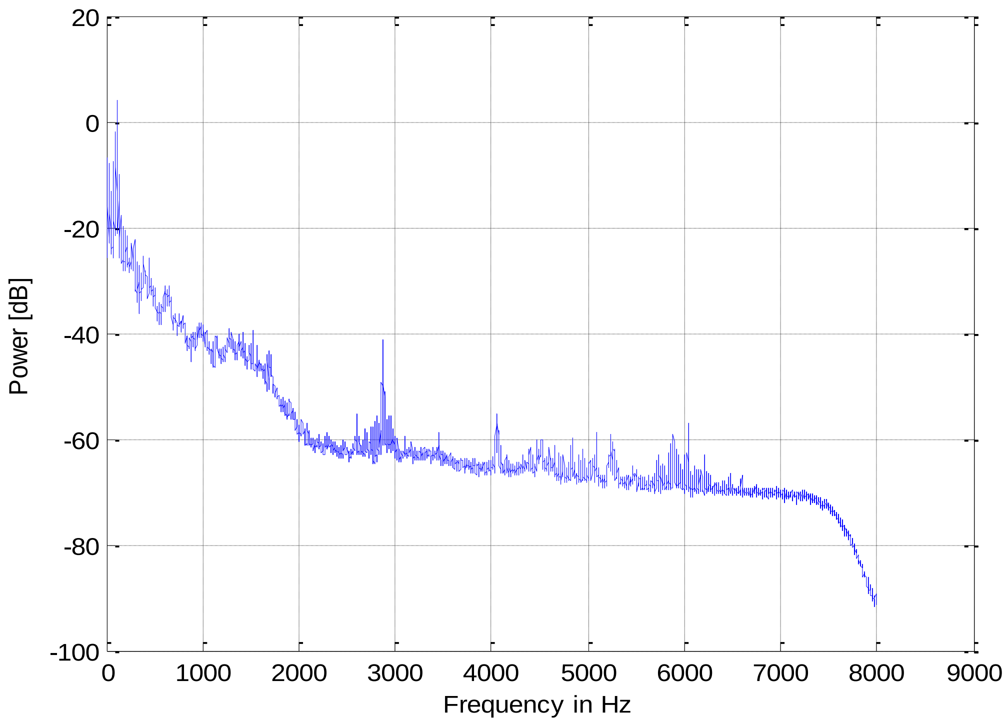

Figure 4.7 Power Spectral Density of Factory Noise

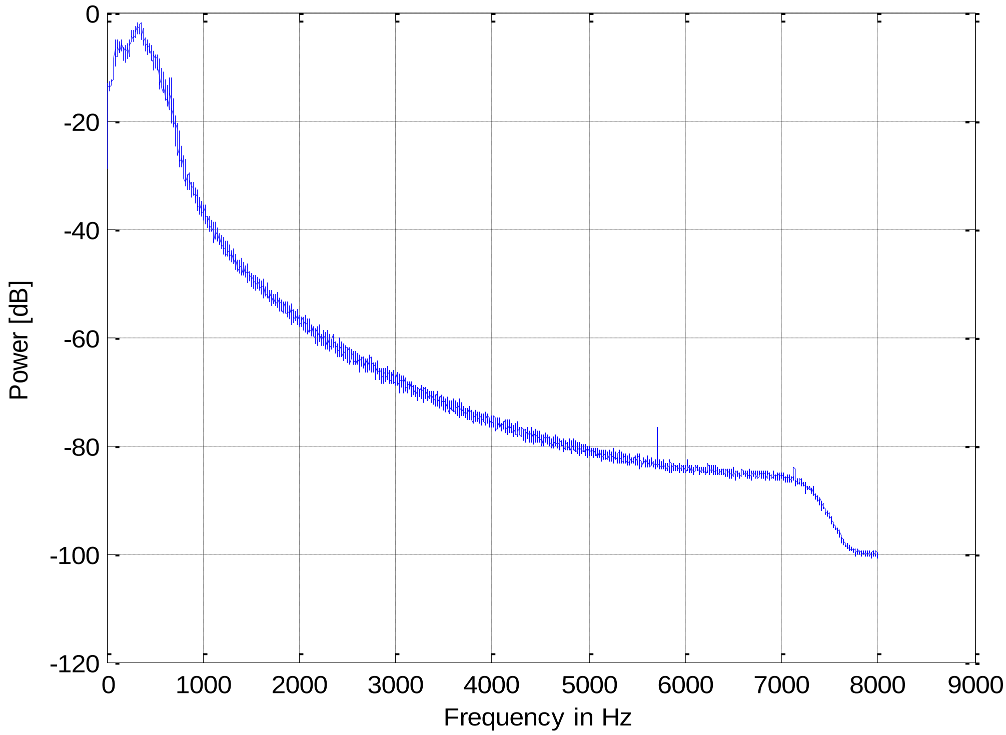

Figure 4.8 Power Spectral Density of Wind Noise

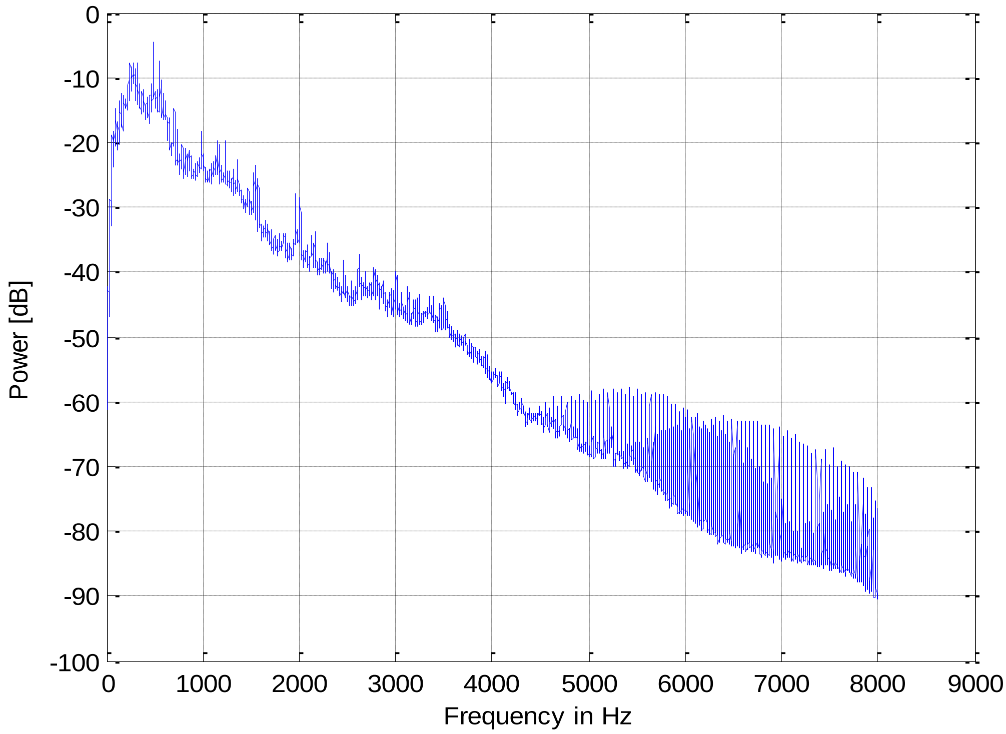

Figure 4.9 Power Spectral Density of Cafeteria Noise

4.3 Results

The SSBMS method gives on an average around 9 dB SNRI and -33 dB SD for all the situations tested for the both male and female speech signal. The SNRI and SD values are shown in Table 4.1 to Table 4.20. It is observed that both SNRI and SD vary a little bit depending on the number of subbands, overlap rates, types of noises and noise levels (-5 dB, 0 dB, 5 dB and 10 dB). SNRI and SD are better for 75% overlap compared to 50% overlap for both male and female speech signals with GN as shown in Table 4.1 to Table 4.4. The SNRI value of around 13 dB is achieved using 512, 256 and 128 number of subbands and a value of around 8 dB for 64 number of subbands for the same signals. The SD values decrease depending on the number of subbands and noise levels for both male and female speech signals with GN. The SD values varied form -29 dB to -37 dB. The SNRI and SD for both male and female speech signals with car noise are shown in Table 4.5 to Table 4.8. The SNRI is around 17 dB for 512, 256 and 128 number of subbands with 75% overlap and 18 dB for 256 and 128 number of subbands with 50% overlap. The SD for both male and female speech signal is around -34 dB for CAN. It can be seen from Table 4.9 to Table 4.20 that the SNRI is much less for the same male and female speeches but mixed with the FN, WN and CN respectively. But the variation in SD values is nearly similar to GN and CAN mixed with that male and female speech. SNRI for both speeches is around 3-6 dB for FN, WN, and CN as shown in Table 4.9 to Table 4.20. It is calculated for -5 dB, 0 dB, 5 dB and 10 dB SNR. In case of wind noise the better SNRI is obtained while using 256 and 128 number of subbands in 75% and 50% overlap. The highest SNRI for WN is about 10 dB when female speech signal at 0dB SNR is processed by 128 number of subbands with 50% of overlap. For cafeteria noise SNRI always gives better result at 50% overlap as compared to 75% overlap. Figure 4.10 and figure 4.12 show average SNRI plots for male and female speech with 75% overlap while the average SNRI plots for 50% overlap are shown in figure 4.11 and 4.13. The average spectral distortion plots for male and female speech with all cases are shown in figure 4.14 and figure 4.15. The computational complexity calculation for SSBMS algorithm is derived in section 4.4.

Input SNR in dB	Number of Subbands	SNRI in dB	SD in dB
-5	512	13.244	-31.854
	256	13.474	-30.702
	128	13.322	-30.169
	64	12.785	-29.615
0	512	13.618	-33.633
	256	13.707	-31.955
	128	12.704	-30.701
	64	11.602	-29.812
5	512	13.141	-35.637
	256	12.690	-33.191
	128	11.148	-31.143
	64	09.434	-30.015
10	512	12.283	-37.300
	256	11.220	-34.092
	128	08.576	-31.471
	64	06.028	-30.142

Table 4.1: SNRI and SD for Male Speech Signal with Gaussian Noise at 75% Overlap

Input SNR in dB	Number of Subbands	SNRI in dB	SD in dB
-5	512	13.811	-31.981
	256	14.722	-30.954
	128	12.268	-29.940
	64	09.984	-29.394
0	512	14.210	-33.601
	256	14.629	-32.445
	128	12.239	-30.669
	64	09.657	-29.628
5	512	13.493	-35.277
	256	13.727	-33.846
	128	10.864	-31.232
	64	07.620	-29.897
10	512	12.697	-36.640
	256	12.291	-34.808
	128	08.606	-31.663
	64	04.558	-30.060

Table 4.2: SNRI and SD for Female Speech Signal with Gaussian Noise at 75% Overlap

Input SNR in dB	Number of Subbands	SNRI in dB	SD in dB
-5	512	10.600	-31.127
	256	11.416	-30.464
	128	12.671	-30.304
	64	12.318	-30.021
0	512	11.429	-32.251
	256	12.462	-31.923
	128	13.041	-31.538
	64	12.639	-30.597
5	512	11.393	-33.295
	256	12.250	-33.399
	128	12.250	-32.792
	64	10.808	-31.124
10	512	10.867	-33.978
	256	11.341	-34.534
	128	10.799	-33.740
	64	08.474	-31.464

Table 4.3: SNRI and SD for Male Speech Signal with Gaussian Noise at 50% Overlap

Input SNR in dB	Number of Subbands	SNRI in dB	SD in dB
-5	512	11.028	-31.200
	256	12.707	-31.060
	128	12.613	-30.354
	64	10.658	-29.548
0	512	11.812	-32.368
	256	13.241	-32.551
	128	13.034	-31.481
	64	11.141	-30.158
5	512	11.709	-33.350
	256	12.686	-34.090
	128	12.148	-32.739
	64	10.047	-30.817
10	512	11.222	-34.127
	256	11.953	-35.390
	128	10.763	-33.707
	64	07.766	-31.272

Table 4.4: SNRI and SD for Female Speech Signal with Gaussian Noise at 50% Overlap

Input SNR in dB	Number of Subbands	SNRI in dB	SD in dB
-5	512	17.083	-34.576
	256	18.456	-32.604
	128	16.238	-30.910
	64	14.954	-29.951
0	512	17.092	-35.858
	256	17.807	-33.391
	128	15.423	-31.217
	64	13.400	-30.521
5	512	16.324	-36.924
	256	16.060	-34.003
	128	13.201	-31.451
	64	10.577	-30.134
10	512	14.742	-37.903
	256	13.397	-34.500
	128	09.741	-31.612
	64	06.614	-30.195

Table 4.5: SNRI and SD for Male Speech Signal with Car Noise at 75% Overlap

Input SNR in dB	Number of Subbands	SNRI in dB	SD in dB
-5	512	17.100	-33.957
	256	18.585	-32.789
	128	16.476	-30.771
	64	13..332	-29.879
0	512	17.029	-35.138
	256	17.557	-33.642
	128	15.345	-31.135
	64	11.784	-30.018
5	512	16.271	-36.224
	256	15.934	-34.391
	128	12.897	-31.463
	64	08.874	-30.155
10	512	14.964	-37.086
	256	13.854	-34.970
	128	09.657	-31.704
	64	05.176	-30.260

Table 4.6: SNRI and SD for Female Speech Signal with Car Noise at 75% Overlap

Input SNR in dB	Number of Subbands	SNRI in dB	SD in dB
-5	512	12.915	-32.905
	256	17.793	-32.651
	128	18.637	-32.207
	64	16.887	-30.794
0	512	13.470	-33.476
	256	17.563	-33.539
	128	17.730	-33.015
	64	15.507	-31.144
5	512	13.323	-33.861
	256	16.462	-34.301
	128	15.866	-32.638
	64	13.052	-31.422
10	512	12.218	-34.171
	256	14.543	-34.939
	128	13.137	-34.133
	64	09.571	-31.613

Table 4.7: SNRI and SD for Male Speech Signal with Car Noise at 50% Overlap

Input SNR in dB	Number of Subbands	SNRI in dB	SD in dB
-5	512	13.261	-33.052
	256	17.874	-32.679
	128	18.336	-31.998
	64	16.429	-30.574
0	512	13.989	-33.788
	256	17.363	-33.606
	128	17.292	-32.708
	64	15.034	-30.940
5	512	14.184	-34.398
	256	15.410	-33.363
	128	15.410	-33.363
	64	12.360	-31.290
10	512	13.504	-34.797
	256	14.641	-35.002
	128	12.795	-33.928
	64	09.017	-31.573

Table 4.8: SNRI and SD for Female Speech Signal with Car Noise at 50% Overlap

Input SNR in dB	Number of Subbands	SNRI in dB	SD in dB
-5	512	3.218	-30.217
	256	2.995	-29.555
	128	2.280	-29.503
	64	2.573	-29.228
0	512	5.822	-31.545
	256	5.879	-30.269
	128	4.152	-29.767
	64	3.397	-29.348
5	512	6.906	-33.382
	256	7.002	-31.421
	128	5.125	-30.303
	64	3.396	-29.608
10	512	6.946	-35.278
	256	6.431	-32.668
	128	4.329	-30.872
	64	2.405	-29.863

Table 4.9: SNRI and SD for Male Speech Signal with Factory Noise at 75% Overlap

Input SNR in dB	Number of Subbands	SNRI in dB	SD in dB
-5	512	4.774	-30.530
	256	4.609	-29.852
	128	2.611	-29.523
	64	2.426	-29.272
0	512	6.521	-31.744
	256	7.036	-30.793
	128	4.399	-29.821
	64	2.555	-29.359
5	512	7.215	-33.397
	256	7.920	-32.076
	128	5.238	-30.374
	64	2.574	-29.576
10	512	7.171	-35.091
	256	7.722	-33.464
	128	4.363	-31.016
	64	1.135	-29.821

Table 4.10: SNRI and SD for Female Speech Signal with Factory Noise at 75% Overlap

Input SNR in dB	Number of Subbands	SNRI in dB	SD in dB
-5	512	0.898	-30.052
	256	0.458	-29.479
	128	1.029	-29.343
	64	1.593	-29.445
0	512	2.888	-30.911
	256	3.489	-30.132
	128	3.447	-29.692
	64	3.123	-29.643
5	512	4.151	-32.085
	256	5.285	-31.279
	128	5.192	-30.622
	64	4.354	-30.175
10	512	4.540	-33.133
	256	5.497	-32.672
	128	4.904	-31.912
	64	3.695	-30.791

Table 4.11: SNRI and SD for Male Speech Signal with Factory Noise at 50% Overlap

Input SNR in dB	Number of Subbands	SNRI in dB	SD in dB
-5	512	2.066	-30.363
	256	2.747	-30.017
	128	1.628	-29.569
	64	1.743	-29.309
0	512	3.515	-31.104
	256	5.026	-30.870
	128	4.002	-29.955
	64	2.636	-29.454
5	512	4.461	-32.169
	256	6.247	-32.224
	128	5.406	-30.774
	64	3.700	-29.903
10	512	4.726	-33.186
	256	6.438	-33.760
	128	5.027	-31.940
	64	2.987	-30.530

Table 4.12: SNRI and SD for Female Speech Signal with Factory Noise at 50% Overlap

Input SNR in dB	Number of Subbands	SNRI in dB	SD in dB
-5	512	3.406	-31.248
	256	7.125	-30.426
	128	9.299	-29.677
	64	4.644	-29.394
0	512	4.931	-32.412
	256	8.552	-31.222
	128	9.595	-29.992
	64	5.778	-29.557
5	512	5.877	-33.725
	256	8.583	-32.178
	128	8.530	-30.422
	64	5.690	-29.757
10	512	6.213	-35.018
	256	7.472	-33.160
	128	6.557	-30.817
	64	4.008	-29.938

Table 4.13: SNRI and SD for Male Speech Signal with Wind Noise at 75% Overlap

Input SNR in dB	Number of Subbands	SNRI in dB	SD in dB
-5	512	4.043	-31.881
	256	8.211	-30.859
	128	9.039	-29.903
	64	3.909	-29.363
0	512	5.221	-33.025
	256	9.292	-31.673
	128	9.385	-30.182
	64	4.368	-29.426
5	512	6.065	-34.372
	256	9.337	-32.706
	128	8.375	-30.614
	64	3.965	-29.592
10	512	6.446	-35.712
	256	8.429	-33.772
	128	6.278	-31.107
	64	2.146	-29.799

Table 4.14: SNRI and SD for Female Speech Signal with Wind Noise at 75% Overlap

Input SNR in dB	Number of Subbands	SNRI in dB	SD in dB
-5	512	0.792	-31.048
	256	3.357	-30.485
	128	8.549	-29.946
	64	8.876	-29.454
0	512	2.261	-31.854
	256	4.997	-31.263
	128	9.444	-30.561
	64	9.007	-29.752
5	512	3.508	-32.654
	256	6.093	-32.233
	128	8.822	-31.505
	64	8.010	-30.228
10	512	4.354	-33.339
	256	6.284	-33.339
	128	7.284	-32.545
	64	6.038	-30.706

Table 4.15: SNRI and SD for Male Speech Signal with Wind Noise at 50% Overlap

Input SNR in dB	Number of Subbands	SNRI in dB	SD in dB
-5	512	1.486	-31.346
	256	4.768	-31.272
	128	9.343	-30.207
	64	8.407	-29.491
0	512	2.683	-32.008
	256	6.195	-31.997
	128	10.138	-30.773
	64	8.316	-29.637
5	512	3.846	-32.796
	256	7.098	-33.031
	128	9.360	-31.628
	64	7.177	-30.018
10	512	4.587	-33.558
	256	7.092	-34.175
	128	7.691	-32.574
	64	5.033	-30.568

Table 4.16: SNRI and SD for Female Speech Signal with Wind Noise at 50% Overlap

Input SNR in dB	Number of Subbands	SNRI in dB	SD in dB
-5	512	3.396	-31.988
	256	3.097	-30.921
	128	0.146	-29.784
	64	-1.661	-29.334
0	512	4.568	-33.399
	256	4.796	-31.987
	128	2.248	-30.224
	64	-0.024	-29.479
5	512	5.213	-34.870
	256	5.661	-33.191
	128	3.524	-30.793
	64	1.210	-29.693
10	512	5.518	-36.228
	256	5.918	-34.277
	128	3.614	-31.328
	64	1.041	-29.910

Table 4.17: SNRI and SD for Male Speech Signal with Cafeteria Noise at 75% Overlap

Input SNR in dB	Number of Subbands	SNRI in dB	SD in dB
-5	512	3.395	-31.987
	256	3.096	-30.920
	128	0.144	-29.784
	64	-1.66	-29.334
0	512	4.577	-33.414
	256	4.809	-31.999
	128	2.267	-30.230
	64	0	-29.481
5	512	5.208	-34.854
	256	5.655	-33.178
	128	3.517	-30.787
	64	1.204	-29.691
10	512	5.518	-36.229
	256	5.918	-34.278
	128	3.614	-31.328
	64	1.040	-29.910

Table 4.18: SNRI and SD for Female Speech Signal with Cafeteria Noise at 75% Overlap

Input SNR in dB	Number of Subbands	SNRI in dB	SD in dB
-5	512	3.348	-31.420
	256	3.098	-30.968
	128	2.573	-30.114
	64	0.748	-29.483
0	512	4.558	-32.192
	256	4.633	-32.071
	128	4.685	-30.850
	64	2.972	-29.846
5	512	5.489	-33.035
	256	5.498	-33.454
	128	5.799	-31.874
	64	4.097	-30.361
10	512	6.030	-33.764
	256	5.809	-34.765
	128	5.938	-32.946
	64	3.908	-30.890

Table 4.19: SNRI and SD for Male Speech Signal with Cafeteria Noise at 50% Overlap

Input SNR in dB	Number of Subbands	SNRI in dB	SD in dB
-5	512	3.348	-31.420
	256	3.096	-30.968
	128	2.572	-30.114
	64	0.747	-29.483
0	512	4.569	-32.200
	256	4.646	-32.084
	128	4.703	-30.859
	64	2.990	-29.851
5	512	5.481	-33.026
	256	5.493	-33.439
	128	5.793	-31.863
	64	4.092	-30.356
10	512	5.481	-33.764
	256	5.493	-34.766
	128	5.793	-32.946
	64	4.092	-30.890

Table 4.20: SNRI and SD for Female Speech Signal with Cafeteria Noise at 50% Overlap

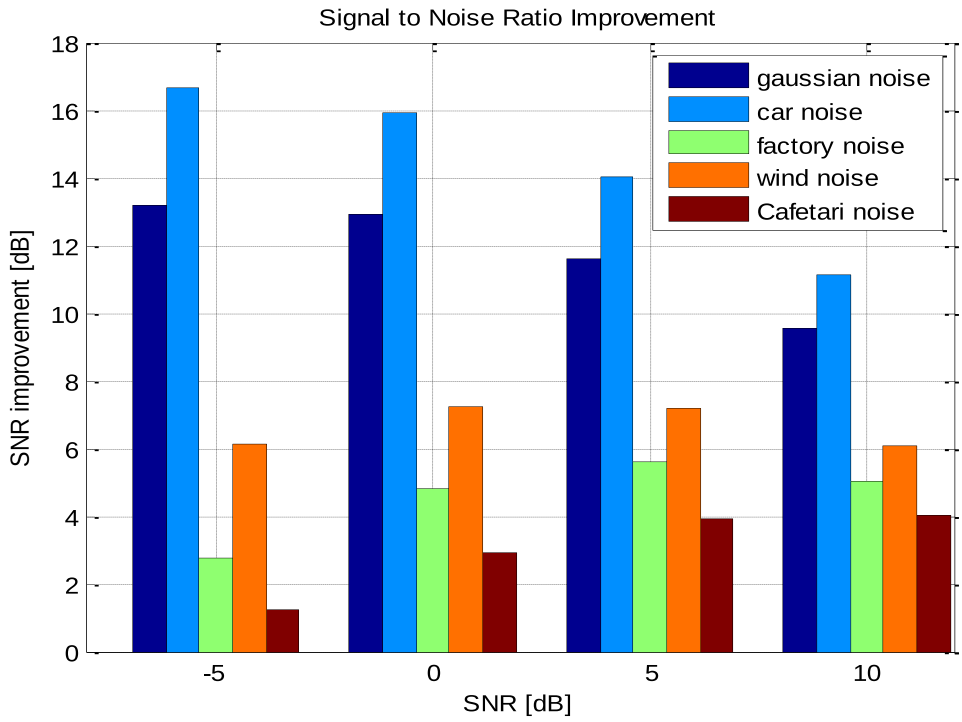

Figure 4.10: Average SNRI Using Male Speech Signal with 75% Overlap.

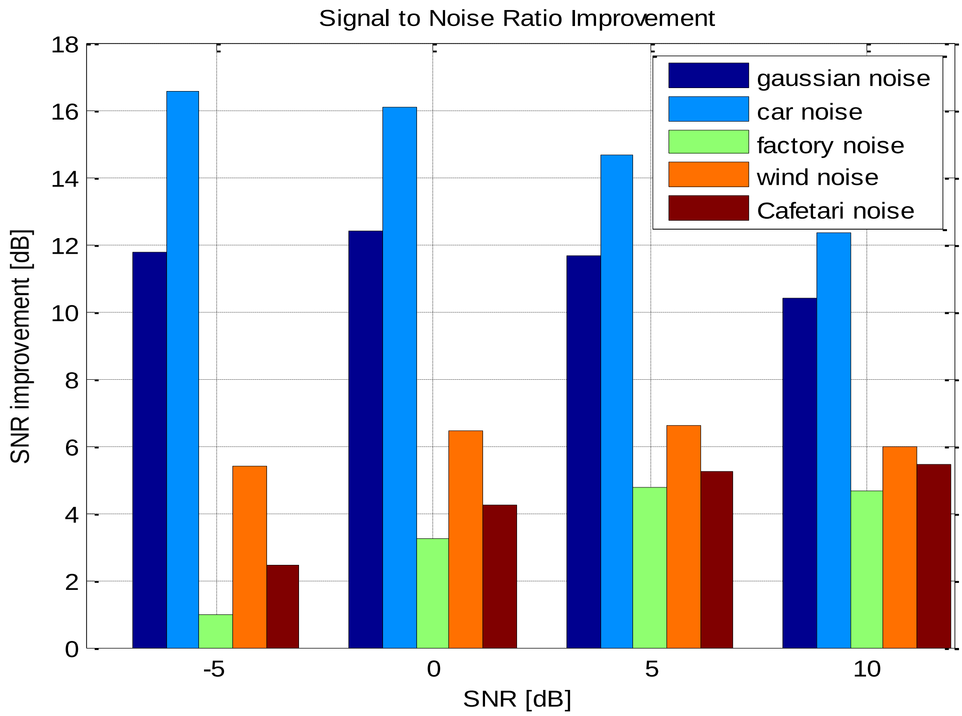

Figure 4.11: Average SNRI Using Male Speech Signal with 50% Overlap.

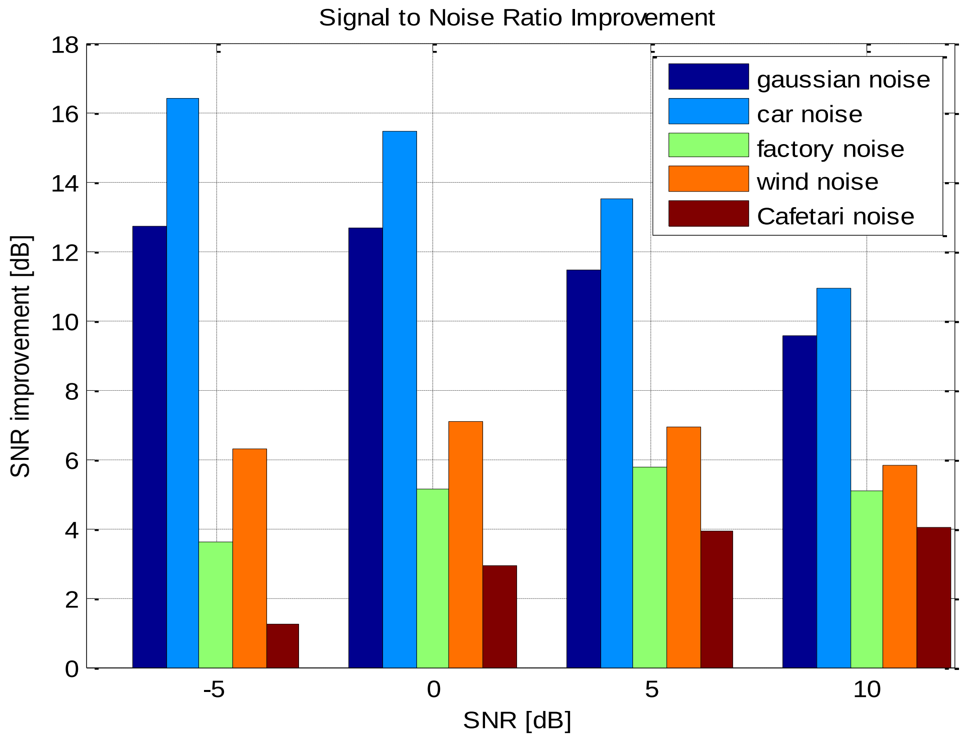

Figure 4.12: Average SNRI Using Female Speech Signal with 75% Overlap.

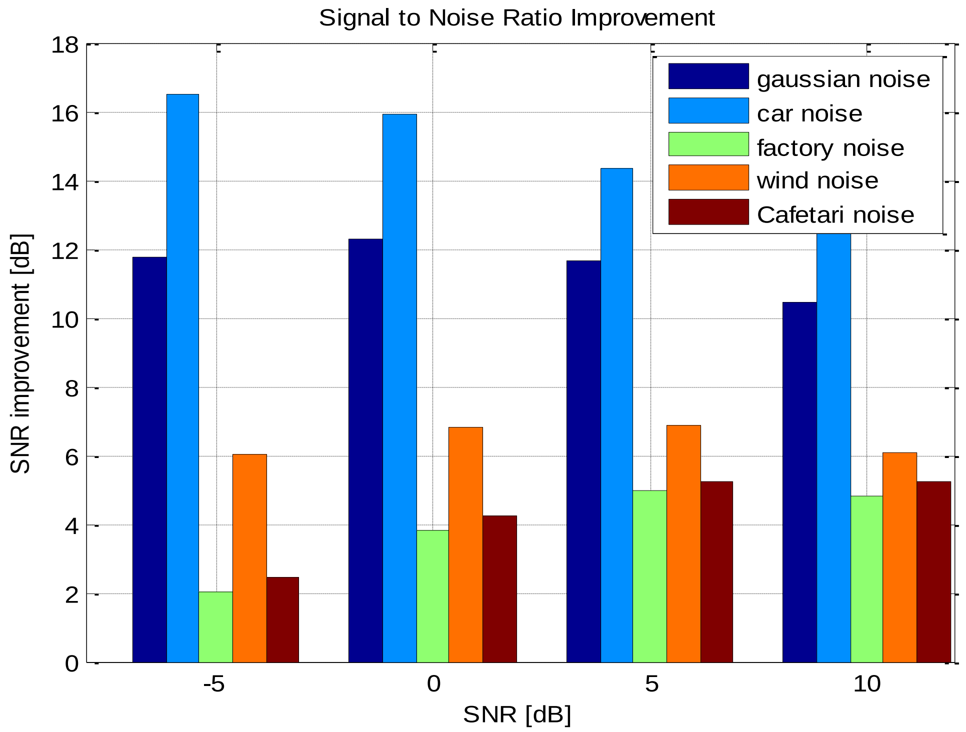

Figure 4.13: Average SNRI Using Female Speech Signal with 50% Overlap.

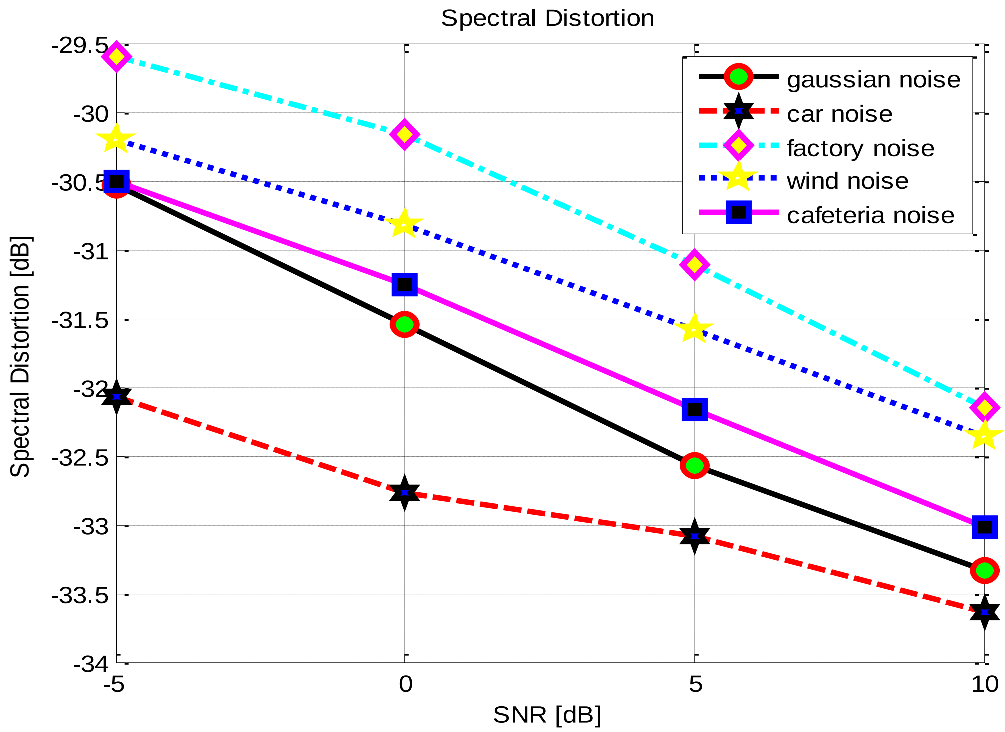

Figure 4.14: Average Spectral Distortion for Male Speech Signal.

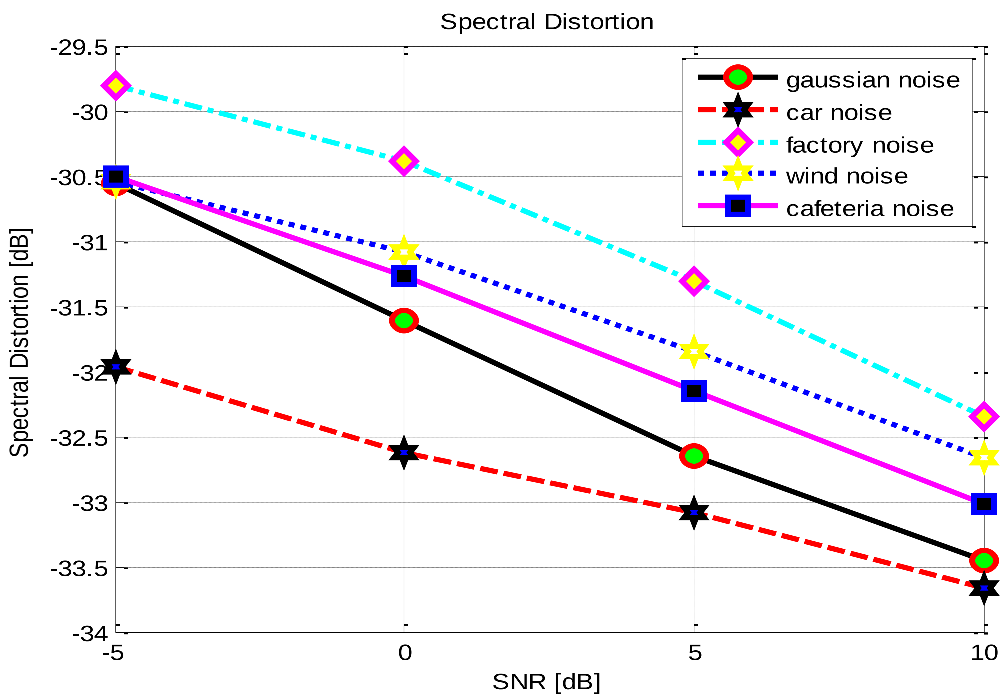

Figure 4.15: Average Spectral Distortion for Female Speech Signal

4.4 Computational Complexity

Consider a signal $x(n)$ of length N. This is then divided into number of short time signal for analysis; each short time signal (each frame) carries some part of the previous signal because of time overlapping. The time overlapping is depending on the decimation/interpolation ratio R. So, the length of the each short time signal is based on the decimation/interpolation ratio and the data window $h(n)$. Suppose, L is the short time signal length with decimation/interpolation ratio R.

Now, the computational complexity of the Spectral Subtraction Based on Minimum Statistics algorithm for each frame is given in Table 4.21

DFT and IDFT Matrix Calculation	Multiplication, $(12 \times L^2)$ Division, $(5 \times L^2)$ Addition, $(10 \times L^2)$
DFT Calculation	Multiplication, $(1 \times L^2)$ Addition, $(L - 1) \times 1 \times L$
Angle Calculation	Addition, $(1 \times L)$ Square, $(2 \times 1 \times L)$ Square Root, $(1 \times L)$
Magnitude Calculation	Division, $(1 \times L)$
Magnitude Square Calculation	Multiplication, $(1 \times L)$
Short Time Signal Power and Short Time Subband Signal Power Calculation	Multiplication, $(8 \times 1 \times L)$ Addition, $(2 \times 1 \times L + 2)$
Minimum Power Calculation	Multiplication, $(2 \times 1 \times L)$ Addition, $(4 \times 1 \times L)$
Noise Power Calculation	Multiplication, $(1 \times L)$
SNR Calculation	Multiplication, $(1 \times L)$ Addition, $(1 \times L)$ Division, $(1 \times L)$
Oversubtraction Factor Calculation	Addition, $((1 \times L)/2)$ Division, $((1 \times L)/2)$ Multiplication, $((1 \times L)/2)$
Q Calculation	Multiplication, $(1 \times L)$ Addition, $(1 \times L)$ Division, $(1 \times L)$

Improve Magnitude Calculation	Multiplication, $(3 \times 1 \times L)$ Square Root, $(2 \times 1 \times L)$
Adding Angle with Improve Magnitude Calculation	Multiplication, $(2 \times 1 \times L)$ Addition, $(1 \times L)$
IDFT Calculation	Multiplication, $(2 \times (1 \times L^2 + 1 + 1 \times L))$ Addition, $((L-1) \times 1 \times L)$
Overlap Add	Multiplication, (9×1) Addition, (4×1)

Table 4.21 Computation Complexity of SSBMS Algorithm

Total numbers of Multiplication, Division, Addition, Square and Square Root for each frame are given below,

$$\text{Multiplication, } 13L^2 + \frac{39}{2}L + 13$$

$$\text{Division, } 5L^2 + 2L$$

$$\text{Addition, } 12L^2 + \frac{17}{2}L + 6$$

$$\text{Square, } 2L$$

$$\text{Square Root, } 3L$$

Now, the total number of computational complexity for SSBMS algorithm in each sample

$$30L + \frac{19}{L} + 35$$

Chapter 5

Conclusion

In this thesis we have worked on noisy speech signal to enhance the speech. The SSBMS algorithm is successfully implemented and performance is observed in five different noisy environments. The performance analysis of the system has focused on its advantages and disadvantages i.e. where it gives high SNRI in slow varying noise as compared to non-stationary noise. It is clear that the selection of α and γ create less effect to the SNRI and SD but the selection of $subf$ has comparatively large effect on results. Generally a better SNRI is accompanied by a more SD signal i.e. the system compromises between high SNRI and low SD. After observing the results it is concluded that the SNRI and SD are comparatively better for both 512 and 256 subbands processed with 75% overlap for both male and female speech signals. It is also concluded that low SNR in the input signals gives high SD. The SD values are within -37dB to -29 dB for all the cases and increases linearly with SNR. The system provides good improvement on car noise for the both male and female speech with better SNRI and low SD. The maximum SNRI is achieved at 18 dB for both male and female speech signal for car noise at -5 dB SNR. The SSBMS algorithm also performs well in Gaussian noise, i.e. around 13 dB SNRI. In case of factory noise, wind noise and cafeteria noise the SNRI for the both male and female speech is around 5 dB. SSBMS algorithm is less complex and computationally efficient. This algorithm is successfully implemented and validated. Tables, plots and graphs that are presented in this thesis give the better view of results.

In our thesis, we have simulated SSBMS algorithm in offline mode, and it can be implemented on real-time in the future. The output of the system contains very little background noise, although this noise is not influencing much the intelligibility of the speech but needs to be improved. The performance of the SSBMS algorithm can be compared with other subtractive type algorithms implemented in the single channel for future work.

References

[1] K. Nitish and J. H. L. Hansen, “Babble Noise: Modelling, Analysis, and Applications’’, in *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 7, pp. 1394-1407, September 2009.

[2] J. C. Junqua, “The Influence of Acoustics on Speech Production: A Noise Induced Stress Phenomenon Known as the Lombard Reflex”, in *ESCA-NATO Workshop on Speech under Stress*, pp 83-90, Lisbon, September 1995.

[3] J. C. Junqua, and J. P. Haton, “Robustness in Automatic Speech Recognition: Fundamentals and Applications,” Kluwer A. Publishers, 1996.

[4] J. S. Lim, and A. V. Oppenheim, “Enhancement and Bandwidth Compression of Noisy Speech”, in *Proc. IEEE*, vol. 67, no.12, pp. 1586-1604, December 1979.

[5] N. Virag, “Speech Enhancement Based on Masking Properties of the Humanauditory System”, Master thesis, Swiss Federal Institute of Technology, 1996.

[6] Y. Ephraim, “Statistical Model Based Speech Enhancement Systems”, in *Proc. IEEE*, vol. 80, no. 10, pp. 1526-1555, October 1992.

[7] Y. Gong, “Speech Recognition in Noisy Environments: A survey”, in *Speech Communication*, Elsevier Science B.V., Amsterdam, Netherlands, vol. 16, no.3, pp 261-291, April 1995.

[8] Z. Yermeche, “Subband Beamforming for Speech Enhancement in Hands-Free Communication”, Ph.D. Dissertation Series No. 2007:14, Blekinge Institute of Technology, 2004.

[9] S. Haykin. “Adaptive Filter Theory”, Prentice-Hall, Englewood Cliffs, NJ, 2nd edition, 1991.

[10] B.Widrow, and S. D. Stearns, “Adaptive signal processing”, Prentice Hall, 1985.

[11] Te. –W. Lee and H. Sawada, “Blind Speech Separation’’, Springer, Netherlands, 2007

[12] R. Martin, “Spectral Subtraction Based on Minimum Statistics”, in Proc. EUSPICO’94, pp. 1181-1185, 1994.

[13] S. F. Boll, “Suppression of Acoustic Noise in Speech Using Spectral Subtraction”, in *IEEE Transactions on Acoustic, Speech, and Signal Processing*, vol. ASSP-27, no. 2, April 1979.

[14] R. Martin, “An Efficient Algorithm to Estimate the Instantaneous SNR of Speech Signals”, in *Proc. EU-ROSPEECH ‘93*, pp. 1093-1096, Berlin, September 21-23, 1993.

[15] R. Crochiere and L. Rabiner, “Multirate Digital Signal Processing”, Prentice Hall, 1983.

[16] X. Serra and J. O. Smith, “Spectral Modelling Synthesis: A Sound Analysis/Synthesis System Based on a Deterministic Plus Stochastic Decomposition”. in *Computer Music Journal*, 1990.

[17] C. Roads, “Microsound”, MIT Press, 2001.

[18] P. D. Welch, “The Use of Fast Fourier Transform for the Estimator of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms”, in *IEEE Trans. Audio and Electroacoustic,* vol. AU-15, pp. 70-73, June 1967.

[19] A. Papoulis, “Probability, Random Variables, and Stochastic Processes”, 2nd ed., McGraw-Hill, 1984.

[20] M. Berouti, R. Schwartz, and J. Makhoul, “Enhancement of Speech Corrupted by Acoustic Noise”, in *Proc. IEEE Conf. ASSP*, pp. 208-211, April 1979.

[21] T. Tolonen, V. V. A. aki, and M. Karjalainen, “Evaluation of Modern Sound Synthesis Methods’’, Technical Report 48, Helsinki Institute of Technology, March 1998.

[22] P. Chris, “Two-Dimensional Fourier Processing of Rasterised Audio’’, June 13, 2008. Available on http://www-users.york.ac.uk/~jjw100/report.pdf

[23] http://www.itu.int/net/itu-t/sigdb/genaudio/Pseries.htm

[24] B. Mukul, "A Modified Spectral Subtraction Method Combined with Perceptual Weighting for Speech Enhancement'', Master thesis, The University of Texas at Dallas, August 2002.

[25] D. K. Sunil, "A Multi-Band Spectral Subtraction Method for Speech Enhancement'', Master thesis, The University of Texas at Dallas, December 2001.

[26] P. Krishnamoorthy, "Combined Temporal and Spectral Processing Methods for Speech Enhancement'', Research Scholar, Indian Institute of Technology Guwahati, Assam, India.

[27] http://a3lab.dibet.univpm.it/research/82

[28] M. Awais, "Multichannel Wiener Filtering for Speech Enhancement in Modulation Domain'', Master thesis, Blekinge Institute of Technology, 2010.