SPEAKER VERIFICATION IN NOISY ENVIRONMENT BY ENHANCING THE SPEECH SIGNAL BY VARIOUS APPROACHES OF SPECTRAL SUBTRACTION

by

Sriram J. 31510104106

Umar Ali S. 31510104117

Varadharajan A. 31510104120

A project report submitted to the

FACULTY OF INFORMATION AND

COMMUNICATION ENGINEERING

in partial fulfillment of the requirements

for the award of the degree of

BACHELOR OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING

SSN COLLEGE OF ENGINEERING

ANNA UNIVERSITY CHENNAI

CHENNAI - 600025

March 2014

CERTIFICATE

Certified that this project report titled **"Speaker Verification in Noisy Environment by Enhancing the Speech Signal by various approaches of Spectral Subtraction"** is the *bonafide* work of "**Sriram J. (31510104106)**, **Umar Ali S. (31510104117)**, and **Varadharajan A. (31510104120)**" who carried out the project work under my supervision. Certified further that to the best of my knowledge, the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or an award was conferred on an earlier occasion on these or any other candidates.

Dr. Chitra Babu
Head of the Department
Professor,
Department of CSE,
SSN College of Engineering,
Kalavakkam - 603 110

Ms. B. Bharathi
Supervisor
Assistant Professor,
Department of CSE,
SSN College of Engineering,
Kalavakkam - 603 110

Place:
Date:

Submitted for the examination held on............

Internal Examiner **External Examiner**

ACKNOWLEDGEMENT

I thank GOD, the almighty for giving me strength and knowledge to do this project.

I would like to thank and deep sense of gratitude to my guide **Ms. B.Bharathi**, Assistant Professor, Department of Computer Science and Engineering, for her valuable advice and suggestions as well as her continued guidance, patience and support that helped me to shape and refine my work.

My sincere thanks to **Dr. CHITRA BABU**, Professor and Head of the Department of Computer Science and Engineering, for her words of advice and encouragement and I would like to thank our project Coordinator **Mrs. J. Bhuvana**, Assistant Professor, Department of Computer Science and Engineering for her valuable suggestions throughout this first phase of project.

I express my deep respect to the founder **Dr. SHIV NADAR**, Chairman, SSN Institutions. I also express my appreciation to our **Dr. S. SALIVAHANAN**, Principal, for all the help he has rendered during this course of study.

I would like to extend my sincere thanks to all the teaching and non-teaching staffs of our department who have contributed directly and indirectly during the course of my project work. Finally, I would like to thank my parents and friends for their patience, cooperation and moral support throughout my life.

ABSTRACT

Speech processing is the study of speech signals and the processing methods of these signals. The signals are usually processed in a digital representation, so speech processing can be regarded as a special case of digital signal processing, applied to speech signal. Aspects of speech processing includes the acquisition, manipulation,storage, transfer and output of digital speech signals. Speech processing systems provide two major applications such as speaker identification and speaker verification. Speaker identification involves processing the voice signal and recognizing the speaker . Speaker verification involves authenticating the speaker. These two methods can be text dependent or text independent. Text dependent systems identifies the speaker based on a particular text while text independent systems identifies the speaker for a wide range of vocabulary. This speech signal is affected by various factors such as noise, channel mismatch, health condition of the speaker, aging, emotion, fatigue etc. In our project, we focus only on noise as the major factor assuming all other factors meet their standards. Speech signals from the uncontrolled environment may contain noise along with required speech components. We propose to use speech signal enhancement methods like spectral subtraction and modified versions of spectral subtraction methods such as :

1. Spectral Subtraction with over subtraction

2. Non linear Spectral Subtraction

3. Multiband Spectral Subtraction

are going to be applied in the speaker verification task and results of which are to be analysed.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

Introduction

1.1 Motivation

Speaker verification is the process of verifying a speaker using his/her voice. The user's voice is taken as the input and authentication is done. In this process, there are various environmental factors affecting the input voice signal of which, one of the most significant factor is the noise. Hence, a detailed research on various approaches which resolves this noise problem is needed.

1.2 Introduction

Speech processing systems provide two major applications such as speaker identification and speaker verification. Speaker identification involves processing the voice signal and recognizing the speaker . Speaker verification involves authenticating the speaker. These two methods can be text dependent or text independent. Text dependent systems identifies the speaker based on a particular text while text independent systems identifies the speaker for a wide range of vocabulary. This speech signal is affected by various factors such as:

1. Noise

2. Channel mismatch

3. Health condition of the speaker
4. Long term variability in people's voices(aging)
5. Emotion
6. Fatigue etc.

In this project, we focus only on noise as the major factor assuming all other factors meet their standards. Speech signals from the uncontrolled environment may contain noise along with required speech components. Speech signal degraded by additive noise, make the listening task difficult for a direct listener, giving poor performance in automatic speech processing tasks like speech recognition speaker verification, hearing aids, speech coders etc. The degraded speech therefore needs to be processed for the enhancement of speech components. The aim of speech enhancement is to improve the quality and intelligibility of degraded speech signal. Improving quality and intelligibility of speech signals reduces listeners fatigue; improve the performance of hearing aids, cockpit communication, videoconferencing, speech coders and many other speech systems. In this project, we are going to enhance the speech signal by using the principles of basic spectral subtraction algorithm, by implementing modified versions of spectral subtraction methods such as :

1. Spectral Subtraction with over subtraction,
2. Non linear Spectral Subtraction
3. Multiband Spectral Subtraction

Once the noise has been eliminated, we are going to build training models and testing models for different speakers and analyse their performances in each of the above methods.

CHAPTER 2

Literature Survey

There are various methods existing for speaker verification in noisy environment. Every method has it's own advantages and disadvantages. Hence, a literature survey is done to choose a method among the existing methods.

[1]Ji Ming, Timothy J. Hazen, James R. Glass, Douglas A. Reynolds, Robust Speaker Recognition in Noisy Conditions, IEEE Trans. Audio, Speech and Language Processing.,vol. 15, no. 5, pp 1711-1723,Jul 2007. [6]

The above mentioned paper investigates the problem of speaker verification in noisy environment but knowledge about the noise characteristics is not available. In this method training data is produced for each model by adding electronically generated noise to the clean speech. To make a model suitable for wide variety of noises, multiple training data set is produced for a model by adding different training noises to the clean speech. As the number of noise conditions increases the size of the model increases and this method does not suit when a noise that is not in training noise set corrupts the speech.

[2]Ramin Halavati, Saeed Bagheri Shouraki,Mina Razaghpour, Hossein Tajik, Arpineh Cholakian,A Novel Noise Immune ,Fuzzy Approach To Speaker Independent,Isolated Word Speech Recognition published in World Automation Congress(WAC),Conference in Budapest, Jul 2006. [7]

The above mentioned paper investigates the problem of isolated word speech recognition using fuzzy modeling which is specifically designed to ignore noise. The task is based on conversion of speech spectrogram into a linguistic fuzzy description and comparison of this representation with fuzzy linguistic descriptions of words. In this approach,instead of conventional acoustic features, fuzzy linguistic variables are used and a fuzzy rule base performs the final recognition of isolated words.

[3]Serajul Haque, Roberto Togneri, A Psychoacoustic Spectral Subtraction Method For Noise Suppression In Automatic Speech Recognition, IEEE International Conference on Acoustics Speech And Signal Processing. ,pp 1618-1621, Mar 2010. [5]

The above mentioned paper investigates the problem of speaker verification in noisy environment by enhancing the speech with spectral subtraction using a over subtraction factor and spectral floor with HMM recognizer. The method shows reduced residual noise and improved word recognition performance in broadband Gaussian noise conditions compared to conventional spectral subtraction method.

[4]W. Alkhaldi, W. Fakhr and N. Hamdy , Automatic Speech/Speaker Recognition In Noisy Environments Using Wavelet Transform , IEEE Conference on Circuits and Systems. ,vol. 1, pp 463-466,Aug. 2002. [8]

The above mentioned paper investigates the problem of speaker verification in noisy environment by using Discrete Wavelet Transform based feature extraction technique for multiband automatic speaker verification. This method shows better verification rates especially at low signal-to-noise ratios. Some motivations of the multi-band technique are:Human speech perception is multi-band by nature, In the case of speech corrupted by an additive noise, only some frequency sub-bands are smeared. It has provided comparable performance with the full-band (conventional) technique, under matched conditions(clean speech for both training and testing). Also, it has been found that combining the acoustic features of two techniques into one speaker recognizer can yield better performance than using either of them, under mismatched conditions((clean speech for training and noisy speech for testing). Moreover, this combining strategy has not degraded the performance under matched conditions.

[5]Anuradha R. Fukane, Shashikant L. Sahare, Different Approaches of Spectral Subtraction method for Enhancing the Speech Signal in Noisy Environments, International Journal of Scientific and Engineering Research, Volume 2, Issue 5, May-2011.ISSN 2229-5518.[1]

The above mentioned paper investigates the various approaches for enhancing the speech by spectral subtraction method starting from the basic approach of spectral subtraction and many modified approaches were used for speech enhancement in noisy environment.

[6]Marwa A. Abd El-Fattah Moawad I. Dessouky Alaa M. Abbas Salaheldin M. Diab El-Sayed M. El-Rabaie Waleed Al-Nuaimy Saleh A. Alshebeili Fathi E. Abd El-samie, Speech enhancement with an adaptive Wiener filter, Journal from Faculty of Electronic Engineering, Menoufia University,Egypt, Aug 2013. [9]

The above mentioned paper investigates the problem of speech enhancement using Adaptive Wiener Filtering Method. This method depends on the adaptation of the filter transfer function from sample to sample based on the speech signal statistics; the local mean and the local variance. A comparison is done between Adaptive wiener filtering method and other traditional methods such as spectral subtraction,wiener filtering and wavelet transform. The results show that the adaptive Wiener filtering method has the best performance as compared to all other speech enhancement methods.

[7]E. Verteletskaya, K. Sakhnov, Voice Activity Detection for Speech Enhancement Applications, Acta Polytechnica Vol. 50 No. 4/2010 [11].

The above mentioned paper investigates the problem of noise estimation by detecting the voiced and unvoiced parts of speech and the unvoiced part(i.e:silence parts) is considered as noise but always unvoiced part cannot be considered as noise because speech part may be affected by noise.

[8]Bc. Jan Kybic, Kalman Filtering and Speech Enhancement, Diploma work jan 1998 [20].

The above mentioned paper performs speech enhancement by recording the noisy speech and noise separately and then subtracting noise from the noisy speech but the problem with this method is that when the noise varies continuously in the environment this method is not feasible.

CHAPTER 3

Proposed System

3.1 Objective :

We propose to develop a speaker verification system which takes speaker's voice as input, under a noisy environment (assuming noise is the only factor which is going to affect the process) and eliminate the noise by spectral subtraction method .

Input: Speaker's voice

Output: Speaker is authenticated.

The proposed system is based on the study on existing system which performs a basic spectral subtraction method. In this project, we are performing the speaker verification in noisy environment by enhancing the speech by using the Spectral Subtraction method and its various approaches such as Spectral Subtraction with over subtraction factor, Non linear Spectral Subtraction , Multiband Spectral Subtraction.

The spectral subtractive algorithm is historically one of the first algorithms proposed for noise reduction. It is based on the principle that one can estimate and update the noise spectrum computed during non speech activity that is during speech pauses , and subtract it from the noisy speech signal to obtain clean speech signal spectrum. Assumption is noise is additive and its spectrum does not change with time.

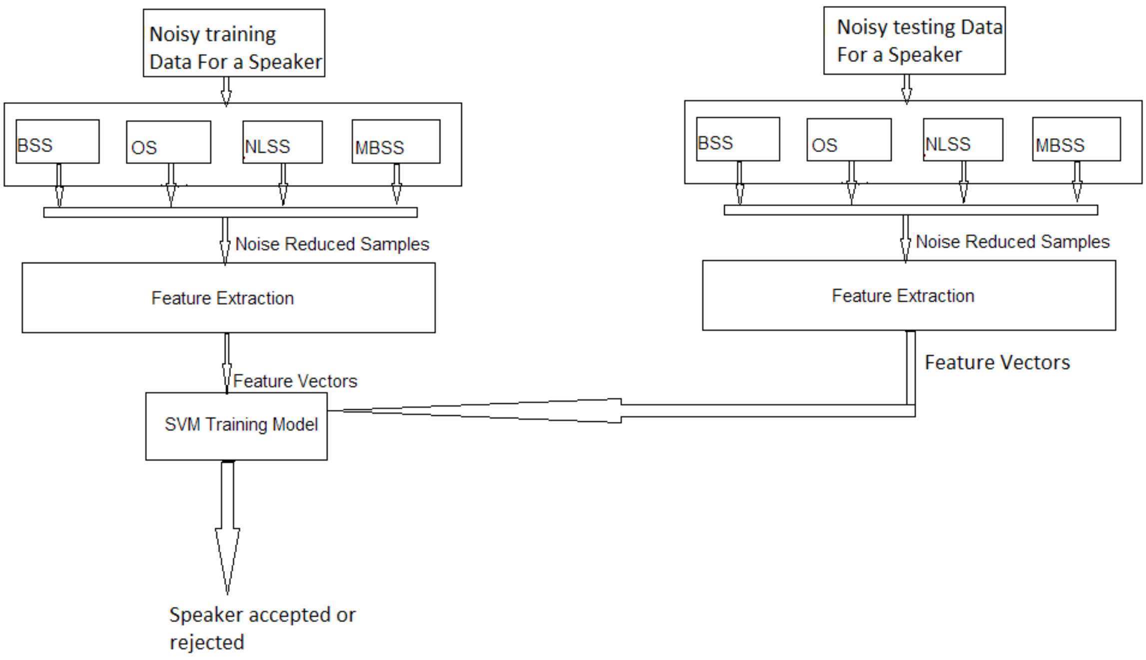

FIGURE 3.1: Overview of System Architecture

Speech Data collection:

Number of speakers : 3

Software used : Wavesurfer 8.5.8

Duration of utterences : 3.5 to 3.8 seconds

Total number of samples : 150

Samples are collected in the environment with **white noise**.

Noise Reduction: Four methods are used for noise reduction in the noisy data. They are:

1. Basic Spectral subtraction.

2. Spectral subtraction with over subtraction.

3. Non-linear spectral subtraction.

4. Multiband spectral subtraction.

3.1.1 Basic Spectral Subtraction(BSS):

System Design

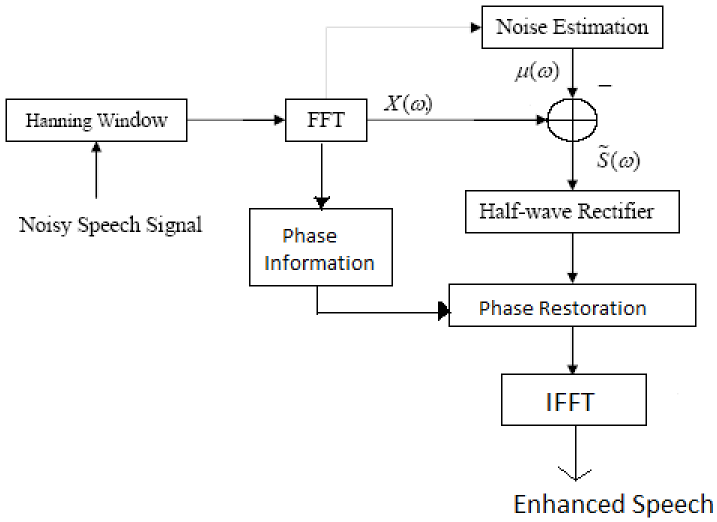

FIGURE 3.2: Basic Spectral Subtraction

Noisy Speech:

The noisy speech is a sample stored in the folder. The required sample is retrieved from folder using the matlab inbuilt function wavread. The return values of that function are analog signal ,number of samples and number of bits per sample.Then the plot function is used to view the waveform of noisy speech.

Windowing:

Speech is non-stationary signal where properties change quite rapidly over timeThis is fully natural it makes the use of DFT impossible. For most phonemes the properties of the speech remain invariant for a short period of time (5-100 ms). Thus for a short window of time, signal processing methods can be applied successfully. Most of speech processing is done in this way by taking short windows (overlapping possibly) and processing them. The short window of signal like this is called frame. In speech recognition the windows are usually overlapping windows, which are analyzed in order to make hypothesis of the current phoneme. The hypotheses are combined over several frames and finally the decision is made to maximize the joint probability. It is a good idea to use overlapping windows summing approximately to 1. **Hanning-window** has the property of summing to 1 when the time difference between successive windows is half of the length of the window.So hanning window has been used in this system.

Fast Fourier Transform(FFT):

The noisy signal in the time domain is converted to frequency domain to analyze the amplitude,frequencies,and phase cleary because in the time domain the values cannot be analyzed clearly. The magnitude of the FFT of the signal is taken using matlab inbuilt function abs which returns two channels the values of both the channels are equal one or both of it can be used further.

Phase Information:

The noisy signal is converted to frequency domain in which phase and magnitude present. Magnitude is affected by noise and phase is not affected by noise hence the phase is stored using the matlab inbuilt function **angle** which is used in later stage for restoration of phase.

Noise Estimation:

The estimation of noise is done using Adaptive noise spectral estimation. In which the noise is estimated in frequency domain, the values of alpha and beta are set to 0.9 and 2 respectively and noise is estimated for each frequency bin in each frame based on condition

noimag = sigmag;

if the sigmag(l, k) > beta∗noimag(l, k-1)

then

noimag(l, k) = noimag(l, k-1)

else

noiest = (1-alpha)*sigmag(l, k) + alpha*noimag(l, k-1)

noimag(l, k) = noiest;

Where

sigmag- magnitude of signal.

noimag- magnitude of noise.

noiest- estimate of noise.

Noise Subtraction and Half wave rectification:

The estimated noise is subtracted from the magnitude of the fft of the signal.If the estimated noise is greater than actual noise then subtraction lead to negative value in the magnitude so if the estimated noise is greater than actual noise then that part is replaced by zero this process is called as half wave rectification.

Restoration of phase information:

After the subtraction of noise and half wave rectification the phase which is already stored is restored to maintain phase information.

IFFT(Inverse Fast fourier transform):

After restoration of phase inverse fourier transform is applied using the matlab inbuilt function ifft to convert frequency domain to time domain to get the enhanced speech.

3.1.2 Spectral subtraction with over subtraction(OS)

In basic spectral subtraction due to half-wave rectification process small, isolated peaks in the spectrum occurs at random frequency locations in each frame. Converted in the timedomain, these peaks sound like tones with frequencies that change randomly from frame to frame. This new type of noise introduced by the half-wave rectification process has been described as warbling and of tonal quality, and is commonly referred as musical noise. Hence over subtraction is a method reduce musical noise in this method if the estimated noise is greater than actual noise then following formula is applied If b is too large,then the residual

if$|Yj(w)|^2 > (a + b)|De(w)|^2$
$|Xej(w)|^2 = |Yj(w)|^2 - |De(w)|^2$
else
$b|De(w)|^2$

With $a >= 1$ and $0 < b <= 1$.

$|Xej(w)|$ denotes the enhanced spectrum
$|De(w)|$ is the spectrum of the noise

Where a is over subtraction factor
b is the spectral floor parameter.

FIGURE 3.3: Over Subtraction

noise will be audible but the musical issues related to spectral subtraction reduces. Parameter a affects the amount of speech spectral distortion. If a is too large then resulting signal will be severely distorted and intelligibility may suffer. If a is too small noise remains in enhanced speech signal. When $a > 1$, the subtraction can remove all of the broadband noise by eliminating most of wide peaks.

So in this system value of is **a is set as 1** and value of **b is set as 0.3**

3.1.3 Non-linear spectral subtraction(NLSS)

NSS is a modification of the method suggested in by making the over subtraction factor frequency dependent and the subtraction process non-linear. In case of NSS assumption is that noise does not affects all spectral components equally. Certain types of noise may affect the low frequency region of the spectrum more than high frequency region. This suggests the use of a frequency dependent subtraction factor for different types of noise.Due to frequency dependent subtraction factor, subtraction process becomes nonlinear. Larger values are subtracted at frequencies with low SNR levels and smaller values are subtracted at frequencies with high SNR levels. The subtraction rule used in the NSS algorithm has the following form Where b is the spectral floor set to 0.1

if|Y(w)| >a(w) N (w) + b|De (w)|
|Xe (w)| = |Y (w)| - a(w) N (w)

else
b|Y(w)|

FIGURE 3.4: NLSS Subtraction Rule

$|Y(w)|$ and $|De(w)|$ are the smoothed estimates of noisy speech and noise respectively

a(w) is a frequency dependent subtraction factor

N(w) is a non-linear function of the noise spectrum where

N(w) = Max ($|De(w)|$)

The N(w) term is obtained by computing the maximum of the noise magnitude spectra $|De(w)|$ over the frames.

The a(w) is given as $a(w) = 1/r + p(w)$

Where r is a scaling factor and p(w) is the square root of the posteriori SNR estimate given as $P(w) = |Y(w)|/|De(w)|$

3.1.4 Multiband spectral subtraction(MBSS)

In MBSS approach the speech spectrum is divided into N overlapping bands and spectral subtraction is performed independently in each band. The processes of splitting the speech signal into different bands can be performed either in the time domain by using band pass filters or in the frequency domain by using appropriate windows. The estimate of the clean speech spectrum in the ith band is obtained by

$|Xei(wk)|^2$ = $|Yi(wk)|^2 - aidi|Di(wk)|^2$

$bi < wk < ei$

Where wk = 2pi k / N, k = 0, 1 ... N 1 are the discrete frequencies

$|Dei(wk)|^2$ is the estimated noise power spectrum obtained during speech absent segment ai is the over subtraction factor of the ith band

di is an additional band.

Subtraction factor can be individually set for each frequency band to customize the noise removal processor

bi and ei are the beginning and ending frequency bins of the ith frequency band. The band specific over subtraction factor is a function of the segmented SNRi of the ith frequency band and is computed as follows

$4.75\ SNRi < -5$

$ai = 3/20(SNRi) - 5 < SNRi < 20$

$1\ SNRi > 20$

The values for di are set to

$1\ fi < 1\ KHz$

$di = 2.5\ 1KHz < fi < (Fs/2)2KHz$

$1.5fi > (Fs/2)2KHz$

Where fi is the upper frequency of the ith band and Fs is the sampling frequency in Hz.

3.2 Feature Extraction

MFCC(mel frequency cepstral coefficients) is used for feature extraction

subtraction rule used in the NSS algorithm has the following form

Speech signal is segmented in to frames of 30 ms and overlapping factor is set to 256 and hamming window is applied.

Then DFT is calculated for each frame.

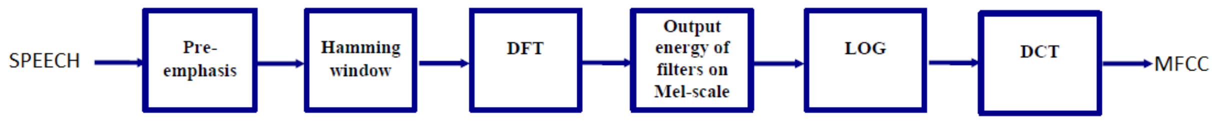

FIGURE 3.5: Feature Extraction

Filterbank Amplitudes is obtained by setting the setting number of filters as 20 in the filterbank and giving the length of fft and sampling frequency of each sample. Then magnitude of fft is multiplied with filterbank amplitudes.

one set of 20 MFCC coefficients is extracted for each and every single frame after taking log and direct cosine transform.

3.3 Training and Testing

SVM(support vector machine):

support vector machines are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model .An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on.

Training and testing are done using the software Libsvm 3.17. From the features of all the samples training and testing files are created for all the three speakers in the system.Training for a speaker in libsvm using gives svm-model as output. Testing is done in libsvm using the testing file and the svm-model obtained from training which gives the speaker verification accuracy and file which contains true and false scores of verification.

CHAPTER 4

Evaluation and results

Spectrum of speech enhanced by basic spectral subtraction for one sample: u18.wav is the noisy speech sample of one speaker and basicss_u18.wav is enhanced speech obtained by basic spectral subtraction.

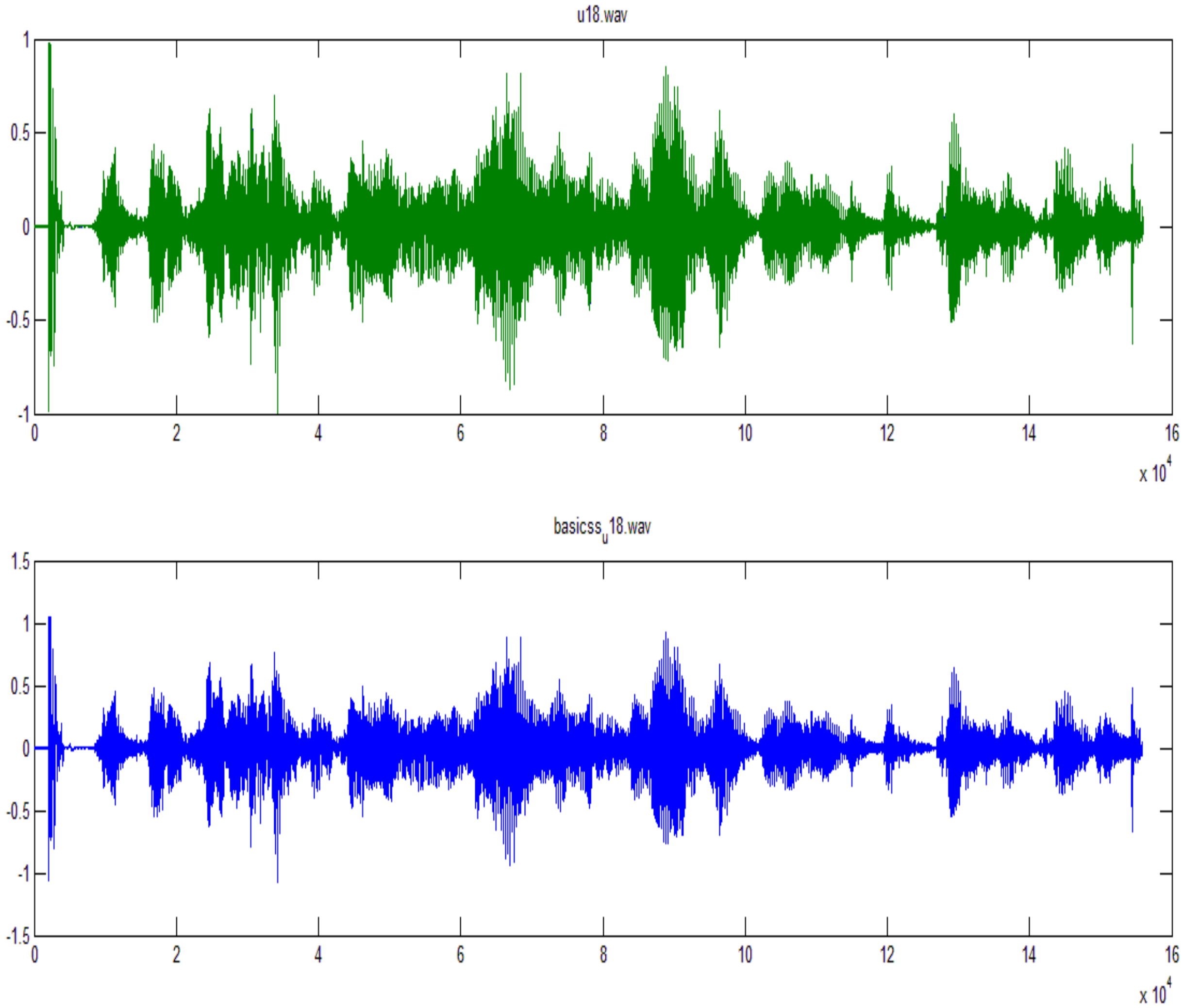

FIGURE 4.1: BSS

Spectrum of speech enhanced by over subtraction for one sample:

u18.wav is the noisy speech sample of one speaker and oversub_u18.wav is enhanced speech obtained by over subtraction.

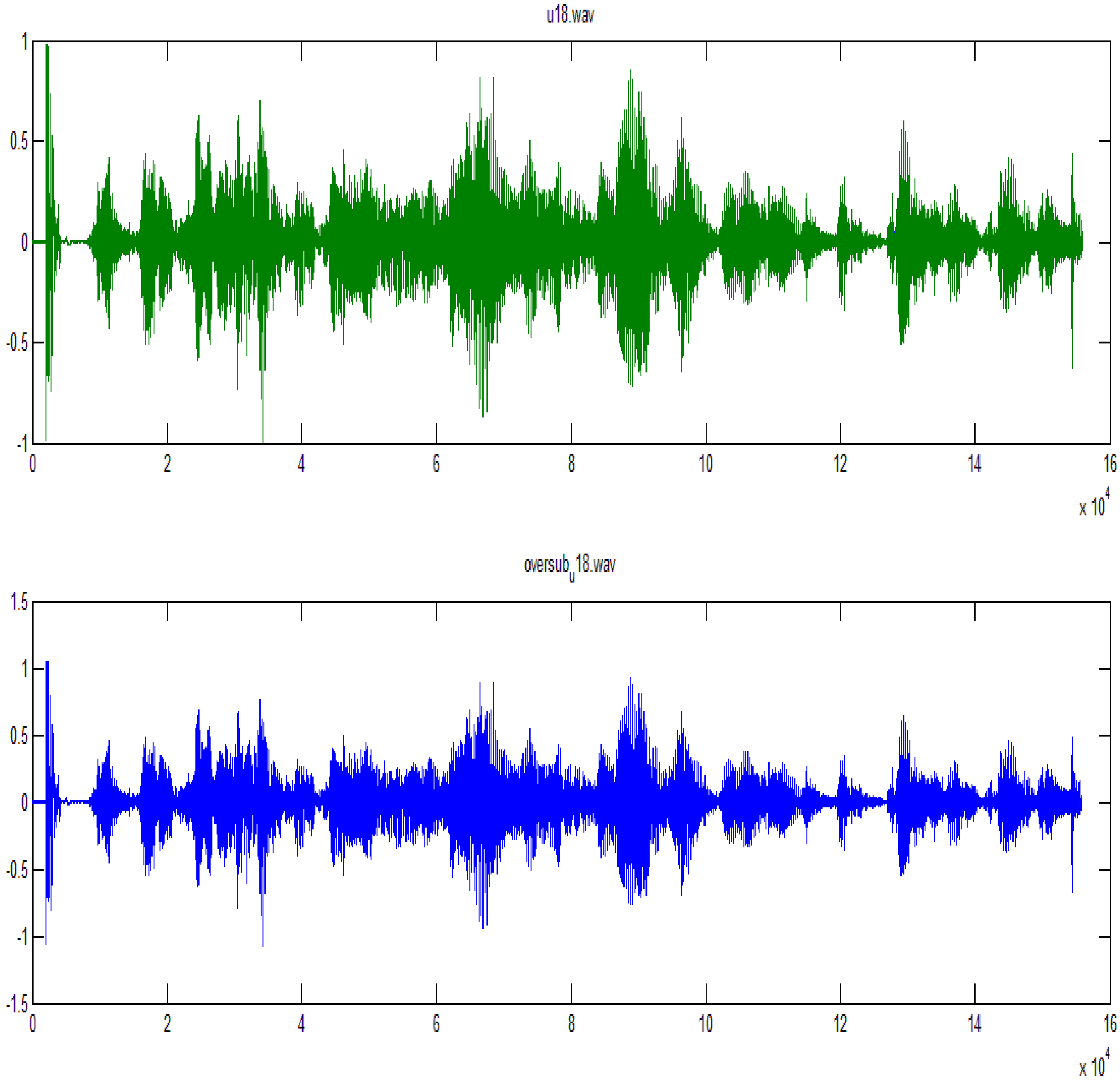

FIGURE 4.2: OS

Spectrum of speech enhanced by non-linear spectral subtraction for one sample:

u18.wav is the noisy speech sample of one speaker and nlss_u18.wav is enhanced speech obtained by non linear spectral subtraction.

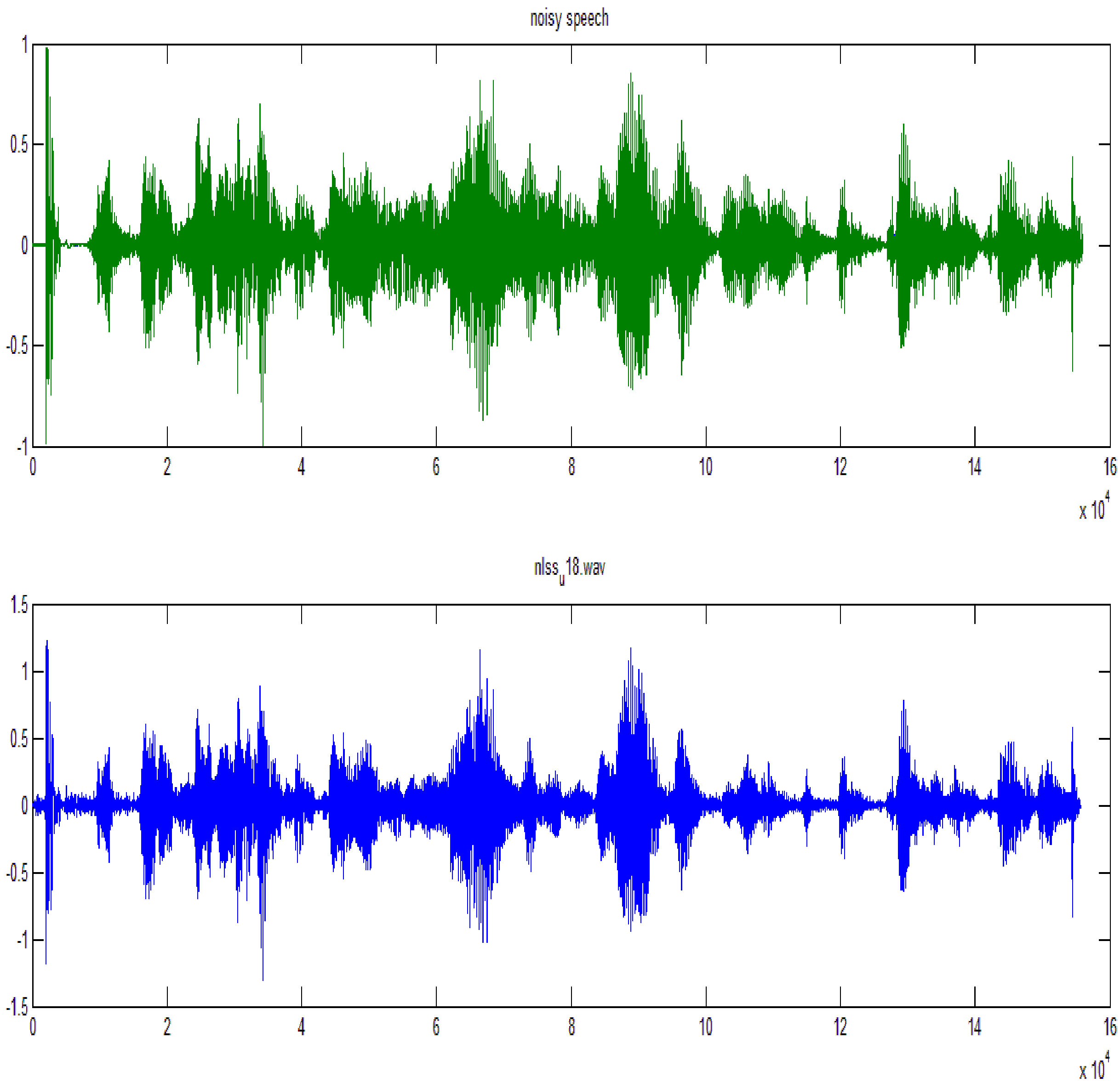

FIGURE 4.3: NLSS

Spectrum of speech enhanced by multiband spectral subtraction for one sample:

u18.wav is the noisy speech sample of one speaker and mbss_u18.wav is enhanced speech obtained by multiband spectral subtraction.

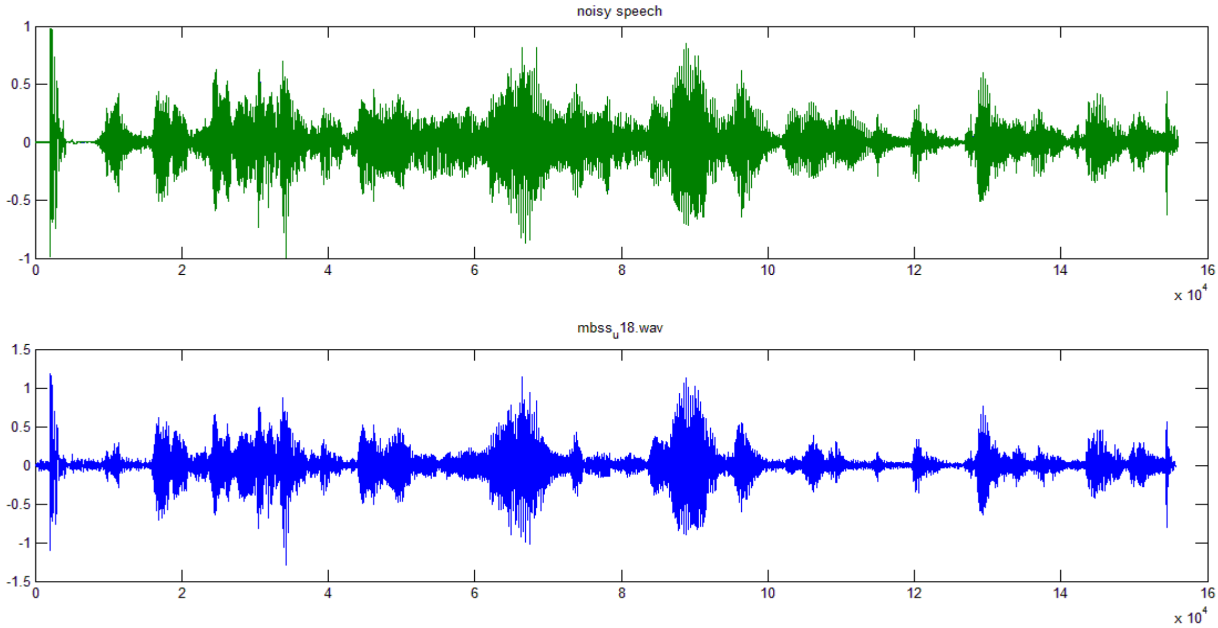

FIGURE 4.4: MBSS

Signal to noise ratio(SNR):

Signal to noise ratio is defined as the ratio of power of signal to power of noise.

SNR= [Power of signal]/[Power of noise]

If the SNR value is low, amount of noise present in the signal is high as they are inversly proportional and vice versa.

Average of SNR values of speech samples of all the speakers in the system for noisy an noise enhanced speech.

Speech	SNR
Noisy	-1.0865304
BSS	-1.069016933
OS	-1.086080467
NLSS	-1.081474867
MBSS	-1.082062133

TABLE 4.1: SNR COMPARISON

Feature Extraction:

Features of a speech sample for a speaker.Number of columns is 20 which indicates the number of features and number of rows is 1633 which indicates frames with overlapping factor 256.

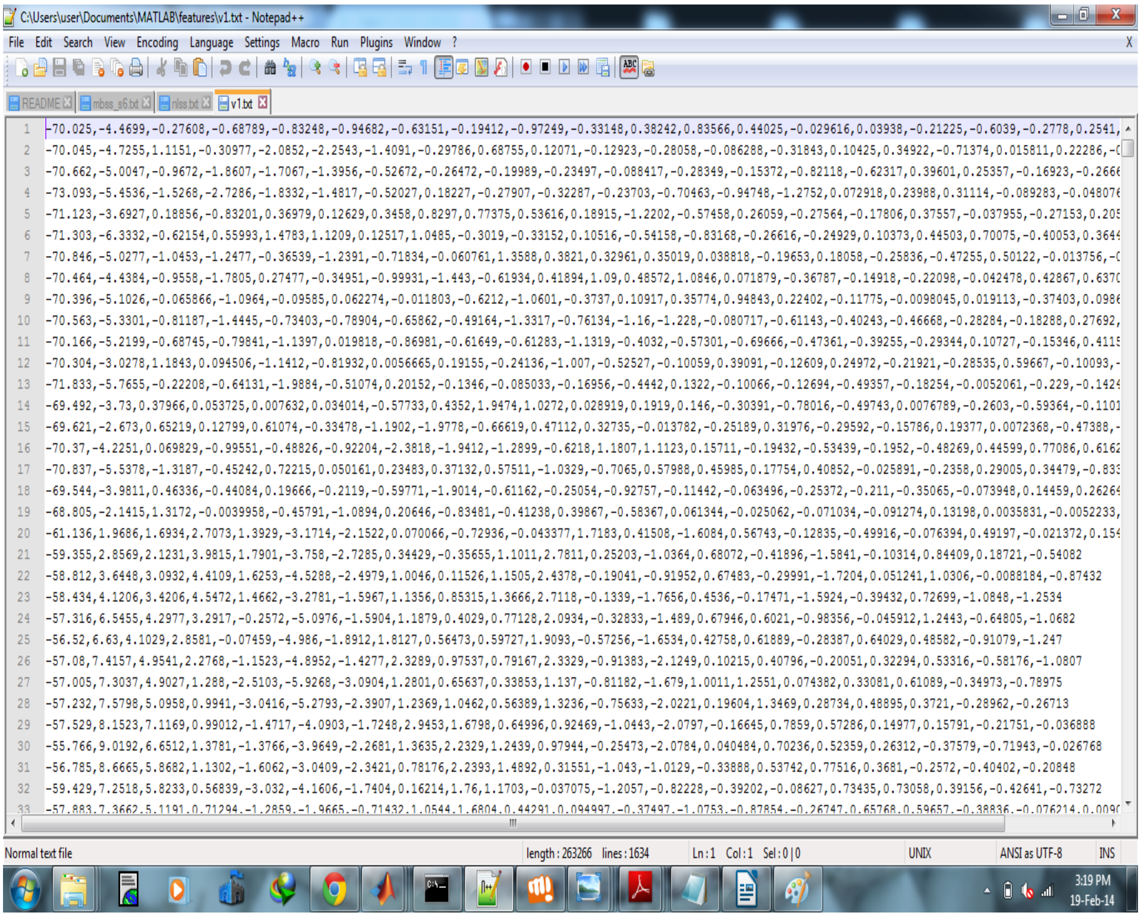

FIGURE 4.5: MFCC

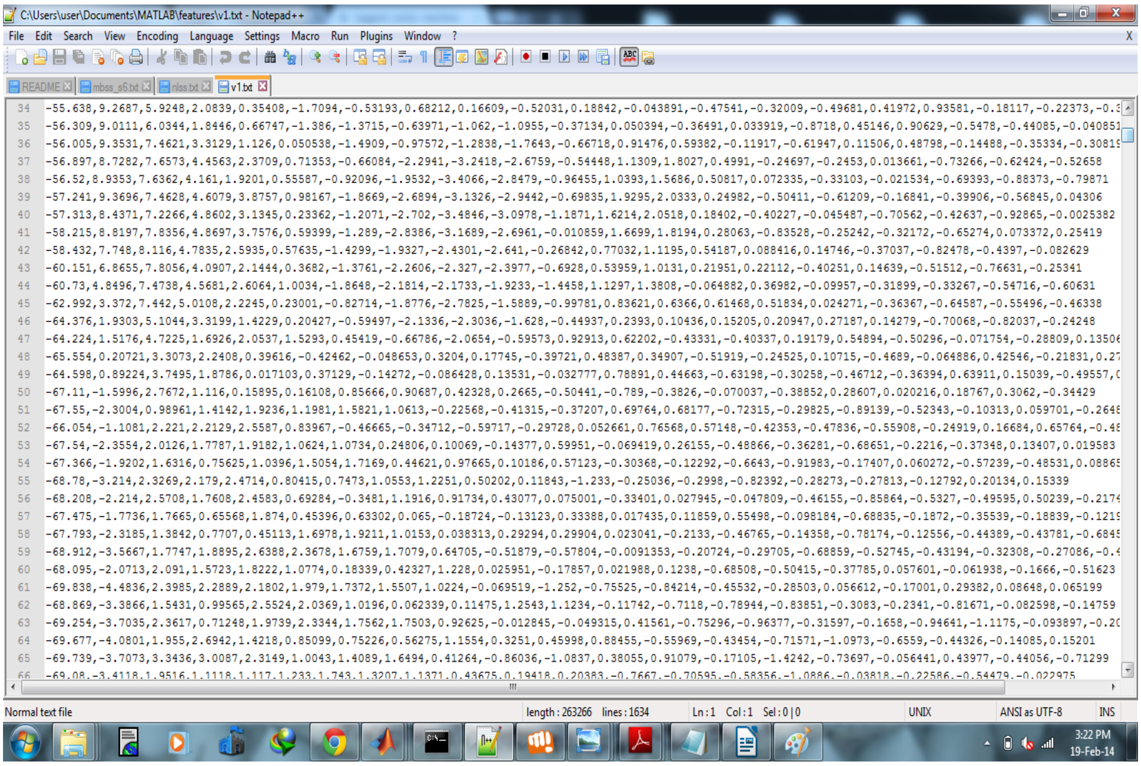

FIGURE 4.6: MFCC

Creating a training model in libsvm by using the training file:

In libsvm training model is created in linear kernel and with probability estimates.

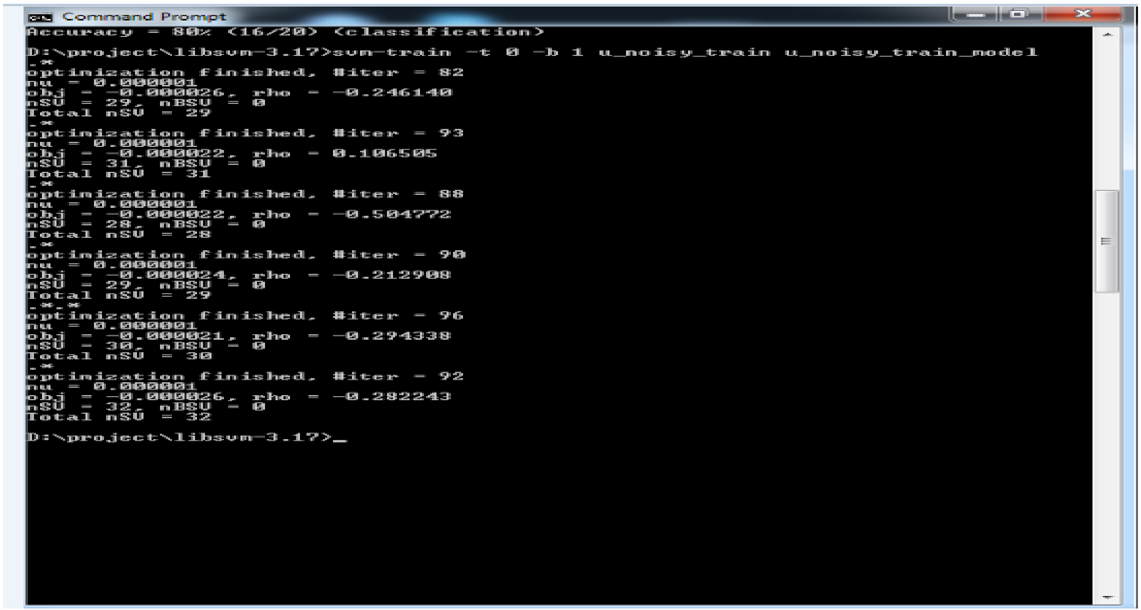

FIGURE 4.7: TRAIN MODEL

Testing for a speaker with noise reduced testing samples in noisy train model

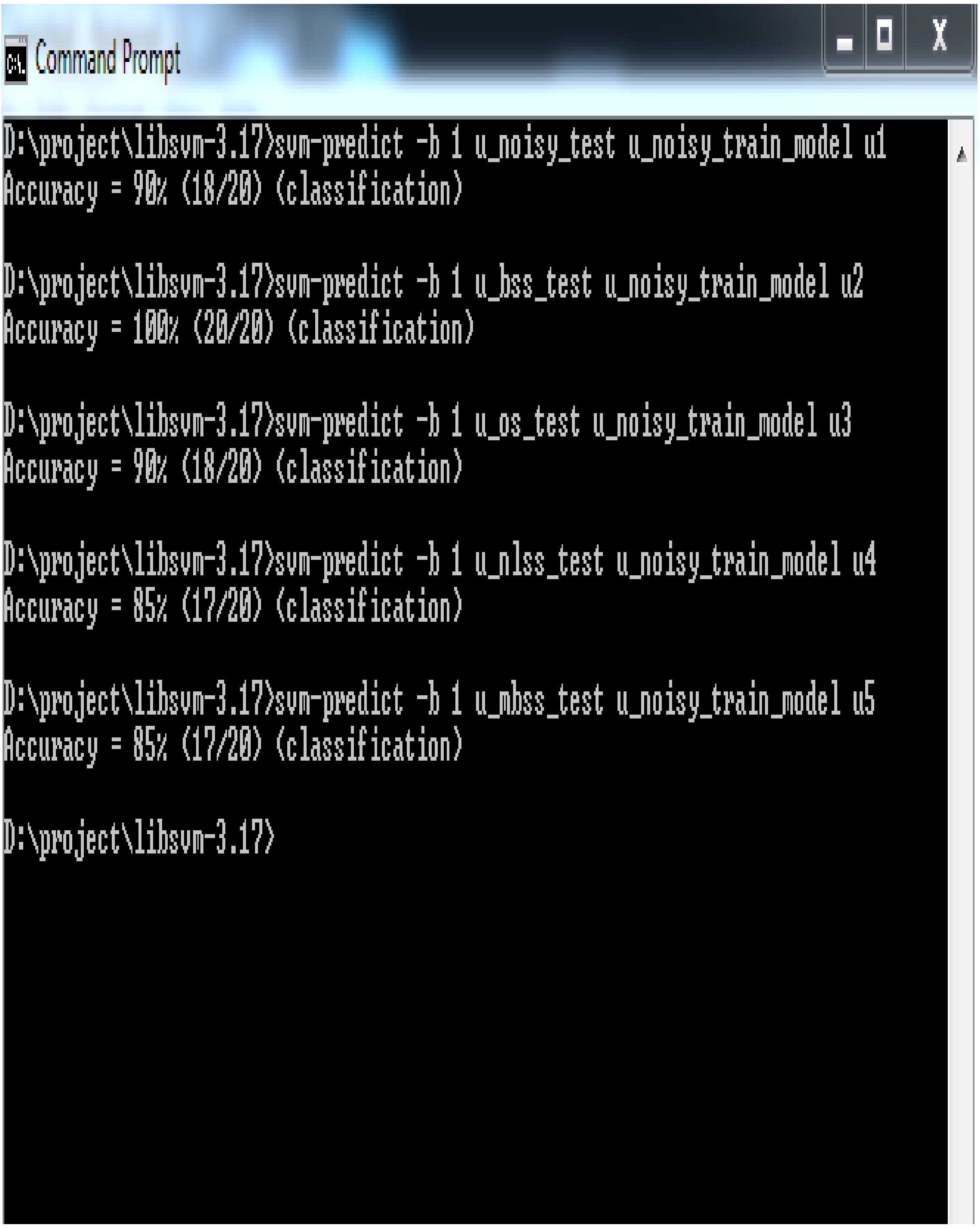

FIGURE 4.8: TESTING

Testing for impostor

That is testing for false speaker in actual speaker training model.

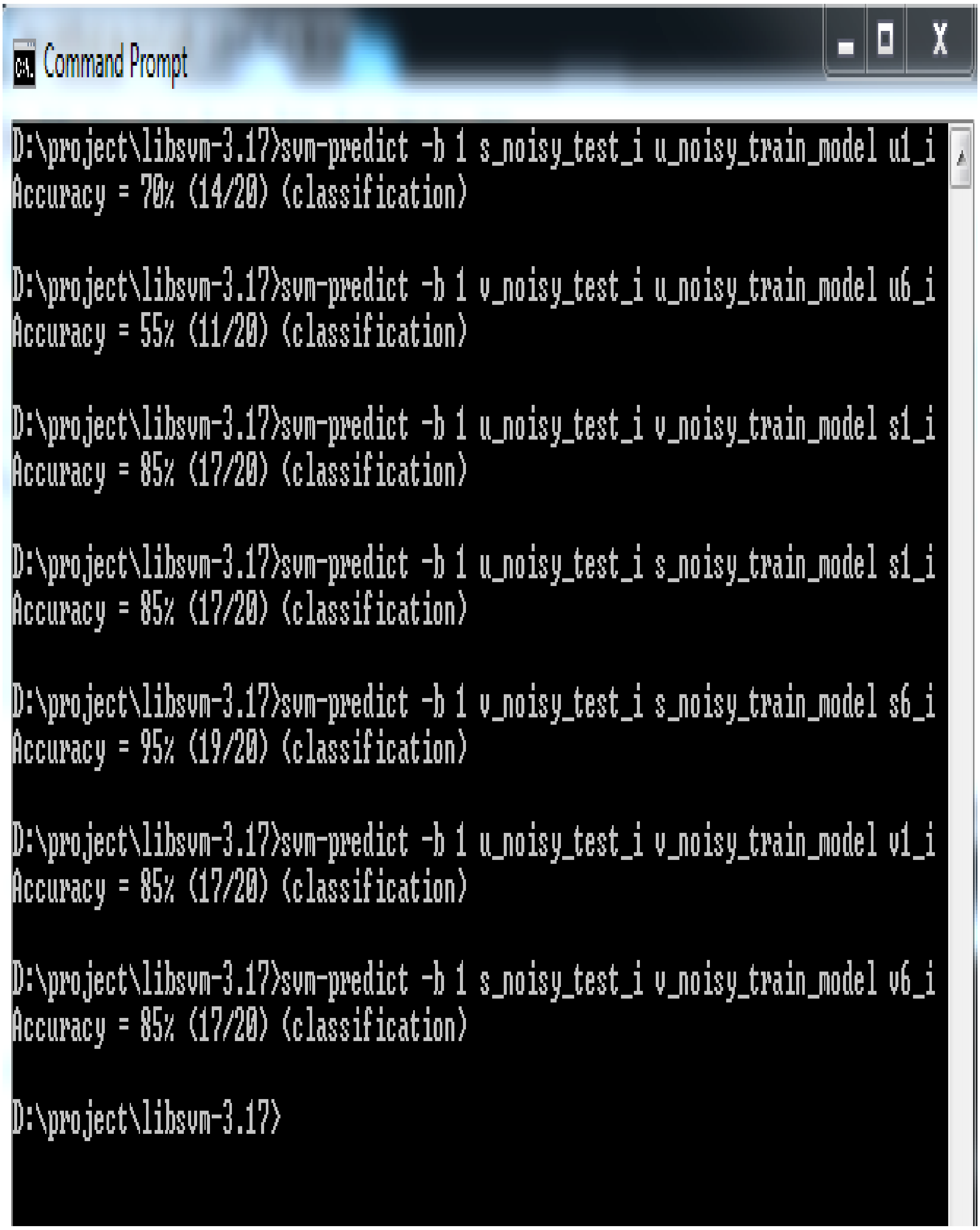

FIGURE 4.9: TESTING IMPOSTOR

Training and testing in noise reduced samples:

Training model is created in noise reduced speech sample and testing is also done in noise reduced speech sample.

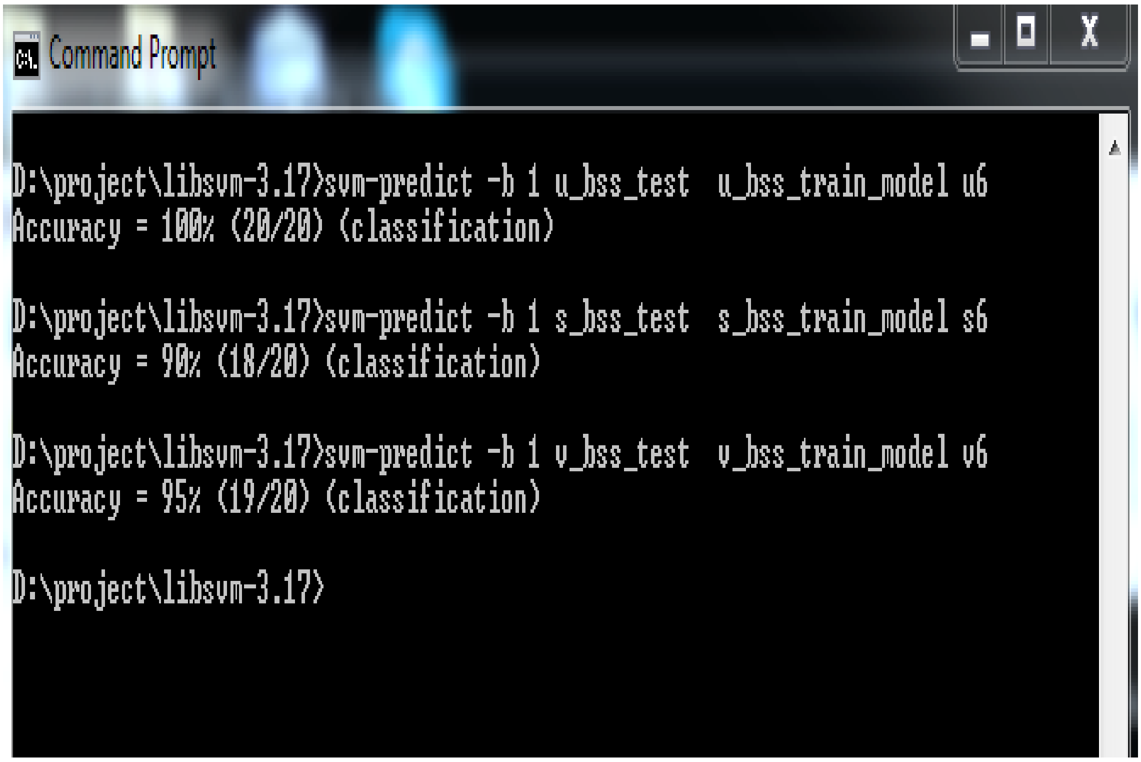

FIGURE 4.10: TRAINING AND TESTING

Testing for impostor:

That is testing for false speaker in actual speaker training model.

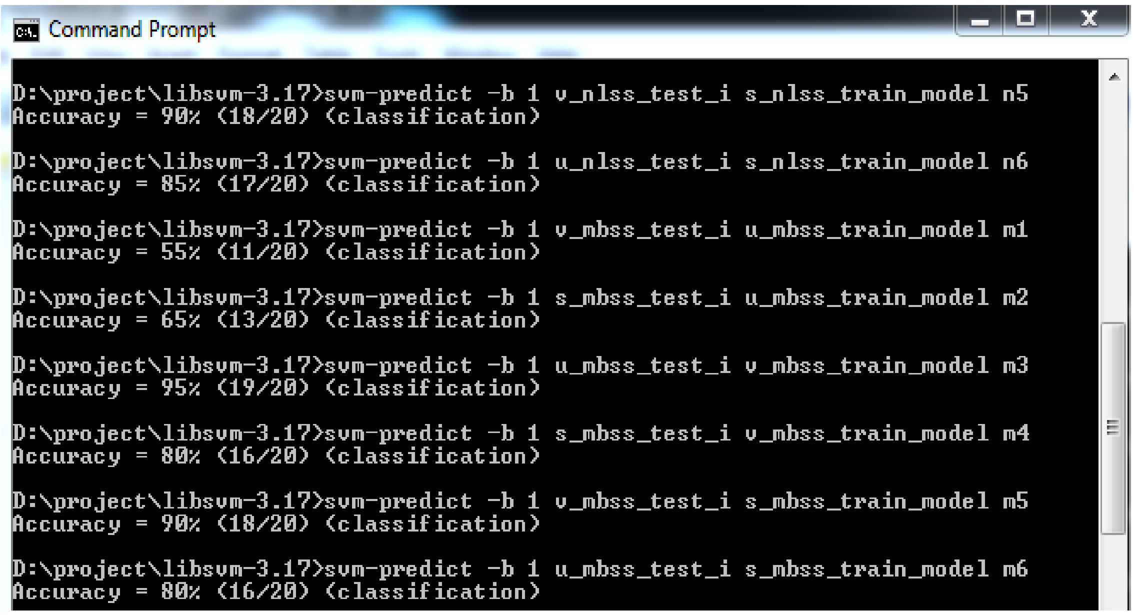

FIGURE 4.11: IMPOSTOR

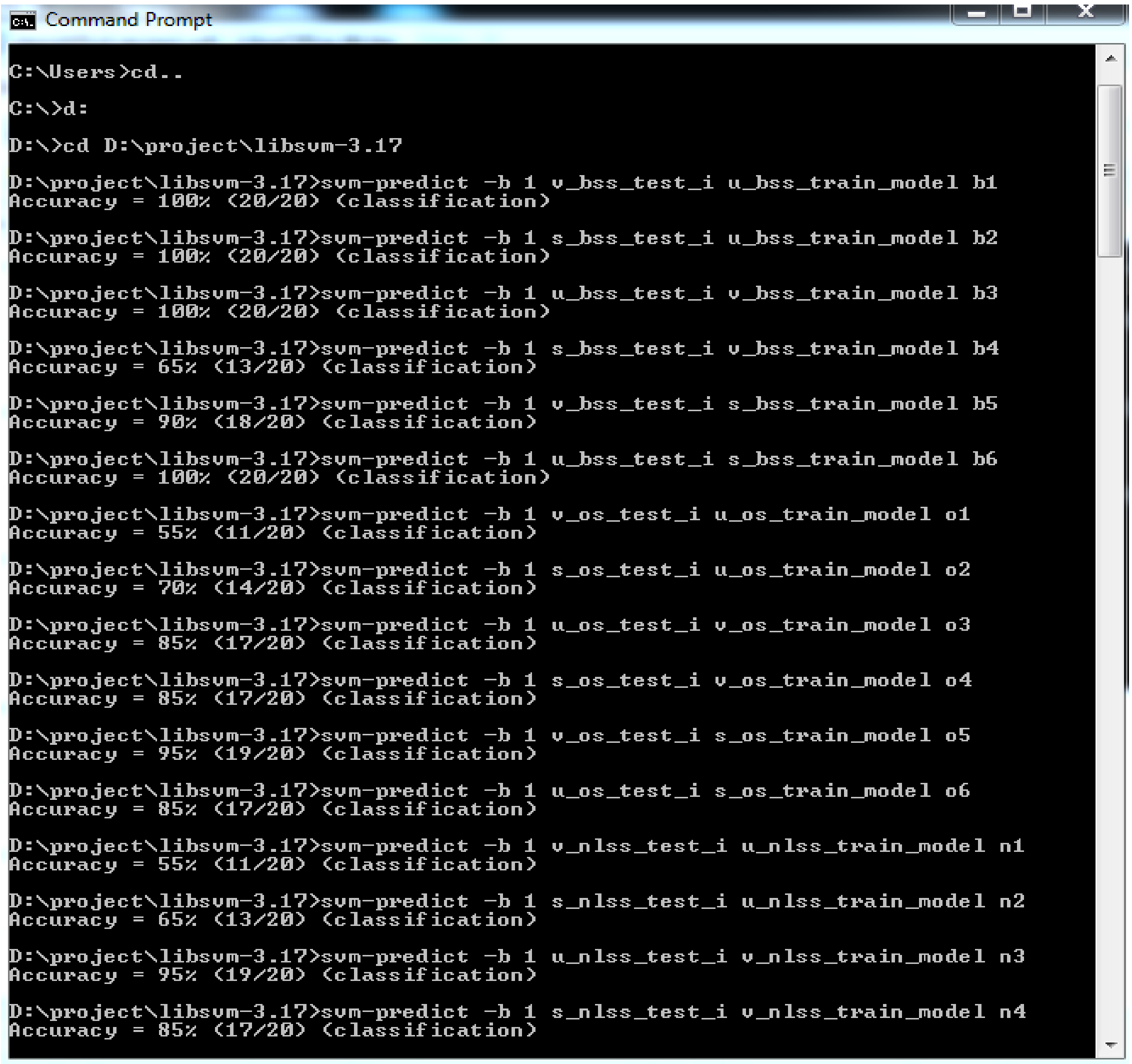

FIGURE 4.12: IMPOSTOR

EQUAL ERROR RATE:

False acceptance rate is the impostor speaker sample getting accepted and false rejection rate is actual speaker sample getting rejected.A biometric security system predetermines the threshold values for its false acceptance rate and its false rejection rate and when the rates are equal the common value is referred to as the equal error rate. The value indicates that the proportion of false acceptances is equal to the proportion of false rejections. The lower the equal error rate value the higher the accuracy of the biometric system. Equal error rate is found by using the matlab interface software DETware_v2.1 which gives graph when the true_scores and impostor_scores of all the speakers in the system are given as input from the graph equal error rate value is calculated.

Equal Error Rate Graphs:

1. Noisy training model and noisy test file-(EER=17.08)
2. For noisy training model and noise reduced test file by basic spectral subtraction-(EER=11.04)
3. For noisy training model and noise reduced test file by over subtraction-(EER=15.6)
4. For noisy training model and noise reduced test file by non-linear spectral subtraction-(EER=15.6)
5. For noisy training model and noise reduced test file by multiband spectral subtraction-(EER=15.5)
6. For noise reduced training and testing by basic spectral subtraction-(EER=5)
7. For noise reduced training and testing by over subtraction-(ERR=17.1)
8. For noise reduced training and testing by non-linear spectral subtraction-(EER=15.4)
9. For noise reduced training and testing by multiband spectral subtraction-(EER=15.4)

SPEAKER VERIFICATION SYSTEM FOR NOISY TRAINING MODEL AND REDUCED NOISE TEST FILE:

NOISY MODEL	EQUAL ERROR RATE
Noisy	17.08
BSS	11.04
OS	15.62
NLSS	15.62
MBSS	15.52

TABLE 4.2: REDUCED NOISE VERIFICATION

NOISY MODEL	EQUAL ERROR RATE
BSS	5.0
OS	17.1
NLSS	15.4
MBSS	15.4

TABLE 4.3: REDUCED NOISE TRAINING

SPEAKER VERIFICATION SYSTEM FOR REDUCED NOISE TRAINING MODEL AND TEST FILE:(4.3)

Conclusion:

For a environment with white noise BASIC SPECTRAL SUBTRACTION(BSS) has lowest EER thus BSS has the higher accuracy in environment with white noise.

Noise Addition:

Babble Noise(group of speakers talking at the background) is added to the speech samples which already contains white noise.Noise is added at the sampling frequency of the the speech.

Spectrum of the noise added sample:

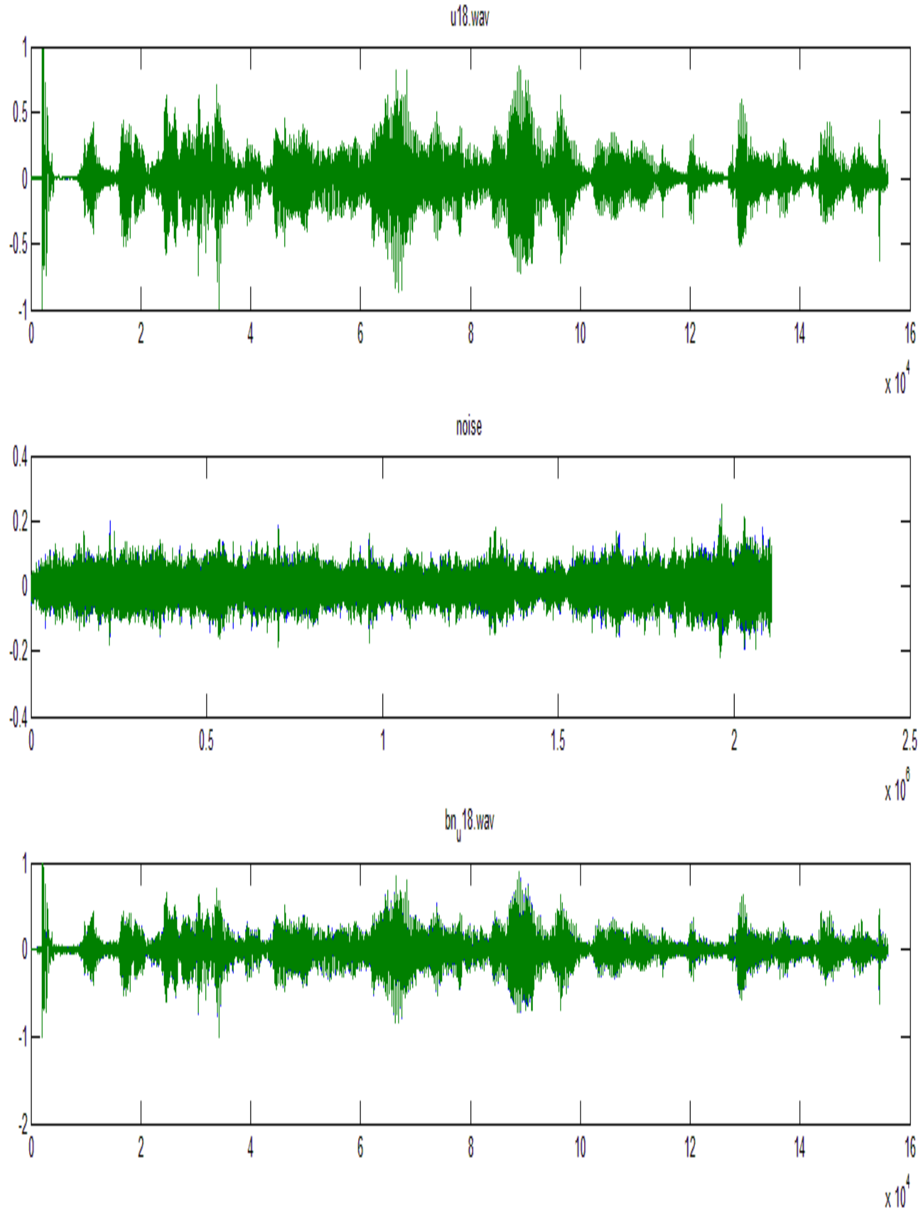

FIGURE 4.13: NOISE ADDITION

Spectrum of speech enhanced by basic spectral subtraction for one sample:

bn_u18.wav is the noisy speech sample of one speaker and basicss_bn _u18.wav is enhanced speech obtained by basic spectral subtraction

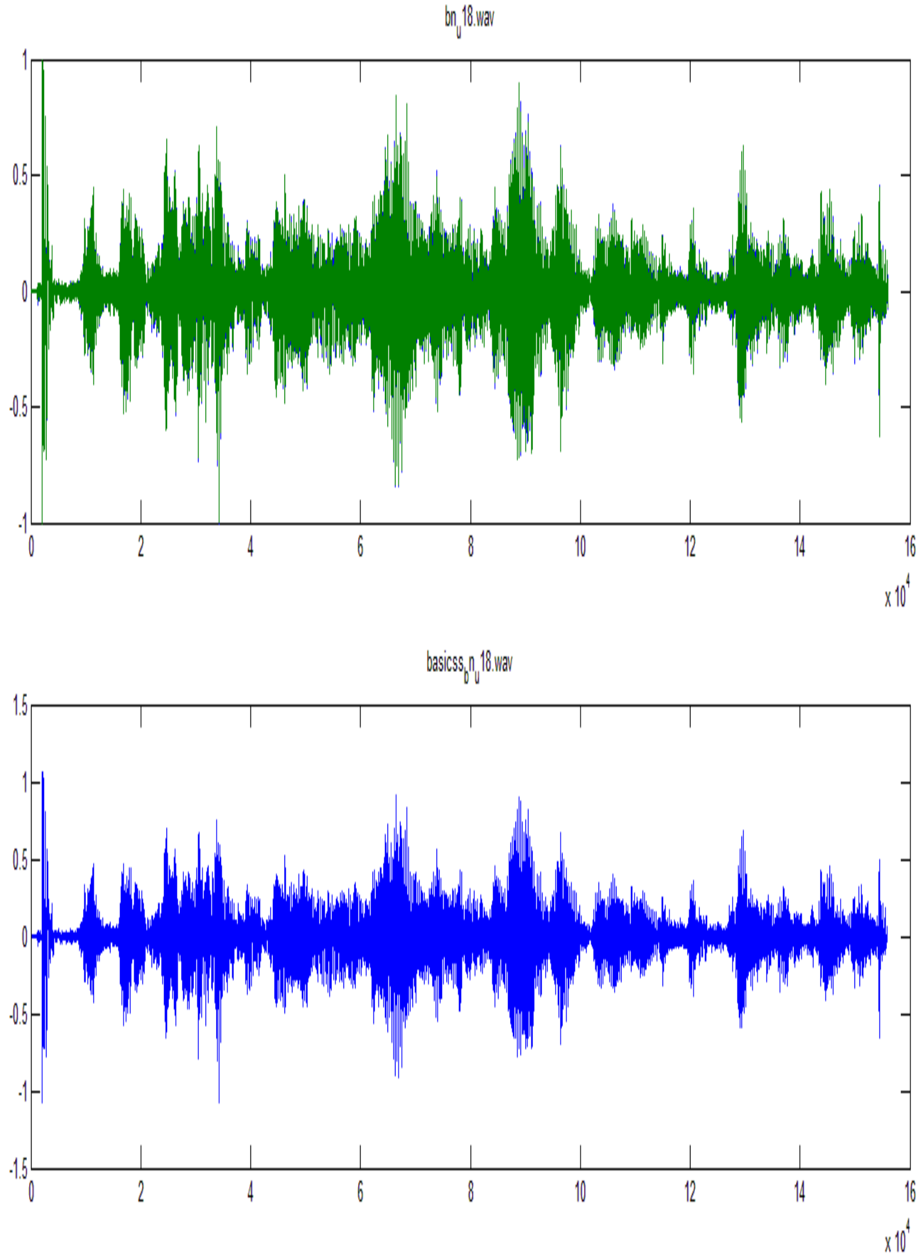

FIGURE 4.14: ENHANCED SPEECH SPECTRUM USING BSS

Spectrum of speech enhanced by over subtraction for one sample:

bn_u18.wav is the noisy speech sample of one speaker and oversub_bn _u18.wav is enhanced speech obtained by over subtraction.

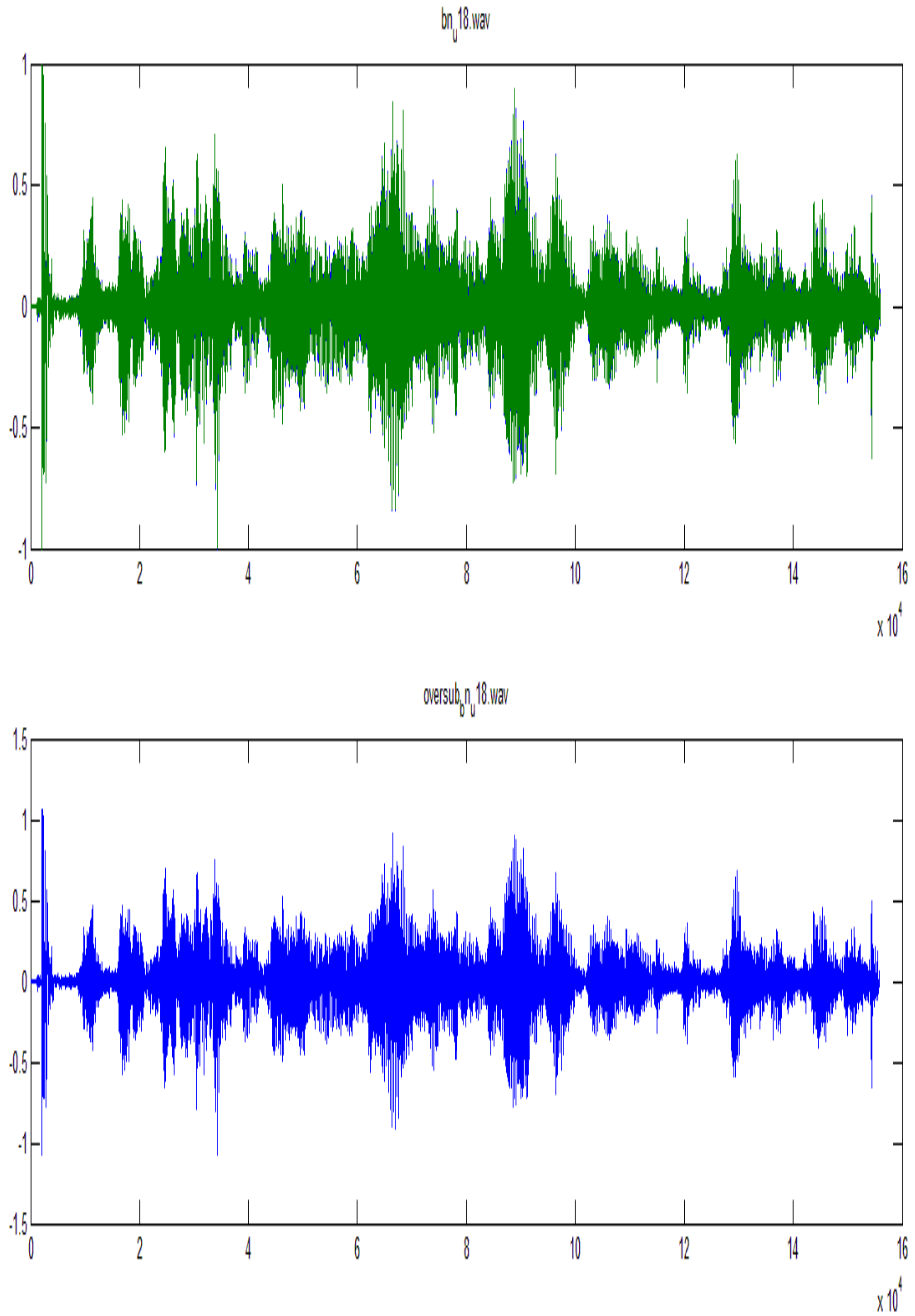

FIGURE 4.15: ENHANCED SPEECH SPECTRUM USING OS

Spectrum of speech enhanced by non-linear spectral subtraction for one sample: bn_u18.wav is the noisy speech sample of one speaker and nlss_bn_u18.wav is enhanced speech obtained by non linear spectral subtraction.

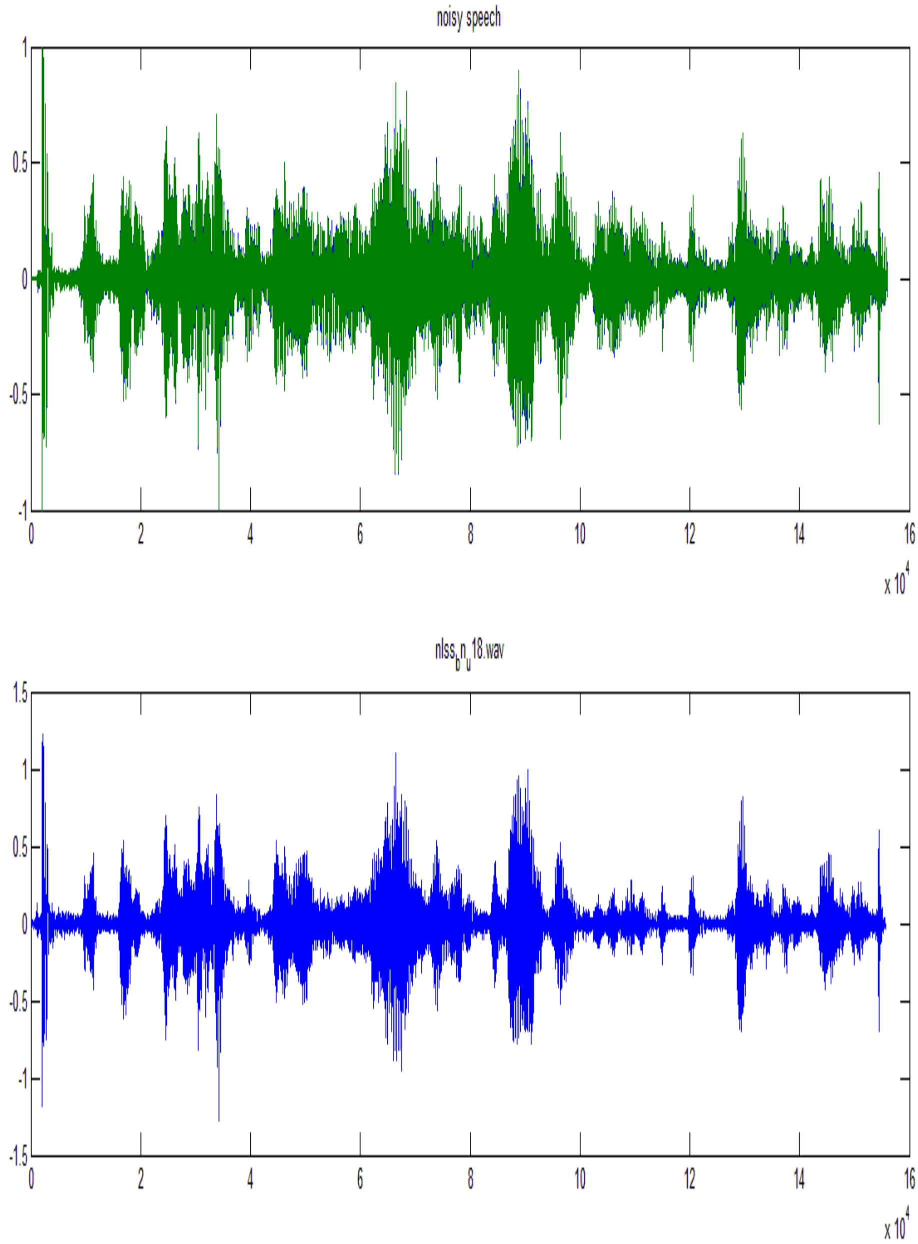

FIGURE 4.16: ENHANCED SPEECH SPECTRUM USING NLSS

Spectrum of speech enhanced by multiband spectral subtraction for one sample: bn_u18.wav is the noisy speech sample of one speaker and mbss_bn_u18.wav is enhanced speech obtained by multiband spectral subtraction.

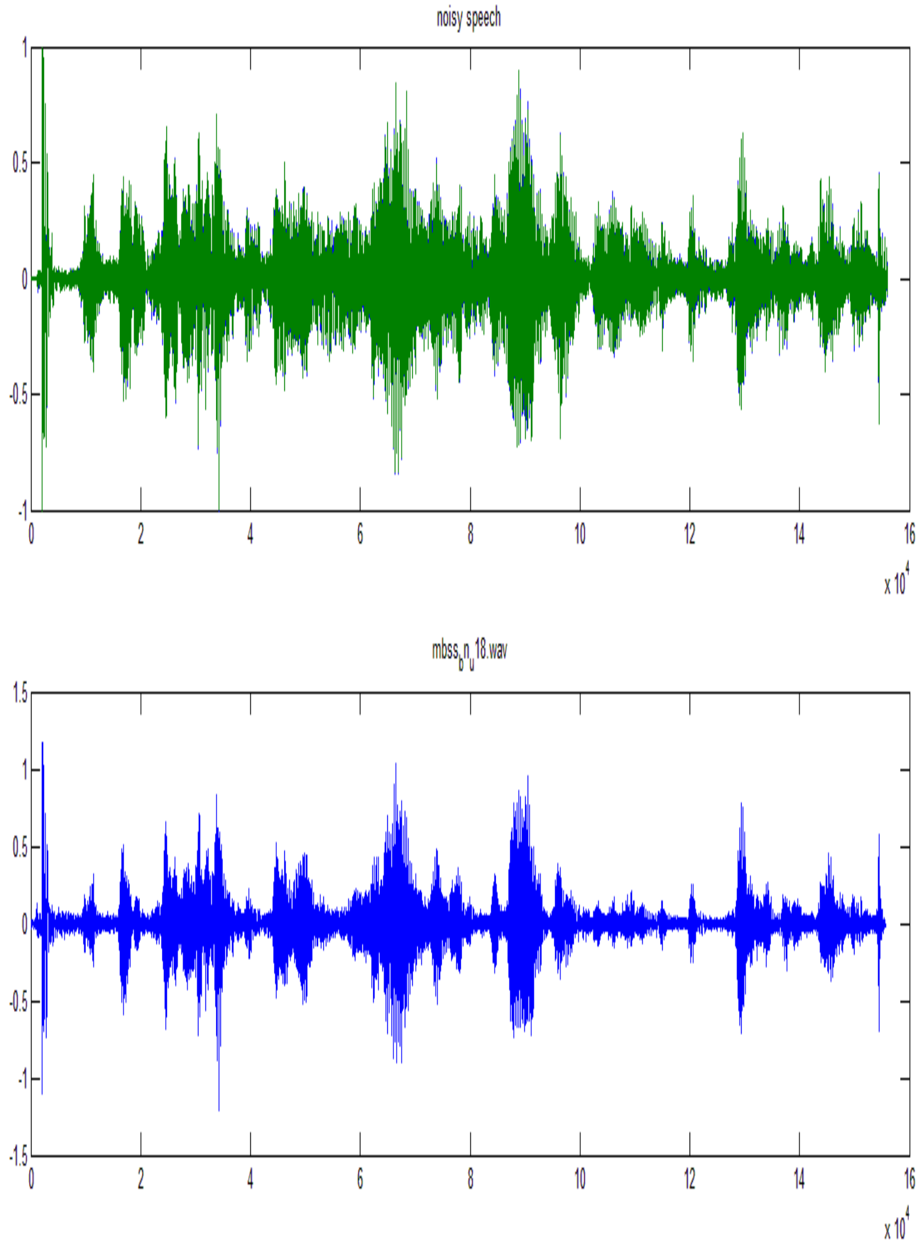

FIGURE 4.17: ENHANCED SPEECH SPECTRUM USING MBSS

Average of SNR values of speech samples of all the speakers in the system for noisy an noise enhanced speech.

After enhancing the speech by various noise elimination methods the speaker verification process is done for two speakers and the performance of the system is

Speech	SNR
Noisy	-1.0937567
BSS	-1.0968016
OS	-1.0968015
NLSS	-1.0915377
MBSS	-1.0932961

TABLE 4.4: SNR VALUES

analyzed using EER(equal error rate).

Equal Error Rate Graphs in babble noise environment:

1. **For Noisy training model and noisy test file-(EER=2.71)**

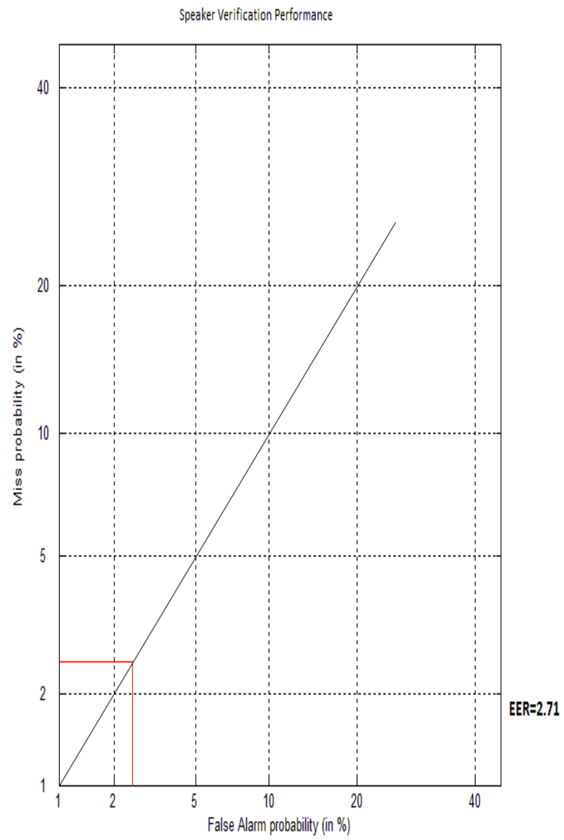

FIGURE 4.18: EER GRAPH FOR NOISY FILE

2. **For noisy training model and noise reduced test file by basic spectral subtraction-(EER=2.71)**

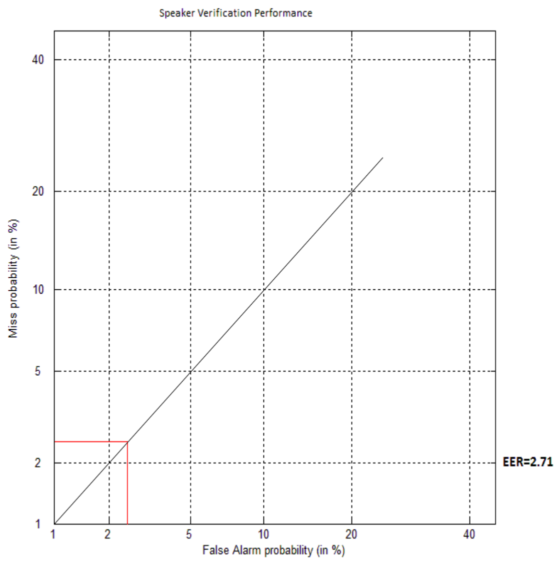

FIGURE 4.19: EER GRAPH FOR BSS

3. **For noisy training model and noise reduced test file by over subtraction-(EER=2.71)**

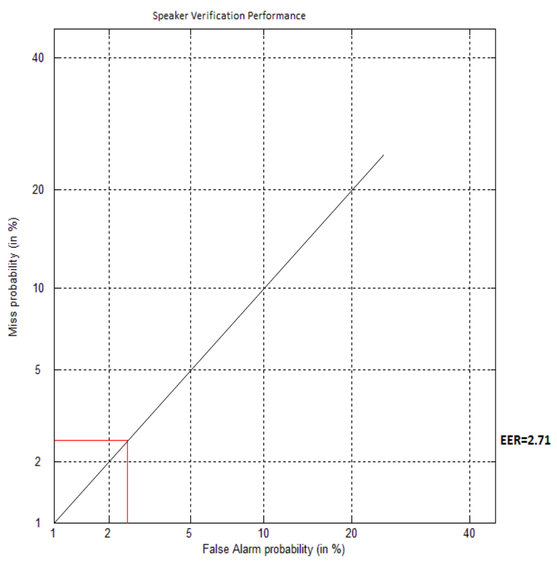

FIGURE 4.20: EER GRAPH FOR OS

4. **For noisy training model and noise reduced test file by non-linear spectral subtraction-(EER=2.64)**

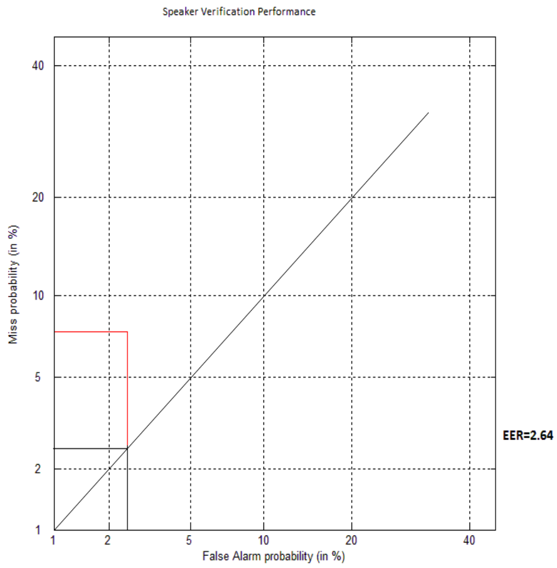

FIGURE 4.21: EER GRAPH FOR NLSS

5. **For noisy training model and noise reduced test file by multiband spectral subtraction-(ERR=2.7)**

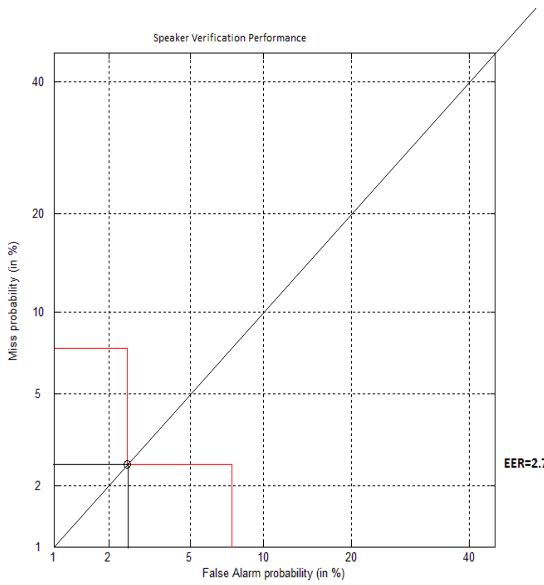

FIGURE 4.22: EER GRAPH FOR MBSS

6. **For noise reduced training and testing by basic spectral subtraction-(ERR=2.71)**

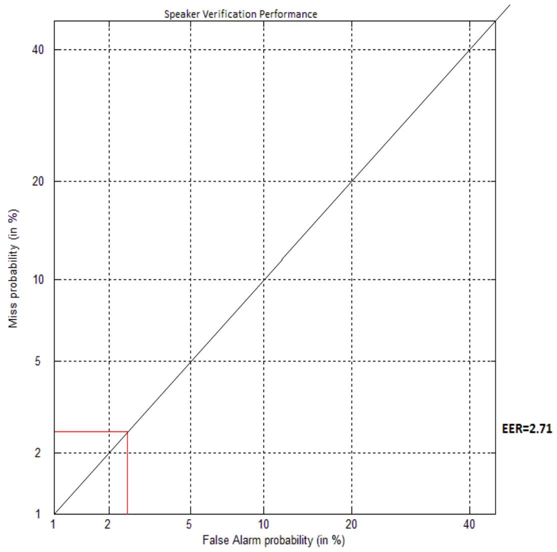

FIGURE 4.23: EER GRAPH FOR NOISE REDUCED FILE BY BSS

7. **For noise reduced training and testing by over subtraction-(ERR=2.71)**

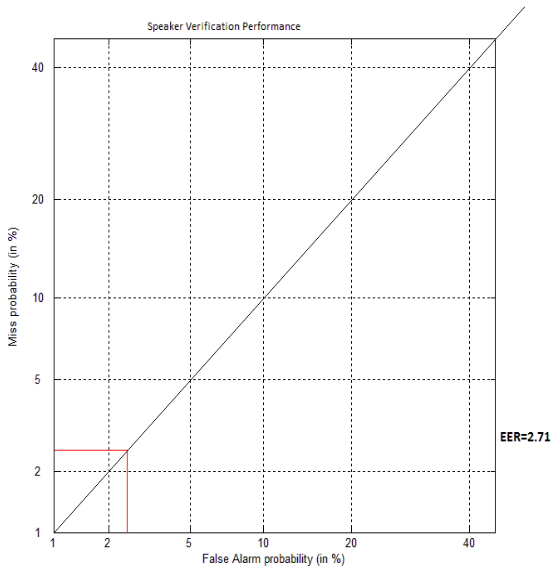

FIGURE 4.24: EER GRAPH FOR NOISE REDUCED FILE BY OS

8. **For noise reduced training and testing by non-linear spectral subtraction-(ERR=2.64)**

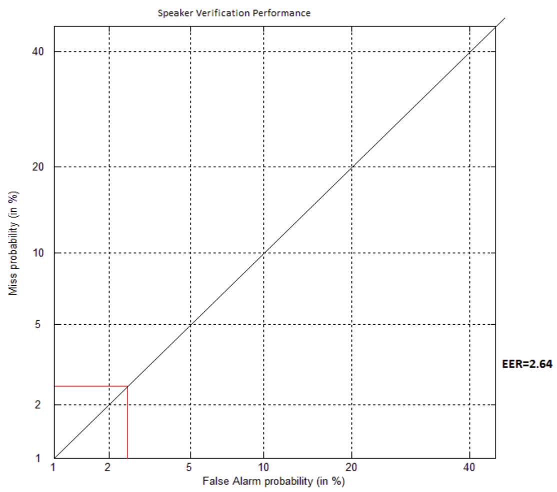

FIGURE 4.25: EER GRAPH FOR NOISE REDUCED FILE BY NLSS

9. **For noise reduced training and testing by multi-band spectral subtraction-(ERR=2.71)**

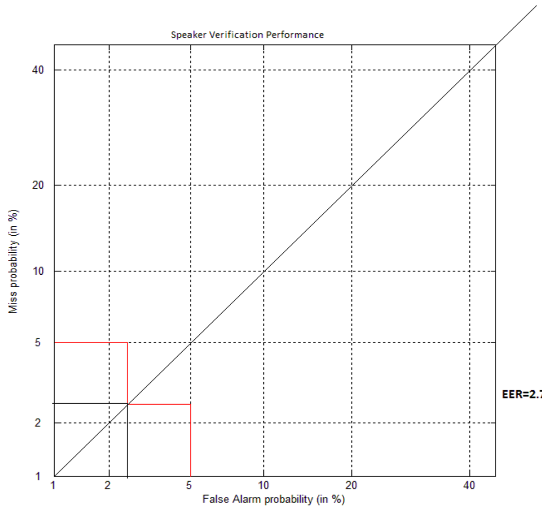

FIGURE 4.26: EER GRAPH FOR NOISE REDUCED FILE BY MBSS

SPEAKER VERIFICATION SYSTEM FOR NOISY TRAINING MODEL AND REDUCED NOISE TEST FILE:

NOISY MODEL	**EQUAL ERROR RATE**
NOISY	2.71
BSS	2.71
OS	2.71
NLSS	2.64
MBSS	2.70

TABLE 4.5: EER COMPARISON

NOISY MODEL	**EQUAL ERROR RATE**
BSS	2.71
OS	2.71
NLSS	2.64
MBSS	2.71

TABLE 4.6: EER VALUES

SPEAKER VERIFICATION SYSTEM FOR REDUCED NOISE TRAINING MODEL AND TEST FILE:(4.6)

Conclusion:

For a environment with white noise and babble noise NON LINEAR SPECTRAL SUBTRACTION(NLSS) has lowest EER thus NLSS has the higher accuracy in environment with white noise and babble noise.

CHAPTER 5

References

[1] Anuradha R. Fukane, Shashikant L. Sahare, Different Approaches of Spectral Subtraction method for Enhancing the Speech Signal in Noisy Environments, International Journal of Scientific and Engineering Research, Volume 2, Issue 5, May-2011.ISSN 2229-5518.

[2] Andrzej Drygajlo and Mounir El-Muliki , Speaker Verification in Noisy Environments with Combined Spectral Subtraction and Missing Feature Theory,IEEE Trans. Acoustics,Speech and Signal Processing.,vol. 1,pp. 121-124, May 1998.

[3] STEVEN F.BOLL, Suppression of Acoustic Noise in Speech Using Spectral Subtraction ,IEEE Trans.Acoustics,Speech and Signal Processing.,vol. 27,no. 2,pp 113-120,Apr 1979.

[4] K. Lebart, J. M. Boucher,A New Method Based on Spectral Subtraction for Speech Dereverberation,acta acustica . ACUSTICA Vol. 87,pp. 359366, 2001.

[5] Serajul Haque, Roberto Togneri, A Psychoacoustic Spectral Subtraction Method For Noise Suppression In Automatic Speech Recognition, IEEE International Conference on Acoustics Speech And Signal Processing. ,pp 1618-1621, Mar 2010.

[6] Ji Ming, Timothy J. Hazen, James R. Glass, Douglas A. Reynolds, Robust Speaker Recognition in Noisy Conditions, IEEE Trans. Audio, Speech and Language Processing.,vol. 15, no. 5, pp 1711-1723,Jul 2007.

[7] Ramin Halavati, Saeed Bagheri Shouraki,Mina Razaghpour, Hossein Tajik, Arpineh Cholakian,A Novel Noise Immune ,Fuzzy Approach To Speaker Independent,Isolated Word Speech Recognition published in World Automation Congress(WAC),Conference in Budapest, Jul 2006.

[8] W. Alkhaldi, W. Fakhr and N. Hamdy , Automatic Speech/Speaker Recognition In Noisy Environments Using Wavelet Transform , IEEE Conference on Circuits and Systems. ,vol. 1, pp 463-466,Aug. 2002.

[9] Marwa A. Abd El-Fattah Moawad I. Dessouky Alaa M. Abbas Salaheldin M. Diab El-Sayed M. El-Rabaie Waleed Al-Nuaimy Saleh A. Alshebeili Fathi E. Abd El-samie, Speech enhancement with an adaptive Wiener filter, Journal from Faculty of Electronic Engineering, Menoufia University,Egypt, Aug 2013

[10] Sahare2 Anuradha R. Fukane1, Shashikant L. Sahare2, Noise estimation Algorithms for Speech Enhancement in highlynon-stationary Environments , International Journal of Computer Science Issues, vol. 8, Issue 2, March 2011.ISSN 1694-0814.

[11] E. Verteletskaya, K. Sakhnov, Voice Activity Detection for Speech Enhancement Applications, Acta Polytechnica Vol. 50 No. 4/2010.

[12]Poonam Sharma and Abha Kiran Rajpoot,Automatic Identification of Silence,Unvoiced and Voiced Chunks in Speech,Computer Science Engineering Department Sharda University, Greater Noida.

[13]Bachu R.G., Kopparthi S., Adapa B., Barkana B.D. , Seperation of Voiced and Unvoiced using Zero Crossing rate and Energy of the Speech signal, Electrical Engineering Department School of Engineering, University of Bridgeport.

[14] Saeed V. Vaseghi, Advanced Digital Signal Processing and Noise Reduction,Second Edition. , ISBNs: 0-471-62692-9.

[15] Hanspeter Schmid,How to use the FFT and Matlabs pwelch function for signal and noise simulations and measurements, University of applied sciences NorthWestern Swizerland School of Engineering,aug 2012.

[16] Ricardo A. Losada,Digital Filters with MATLAB, The MathWorks,inc, may 2008.

[17] F. Capman, J. Boudy, P. Lockwood,Acoustic Echo Cancellation and Noise Reduction in the Frequency-Domain: A Global Optimisation, Matra Communication, Speech Processing Department.

[18]Mohammad Sadgh Talebi, Frequency Domain Signal Processing Using MATLAB, Sharif University of Technology.

[19] Rafael C. Gonzalez Richard E. Woods, AN INTRODUCTION TO MATRIX MANIPULATION IN MATLAB, Digital Image Processing, Second Edition.

[20]Bc. Jan Kybic, Kalman Filtering and Speech Enhancement, Diploma work jan 1998.

[21] H.T. Hu and C. Yu, Adaptive noise spectral estimation for spectralsubtraction speech enhancement sep2007.

References