

COMPARISON OF NOISE REDUCTION FILTERS FOR SPEECH ENHANCEMENT

A Thesis Submitted in Fulfillment of the Requirement for the Award of the Degree of

MASTER OF ENGINEERING

In

Wireless Communication

Submitted By

Neha

801563016

Under Supervision of

Dr. Ashutosh Singh

Assistant Professor



ELECTRONICS AND COMMUNICATION ENGINEERING DEPARTMENT

THAPAR UNIVERSITY, PATIALA, PUNJAB

JULY, 2017

DECLARATION

I, Neha, hereby declare that the work presented in this thesis entitled “**Comparison of Noise Reduction Filters for Speech Enhancement**” in fulfillment of the requirement for the award of degree of **Master of Engineering in Wireless Communication** submitted at Electronics and Communication department, Thapar University, Patiala is an authentic record of work carried out under supervision of **Dr. Ashutosh Kumar Singh**, Assistant Professor, **Department of Electronics and Communication Engineering, Thapar University, Patiala, Punjab**. The matter presented in this thesis has not been submitted either in part or full to any other university or institute for the award of any other degree.

Date: 08-Sept-2017

Neha

Neha

801563016

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

Date: 08-Sep-2017

Singh

Dr. Ashutosh Kumar Singh

Assistant Professor

ECED, TU, Patiala

ACKNOWLEDGEMENT

With deep sense of gratitude I express my sincere thanks to my esteemed and worthy supervisor, **Dr. Ashutosh Singh, Assistant Professor**, Department of Electronics and Communication Engineering, Thapar University, Patiala for his valuable guidance in carrying out work under his effective supervision, encouragement, enlightenment and cooperation. Most of the novel ideas and solutions found in this dissertation are the result of our numerous stimulating discussions.

I shall be failing in my duties if I do not express my deep sense of gratitude towards **Dr. Alpana Agarwal**, Professor and Head of the Department of Electronics and Communication Engineering, Thapar University, Patiala who has been a constant source of inspiration for me throughout this work, and for the providing us with adequate infrastructure in carrying the work.

I am also thankful to **Dr. Hem Dutt Joshi**, Assistant Professor and P.G. Coordinator, Electronics and Communication Engineering Department, for the motivation and inspiration that triggered me for this work.

I am greatly indebted to all my friends who constantly encouraged me and also would like to thank the entire faculty and staff members of Electronics and Communication Engineering Department for their unyielding encouragement.

At last but not the least my gratitude towards my parents, who always supported me in doing the things my way and whose everlasting desires, selfless sacrifice, encouragement, affectionate blessings and help made it possible for me to complete my degree.

NEHA

801563016

ABSTRACT

Speech is one of the best and easy ways for communication. In speech communication, the speech signal is affected by some background noises, additive white Gaussian noise (AWGN) as well as distortions from Rayleigh and Rician channel. The performance of speech processing is reduced by background noise which makes it difficult to listen to the speech and is the major reason for speech quality degradation.

Noise reduction techniques play a key role in removing the artifacts. These noise reduction techniques are used for separating the clean speech from noisy observations. In this thesis, two algorithms are used to eliminate the noises in single channel speech enhancement such as; spectral subtraction and Wiener filter. The main motive of these algorithms is to reduce the level of noise and improve the corresponding signal-to-noise ratio (SNR) at different power levels of input SNR. This paper involves the theoretical and practical analysis of these algorithms.

In which wiener filter is compared with the spectral subtraction algorithm based on SNR improvement introduced by them. These noise reduction techniques show that output SNR is more than the input SNR. This means, filters are able to remove the noise from noisy speech. In whole, experimental results show that SNR improvement of the Wiener filter is much better than the implementation of spectral subtraction.

TABLE OF CONTENTS

Sr. No	Name of the Chapters	Page No
	<i>Declaration</i>	<i>II</i>
	<i>Acknowledgement</i>	<i>III</i>
	<i>Abstract</i>	<i>IV</i>
	<i>Table of contents</i>	<i>V</i>
	<i>List of tables</i>	<i>VII</i>
	<i>List of figures</i>	<i>VIII</i>
	<i>List of Abbreviations</i>	<i>X</i>
<i>Chapter 1</i>	<i>Introduction</i>	<i>1-9</i>
	1.1 <i>Speech processing</i>	<i>1</i>
	1.1.1 <i>The Speech Chain</i>	<i>1</i>
	1.1.2 <i>Application of Speech Processing</i>	<i>2</i>
	1.2 <i>Speech Enhancement</i>	<i>4</i>
	1.3 <i>Application of Speech Enhancement</i>	<i>6</i>
	1.4 <i>Single Channel Enhancement Method</i>	<i>8</i>
	1.4.1 <i>Spectral Subtraction</i>	<i>8</i>
	1.4.2 <i>Wiener Filter</i>	<i>8</i>
	1.5 <i>Organization of Thesis</i>	<i>9</i>
<i>Chapter 2</i>	<i>Literature Survey</i>	<i>10-17</i>
	2.1 <i>Speech Enhancement Using Spectral Subtraction</i>	<i>10</i>
	2.2 <i>Optimal Linear Filtering Technique</i>	<i>11</i>
	2.3 <i>Improved SNR Estimation For Speech Enhancement</i>	<i>12</i>
	2.4 <i>Causal and Non-Causal Filters for Noise Reduction in Time-Domain</i>	<i>12</i>
	2.5 <i>A Priori Estimation</i>	<i>13</i>
	2.6 <i>Power Spectral Density (PSD) Estimation</i>	<i>14</i>
	2.7 <i>Estimator Analysis</i>	<i>14</i>
	2.8 <i>Channel Characteristics</i>	<i>14</i>
	2.8.1 <i>AWGN Channel</i>	<i>15</i>
	2.8.2 <i>Rayleigh Fading Channel</i>	<i>16</i>
	2.8.3 <i>Rician Fading Channel</i>	<i>16</i>
	2.9 <i>Short Time Fourier Transform (STFT) Analysis</i>	<i>17</i>
	2.10 <i>Problem Statement</i>	<i>17</i>

<i>Chapter 3</i>	Noise Reduction Technique Based on Spectral Subtraction.....	18-26
3.1	Process of Spectral Subtraction.....	18
3.2	Additive Noise Model.....	21
3.3	Spectral Subtraction Estimator and Spectral Error.....	21
3.4	Half-Wave rectification and Residual Noise Reduction.....	22
3.5	Synthesis.....	22
3.6	Simulation Results.....	22
<i>Chapter 4</i>	Noise Reduction Filter Based on Wiener Filter.....	27-36
4.1	Signal Model.....	27
4.3	Introduction To Noise Reduction Parameters.....	28
4.4	Decision Directed (DD) Approach.....	28
4.5	Two Step Noise Reduction (TSNR) Approach.....	29
4.6	Harmonic Regeneration Noise Reduction (HRNR) Approach.....	31
4.7	Simulation Results.....	32
<i>Chapter 5</i>	Results and Discussion.....	37-41
5.1	Performance Measure.....	36
5.2	Comparative Analysis of Both Filters in Different Channels.....	36
<i>Chapter 6</i>	Conclusion.....	42
	List of publications.....	43
	References.....	44-47

LISTS OF TABLES

Sr. No	Table Details	Page No
Table 5.1	<i>Performance of Spectral subtraction and Wiener filter in AWGN channel.....</i>	38
Table 5.2	<i>Performance of Spectral subtraction and Wiener filter in Rayleigh channel.....</i>	39
Table 5.3	<i>Performance of Spectral subtraction and wiener filter in Rician channel.....</i>	40

LISTS OF FIGURES

Sr. No	Figure Details	Page No
<i>Figure 1.1</i>	<i>Speech Chain.....</i>	2
<i>Figure 1.2</i>	<i>A Block diagram of Speech Coding.....</i>	3
<i>Figure 1.3</i>	<i>Block diagram of speech Synthesis.....</i>	3
<i>Figure 1.4</i>	<i>A Block Diagram of Speech Recognition.....</i>	4
<i>Figure 1.5</i>	<i>Sources of Noise.....</i>	5
<i>Figure 1.6</i>	<i>Speech Enhancement Applications.....</i>	7
<i>Figure 1.7</i>	<i>Optimal Wiener Filter Model.....</i>	8
<i>Figure 2.1</i>	<i>AWGN Model.....</i>	15
<i>Figure 3.1</i>	<i>Block diagram of Spectral Subtraction.....</i>	19
<i>Figure 3.2</i>	<i>Flow chart of spectral subtraction method.....</i>	20
<i>Figure 3.3</i>	<i>Spectrograms of Spectral subtraction algorithm using male voice in AWGN channel with SNR 2dB (a) Original signal (b) Noisy speech signal (c) reconstructed signal using Spectral subtraction.....</i>	23
<i>Figure 3.4</i>	<i>Spectrograms of Spectral subtraction algorithm in Rayleigh channel using male voice with SNR 5dB (a) Original signal (b) Noisy speech signal (c) reconstructed signal using Spectral subtraction.....</i>	24
<i>Figure 3.5</i>	<i>Spectrograms of Spectral subtraction algorithm in Rician channel using female voice with SNR 7dB (a) Original signal (b) Noisy speech signal (c) reconstructed signal using Spectral subtraction.....</i>	25
<i>Figure 3.6</i>	<i>Spectrograms of Spectral subtraction algorithm using female voice in AWGN channel with SNR 10dB (a) Original signal (b) Noisy speech signal (c) reconstructed signal using Spectral subtraction.....</i>	26
<i>Figure 4.1</i>	<i>Spectrograms of Wiener filter using male voice in AWGN channel with SNR 2dB (a) Original speech signal (b) Noisy speech signal (c) Enhanced speech signal using TSNR technique (d) Enhanced speech signal using HRNR technique.....</i>	33
<i>Figure 4.2</i>	<i>Spectrograms of Wiener filter using male voice in Rayleigh Channel with SNR 5dB (a) Original speech signal (b) Noisy speech signal (c) Enhanced speech signal using TSNR technique (d) Enhanced speech signal using HRNR technique.....</i>	34
<i>Figure 4.3</i>	<i>Spectrograms of Wiener filter using female voice in Rician Channel with SNR 7dB (a) Original speech signal (b) Noisy speech signal (c) Enhanced</i>	

speech signal using TSNR technique (d) Enhanced speech signal using HRNR technique..... 35

Figure 4.4 Spectrograms of Wiener filter using female voice in AWGN Channel with SNR 10dB (a) Original speech signal (b) Noisy speech signal (c) Enhanced speech signal using TSNR technique (d) Enhanced speech signal using HRNR technique..... 36

LIST OF ABBREVIATIONS

<i>DSP</i>	<i>Digital signal Processing</i>
<i>SNR</i>	<i>Signal to Noise Ratio</i>
<i>TTS</i>	<i>Text-to-Speech</i>
<i>STSA</i>	<i>Short Time Spectral Amplitude</i>
<i>MSE</i>	<i>Mean Square Error</i>
<i>DD</i>	<i>Decision Directed</i>
<i>TSNR</i>	<i>Two-Step Noise Reduction</i>
<i>HRNR</i>	<i>Harmonic regeneration Noise Reduction</i>
<i>AWGN</i>	<i>Additive White Gaussian Noise</i>
<i>ML</i>	<i>Maximum Likelihood</i>
<i>STFT</i>	<i>Short Time Fourier Transform</i>
<i>LOS</i>	<i>Line-of-Sight</i>
<i>MMSE</i>	<i>Minimum Mean Square Estimator</i>

CHAPTER 1

INTRODUCTION

In this chapter, a brief introduction of speech processing is provided. It deals with the transmission, processing and reception of speech signals. Further, the applications and advantages have been described in brief. In practical situations, a speech signal is prone to noise and interference while passing through a channel. It is necessary to remove any kind of distortion in the speech signal. To achieve this, various techniques like speech enhancement etc. are used. Such techniques and methods have been discussed here briefly but comprehensively.

1.1. SPEECH PROCESSING

Speech is one of the most elementary forms of information. Human interact with everyone in many ways like eye contact, facial expression, gesture, most importantly speech. Therefore, speech is the simple way to communicate with one another and systematic form of exchanging information. There are many different ways to communicate by speech in the form of sentence, words and phrases by using some semantic rules [1]. With the help of telephone, it becomes easiest to communicate with each other over long distance using their original speech instead of text conversations. When the speech information is transmitted over the transmission medium, it is necessary to process the acoustic data. Speech processing is defined as the analysis of speech signals and the processing method of these signals. Speech processing can be considered as the specific case of digital signal processing (DSP). Then, speech processing inherits pictures. Speech processing changes the speech signal into digital signal. Digital signal is the discrete signal which has series of numbers. Therefore, the signal can be employed in various applications. Speech processing is a field of digital signal processing where the signals are processed in their digital format. In general, the speech processing requires the acquisition of the speech signal, operating the signal, storing it for further processing; transmit to the target point and finally generates the output [2]. The exchange information can be examined at distinct levels. At the phonetics level, it deals with the properties of sound signals. The acoustic level is a branch of physical science. Morphology studies the formation and composition of words. The phonological level is the essential interface between phonetics and linguistically representation of higher levels. The semantically level is related with how meaning is deduced from words and concepts. At last, pragmatic studies about how meaning is deduced from context [3]. Therefore, speech is one of the most intriguing signals that humans work daily.

1.1.1. The Speech Chain

To acknowledge the nature of the speech signal and how DSP methods, with communication technologies, can be used to help resolve the variety of application scenarios. However, the speech chain describes the levels in speech communication when a message moves between the speaker and the listener. The speech chain contains; speech production, auditory feedback to the speaker, speech

transmission (over electronic communication system or through air), speech perception and understanding by listener as shown in Figure 1.1. The levels of speech chain are [4]

- **Linguistic level** (from speaker side) - where transmission of messages with the selection and organizing of appropriate words, phrases and sentences.
- **Physiological level** (from speaker side) – where the vocal track elements generate the sounds along with the linguistic units of the utterance.
- **Acoustic level** – in which sound is freed from the mouth and transferred to both the speaker and to the listener.

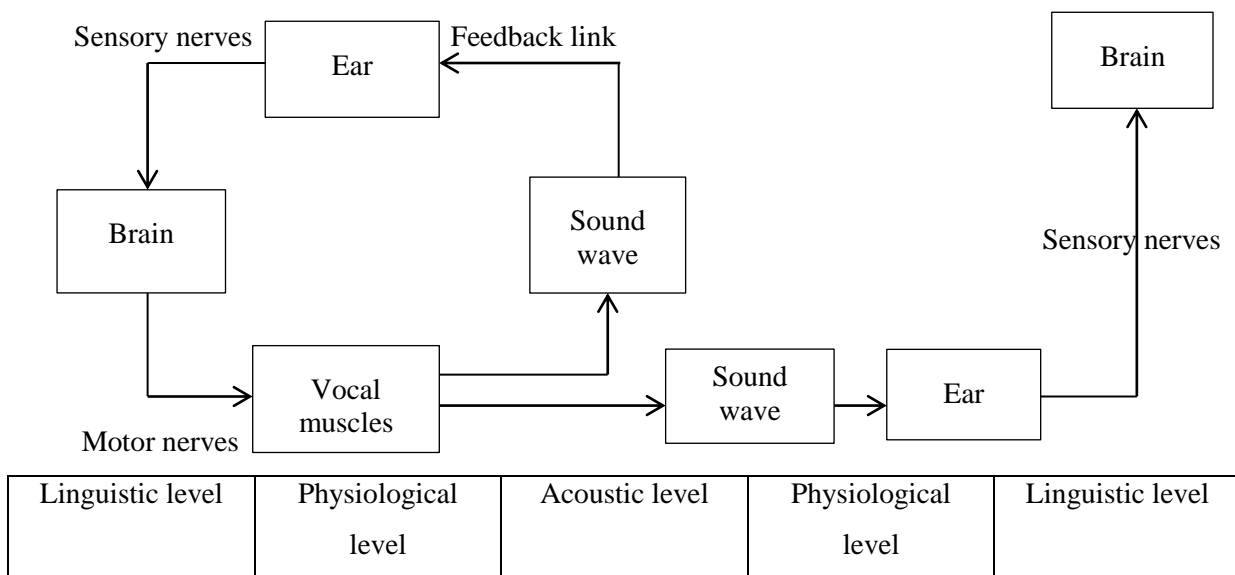


Figure 1.1 Speech chain [4]

- **Physiological level** (from listener side) – where the sound is examined by the ear and the auditory nerves.
- **Linguistic level** (from listener side) – the speech is received as a series of linguistic units and then acknowledged the words, phrases and sentence.

1.1.2. Application of Speech Processing

The speech processing has a variety of applications are given below.

- **Speech coding** – It is the process of compressing the digital audio data without any loss of quality and intelligibility. The main objective of the speech coding is to reduce the bit rate and preserve the level of speech signal. It uses the speech specific parameter with data compression algorithm to produce the modeled parameters in compress bit stream. It takes the benefits of properties of speech production and perception. Speech coding is mainly used in security, teleconferences, reduction in bit rate, removal of noise, in narrowband communication, channel cellular communication etc. [5].

The Figure 1.2 shows the block diagram of speech coding. According to this diagram, first analog to digital converter (ADC) is employed to convert an analog signal to digital representation. Then digital signal is analyzed and encoded. This encoded data transmitted over channel and give the final encoded signal. Using this inverse of analysis processing, the digital signal is synthesized and converted into analog signal. And then decoded signal is takes place [6].

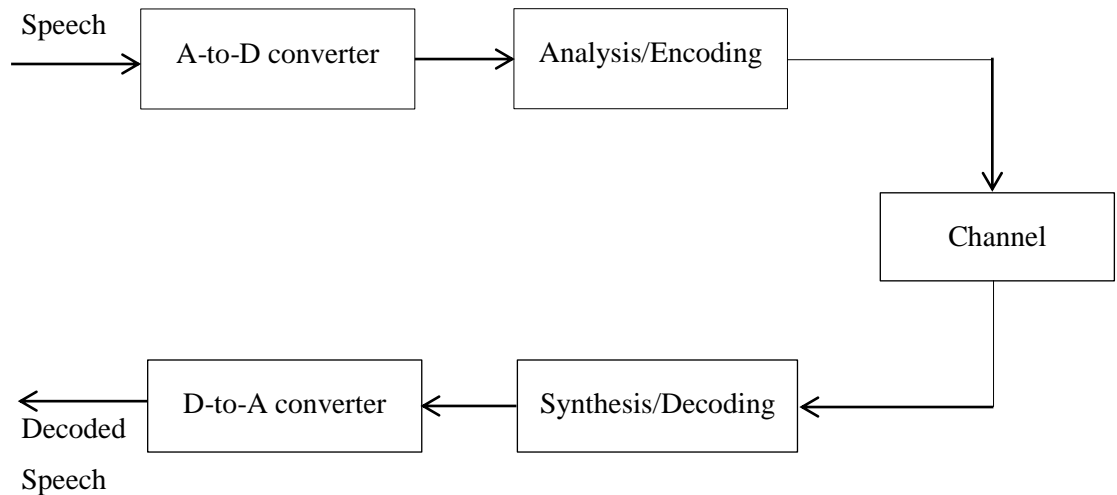


Figure 1.2 A block diagram of speech coding [6]

- **Speech synthesis** – It is also called as Text-to-Speech (TTS). Speech synthesis is the process of producing a speech signal of an unknown text. In other words, it is the artificial production of human speech. It converts the linguistic representation into a complete speech. These systems are important part of human machine communication systems such as email access, alarm system, telephone based information in air travel, automatic transactions, handheld devices etc. [5].

The Figure 1.3 illustrates the block diagram of speech synthesis. In the first stage, it maps the input text into a basic form, and evaluates the structure of input. And then determine the pronunciation of the words. In the last stage, it converts the symbolic representation into real speech signal [6].

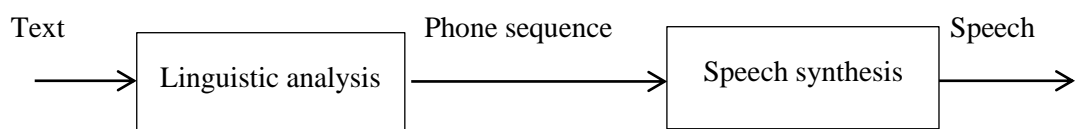


Figure 1.3 Block diagram of speech synthesis [6]

- **Speech recognition** – It is the process of extracting usable linguistic information from speech. It is a duty for machine to recognize and understand the speech signals [5]. The speech recognition applications include command and control (C&C) applications, agent services such as calendar entry and update, voice dictation to creates memos, documents etc.

The Figure 1.4 represents the block diagram of speech recognition. Using A-to-D converter, the system converts analog signal into digital form and then feature analysis converts the digital signal into a set of features. The last block named as pattern matching, dynamically aligns feature vector with a continual set of recorded patterns, and selects the identity related with the pattern which is the perfectly matched to the dynamically aligned feature vector of the speech signal [6]. The process is known as speech recognition.

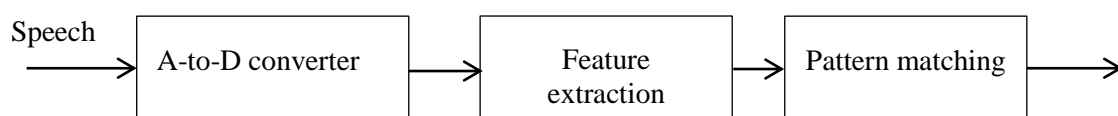


Figure 1.4 A block diagram of speech recognition [6]

- **Speaker recognition** – speaker recognition is used for recognize a people from a speaking expression. It includes the two subfields: speaker verification and speaker identification [5], [6].
- **Speech enhancement** – Speech enhancement is used in noisy environment. The goal of the speech enhancement is to decrease the level of noise from noisy speech signal and enhance the quality and intelligibility of the speech. There are many techniques are used for speech enhancement in single channel and multiple channel [5], [6].
- **Language translation** – It converts the one form of language into another form of language [5], [6].

1.2. SPEECH ENHANCEMENT

Speech signals are one of the necessary and easiest ways for communication between human beings. Speech communications are used in our daily lifestyle. The speech communication includes a speaker, a listener, and a variety of communication conveniences. Speech communications occurs all over, like domestic homes, seminars, cocktail parties, work, medical appointments, and school conferences. When both speaker and listener are close to each other in a silent medium, communication is generally secure and precise. Usually, at a distance, random environmental noises degrade the quality of speech communication between them. Noises exist all over, and are developed by many factors; therefore it is difficult to recognize them all [7]. Noise can be categorized into additive noise and non-additive noise. The additive noise contains broadband noise, pulse noise and periodical noise. The periodical noise generated from engine and pulse noise generated from discharge and explosion. Also broadband noise contains heat noise, quantization noise, and random

noises like pink noise and white noise. On the other hand, non-additive contains convolution noise and multiplier noises which can be converted into additive noise by homomorphic transform [8]. The features of the noises are either well known or unknown; though, they all can disrupt, distort, or degrade the status of speech signals. Therefore, noisy background is probable to affect people with hearing loss. The field of research that probes eliminating noise from distorted speech with using distinct signal processing techniques is known as speech processing [7].

Noise reduction/speech enhancement, is a problem of retrieve a clean speech spectrum from the speaker examination corrupted by some degree of background noise such as fan noise, car noise and other interfering speakers through improving the SNR to produce the speech signals with a greater perceptual quality [9]. Figure 1.5 shows the block diagram of different sources of noise. In this figure, the noise speech signal is combination of background noise, desired speaker, transmission noise and other speaker noise. The noisy speech signal passes through speech enhancement algorithm to obtain a clean speech signal, which is equal to the original speech signal and then easily audible to listener.

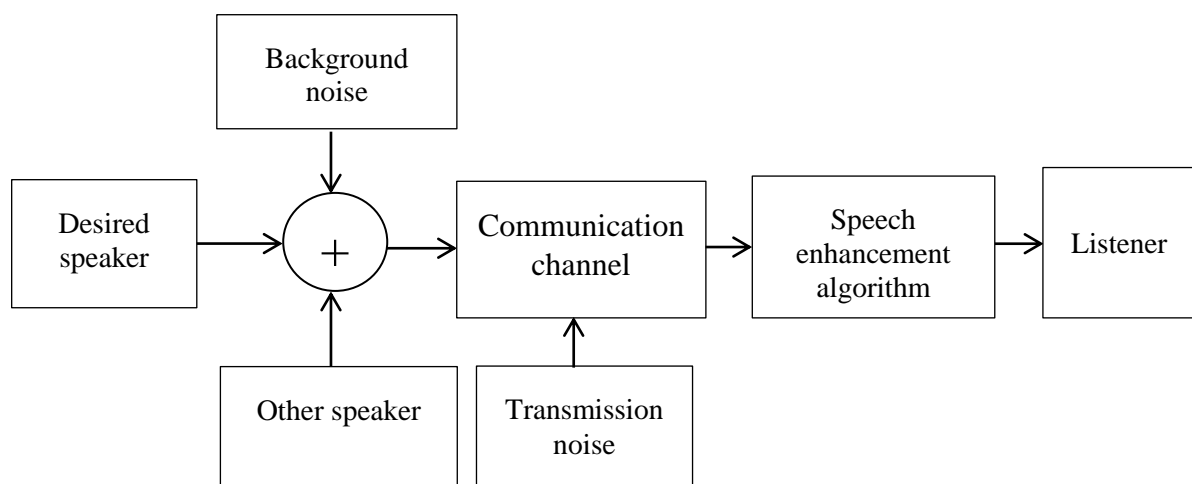


Figure 1.5 Sources of noise [10]

Depending on the SNR, disturbances make a conversation uncomfortable in the worst case. In order to manage with such acoustic environments, speech enhancement algorithms helps in reducing background noises, echoes, Amplifier noise, quantization noise, or reverberation without affecting the speech signal by means of digital signal processing [11]. The improvement of SNR, intelligibility, and computational complexity are key criterion for the execution of the noise reduction system. However, no speech enhancement systems can refine both speech intelligibility and quality. Speech intelligibility can be considered as a feature of quality, after all good quality speech regularly provides good intelligibility and unintelligible speech would not be categorized as having good quality [12].

Generally, the speech relevant method of noise reduction divided into different classes, i.e., spectral subtraction algorithms, subspace algorithms, and statistical based algorithms [13].

- **Spectral Subtraction algorithm** is one of the standard methods used for suppress the noise. It operates in the frequency domain by removing the background noise from each spectral band which coincides to the noise contribution [12].
- **Subspace algorithms** operate in the autocorrelation domain. In this algorithm, speech and noise components are considered as orthogonal, so that they can be freely separated. But in this algorithmic design, to finding the orthogonal components have become computationally more costly [14].
- **In statistical based algorithm** minimum mean squared error algorithms include Wiener filter and Short time spectral amplitude (STSA) estimator. In which Wiener filter method works on removal of noise component. On the other hand, in the time-domain, the noise reduction problem can be produced as a linear filtering technique either on a block basis or on a sample basis. In both conditions, the greatest crucial issue of noise reduction is to design an optimal filter that can notably reduce the impact of noise with keeping the filtered speech signal close to its original speech signal [9].

1.3 APPLICATION OF SPEECH ENHANCEMENT

Noise reduction, sometimes called as Speech enhancement, is an important vital signal processing problem and it deals with processing of noisy signals, aiming at improving their quality and intelligibility [9]. Clean speech signals corrupted by different types of noises, which causes degrade in the speech quality and intelligibility. And then cause listeners to suffer from fatigue. Furthermore, speech processing systems are often designed under the belief that only a single, clean signal is present at the time. That's why noise reduction plays an important act in many communication and speech processing systems. On the other hand, the quality of speech indicates how a speaker communicates an announcement and consists of such feature like naturalness and speaker recognisability. Intelligibility is concerned with what the speaker had said, i.e., the meaning or information content beyond the words. That's why a noisy domain degrades the speaker and listeners capability to communicate. To decrease the influence of this issue speech enhancement can be executed [10].

The noise reduction filters are effective for upgrading any type of desired signal. The desired signal is speech which is employed in many applications like telecommunications, mobile phones, multiparty teleconferencing, gearing aids, automatic speech recognition, and human-machine interfaces and hearing-aids [15]. Figure 1.6 represents the applications of speech enhancement. The speech has to be recorded using number of speakers, and the speech will be corrupted by noise such as, fan noise, car noise and other interfering speakers etc. then the noise will degrade the speech quality and intelligibility, which has a harmful impact on speech applications.

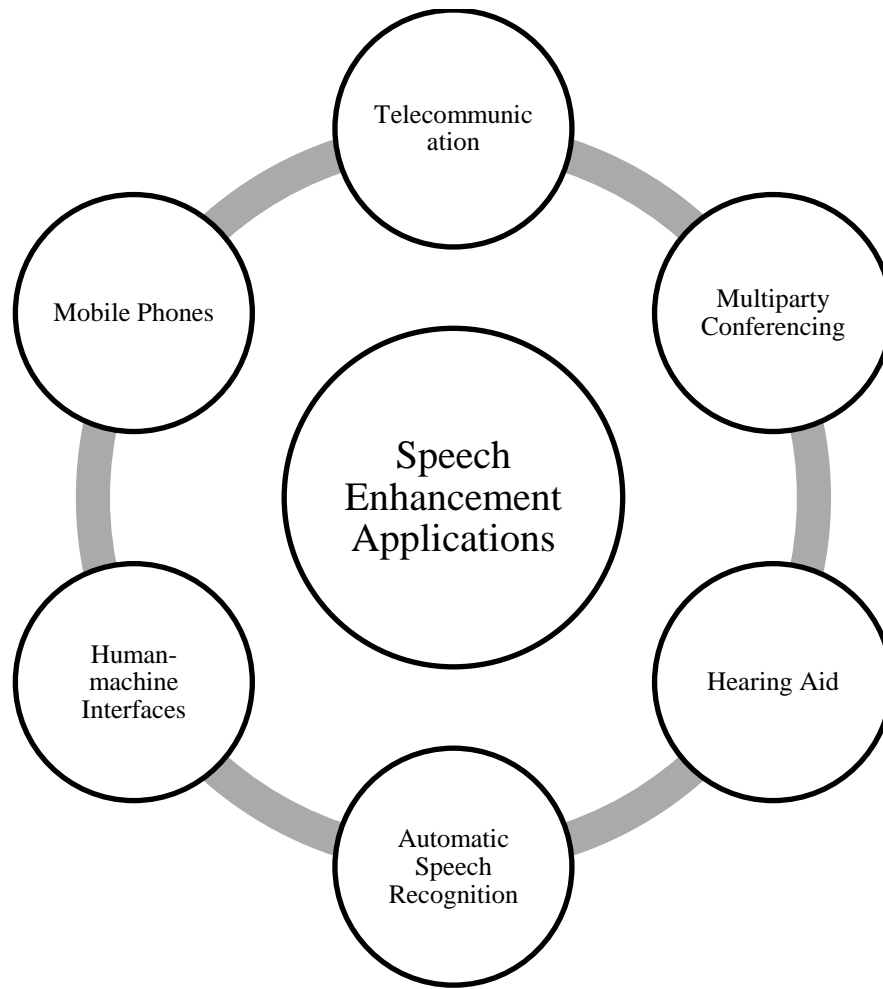


Figure 1.6 Speech Enhancement Applications

In multiparty conferencing, e.g., the background noise received by mike at each position of the conference merges at the main bridge with other signals. Then the loudspeaker at the each position of the conference produces sum of the all noises. That's why, without noise reduction, no communication takes place [16]. In hearing-aids, e.g., degraded speech quality can cause listener fatigue. Therefore, it is necessary to establish noise reduction techniques to decrease the influence of the noise in signal processing applications [15]. Therefore speech enhancement is essential to avoid the reduction of speech quality and outdo the constraints of human auditory systems.

1.4 SINGLE CHANNEL SPEECH ENHANCEMENT METHOD

The single channel speech enhancement methods formulate two notions on the observations of noisy signal:

- The clean speech signal and the additive noise signal are uncorrelated.
- Noise statistics change moderate than speech statistics.

Single channel speech enhancement method mainly deals with the applications where a single speaker is used for recording purpose like mobile telephony. In this method, two filter algorithms are used i.e. spectral subtraction algorithm and wiener filter. These techniques provide enhancement in the

quality of corrupted speech. In these filtering techniques, the noise reduction problem is defined as filter design problem where the objective of such filter is to mitigate the noise as much as possible [17].

1.4.1 SPECTRAL SUBTRACTION

Spectral Subtraction is one of the first and easy approaches for single channel speech enhancement. In this technique, the estimated noise spectrum is subtracted from the noisy speech spectrum (clean speech and noise) and then estimated clean speech spectrum. Spectral subtraction suffers from background noise, fan noise, musical noise and speech distortion. Spectral Subtraction technique is deal with transform domain approach, in which signal is converted into frequency domain [7], [12]. The advantage of this transform domain is that it is easy to classify speech and noise components and hence easily noisy components can remove.

1.4.2 WIENER FILTER

In signal processing, the Wiener filter is another approach to reduce the noise in speech signals. This filter method is based on the reducing the mean square error (MSE) among original signal and estimated signal [12]. Using this optimal linear technique, the a posteriori SNR is always greater than or equal to a priori SNR, which means that this technique is useful for clean the noisy speech signals [16]. The block diagram of optimal Wiener filter model is illustrated in Figure 1.8. In this process, the input signal is obtained by sum of original signal and different noise and then passed through Wiener filter. The enhanced signal is obtained with less amount of noise.

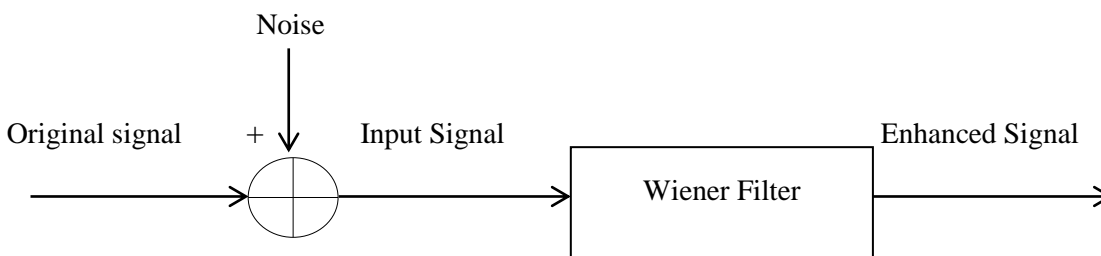


Figure 1.7 Optimal Wiener Filter model [18]

1.5 ORGANIZATION OF THESIS

The organization of this thesis is as follows:

Chapter 2 of this report covers the study of literature by highlighting the points summarized in each paper.

Chapter 3 introduces a brief description on the spectral Subtraction, which is used for reducing the noise in speech spectrum with proper model of spectral subtraction algorithm. In the last section of this chapter introduces the input SNR and Output SNR which are included in performance measure.

Chapter 4 presents another filtering method for reducing the noise and helps in improving the SNR i.e. Wiener filter. This chapter also introduces the noise reduction parameters such as a priori SNR and a

posteriori SNR. In this chapter, we discuss the DD approach with TSNR method and HRNR method for removing reverberation effect and harmonic distortion.

Chapter 5 introduces the results and discussion on spectral subtraction and Wiener filter.

In chapter 6 presents the conclusion of this thesis report.

CHAPTER 2

LITERATURE SURVEY

This chapter briefly summarizes the previously published work in the field of speech enhancement. The motive behind studying the literature is to gain in-depth knowledge related to the subject and to gain the motivation for advanced research. This helps us to identify any gap in the research methodology and work on it.

2.1 SPEECH ENHANCEMENT USING SPECTRAL SUBTRACTION

The Spectral subtraction method is used to reduce the noise and enhance the speech with good quality and superior intelligibility. This is the first and easiest method in speech processing for speech enhancement. In this method, the magnitude spectrum of the clean speech is estimated by subtracting the magnitude spectrum of the noise from the noisy speech spectrum with assuming additive background noise [12]. In the spectral subtraction process, the noisy speech signal is categorized into short frames using window. This window is used as hanning window. It is a bell curve shape which is multiplies to the noisy speech of the 1/2 overlapped data buffers [7]. Some part of the noisy speech is lie outside the window which is assumed as zero and rest of the part are calculated for further processing.

The window $h_w[n]$ can be defined as

$$h_w[n] = \begin{cases} 0.5 - 0.5 \cos\left(\frac{2\pi}{L}n\right) & 0 \leq n \leq L - 1 \\ 0 & \text{otherwise} \end{cases} \quad (2.3.1)$$

Where, L is the length of hanning window.

When the clean speech is passing through spectral subtraction filter then spectral subtraction estimator is obtained. The difference among clean speech and estimator is known as spectral error .

For reducing the spectral error, the further processing include –

- Half-wave rectification
- Reducing residual noise
- Signal attenuation during non-speech

In half-wave rectification, for each value of k, if the magnitude of estimated noise is greater than the magnitude of estimated noisy speech, then magnitude of estimated speech is made zero. Moreover, the difference $W_R = W - \mu e^{j\theta_w}$, is called residual noise. After this residual noise processing, noisy speech's phase was added to the modified magnitude spectrum and inverse FFT was done to produce short time domain signal and then the synthesized short time frames are added with 40% overlap to reconstruct the speech signal [13].

Moreover, using different collaboration of 1/4 windows and 3 or 6 frames averaging did not enhance the potential of the denoising algorithm. However, examine the different methods, using 1/4 overlapped data buffers along with 256 points hanning window and produced the maximum improvement differential SNR and greatest noise removal from noisy speech [7].

Also, for reducing the stationary noise, a new algorithm is introduced named as stand-alone suppression algorithm. It is used for reduce the background noise which is digitally added to the speech. This algorithm is used for speech recognition, narrowband voice communication, or speaker authentication. When background noise becomes non stationary, then this algorithm requires a detector named as speech activity detector for estimate the bias of new noise. This estimate is produced by averaging the magnitude of signal spectrum during non-speech activity. It maintains the equality between the characteristics and magnitude of the noise which are observed during speech activity and during absence of speech. The performance of this method expressed with or without noise reduction using short time spectra and enhancement in quality and intelligibility using Diagnostic Rhythm Test (DRT) [19].

Using short term energy (STE) for the evaluation of noise in speech signal, works for both stationary and non-stationary noise. In which spectrum of noise is continuing revised throughout the speech by comparing STE values of the current frame to the average STE values of all the coming frames. Using this estimation, it obtain clean speech signal with less speech distortion [20].

2.2 OPTIMAL LINEAR FILTER TECHNIQUE

In optimal linear filtering techniques, the original clean speech signal is attained by passing the noisy speech signal through an optimal linear transformation. This filtering technique helps in reducing noise level and improves the SNR. However, there are variety of algorithms are introduced such as Wiener filter, signal subspace method, spectral restoration, and parametric method [16]. Wiener filter is a linear filter to retain the clean speech from noisy speech by reducing the MSE between original signal magnitude spectrum and estimated signal magnitude spectrum [12]. In which, the clean signal and noise are not correlated to each other. This filter is introduced a speech distortion index and noise reduction factors to measure the amount of quantity of noise being reduced. In which the amount of noise reduction is equivalent to the amount of speech distortion. However, more noise reduction suffers more speech distortion. In single channel, the a posteriori SNR is always greater than or equal to a priori SNR, which means that this technique is useful for clean the noisy speech signals. Although, there are another approaches like if no any a priori knowledge is there, then a free variable is used to control the compromised among noise reduction and speech distortion and when we have number of microphones then multiple observations of the speech signal can be used to decrease the noise with small speech distortion [21].

2.3 IMPROVED SNR ESTIMATION FOR SPEECH ENHANCEMENT

In single channel speech enhancement technique, the noise reduction processes rely on the evaluation of a short time spectral gain. This gain is a function of the a priori SNR and/or the a posteriori SNR.

The decision directed (DD) approach is used for restrain the quality of musical noise in single microphone in noisy environment and estimates the a priori SNR [22]. In DD approach, the a priori SNR rely on the estimation of the speech spectrum in the previous frame and also it chases the structure of the a posteriori SNR along with a frame delay. A priori SNR allow notable noise power reduction without introducing music noise affects. The main drawback of DD approach is that in which gain function does not complement with the current frame. Then bias is introduced by this frame delay which generates the irritating reverberation effect [23].

The two-step noise reduction (TSNR) technique is employed to remove the bias of DD approach. To increase the performance of noise reduction technique, estimation of a priori SNR is taken in two step strategy. In DD approach, it produces the frame delay. Accordingly, the spectral gain calculated at present frame p matches the proceeding frame $(p - 1)$. Based on this, the spectral gain is calculated for the succeeding frame $(p + 1)$ using DD algorithm and implement it to the present frame because of the frame delay. Hence, this approach is taken into two steps. In first step, the spectral gain is calculated using DD algorithm and in the second step, the spectral gain is calculated for the next frame $(p + 1)$. This approach helps in removing additional processing delay which is introduced by DD approach. According to this, the gain matches to the current frame. Hence, this approach removes the irritation reverberation effect which is introduced by DD algorithm [24].

Despite that, classic approach, inclusive of TSNR, introduces harmonic distortion due to estimation error by noise PSD estimator. Speech enhancement based on minimum statistics noise estimator which has potential to conserve weal signals and provides a better intelligibility. The artificial signal is obtained to maintain the speech harmonics by the Harmonic regeneration Noise reduction (HRNR) technique. The speech harmonics are restored by applying non-linear function. In which, the artificial signal is restored at the similar locations as the clean speech ones [25].

Hence, this approach has capability to preserve the harmonics and remove distortion.

2.4 CAUSAL OR NON-CAUSAL FILTERS FOR NOISE REDUCTION IN TIME DOMAIN

Causal filters are used for the reduction of noise in speech processing and they can be implemented directly. In order to improve the performance of noise degradation and speech distortion, then non-causal filters are used. But non-causal filter depends on future values, that's why not directly implementable. This filter is based on orthogonal signal and harmonic signal decomposition. Orthogonal decomposition based filters can be utilized for upgrading any type of desired signal but

they are reasonable to non-stationary noise. Therefore, it's hard to calculate the noise statistics in the presence of desired signal. On the other hand, in harmonic decomposition filters, they are based on desired signal statistics but it causes distortion of desired signal. Hence, in which orthogonal based Wiener (ODW) filter is used to minimize the MSE and increasing the output SNR. Orthogonal based minimum variance distortion less response (ODMVDR) filter is used as temporal filtering and designed for reducing variance of residual interference and noise which are obtained after filtering. Harmonic distortion based linear-constraint minimum variance (HDLICMV) filters used for extracting the periodic signals such as musical instruments and voiced signals. This filter helps in reducing the noise without any speech distortion. Moreover, some of the filters can be upgraded recursively. In which output SNR of filter increases as filter length increases assuming desired signal is stationary. However, more noise reduction can be achieved using non-causal filters than causal filters in respect of output SNR, PESQ scores [26].

2.5 A PRIORI ESTIMATION

In the estimation of a priori SNR, a statistical model can be used for speech enhancement that considers the time correlation among successive speech spectral components. The series of speech spectral variances is an arbitrary process, which is normally related to the series of speech spectral magnitudes. Causal and non-causal estimators for a priori SNR are obtained with the evaluation of the speech spectral components. Causal estimator deals with decision directed algorithm for a priori SNR. A priori SNR has two phases. One is a priori SNR values from the previous frame and other is estimation of maximum likelihood (ML) for a priori SNR values based on current frame. This estimator cannot react fast to a sudden rise of SNR. Thus, the level of musical noise increases [27]. But non-causal estimator has superior estimation between speech onsets and noise irregularities. Onsets of speech are best conserved and after that degraded of musical noise takes place. That's why non-causal estimator provides the reduced amount of signal distortion and musical noise than decision directed algorithm [28].

A detector is used along with an estimator for presence of speech in short time Fourier transform (STFT) domain. They minimize the cost function that considers the both detector and estimator errors and these cost parameters constraint the trade-off among residual musical noise, speech quality and noise reduction. In distinction with DD estimator, non-causal estimator is able to distinguish the speech onsets and noise irregularities. Therefore, a modified DD a priori SNR estimation is used for suppression of transient noise with less amount of speech distortion [29].

Also, non-causal estimator is used for Laplacian and Gamma speech models. In which spectral gain for Laplacian and Gamma speech models are monotonically increasing functions of the a posteriori SNR with constant a priori SNR. Using this non-causal estimator, native bursts of noise and speech onsets are assigned a lower a priori SNR and higher a priori SNR respectively. Thus, speech onsets are better maintained and the musical noise effect is decreased [30].

2.6 POWER SPECTRAL DENSITY (PSD) ESTIMATION

In single channel speech enhancement, it is important to estimate the power spectral density due to effect on quality and intelligibility of the enhanced speech. But in inauspicious environment, estimation of noise PSD is a challenging problem [31]. In this approach a novel noise estimation algorithm is introduced, which is based on minimum statistics and on an optimal signal PSD smoothing method. By minimizing the mean square error criterion, optimal PSD smoothing parameter is used for tracing the non-stationary noise which is superimposed by speech in each frequency band. Also minimum statistics (MS) approach [32] is used for improve the performance of spectral subtraction method in non-stationary noise. This approach tracks the minimal values of smoothed power estimate of the noisy signal. By balancing the succeed bias, an estimate of the noise PSD is obtained. This algorithm is also based on maximum a posteriori (MAP) estimation of the noise variance of non-stationary white Gaussian process. This method produces lesser estimation error in low SNR conditions and lesser fluctuations (called musical noise) of the estimated values in all environmental conditions. Resulting, these approaches have ability to preserve the weak signals and provide better quality or intelligibility as measured by PESQ scores [33].

2.7 ESTIMATOR ANALYSIS

In log short-time spectral amplitude (STSA) speech enhancement, a priori SNR estimator has notable influence on reducing noise. There are three estimator rules for estimate the parameters of speech model such as maximum likelihood (ML), MAP, and minimum mean square estimator (MMSE). In which DD type estimator is based on the previous frame amplitude estimate. This estimator has two terms; SNR values from the previous frame and estimation of ML SNR from the current frame. In combination with log STSA speech enhancement method, in ML approach, it estimates the unknown parameter without any a priori knowledge. Therefore, this method causes smoothly increasing noise attenuation in the absence of speech of speech segments of the signal. Moreover, MMSE estimator is introduced, which includes the smoothing deportment of the DD algorithm [34]. The main goal of MMSE estimator is minimizes the mean square error (MSE) among clean speech and estimated clean speech. It can be written as

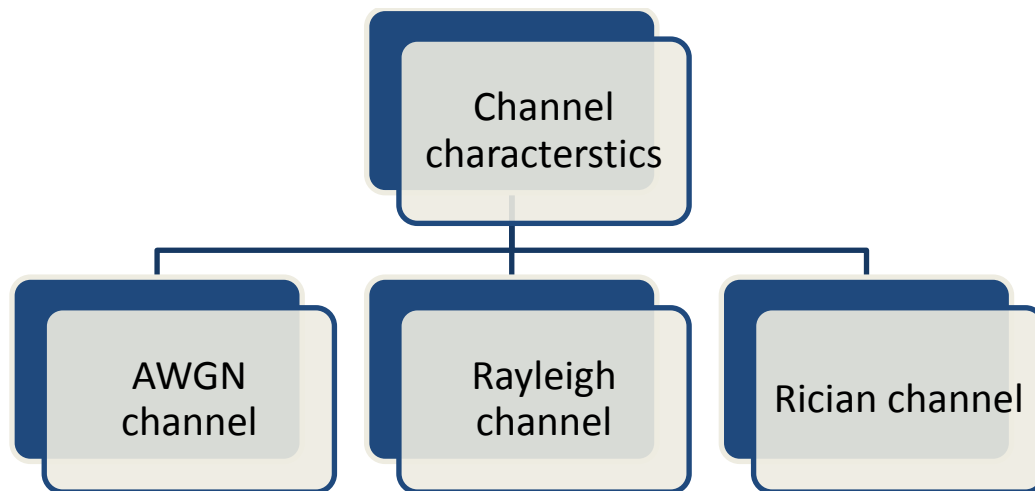
$$MSE = E[(y - \hat{y})^2] \quad (2.7.1)$$

Where, y denoted as clean speech and \hat{y} denoted as estimated speech. However, this approach helps in reducing musical noise and preserves the week signals.

2.8 CHANNEL CHARACTERSTICS

Wireless communication, is one of the most vivid areas in the communication field due to convergence of some factors. In telecommunication, it may be used to convey information across short and long distances. Channel is the physical way which is used to transfer the signal from the

transmitter to the receiver [35]. Interference is a basic nature of wireless communication systems, in which multiple transmissions often take place simultaneously over a common communication medium [36]. When a signal travels from a transmitter to receiver, it goes through reflection, scattering and diffraction from the trees, buildings, mountains, etc. Thus, multipath signals will come at a receiver with different phases and amplitudes. This phenomenon is known as multipath fading [37]. There are three types channel characteristics explained as –



2.8.1 AWGN channel

In communication channel, AWGN is the statistically random radio noise described by a broad range of frequency respecting to a signal. In which channel is linear and time-invariant. This model has set of assumptions which are

- The noise is additive which means received signal is the sum of transmitted signal and noise, where noise is statistically not dependent to the signal.
- The noise is white which means the power spectrum density is flat.
- The noise samples have a Gaussian distribution.

AWGN is generally used to transmit the signals whereas signals travel from the channel and imitate background noise of channel [38].

The mathematical representation in the received signal $p(t) = q(t) + c(t)$ that goes through AWGN channel where $q(t)$ is the transmitted signal and $c(t)$ is the noise [32] shown in Figure 2.1.

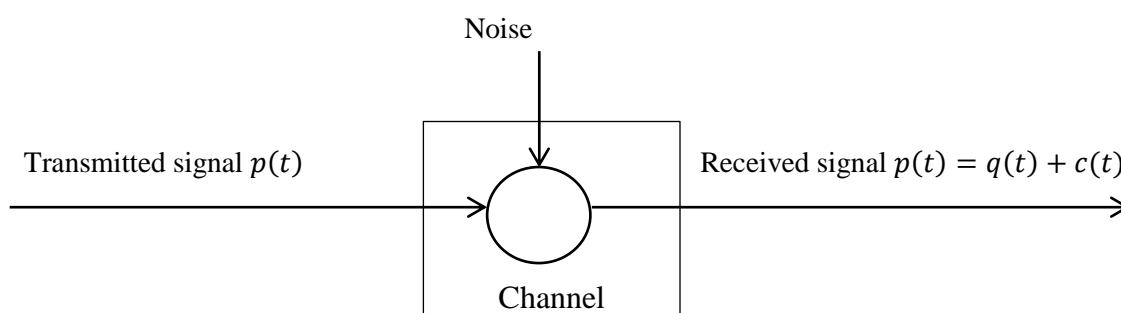


Figure 2.1 AWGN model [39]

2.8.2 Rayleigh fading channel

In Rayleigh fading model, the fading is produced by multiple receptions. Rayleigh fading model supposes that the magnitude of a signal goes through the transmission medium will change randomly, or fade, in accordance with Rayleigh distribution. This fading is a sensible model when there are many entities in the environment that separate the radio signal before it comes at the receiver. This fading is used when there is no dominant Line-of-Sight (LOS) path among transmitter and receiver and it is also considered as a more generalized case of Rician channel model [37].

In Rayleigh Fading model, it is supposed to have only two multipath receptions. This Fading can be obtained from zero-mean complex Gaussian processes. In which, the two Gaussian Random variables are added and taking the envelope of these variables gives a Rayleigh distributed process. The probability density function (PDF) of Rayleigh distribution with envelope of received signal is

$$p(y) = \frac{y}{\sigma^2} \exp\left(-\frac{y^2}{2\sigma^2}\right) \quad 0 \leq t \leq \infty \quad (2.8.2.1)$$

Where y is the envelope of received signal.

$\frac{y}{\sigma^2}$ is the instantaneous power.

σ is the rms value of received signal.

2.8.3 Rician fading channel

Rician fading happens when the signal comes at the receiver from two separate paths in which one of the paths is changing. It is a stochastic model for radio propagation. This fading is happens when one of the paths is LOS and much robust than others. In which, the amplitude gain is specified by a Rician distribution [37]. The pdf of Rician channel is given by

$$p(y/b, \sigma) = \frac{y}{\sigma^2} \exp\left(-\frac{y^2 + b^2}{2\sigma^2}\right) I_0\left(\frac{yb}{\sigma^2}\right) \quad \text{for } y \geq 0, b \geq 0 \quad (2.8.3.1)$$

Where b is the highest amplitude of the dominant signal. $I_0(\cdot)$ is zero order Bessel function.

Sometimes Rician distribution is expressed in terms of a parameter k known as Rician factor which is represented by:

$$k = 10 \log \frac{y^2}{2\sigma^2} \quad (2.8.3.2)$$

As γ approaches to 0, k approaches ∞ dB and as the dominant path reduces in amplitude, the Rician distribution degenerates to a Rayleigh distribution.

2.9 SHORT TERM FOURIER TRANSFORM (STFT) ANALYSIS

Speech is non-stationary signal where properties vary rapidly over time. In time domain representation, waveform gives some point of the dynamics and periodicity of audio, but excluding this, it is not clear example frequency distribution. The Fourier transform of discrete signals is known as DFT. FFT is fast computation of DFT. Short term of FT is called as STFT. In STFT domain, the noise is calculated in each STFT sub band and the eliminated noise is eliminated from the noisy STFT coefficients to achieve reduced noise. To obtain the better estimate, the coefficients are combining with the subspace method and the optimal filtering methods through the joint diagonalization of the clean speech and noise signal correlation matrices. This filter maintains the tradeoff between output SNR and speech distortion. The spectrogram of a speech signal can be calculated as magnitude squared STFT. The spectrogram represents the spectral distribution of power within the signal in the time-frequency domain. The spectrogram is more superior to waveform to represent speech example harmonics and it has fixed design that depends on the chosen window size. However, the wider window provides superior frequency resolution but inferior time resolution. Along this, longer window provides higher frequency resolution, if too long then spectral characteristics would change. For smooth transitions, the windows are selected to be overlapping. When appending copies of the segments succeeding, then discontinuous in waveform takes place. This results in spectral leakage. To decrease the spectral leakage, we can multiply the segments with hamming window that reaches zero at the ends. Moreover, zero padding is done for required length by appending the short segments with zeros [40].

2.10 PROBLEM STATEMENT

The purpose of this thesis is to:

- Analyze and equate the performance of speech enhancement algorithms such as spectral subtraction and Wiener filter in terms of SNR in the presence of background noises in AWGN channel, Rayleigh channel and Rician channel.
- Estimate the a priori SNR which has a significant function in speech enhancement for reducing the noise and gives the best performance at receiver.
- In order to improve the quality and intelligibility of speech, it is essential to estimate the PSD. PSD is a smoothing parameter that helps in tracking the non-stationary noise which is overlaid by speech in each frequency band.
- Examine the differences and dependencies between spectral subtraction and DD approach for reducing the noise in single channel. Recognize techniques that can be used to improve performance of the system in the single channel speech enhancement.

CHAPTER 3

NOISE REDUCTION TECHNIQUE BASED ON SPECTRAL SUBTRACTION

The main motive of speech enhancement is to enhance the behavior of speech communication system in different environment conditions. In single channel, there are many algorithms are utilized for speech enhancement. Spectral subtraction is one of the simplest methods used for reducing the noise when original signal is passing through different channels such as AWGN, Rayleigh and Rician channel. This process also helps in improving the SNR value as well as enhancing the speech quality. This technique operates in the frequency-domain and supposes that input noisy signal spectrum is the sum of noise spectrum and speech spectrum [8].

3.1 PROCESS OF SPECTRAL SUBTRACTION

Nowadays, speech processing is generally used in practical applications such as speaker identification, speech pre-processing for aids, speech recognition and gender classification system. Speech signals need to be clean for greater achievement of the speech processing system. Practically, background environment noise is acoustically mixed with the original clean speech signal and reduces the performance of the system. To boost the performance of the system, speech enhancement techniques are used [20]. In general, single channel speech enhancement problem deals with such applications in which one speaker is used for transcript purpose. These speech enhancement techniques degrade the noise from noisy speech signal. There are numerous speech enhancement techniques are used to degrade the noise and provide speech signal with superior quality [17].

Basic Spectral subtraction is one of the first and easy approaches for single channel enhancement. This method is known for its minimal complexity. Therefore, it is easier to implement. This method is used for restoring fundamental parameters like magnitude spectrum and power spectrum when the clean speech signal is distorted by additive noise. The main motive of spectral subtraction is to determine the noisy spectrum [12].

As it can be seen from the block diagram depicting Spectral subtraction in Figure 3.1, a noisy spectrum is subtracted from noisy speech spectrum to obtain the estimate of clean speech spectrum. The estimate of noise spectrum is updated at the instances when no useful speech signal is present and only noise exists [41]. And then spectral error is calculated. Further processing is performed on spectral error takes place once it is estimated. We assume that noise is stationary i.e., does not change with time.

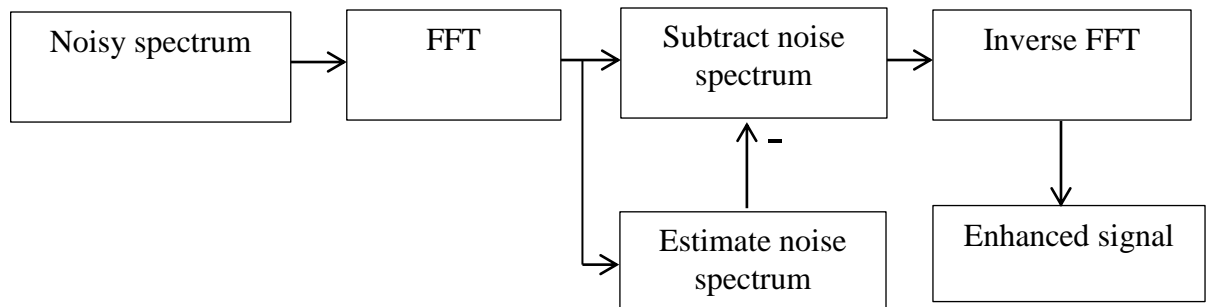


Figure 3.1 Block diagram of Spectral Subtraction[13]

The Spectral subtraction procedure consists of two fundamental principles:

- Estimation of noisy spectrum of background noise.
- Speech spectrum subtraction from the noisy speech spectrum.

Generally, Spectral subtraction process operates in the frequency-domain. In this process, it's important to divide the continuous time-domain signal into overlapping segments. The speech signal is generated at a sample rate of 8 kHz and 256-point Fourier transform is implemented on the 25msec frame of input speech signal.

The flow chart of Spectral subtraction method is shown in Figure 3.2. In this process, the noisy speech spectrum is segmented into short time frames using hamming window and then these time segments are overlapped to an extent of 40%. Using FFT transform, the domain of noisy speech spectrum is changed from time-domain to frequency-domain. The average of spectrum is calculated; and an estimate of noise spectrum is subtracted from the segments where speech is absent. After that spectral error is calculated and output spectrum is biased down. Further processing of this procedure is performed to reduce the residual noise. The noisy speech phase is added to the modified spectrum. And then time-domain waveform is synthesized from it. Subsequently, 40% overlapping is done on adjacent frames in the synthesized waveforms to achieve the enhanced speech. Thus reconstructed signal obtained is of superior quality [13].

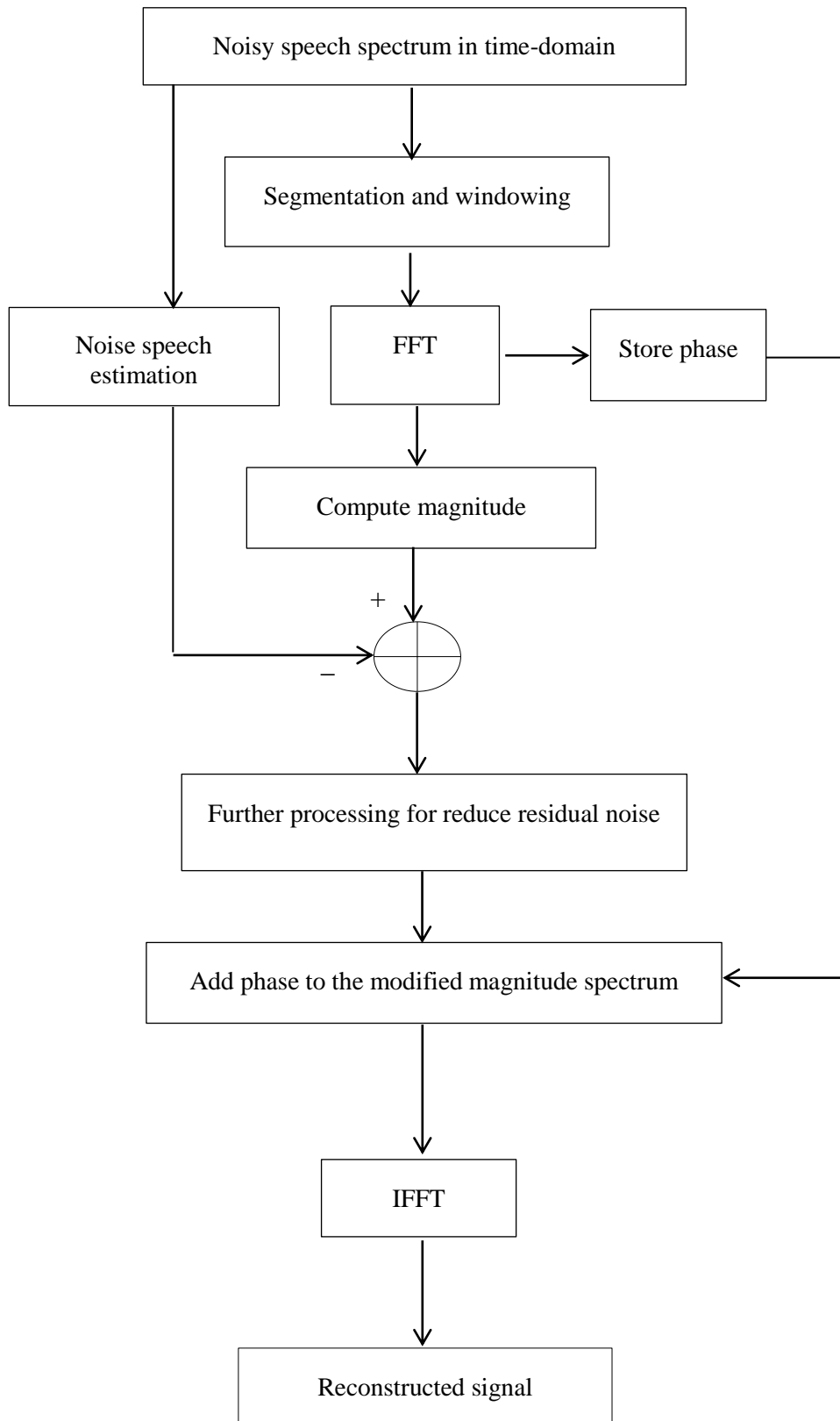


Figure 3.2 Flow chart of spectral subtraction method [13]

3.2 ADDITIVE NOISE MODEL

Let $y(n)$ be the noisy speech signal can be written as

$$y(n) = x(n) + c(n) \quad (3.2.1)$$

Where, $x(n)$ is the clean speech signal and $c(n)$ is additive background noise. Also, n represents the discrete time index.

In spectral subtraction algorithm, we assume that the noise $c(n)$ which is uncorrelated with the clean speech signal $x(n)$ to estimate the noise spectrum. The short time Fourier transforms of $y(n)$ is,

$$Y(l, m) = X(l, m) + C(l, m) \quad (3.2.2)$$

Where,

$$Y(l, m) = \sum_{p=-\infty}^{\infty} y(p)w(l-p)e^{-j\omega p} \quad (3.2.3)$$

and $w(n)$ = hamming window. Where, $Y(l, m)$, $X(l, m)$ and $C(l, m)$ represents the m^{th} spectral component of l^{th} time window of $y(n)$, $x(n)$ and $c(n)$ respectively.

3.3 SPECTRAL SUBTRACTION ESTIMATOR AND SPECTAL ERROR

When clean speech signal is passed through spectral subtraction filter $H(l, m)$, then the spectral subtraction estimator $\hat{X}(l, m)$ is formed. It can be represented as

$$\hat{X}(l, m) = H(l, m)Y(l, m) \quad (3.3.1)$$

Where,

$$H(l, m) = 1 - \frac{\mu(m)}{|Y(l, m)|} \quad (3.3.2)$$

And

$$\mu(m) = E\{|C(l, m)|\} \quad (3.3.3)$$

By substituting values of $H(l, m)$ and $\mu(m)$ in Equation (3.3.1), then new equation becomes

$$\hat{X}(l, m) = [|Y(l, m)| - \mu(m)] e^{j\theta_y(l, m)} \quad (3.3.4)$$

Where, $\theta_y(l, m)$ represents the phase of noisy signal. In which the noisy phase is use for an estimate of the clean speech signal. The difference between the clean speech and the estimator is known as spectral error which is represented as $\epsilon(l, m)$ and is given by

$$\epsilon(l, m) = \hat{X}(l, m) - X(l, m) \quad (3.3.5)$$

$$\epsilon(l, m) = C(l, m) - \mu(m)e^{j\theta_y} \quad (3.3.6)$$

Further processing done for decreasing the remaining noise.

3.4 HALF WAVE RECTIFICATION AND REDUCTION OF RESIDUAL NOISE

For each m values, if the estimated noise is more than the noisy speech, then the estimated speech magnitude is made zero. Hence estimator becomes,

$$|\hat{X}(l, m)| = \begin{cases} |Y(l, m)| - \mu(m) & \text{for } |Y(l, m)| \geq \mu(m) \\ 0 & \text{otherwise} \end{cases} \quad (3.4.1)$$

In which, the problem is occur when estimated magnitude value is negative. In order to cope up with this problem, all negative magnitude values are set to zeros. That's why this modification is known as half wave rectification, since the negative half of wave is corrected. It has the advantage of eliminating any noise with magnitude smaller than $\mu(m)$, additionally; it also inaccurately removes the speech if the sum of noise and speech is lower than $\mu(m)$. Therefore, the quality of estimated signal is decreased when the noise level is high as compare to speech. The difference $C_R = C - \mu e^{j\theta_c}$, which is known as residual noise. The noise will be present in the spectrum as narrow bands of magnitude spikes spaced randomly. The residual noise can be removed by modifying the short time spectrum. It can be represented as

$$|\hat{X}(l, m)| = \begin{cases} |\hat{X}(l, m)|; & |\hat{X}(l, m)| \geq \max |C_R| \\ \min[|\hat{X}(l, m)|_{l=i-1, i, i+1}] & |\hat{X}(l, m)| < \max |C_R| \end{cases} \quad (3.4.2)$$

Where, $|C_R|$ = noise removal computed during absence of speech.

3.6 SYNTHESIS

After this residual noise processing, noisy speech's phase was added to the modified magnitude spectrum and inverse FFT was done to produce short time domain signal and then the synthesized short time frames are added with 40% overlap to reconstruct the speech signal.

3.7 SIMULATION RESULTS

This section represents the experimental results acquired by evaluating the performance of Spectral subtraction algorithm. In which, a clean speech signal with 8 kHz sampling frequency was used; contains female voice in it. The experiment is conducted, by corrupting the speech by different types of channels with intensity varied from 2 to 10dB. The three channels are AWGN channel, Rayleigh channel and Rician channel.

To illustrate the behavior and performance of this technique in the presence of different channels are shown in Figure 3.3 to 3.6. The Figure 3.3(a) to 3.6(a) are shows the spectrogram of the original speech signal and Figure 3.3(b) to 3.6(b) are shows the spectrogram of noisy speech signal in the presence of different channels using male and female voice signal. The Figure 3.3(c) to 3.6(c) represents the reconstructed speech using spectral subtraction in different environment conditions.

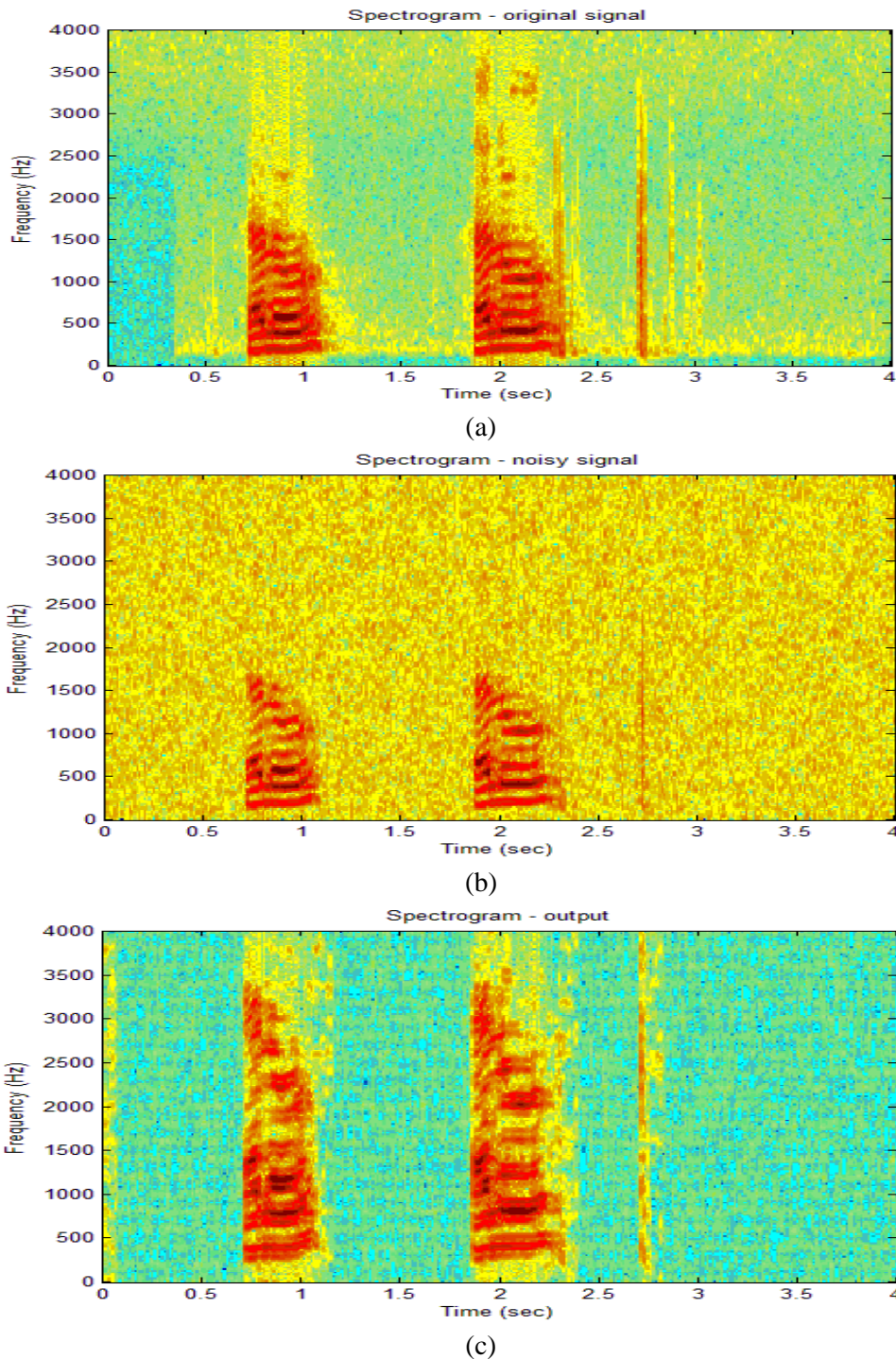


Figure 3.3 Spectrograms of Spectral subtraction algorithm using male voice in AWGN channel with SNR 2dB (a) Original speechsignal (b) Noisy speech signal (c) reconstructed speech signal using Spectral subtraction.

The spectrogram in Figure 3.1 illustrates the spectrogram of reconstructed signal using spectral subtraction filter. For simulations, a clean speech signal with 8 kHz sampling frequency has been used, which contains male voice. An AWGN of power level 2dB SNR is added to it, which degrades the quality of speech signal and distorts it. Using the subtraction filter algorithm, noise is removed and the recovered signal is approximately equal to the original signal.

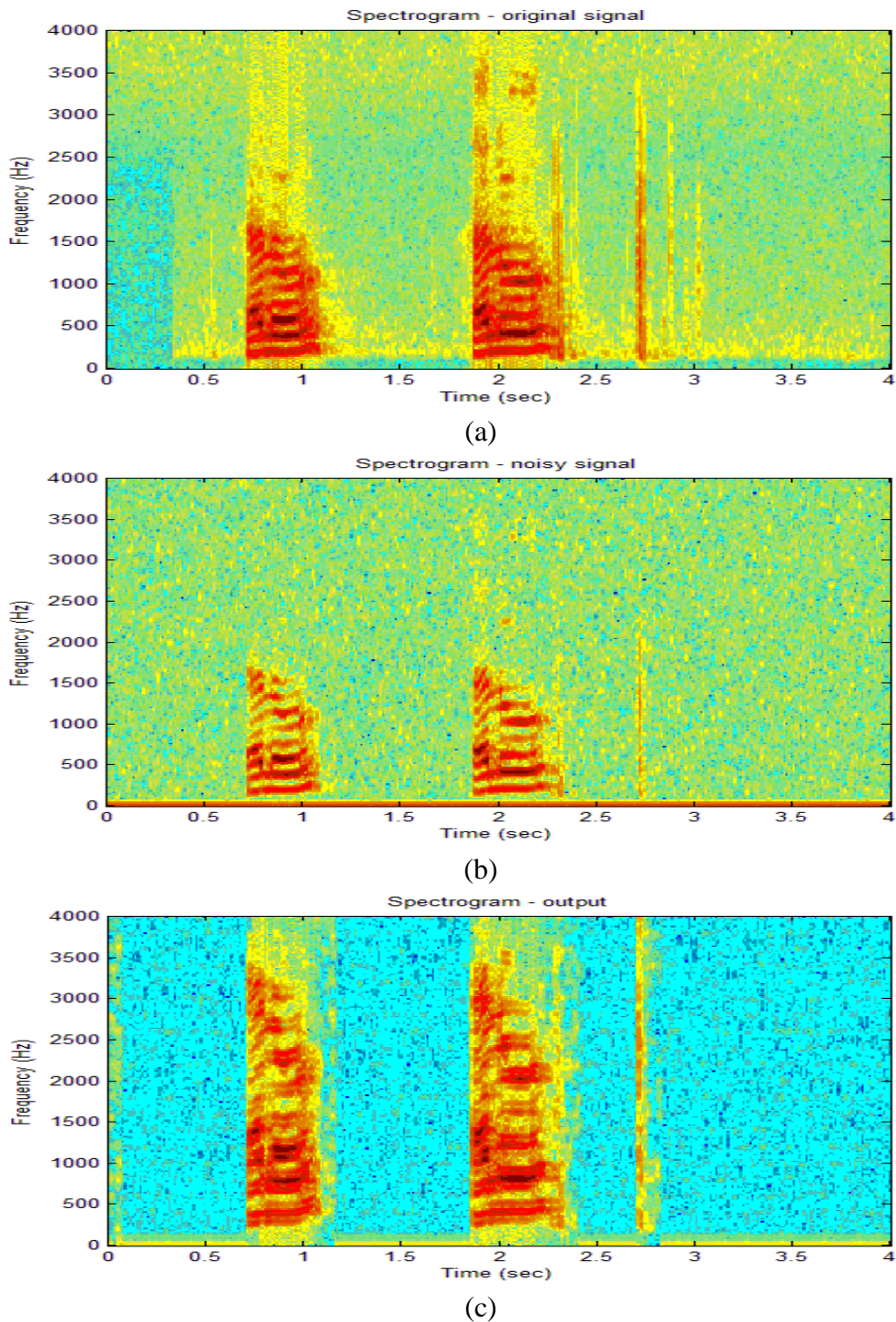
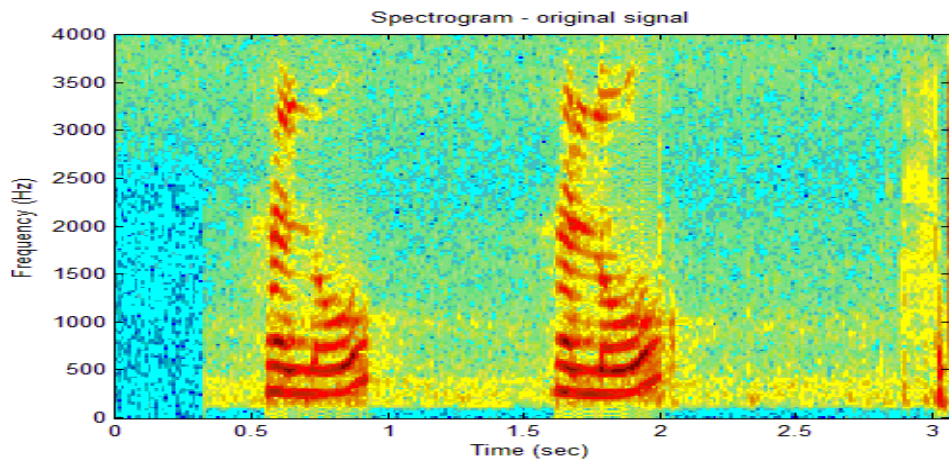
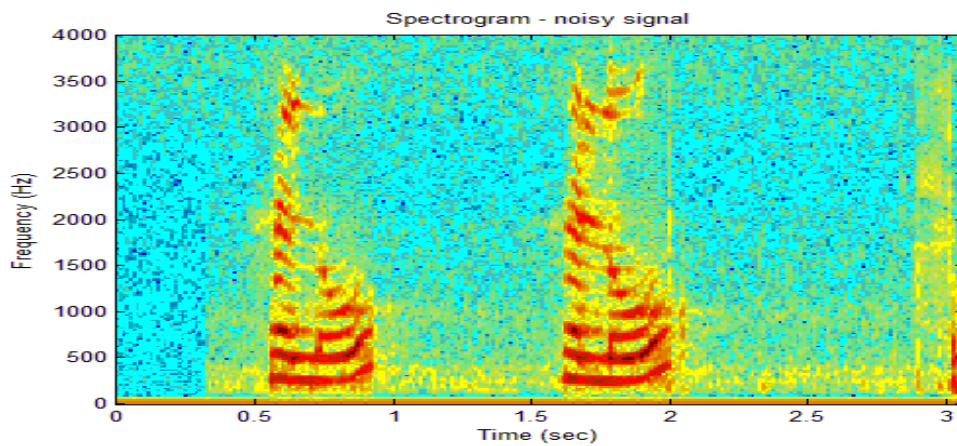


Figure 3.4 Spectrograms of Spectral subtraction algorithm in Rayleigh channel using male voice with SNR 5dB (a) Original speech signal (b) Noisy speech signal (c) reconstructed speech signal using Spectral subtraction.

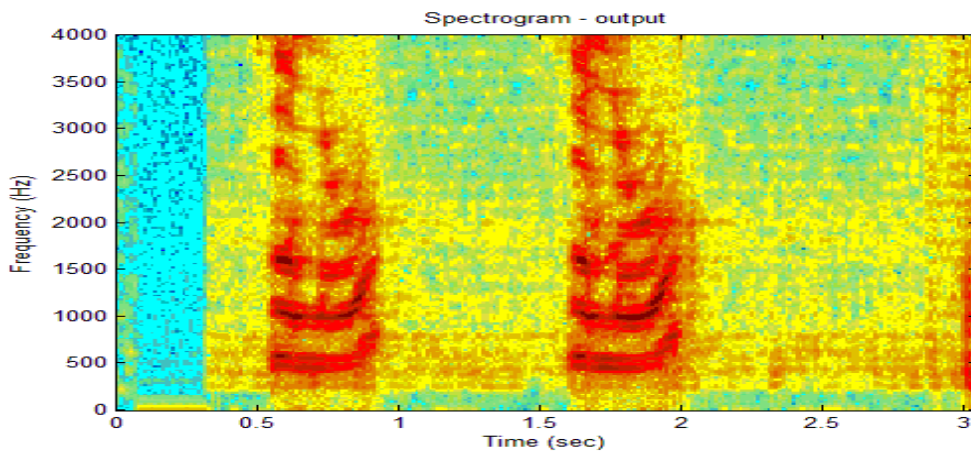
The spectrogram in Figure 3.2 shows the result of clean, noisy and recovered speech signal using spectral subtraction algorithm. For simulation, a clean male voice with 8 kHz sampling frequency is used. In communication process, when clean speech signal is passed through Rayleigh channel with power level of 5dB SNR. Then original signal gets distorted. Through spectral subtraction algorithm, output speech signal is enhanced as compared to input speech signal.



(a)



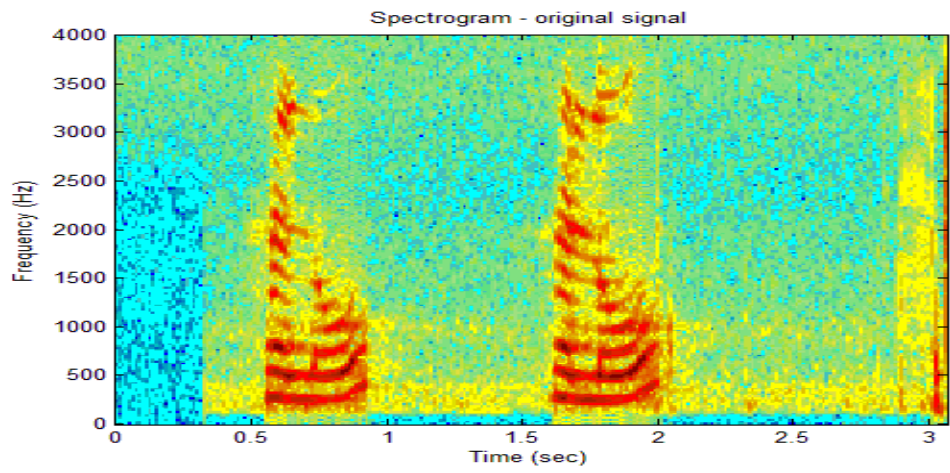
(b)



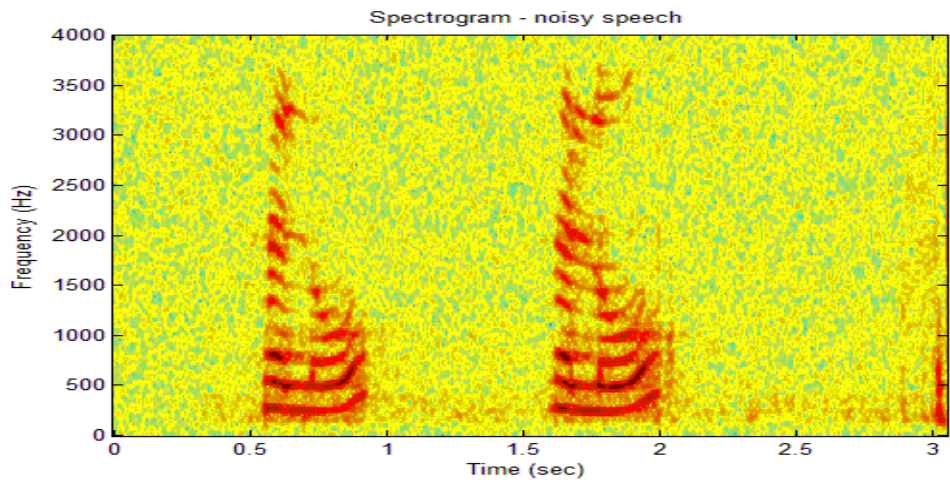
(c)

Figure 3.5 Spectrograms of Spectral subtraction algorithm in Rician channel using female voice with SNR 7dB (a) Original speech signal (b) Noisy speech signal (c) reconstructed speech signal using Spectral subtraction

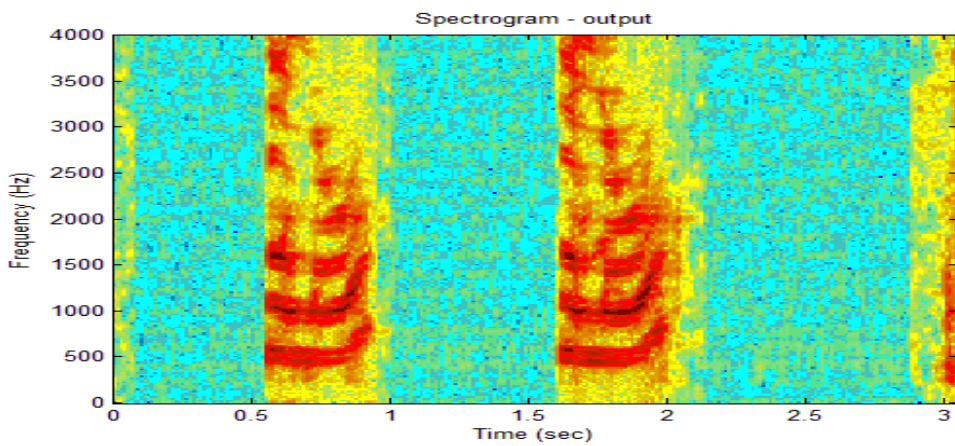
The spectrograms in Figure 3.3 represent the result of clean, noisy and recovered speech signal using subtraction filter. For simulation, a clean female voice with 8 kHz sampling frequency is used. When speech signal is passed through Rician channel at power level of 7dB SNR, then this channel distorts the original speech signal. Using spectral subtraction algorithm, speech signal is recovered.



(a)



(b)



(c)

Figure 3.6 Spectrograms of Spectral subtraction algorithm using female voice in AWGN channel with SNR 10dB (a) Original speech signal (b) Noisy speech signal (c) reconstructed speech signal using Spectral subtraction

The spectrogram in Figure 3.4 shows the result of reconstructed signal using spectral subtraction algorithm. For simulation, a female voice with 8 kHz sampling frequency is used. This clean signal is mixed with AWGN of power level 10dB SNR, which decreases the speech quality. However, noise is removed using spectral subtraction algorithm; obtained signal is close to the clean signal.

CHAPTER 4

NOISE REDUCTION TECHNIQUE BASED ON WIENER FILTER

A number of methods are accessible for estimating the clean speech, one of the method based on MMSE estimation i.e. Wiener filter. In which the noise is unrelated and additive with regard to speech, the minimization of $E\{(\hat{X}(l, m) - X(l, m))^2\}$ leads to,

$$G(l, m) = \frac{E\{|X(l, m)|^2\}}{E\{|X(l, m)|^2\} + E\{|C(l, m)|^2\}} = \frac{S\hat{N}R_{prio}(l, m)}{1 + S\hat{N}R_{prio}(l, m)} \quad (4.1)$$

The noise reduction process depends on the estimation of $G(l, m)$. This gain is a function of a priori SNR and a posteriori SNR. However, $S\hat{N}R_{prio}(l, m)$ is estimated by using DD approach.

In the analysis of the estimator, it was indicated that the a priori SNR chases the shape of a posteriori SNR. But in which gain function $G(l, m)$ does not match with the current frame then it initiates the frame delay which causes irritating reverberation effect. TSNR method is used to remove this reverberation effect of DD approach while maintain its advantage.

However, the classic short-time suppression techniques, introduces harmonic distortion due to error in estimating noise PSD. To get the Better of this problem, an artificial signal is produced to maintain the speech harmonics by HRNR technique. A priori SNR was improved by using the artificial signal and used for calculating a spectral gain which can replace the speech harmonics [23].

4.1 SIGNAL MODEL

Consider the noisy speech signal $y(n)$ is given by,

$$y(n) = x(n) + c(n) \quad (4.1.1)$$

Where, $x(n)$ and $c(n)$ denote the clean speech signal and noise signal respectively. The STFT of $y(n)$ is,

$$Y(l, m) = X(l, m) + C(l, m) \quad (4.1.2)$$

Where, $Y(l, m)$, $X(l, m)$ and $C(l, m)$ represents the m^{th} spectral component of l^{th} time window of $y(n)$, $x(n)$ and $c(n)$ respectively.

4.2 INTRODUCTION TO NOISE REDUCTION PARAMETERS

The estimator of SNR is calculated by using noisy features to get a spectral gain $G(l, m)$. An estimate of $X(l, m)$ is formed by apply this spectral gain $G(l, m)$ to each $Y(l, m)$, the speech enhancement techniques are implemented to compute the two criterions: a priori SNR and a posteriori SNR given by,

$$SNR_{prio}(l, m) = \frac{E[|X(l, m)|^2]}{E[|C(l, m)|^2]} \quad (4.2.1)$$

$$SNR_{post}(l, m) = \frac{|Y(l, m)|^2}{E[|C(l, m)|^2]} \quad (4.2.2)$$

Where, $E[.]$ is the expectation operator. Another criterion, the instantaneous SNR is

$$SNR_{inst}(l, m) = \frac{|Y(l, m)|^2 - E[|C(l, m)|^2]}{E[|C(l, m)|^2]} \quad (4.2.3)$$

$$SNR_{inst}(l, m) = SNR_{post}(l, m) - 1 \quad (4.2.4)$$

In spectral subtraction approach, it can be explained as a direct evaluation of the local a priori SNR. But it is helpful to assess the accuracy of the a priori SNR estimator. In practical implementation, the PSDs of speech $E[|X(l, m)|^2]$ and noise $E[|C(l, m)|^2]$ are unknown [18]. Hence, it is important to estimate a priori SNR and a posteriori SNR. The noise PSD estimate $E[|C(l, m)|^2]$, noted $\hat{\gamma}_c(l, m)$, can be estimated using classic recursive relation. The spectral gain $G(l, m)$ is acquired by,

$$G(l, m) = g(SNR_{post}(l, m), SNR_{prio}(l, m)) \quad (4.2.5)$$

Where, Function g is chosen for Wiener filtering. The estimated speech signal is obtained by applying the spectral gain to the noisy spectrum.

$$\hat{X}(l, m) = G(l, m)Y(l, m) \quad (4.2.6)$$

4.3 DECISION DIRECTED (DD) APPROACH

Based on the definition, the derivation and relation of $SNR_{prio}(l, m)$ with $SNR_{post}(l, m)$ is represented as,

$$SNR_{prio}(l, m) = \frac{E[|X(l, m)|^2]}{E[|C(l, m)|^2]} \quad (4.3.1)$$

$$SNR_{prio}(l, m) = E[SNR_{post}(l, m) - 1] \quad (4.3.2)$$

Adding Equations (4.3.1) and (4.3.2), we can write

$$SNR_{prio}(l, m) = E \left\{ \frac{1}{2} \frac{[|\hat{X}(l, m)|^2]}{E[|C(l, m)|^2]} + [SNR_{post}(l, m) - 1] \right\} \quad (4.3.3)$$

And,

$$SNR_{prio}^{DD}(l, m) = \beta \frac{[|\hat{X}(l-1, m)|^2]}{E[|C(l, m)|^2]} + (1 - \beta)P[SNR_{post}(l, m) - 1], \text{ for } 0 \leq \beta \leq 1 \quad (4.3.4)$$

Where P denotes the half wave rectification and $\hat{X}(l-1, m)$ is the estimated speech spectrum at previous frame. Moreover, $P[.]$ can be defined as,

$$P[x] = \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (4.3.5)$$

According to Equation (4.3.4), the a priori SNR estimator with DD approach, is known as $S\hat{N}R_{prio}^{DD}(l, m)$. The behavior of $S\hat{N}R_{prio}^{DD}(l, m)$ is controlled by the parameter β (value $\beta = 0.98$).

According to Equation (4.2.6),

$$G_{DD}(l, m) = \frac{S\hat{N}R_{prio}^{DD}(l, m)}{1 + S\hat{N}R_{prio}^{DD}(l, m)} \quad (4.3.6)$$

In DD algorithm, there are two effects which can be explained as,

- When $S\hat{N}R_{inst}(l, m) \gg 0dB$, then $S\hat{N}R_{prio}(l, m)$ is equal to the $S\hat{N}R_{inst}(l, m)$ by a frame delay.
- When $S\hat{N}R_{inst}(l, m)$ is lower or close to 0dB, then $S\hat{N}R_{prio}(l, m)$ is equivalent to plane and delayed version of the $S\hat{N}R_{inst}(l, m)$. Hence, the variance of $S\hat{N}R_{prio}(l, m)$ is minimized as compared to $S\hat{N}R_{inst}(l, m)$.

The main issue of the DD algorithm is the delay inherent in the speech transition, i.e. speech onset and offset [23]. Moreover, this delay initiate to a bias in gain estimation which generate the reverberation effect.

4.4 TWO STEP NOISE REDUCTION (TSNR) APPROACH

In order to improving the noise reduction performance, the a priori SNR can be evaluated in two steps, referred as TSNR method. In DD algorithm, initiate a frame delay when the

parameter β is near to 0.98. In the second step, the frame delay is eliminated which is the main issue of DD algorithm.

In the first step, the spectral gain is calculated for the frame $(l + 1)^{th}$ using DD algorithm and then applied to the l^{th} frame of noisy speech. In second step, the gain is used to evaluate the a priori SNR at $(l + 1)^{th}$ frame.

$$\begin{aligned} S\hat{N}R_{prio}^{TSNR}(l, m) &= S\hat{N}R_{prio}^{DD}(l + 1, m) \\ &= \beta' \frac{|G_{DD}(l, m)Y(l, m)|^2}{\hat{\gamma}_d(l, m)} + (1 - \beta')P[S\hat{N}R_{post}(l + 1, m) - 1] \end{aligned} \quad (4.4.1)$$

Where, the role of β' is same as β but with different value. In calculating $S\hat{N}R_{post}(l + 1, m)$, $Y(l + 1, m)$ produces additional delay. Therefore, β' can be chosen as 1. According to this, the equation (4.4.1) becomes:

$$S\hat{N}R_{prio}^{TSNR}(l, m) = \frac{|G_{DD}(l, m)Y(l, m)|^2}{\hat{\gamma}_c(l, m)} \quad (4.4.2)$$

Therefore, the additional delay can be avoided without any knowledge of future frames. In addition, $\beta' = 1$, the musical noise level is decreased to minimum level produced by DD approach.

At last, we calculate the gain as,

$$G_{TSNR}(l, m) = \frac{S\hat{N}R_{prio}^{TSNR}(l, m)}{S\hat{N}R_{prio}^{DD}(l, m)} \quad (4.4.3)$$

To get the estimate of speech spectrum, the gain is multiplied by noisy speech spectrum,

$$\hat{X}(l, m) = G_{TSNR}(l, m)Y(l, m) \quad (4.4.4)$$

The behaviour of the TSNR algorithm can be explained as,

- When $S\hat{N}R_{inst}(l, m) \gg 0dB$, then a priori SNR is equal to the inst. SNR without delay.
- When $S\hat{N}R_{inst}(l, m)$ lower or closer to 0dB is, then $S\hat{N}R_{prio}^{TSNR}(l, m)$ is further decreased compared to $S\hat{N}R_{prio}^{DD}(l, m)$.

To summarize this, the noise reduction performance was enhanced by the TSNR algorithm [19], after all at particular frame gain equals to the current frame. It helps in preserve the speech transition i.e. speech onset and outset and eliminating the reverberation effect produced by DD algorithm.

4.5 HARMONIC REGENERATION NOISE REDUCTION (HRNR) APPROACH

In single channel noise reduction techniques, it is difficult to estimate accurate noise due to estimation error. The output spectrum $\hat{X}(l, m)$, or $\hat{x}(t)$ in the time domain, produced by approaches like DD, TSNR. They are still suffering from distortions. These distortions are harmonic in nature in which only noise components and are to be suppressed. For removing the artificial these distortions, all missing harmonics have to be regenerated. This signal is used to estimate the gain and that helps in restore the harmonics [25].

The artificial restored signal can be formed by non-linear function to the time signal $\hat{x}(t)$,

$$X_{harmono}(t) = NL(\hat{S}(t)) \quad (4.5.1)$$

In which, the restored harmonics $X_{harmono}(t)$ are present at the similar location as that of the clean speech ones. But in which amplitude of harmonics of artificial signal are biased. Therefore, it is used for enhance the a priori SNR.

$$S\hat{N}R_{prio}^{HRNR}(l, m) = \frac{\rho(l, m)|\hat{X}(l, m)|^2 + (1 - \rho(l, m))|X_{harmono}(l, m)|^2}{\hat{Y}_c(l, m)} \quad (4.5.2)$$

Where,

$$\rho(l, m) = G_{TSNR}(l, m)$$

This parameter is used to control the mixing level of $|\hat{X}(l, m)|^2$ and $|X_{harmono}(l, m)|^2$.

$S\hat{N}R_{prio}^{HRNR}(l, m)$ is used for calculate the gain to maintain the harmonics.

The spectral gain is given as,

$$G_{HRNR} = \frac{S\hat{N}R_{prio}^{HRNR}(l, m)}{1 + S\hat{N}R_{prio}^{HRNR}(l, m)} \quad (4.5.3)$$

And $\hat{X}(l, m)$ was calculated as,

$$\hat{X}(l, m) = G_{HRNR}(l, m)Y(l, m) \quad (4.5.4)$$

4.6 SIMULATION RESULTS

This section represents the experimental results acquired by evaluating the performance of Wiener filter using DD algorithm. A clean speech signal with 8 kHz sampling frequency was used; contains female voice in it. Therefore, the following parameters have been selected: frame size $L = 256$ (32ms), FFT's size $N_{FFT} = 512$. In which, the TSNR approach output is used as an input of the HRNR approach. The experiment is conducted, by corrupting the speech by different types environment conditions with intensity varied from 2dB to 10dB. The channels used in Wiener filter are AWGN channel, Rayleigh channel and Rician channel. To illustrate the behavior and performance of this technique using male and female voice signal in the presence of different channels are shown in Figure 4.1 to 4.4 (a-d). The Figure 4.1 (a) to 4.4 (a) are shows the original clean speech signal and Figure 4.2 (b) to 4.4 (b) are shows the noisy speech signal. The Figure 4.1 (c) to Figure 4.4 (c) are represents the enhanced speech using TSNR approach and Figure 4.1 (d) to Figure 4.4 (d) represents the enhanced speech using HRNR approach. By comparing the cases (c) and (d) in each figure, it seems that numerous harmonics are maintained using HRNR approach since they are compressed when using TSNR approach.

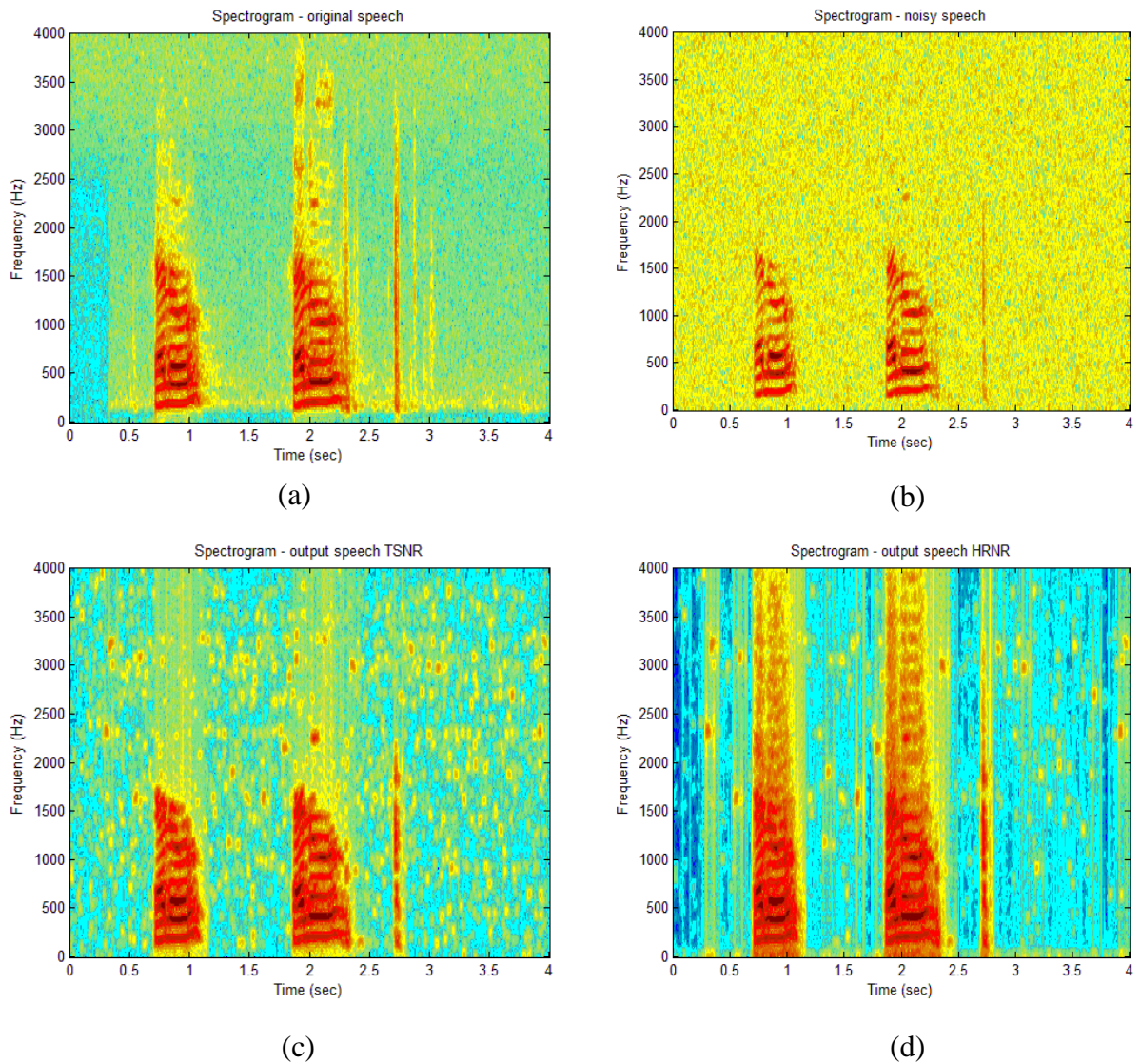


Figure 4.1 Spectrograms of Wiener filter using male voice in AWGN channel with SNR 2dB (a) Original speech signal (b) Noisy speech signal (c) Enhanced speech signal using TSNR technique (d) Enhanced speech signal using HRNR technique.

The spectrogram in Figure 4.1 illustrates the result of clean, noisy, TSNR and HRNR algorithm respectively. For simulation, a clean male voice signal with 8 kHz sampling frequency is added with AWGN of power level of 2dB. In DD approach, it produces the delay which generates reverberation effect. This effect is eliminated by TSNR algorithm. But signal is still suffers from harmonic distortion which are removed by HRNR. Therefore recovered signal is obtained.

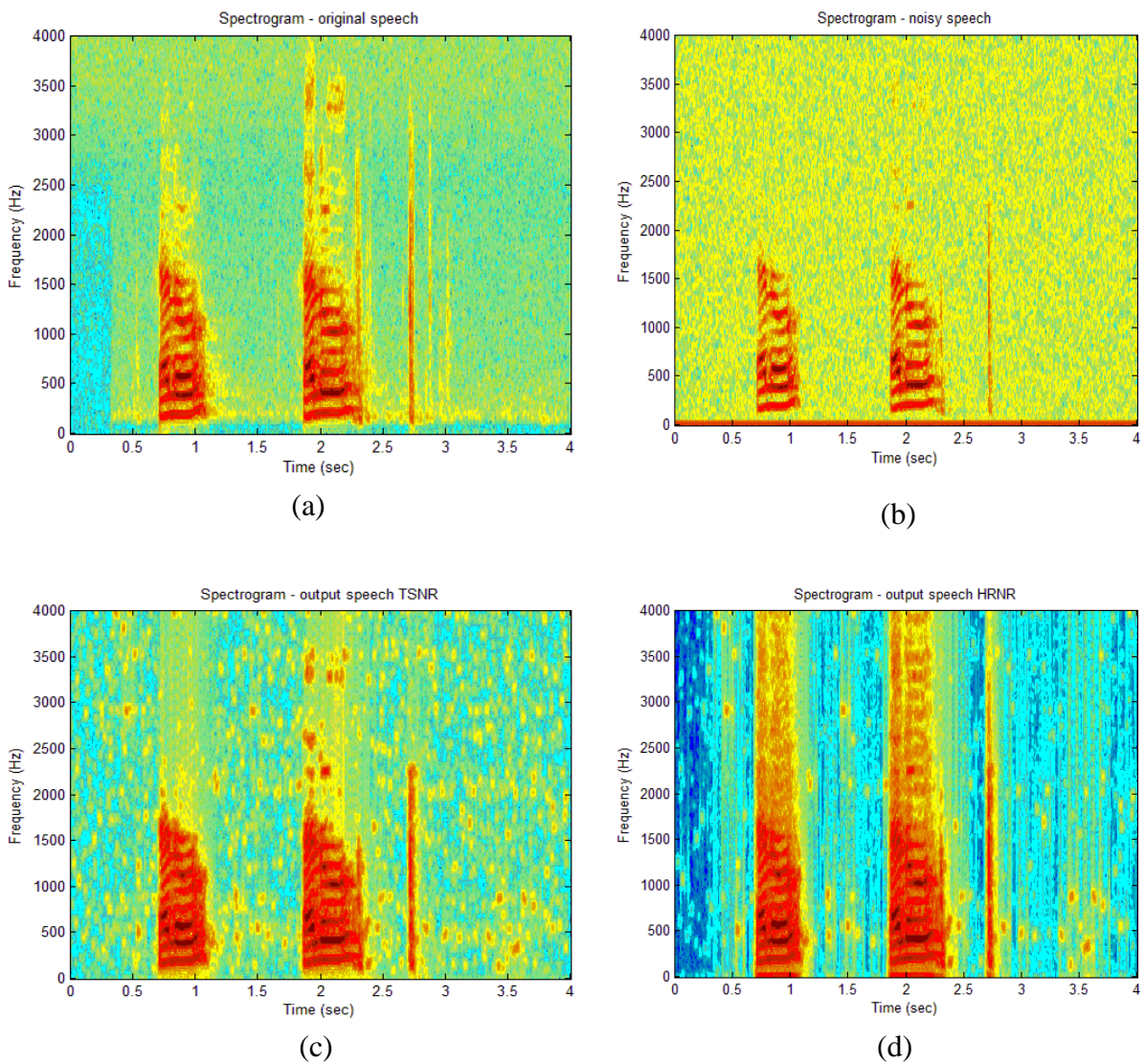


Figure 4.2 Spectrograms of Wiener filter using male voice in Rayleigh Channel with SNR 5dB (a) Original speech signal (b) Noisy speech signal (c) Enhanced speech signal using TSNR technique (d) Enhanced speech signal using HRNR technique.

The spectrogram in Figure 4.2 represents the result of clean, noisy, TSNR and HRNR algorithm using Wiener filter. For simulation, a clean male voice with 8 kHz sampling frequency is used. In communication process, when clean speech signal is passed through Rayleigh channel with power level of 5dB SNR. Then original signal gets distorted. Using TSNR algorithm, the reverberation effect is removed which produced by Rayleigh channel. But signal is still suffers from harmonic distortion which are removed by HRNR. Therefore recovered signal is obtained. But signal is still suffers from harmonic distortion which are removed by HRNR. Through Wiener filter, speech signal is recovered.

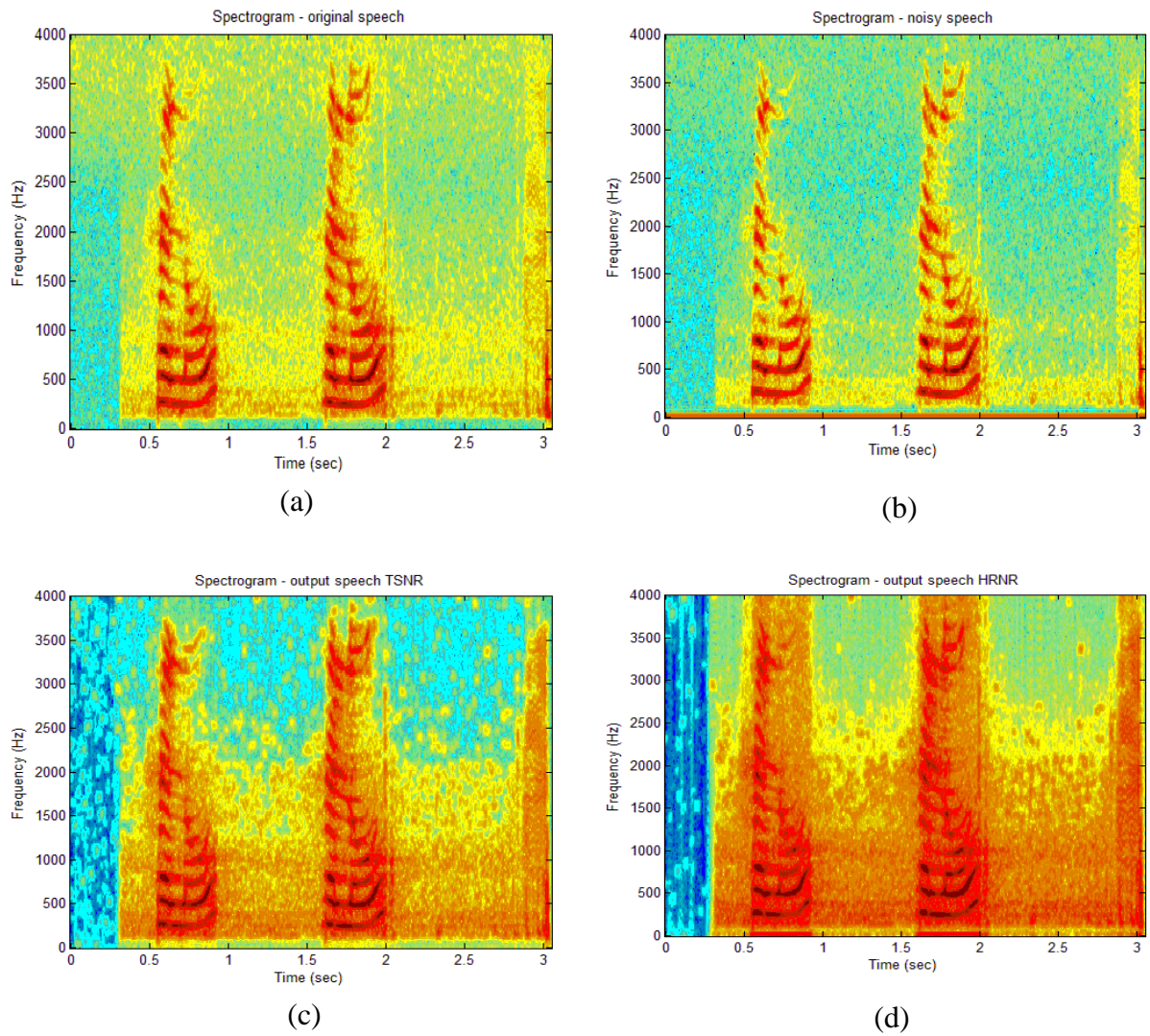


Figure 4.3 Spectrograms of Wiener filter using female voice in Rician Channel with SNR 7dB (a) Original speech signal (b) Noisy speech signal (c) Enhanced speech signal using TSNR technique (d) Enhanced speech signal using HRNR technique

The spectrogram in Figure 4.3 represents the result of noisy and recovered speech signal using Wiener filter. For simulation, a clean female voice with 8 kHz sampling frequency is used. When speech signal is passed through Rician channel at power level of 10dB SNR, then this channel distorts the original speech signal. Using Wiener algorithm, speech signal is recovered but in this more amount of speech is enhanced.

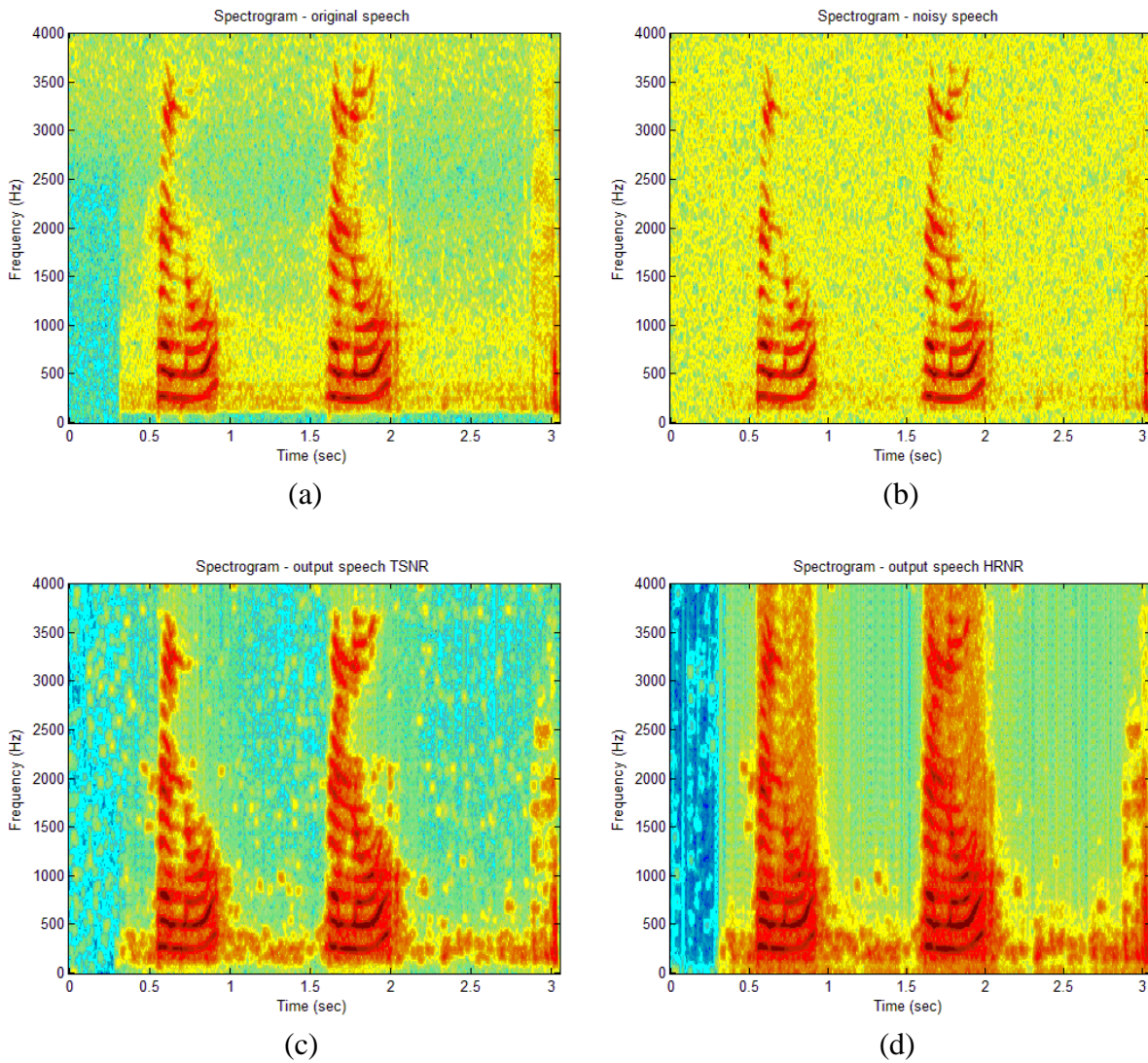


Figure 4.4 Spectrograms of Wiener filter using female voice in AWGN Channel with SNR 10dB (a) Original speech signal (b) Noisy speech signal (c) Enhanced speech signal using TSNR technique (d) Enhanced speech signal using HRNR technique

The spectrogram in Figure 4.4 shows the reconstructed signal using Wiener filter. For simulations, a clean speech signal with 8 kHz sampling frequency has been used, which contains female voice. An AWGN of power level 10dB SNR is added to it, which degrades the quality of speech signal and distorts it. Using the Wiener filter algorithm, noise is removed and the recovered signal is approximately equal to the original signal.

CHAPTER 5

RESULTS AND DISCUSSION

This chapter includes the performance analysis of speech enhancement algorithms i.e. spectral subtraction and Wiener filter which are described in the previous chapters. In both filters, speech is degraded by adding different noises such as additive white Gaussian noise (AWGN), and passing through Rayleigh channel and Rician channel with different SNR levels. Also compares the both algorithms corresponding to their improved SNR values.

5.1 PERFORMANCE MEASURES

The performance of the implemented technique is calculated by the objective measurements. The metrics used in this thesis is SNR. SNR is defined as ratio of desired signal to additive noise, using a logarithmic scale as

$$SNR_{dB} = 10 * \log_{10} \left(\frac{P_{signal}}{P_{noise}} \right) \quad (5.1.1)$$

Where, P_{signal} represents the power of clean speech signal and P_{noise} represents the power of additive noise.

The SNR improvement is acquired by calculating the SNR at the input denoted as iSNR and also at the output which is denoted as oSNR of the enhancement system and then subtracting the iSNR from the oSNR, i.e.

$$SNR \text{ improvement} = oSNR - iSNR$$

5.2 COMPARATIVE ANALYSIS OF BOTH FILTERS IN DIFFERENT CHANNELS

The spectral subtraction and Wiener filter are estimated under different noise environments such as AWGN, Rayleigh and Rician channel with the corresponding information. The Table 5.1, 5.2, and 5.3, showing the results of improved SNR values for varying input SNR levels.

In the Table 5.1, shows improved SNR for the AWGN channel at different input SNR levels i.e. 2 to 10 dB using spectral subtraction filter and Wiener filter. Table 5.2 represents the improved SNR for the Rayleigh channel at different input SNR levels i.e. 2 to 10 dB using spectral subtraction filter Wiener filter. Also Table 5.3, presents the for the Rician channel at different input SNR levels i.e. 2 to 10 dB using spectral subtraction filter and Wiener filter.

Signals	Input SNR (dB)	Improved SNR (dB)	
		Spectral subtraction filter	Wiener filter
Male	2	7.2094	12.6389
	5	7.3394	15.3289
	7	7.3549	17.2814
	10	7.3724	20.2443
Female	2	9.5060	17.4705
	5	9.5656	20.1602
	7	9.5250	22.1613
	10	9.7265	25.1026

Table 5.1 Performance of spectral subtraction and Wiener filter in AWGN channel

In this experiment, the improvement in SNRs has been compared with the spectral subtraction and Wiener filter using a tabular form. The Table 5.1 presents the comparison of SNR for male and female speech signal. At the input of 10dB SNR, Wiener filter gives improved SNR 20.2443dB as long as at the same input SNR level improved SNR in spectral subtraction filter is 7.3724dB. On the other hand using female voice, the improved SNR in Wiener filter is 25.1026dB at the 10dB SNR input level which is more than the spectral subtraction algorithm.

Signals	Input SNR (dB)	Improved SNR (dB)	
		Spectral subtraction filter	Wiener filter
Male	2	4.3131	6.3266
	5	4.3637	9.2289
	7	4.4381	11.2375
	10	4.4184	14.0926
Female	2	6.5723	11.2384
	5	6.6463	14.1043
	7	6.6063	16.0209
	10	6.7431	18.9659

Table 5.2 Performance of spectral subtraction and Wiener filter in Rayleigh channel

Table 5.2 contains the values considering the case where male and female speech signals are mixed with Rayleigh fading channel. In this Experiment, at the highest 10dB input SNR level, spectral subtraction filter gives the improved SNR of 6.7431dB though Wiener filter gives the 18.9659dB improved SNR at the same sampling frequency. It can be recognized that there is a significant improvement in SNR using Wiener filter at the different input SNR levels takes place.

Signals	Input SNR (dB)	Improved SNR (dB)	
		Spectral subtraction filter	Wiener filter
Male	2	6.8678	10.6887
	5	6.8733	13.6047
	7	6.8775	15.5663
	10	6.8845	18.5468
Female	2	9.2097	15.6683
	5	9.2170	18.6448
	7	9.2174	20.6222
	10	9.2192	23.6023

Table 5.3 Performance of Wiener filter and spectral subtraction filter in Rician channel

Table 5.3 shows the comparison of spectral subtraction and Wiener filter in terms of improved SNR using male and female voice in Rician channel. Using male voice at the highest 10dB input SNR level, Wiener filter gives the 18.5468dB improved SNR and spectral subtraction filter gives the 6.8845dB improved SNR at the same sampling frequency. On the other hand using female voice, the improved SNR in Wiener filter is 23.6023dB and in spectral subtraction filter is 9.2192dB at the 10dB SNR input level. Overall comparison of above Tables, it shows that Wiener filter gives the best performance than spectral subtraction filter at the different input SNR levels.

DISCUSSION

In this chapter, the comparison of the noise reduction techniques is discussed in different conditions such as in the presence of AWGN channel, Rayleigh channel and Rician channel at different input SNR. According to the comparison table, it shows that both techniques are useful for reducing the noise in terms of SNR but Wiener noise reduction technique gives the best result than spectral subtraction noise reduction technique in the presence of all the channels using both male and female voices at the different input SNR. Using Wiener noise reduction technique, as input SNR increases, improvement in SNR also increases. On the other hand, using spectral subtraction noise reduction technique, as input SNR increases, improvement in SNR values remains same. Overall comparison of both noise reduction technique in AWGN, Rayleigh and Rician channel, Wiener filter gives the better performance in AWGN channel. Therefore, more noise is reduced in AWGN channel than the other channels. Since, at the highest SNR, less amount of noise is present in the speech in Wiener filter. Hence, Wiener filter is the best filter to remove the noise as compared to the spectral subtraction.

CHAPTER 6

CONCLUSION AND FUTURE SCOPE

When microphone is present at distance, then obtained speech is generally suffering from background noise, reverberation effect and other interferences therefore degradation of intelligibility and fidelity of speech takes place. In this thesis, two noise reduction techniques are used to remove the noise in single channel speech enhancement. These noise reduction techniques are spectral subtraction and Wiener filter. These filter algorithms helps in improving the SNR.

In spectral subtraction a noisy spectrum is subtracted from noisy speech spectrum to obtain the estimate of clean speech spectrum. Further, estimate of noise spectrum is updated at the instances when no useful speech signal is present and only noise exists. In the analysis of spectral subtraction algorithm SNR is improved with the linear increase of input SNR but not as much as required by the user. It means noise is removed in fewer amounts. Therefore, this noise reduction technique is not useful.

Another noise reduction technique is Wiener filtering where a priori SNR is a function of multiplicative gain. In DD method, the gain function does not match with the current frame then it initiates the frame delay which causes irritating reverberation effect. The TSNR technique is used to remove this reverberation effect while preserving the speech transitions however error in noise PSD estimation leads to harmonic distortions. In order to cope up with this problem, an artificial signal is used for improving the a priori SNR and used for evaluating a spectral gain which can replace the speech harmonics. According to the analysis of Wiener filter, the SNR is improved as the power level of input SNR increases. This increase in SNR leads to the noise reduction. Hence, Wiener filter gives the best performance in terms of SNR as compared to the spectral subtraction algorithm.

In future, the following work may be accomplished. The noise reduction filters can be utilized with the optimization techniques like Particle Swarm Optimization (PSO). This algorithm is used to enhance and optimize the performance of these noise reduction filters. This optimization strategy produces a better SNR ratio in comparison of the existing methods. As the SNR increases, the precision of the speech quality will increase. By enhancing the speech quality of speech, the hearing aid impaired person will hear the words more easily. Further these noise reduction filters can also be tested for musical noise.

LIST OF PUBLICATIONS

Performance of Filters in noise Separation from noise, *Recent Advances in Electrical and Electronics Engineering* (Communicated).

REFERENCES

- [1] Khilari Prachi and Bhope VP (2015). A Review on Speech to Text Conversion Methods, *International Journal of Advanced Research in Computer Engineering & Technology*, 4(7), 3067-3072.
- [2] K Sathiarekha and Kumaresan S (2016). A Survey on the Evolution of Various Voice Conversion Techniques, *International Conference on Advanced Computing and Communication Systems* [3rd: Coimbatore, INDIA: 2016].
- [3] Drugman Thomas. *Advances in Glottal Analysis and its Applications*, University of Mons, Faculty of Engineering, TCTS Lab, 2011.
- [4] Professor Lawrence Rabiner. Digital Speech Processing. Available at http://www.ece.ucsb.edu/Faculty/Rabiner/ece259/digital%20speech%20processing%20course/lecturesnew/Lecture%204winter_2012_6tp.pdf (Accessed on Jan-March 2015)
- [5] Sehgal Prayank and Jain RK (2013). Speech Processing, *International Journal of Engineering Sciences & Emerging Technologies*, 5(2), 83-87.
- [6] Lawrence Rabiner and Ronald Schafer. *Theory and Application of Digital Speech Processing*. US: Prentice Hall Press, 2010, 1-1056.
- [7] Karam Marc et al. (2008). Noise removal in speech processing using spectral subtraction, *Journal of Signal and Information Processing*, 5, 1146-1147.
- [8] Chen Zhixin (2011). Simulation of Spectral Subtraction Based Noise Reduction Method, *International Journal of Advanced Computer Science and Applications*, 2(8), 30-32.
- [9] Huang Gongping et al. (2014). A Family of maximum SNR filters for noise reduction, *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 22(12), 2034-2046.
- [10] Sumithra M G, "Certain explorations on speech enhancement techniques for automatic speaker recognition in noisy environment".
- [11] S Lakshmikanth , KR Natraj and KR Rekha (2014). Noise Cancellation in Speech Signal Processing-A Review, *International Journal of Advanced Research in Computer and Communication Engineering*, 3(1), 5175-5186.
- [12] Abd El-Fattah Marwa et al. (2014). Speech enhancement with an adaptive Wiener filter. *International Journal of Speech Technology*, 17(1), 53-64.
- [13] Vihari Siddala et al. (2016). Comparison of speech enhancement algorithms. *Twelfth International Multi-Conference on Information Processing*, [12th: Bangalore, India: 2016], 666-676.

- [14] Loizou Philipos C., Lobo Arthur and Hu Yi (2006). Subspace Algorithms for Noise Reduction in Cochlear Implants. *The Journal of the Acoustical Society of America*, 118(5), 2791–2793.
- [15] Jensen Rindom Jesper *et al.* (2012). Non-Causal Time-Domain Filters for Single Channel Noise Reduction, *IEEE Transactions on Audio, Speech, and Language Processing*, 20(5), 1526 – 1541.
- [16] Chen Jingdong, Benesty Jacob and Huang Yiteng (2007). On the optimal linear filtering techniques for noise reduction. *Speech Communication*, 49(4), 305-316.
- [17] K Ravi Kumar and Subbaiah PV (2016). A Survey on Speech Enhancement Methodologies, *I.J. Intelligent Systems and Applications*, 12(05), 37-45.
- [18] Noise Reduction Filter Available at <http://www.lectrosonics.com/Support/ASPEN/noise-reduction-filter.html> (Accessed on June 2014).
- [19] F. Boll Steven *et al.* (1979). Suppression of acoustic noise in speech using spectral subtraction, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 27(2), 113-120.
- [20] Bharti Shambu Shankar, Gupta Manish and Agarwal Suneeta (2016). A New Spectral Subtraction Method for Speech Enhancement Using Adaptive Noise Estimation, *International Conference on Recent Advances in Information Technology*, [3rd: Dhanbad, India: 2016].
- [21] Chen Jingdong *et al.* (2006). New insights into the noise reduction Wiener filter, *IEEE Transactions on Audio, Speech, and Language Processing*, 14(4), 1218-1234.
- [22] Scalart Pascal and Filho Vieira Jozue (1996). Speech Enhancement based on a priori Signal to Noise Estimation, *IEEE International Conference on Acoustics, Speech, and Signal Processing*, [Atlanta, GA, USA: 2016], 629-632.
- [23] Plapous Cyril, Marro Claude and Scalart Pascal (2006). Improved Signal-to-Noise Ratio Estimation for Speech Enhancement, *IEEE Transactions on Audio, Speech, and Language Processing*, 14(6), 2098 – 2108.
- [24] Plapous Cyril Plapous *et al.* (2004). A two-step noise reduction technique, *IEEE International Conference on Acoustics, Speech, and Signal Processing*, [4th: Montreal, Que. Canada 2004], 289-292.
- [25] Plapous Cyril, Marro Claude and Scalart Pascal (2005). Speech enhancement using harmonic regeneration, *IEEE International Conference on Acoustics, Speech, and Signal Processing*, [Shanghai, USA: 2005], 157-160.

- [26] Jensen Rindom Jesper *et al.* (2012). Non-causal time-domain filters for single channel noise reduction, *IEEE Transactions on Audio, Speech, and Language Processing*, 20(5), 1526 – 1541.
- [27] Cohen Israel (2004). Speech Enhancement Using a Non-causal a Priori SNR Estimator, *IEEE Signal Processing Letters*, 11(9), 725-728.
- [28] Abramson Ari and Cohen Israel (2007). Simultaneous Detection and Estimation Approach for Speech Enhancement, *IEEE Transactions on Audio, Speech, and Language Processing*, 15(8), 2348-2359.
- [29] Abramson Ari and Cohen Israel (2007). Simultaneous Detection and Estimation Approach for Speech Enhancement, *IEEE Transactions on Audio, Speech, and Language Processing*, 15(8), 2348-2359.
- [30] Cohen Israel (2005). Speech Enhancement Using Super-Gaussian Speech Models and Non-Causal a priori SNR Estimation, *Speech Communication*, 47(3), 336-350.
- [31] Kuklasinski Adam (2016). Maximum Likelihood PSD Estimation for Speech Enhancement in Reverberation and Noise, *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 24(9), 1599-1612.
- [32] Taghia Jalal *et al.* (2011). An Evaluation of noise power spectral density estimation algorithms in adverse acoustic environments, *IEEE International Conference on Acoustics, Speech and Signal Processing*, [36th: Prague, Czech Republic: 2011], 4640-4643.
- [33] Martin Rainer *et al.* (2001). Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum Statistics, *IEEE Transactions on Speech and Audio Processing*, 9(5), 504-512.
- [34] Ephraim Yariv and Malah David (1984). Speech Enhancement Using a- Minimum Mean- Square Error Short-Time Spectral Amplitude Estimator, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 32(6), 1109-1121.
- [35] Adeyemo ZK *et al.* (2012). Comparative evaluation of fading channel model selection for mobile wireless transmission system. *International Journal of Wireless & Mobile Networks*, 4(6), 127-138.
- [36] Yan Xin. *Interference Management in Wireless Communication Systems: Theory and Applications*. EURASIP Journal on Wireless Communications and Networking, 2011.
- [37] Chavan MS, Chile RH and Sawant SR (2011). Multipath Fading Channel Modelling and Performance Comparison of Wireless Channel Models. *International Journal of Electronics and Communication Engineering*, 4(2), 189-203.

- [38] ATM Shafiul Alam, “Effect of Additive White Gaussian Noise (AWGN) on the Transmitted Data”.
- [39] Babu A. Sudhir and Rao K.V Sambasiva (2011). Evaluation of BER for AWGN, Rayleigh and Rician Fading Channels under Various Modulation Schemes, *International Journal of Computer Applications*, 26(9), 23-28.
- [40] Wenwu Wang. Media Processing. Available at <http://personal.ee.surrey.ac.uk/Personal/W.Wang/teaching/EEE2044/third.ppt> (Accessed on 2013).
- [41] Dixit Sunita, Yusuf Mulge MD (2014). Review on Speech Enhancement Techniques. *International Journal of Computer Science and Mobile Computing*, 3(8), 285-290.

ORIGINALITY REPORT

% **17**
SIMILARITY INDEX

% **3**
INTERNET SOURCES

% **15**
PUBLICATIONS

% **4**
STUDENT PAPERS

PRIMARY SOURCES

1 C. Plapous. "Improved Signal-to-Noise Ratio Estimation for Speech Enhancement", IEEE Transactions on Audio Speech and Language Processing, 11/2006 % **1**
Publication

2 osaptesting.osa.org % **1**
Internet Source

3 Karam, Marc, Hasan F. Khazaal, Heshmat Aglan, and Cliston Cole. "Noise Removal in Speech Processing Using Spectral Subtraction", Journal of Signal and Information Processing, 2014. % **1**
Publication

4 Jensen, Jesper Rindom, Jacob Benesty, Mads Græsbøll Christensen, and Søren Holdt Jensen. "Non-Causal Time-Domain Filters for Single-Channel Noise Reduction", IEEE Transactions on Audio Speech and Language Processing, 2012. % **1**
Publication

5

Submitted to Higher Education Commission
Pakistan

Student Paper

% 1

6

Kirubagari, B., and S. Palanivel. "Three-stage hybrid system for speech signal enhancement", International Journal of Signal and Imaging Systems Engineering, 2015.

Publication

% 1

7

J. Chen. "New Insights Into the Noise Reduction Wiener Filter", IEEE Transactions on Audio Speech and Language Processing, 7/2006

Publication

<% 1

8

P. Scalart. "Reliable a posteriori signal-to-noise ratio features selection", IEEE Workshop on Applications of Signal Processing to Audio and Acoustics 2005, 2005

Publication

<% 1

9

Submitted to Swinburne University of
Technology

Student Paper

<% 1

10

Chen, J.. "On the optimal linear filtering techniques for noise reduction", Speech Communication, 200704

Publication

<% 1

11

C. Plapous, C. Marro, P. Scalart. "Improved Signal-to-Noise Ratio Estimation for Speech

<% 1

Enhancement", IEEE Transactions on Audio, Speech and Language Processing, 2006

Publication

12

Huang, Gongping, Jacob Benesty, Tao Long, and Jingdong Chen. "A Family of Maximum SNR Filters for Noise Reduction", IEEE/ACM Transactions on Audio Speech and Language Processing, 2014.

Publication

13

Sid-Ahmed Selouani. "Speech enhancement employing a sigmoid -type gain function with a modified a priori signal-to-noise ratio (SNR) estimator", 2008 Canadian Conference on Electrical and Computer Engineering, 05/2008

Publication

14

SpringerBriefs in Electrical and Computer Engineering, 2016.

Publication

15

I. Cohen. "Relaxed Statistical Model for Speech Enhancement and a Priori SNR Estimation", IEEE Transactions on Speech and Audio Processing, 9/2005

Publication

16

Bharti, Shambhu Shankar, Manish Gupta, and Suneeta Agarwal. "A new spectral subtraction method for speech enhancement using adaptive noise estimation", 2016 3rd

<% 1

<% 1

<% 1

<% 1

<% 1

International Conference on Recent Advances in Information Technology (RAIT), 2016.

Publication

17

www.ijetr.org

Internet Source

<% 1

18

acm2008.cct.lsu.edu

Internet Source

<% 1

19

Upadhyay, Navneet, and Abhijit Karmakar. "Spectral Subtractive-Type Algorithms for Enhancement of Noisy Speech: An Integrative Review", International Journal of Image Graphics and Signal Processing, 2013.

Publication

<% 1

20

Submitted to Afyon Kocatepe University

Student Paper

<% 1

21

Ykhlef, Farid, Hadjer Ykhlef, and Abdelkader Aissat. "Influence of Dolph-Chebyshev window on speech enhancement", 2012 International Conference on Multimedia Computing and Systems, 2012.

Publication

<% 1

22

V GARG. "Radio Propagation and Propagation Path-Loss Models", Wireless Communications & Networking, 2007

Publication

<% 1

23

Tasmaz, H.. "Speech enhancement based on

undecimated wavelet packet-perceptual filterbanks and MMSE-STSA estimation in various noise environments", Digital Signal Processing, 200809

Publication

<% 1

24

Submitted to Jawaharlal Nehru Technological University

Student Paper

<% 1

25

Taghia, Jalal, Jalil Taghia, Nasser Mohammadiha, Jinqiu Sang, Vaclav Bouse, and Rainer Martin. "An evaluation of noise power spectral density estimation algorithms in adverse acoustic environments", 2011 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), 2011.

Publication

<% 1

26

Abe, Tomomi. "A study on single channel noise reduction =Tan'itsu chaneru zatsuon yokusei shuho ni kansuru kenkyu", DSpace at Waseda University, 2013.

Publication

<% 1

27

Wang, Xianghui, Jacob Benesty, and Jingdong Chen. "A single-channel noise cancelation filter in the short-time-fourier-transform domain", 2016 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), 2016.

Publication

<% 1

28

V, Sunnydayal., N. Sivaprasad, and T. Kishore Kumar. "A Survey on Statistical Based Single Channel Speech Enhancement Techniques", International Journal of Intelligent Systems and Applications, 2014.

Publication

<% 1

29

Yong, Pei Chee, Sven Nordholm, and Hai Huyen Dam. "Trade-off evaluation for speech enhancement algorithms with respect to the a priori SNR estimation", 2012 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), 2012.

Publication

<% 1

30

P. Scalart. "Speech Enhancement Using Harmonic Regeneration", Proceedings (ICASSP 05) IEEE International Conference on Acoustics Speech and Signal Processing 2005, 2005

Publication

<% 1

31

Upadhyay, Navneet, and Abhijit Karmakar. "The spectral subtractive-type algorithms for enhancing speech in noisy environments", 2012 1st International Conference on Recent Advances in Information Technology (RAIT), 2012.

Publication

<% 1

32

Chinaev, Aleksej, Alexander Krueger, Dang Hai Tran Vu, and Reinhold Haeb-Umbach.

<% 1

"Improved noise power spectral density tracking by a MAP-based postprocessor", 2012 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), 2012.

Publication

33

A. S. Madhukumar. "Single Carrier Cyclic Prefix-Assisted CDMA System with Frequency Domain Equalization for High Data Rate Transmission", EURASIP Journal on Wireless Communications and Networking, 2004

Publication

34

Israel Cohen. "Spectral Enhancement Methods", Springer Handbook of Speech Processing, 2008

Publication

35

Chin-Teng Lin. "Single-channel speech enhancement in variable noise-level environment", IEEE Transactions on Systems Man and Cybernetics - Part A Systems and Humans, 1/2003

Publication

36

Submitted to Universiti Utara Malaysia

Student Paper

37

Adam R. Gardner, Arnold D. Kim, Vasan Venugopalan. "Radiative transport produced by oblique illumination of turbid media with collimated beams", Physical Review E, 2013

38

www.mdpi.com

Internet Source

<% 1

39

Vaseghi. "Speech Enhancement in Noise",
Advanced Digital Signal Processing and Noise
Reduction, 11/11/2005

Publication

<% 1

40

Hadhami, Issaoui, and Bouzid Aicha. "Speech
Signal Enhancement Using Empirical Mode
Decomposition and Adaptive Method Based on
the Signal for Noise Ratio Objective
Evaluation", International Review on
Computers and Software (IRECOS), 2014.

Publication

<% 1

41

techreports.cs.queensu.ca

Internet Source

<% 1

42

www.ijclnlp.org

Internet Source

<% 1

43

Submitted to University of the West Indies

Student Paper

<% 1

44

Yu Fang. "Speech enhancement based on
modified a priori SNR estimation", Frontiers of
Electrical and Electronic Engineering in China,
12/2011

Publication

<% 1

Claudia Manfredi. "A simple subspace

- 45 approach for speech denoising", Logopedics
Phoniatics Vocology, 12/1/2001 <% 1
Publication
-
- 46 "IEEE-ICDCS conference proceeding", 2012
International Conference on Devices Circuits
and Systems (ICDCS), 03/2012 <% 1
Publication
-
- 47 ocw.mit.edu <% 1
Internet Source
-
- 48 R. Martin. "Noise power spectral density
estimation based on optimal smoothing and
minimum statistics", IEEE Transactions on
Speech and Audio Processing, 7/2001 <% 1
Publication
-
- 49 Upadhyay, Navneet, and Abhijit Karmakar.
"Speech Enhancement using Spectral
Subtraction-type Algorithms: A Comparison
and Simulation Study", Procedia Computer
Science, 2015. <% 1
Publication
-
- 50 Submitted to Indian Institute of Technology,
Kharagpure <% 1
Student Paper
-
- 51 Hongqing Miao. "Implementation of bi-channel
G.726 speech codec on fixed-point DSP",
International Conference on Neural Networks
and Signal Processing 2003 Proceedings of the <% 1

2003 ICNNSP-03, 2003

Publication

52

M Wilde, Mark, Patrick Hayden, Francesco Buscemi, and Min-Hsiu Hsieh. "The information-theoretic costs of simulating quantum measurements", *Journal of Physics A Mathematical and Theoretical*, 2012.

Publication

53

Submitted to Imperial College of Science, Technology and Medicine

Student Paper

54

Hansen, F. K., and K. M. Gorski. "Fast cosmic microwave background power spectrum estimation of temperature and polarization with Gabor transforms", *Monthly Notices of the Royal Astronomical Society*, 2003.

Publication

55

interesjournals.org

Internet Source

56

Submitted to SUNY, Binghamton

Student Paper

57

Gang Wang, , Chunguang Li, and Le Dong. "Noise Estimation Using Mean Square Cross Prediction Error for Speech Enhancement", *IEEE Transactions on Circuits and Systems I Regular Papers*, 2010.

Publication

<% 1

<% 1

<% 1

<% 1

<% 1

<% 1

58

Rao, Venkata Rama, Rama Murthy, and K. Srinivasa Rao. "Speech Enhancement Using Cross-Correlation Compensated Multi-Band Wiener Filter Combined with Harmonic Regeneration", Journal of Signal and Information Processing, 2011.

Publication

<% 1

59

Hong-Jiang Zhang. "Automated extraction of music snippets", Proceedings of the eleventh ACM international conference on Multimedia - MULTIMEDIA 03 MULTIMEDIA 03, 2003

Publication

<% 1

60

Love, . "Supersymmetric dark matter", Graduate Student Series in Physics, 2004.

Publication

<% 1

61

S. Sridharan. "Adaptive parameter compensation for robust hands-free speech recognition using a dual beamforming microphone array", Proceedings of 2001 International Symposium on Intelligent Multimedia Video and Speech Processing ISIMP 2001 (IEEE Cat No 01EX489) ISIMP-01, 2001

Publication

<% 1

62

Xia, Bingyin, and Changchun Bao. "Compressed domain speech enhancement method based on ITU-T G.722.2", Speech

<% 1

Communication, 2013.

Publication

63

Brill, Michael H.. "Doppler effect: surprises from the time domain", Journal of Nanophotonics, 2010.

Publication

<% 1

64

Upadhyay, Navneet, and Abhijit Karmakar. "Single channel speech enhancement utilizing iterative processing of multi-band spectral subtraction algorithm", 2012 2nd International Conference on Power Control and Embedded Systems, 2012.

Publication

<% 1

65

Robust Speech Recognition of Uncertain or Missing Data, 2011.

Publication

<% 1

66

Djendi, Mohamed, and Pascal Scalart. "Reducing over- and under-estimation of the a priori SNR in speech enhancement techniques", Digital Signal Processing, 2014.

Publication

<% 1

EXCLUDE QUOTES ON

EXCLUDE MATCHES < 10 WORDS

EXCLUDE BIBLIOGRAPHY ON