PROJECT REPORT-EC306

Report on: Speech Enhancement using Adaptive Wiener filtering technique

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY

Submitted to:

Prof. Vijaya C

Dept. of E&C Engineering

Visiting faculty, IIIT Dharwad

Submitted by:

Kummara Bhargavi (18BEC024)

Kukkala Bharath (18BEC023)

Contents

I. Theoretical

II. Experimental

Acknowledgement

We would like express our sincere thanks to Dr. Vijaya C ma'am and Supritha Shetty ma'am for giving us an opportunity to do this interesting project on **Speech Enhancement**, which helped us to explore about new things amidst this pandemic.

Also, we would like thank them for their valuable suggestions, comments and guidance throughout the course of the project. Although, this report has been prepared with utmost care and interest, we take our responsibility and accept if there is any small imperfection.

Introduction

Speech plays a vital role in communication in our daily life. Speech signals are the most widely used signals between humans to convey messages, because they convey information with the emotion of a human voice. Hence, many researchers gave a large attention to speech processing.

Certain properties of the speech signal are:

- Speech has a bandwidth of only 4 KHz
- It is one-dimensional signal, with time as independent variable
- Frequency spectrum of speech signal is not constant in time
- Although human beings have an audible frequency range of 20 Hertz to 20 KHertz, the human speech has significant frequency components only up to 4 KHz.

The most common problem in speech processing is the effect of interference noise in the signals. This noise masks the signal and reduces its quality (intelligibility). Speech communication is greatly affected by the presence of background noise, which is always present in any location. Therefore, speech signals get distorted mostly by **additive noise** and such signals which are degraded are referred as noisy speech signals.

We cannot restore the original noiseless signal completely, so there is a need to enhance these degraded /noisy speech signal to improve the quality. With the aim of reducing listener fatigue, many speech enhancement techniques and algorithms are proposed and introduced.

Literature Survey

A brief survey on techniques for enhancing speech is crystal clearly explained in the report by Tayseer M.F. Taha and Amir Hussain published in the International Journal of Computer Applications. These both authors have taken us through all the available techniques for enhancing speech like the Conventional methods, Adaptive filtering methods, Machine Learning models and multi-modal methods. In other words, they gave us an understanding about the types of audio enhancement categories along with the explanation of various kinds of noises that can be added to the speech or audio.

Another paper we have gone through is Speech Enhancement by Jae S. Lim published in *ICASSP*. This paper is more focused on enhancing speech degraded by echoes. Also this paper explained about the spectral subtraction and introduced adaptive noise cancelling algorithms.

From the above two cited reports, we are thoroughly introduced to the basics of speech enhancement techniques like Spectral subtraction method, Traditional Wiener Filtering technique, Adaptive noise cancellation techniques with different algorithms and MMSE methods. The authors have mainly focused on speech signal corrupted by additive noise and on signal processing techniques rather than machine learning and related stuff.

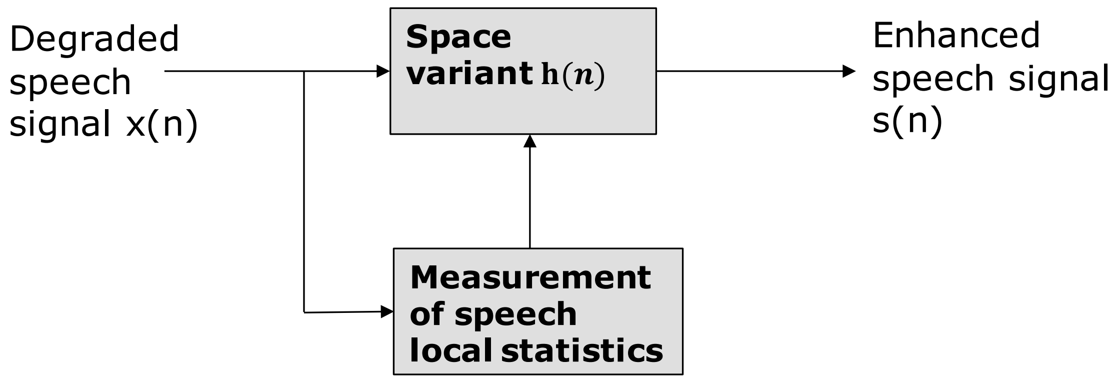
We have also gone through the papers that have explained about MMSE and MAP estimation, but found it challenging to understand many technical details.

Then we came across a paper published by *IEEE* on Adaptive Wiener Filter for Speech Enhancement by Aishwarya Yelwande, Ansha Dixit and Saritha Kansal. This gave an overview about the adaptive filtering using NLMS (Normalized Least Mean Square) and LMS (Least Mean Square) algorithms as adaptive noise cancellation algorithms. Also they have quoted that the implementation of adaptive wiener filter is done in time domain rather than in frequency domain for accommodation of varying nature of speech signal. This paper has enhanced speech using time domain while the other papers implementation is done in frequency domain.

The final paper which we found interesting and worth giving a try is *Speech Enhancement with an Adaptive Wiener Filter* by 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 published by *Springer*. This paper describes adaptive wiener filter technique which is based on the adaption of filter transfer function from sample to sample and enhancing the signal by using simple statistical measures like mean and variance. As quoted before, this method is implemented in *time domain*.

Overview of Adaptive Winer Filtering method

Implementation of this technique is done in **time domain** rather than frequency domain, for accommodation of varying nature of speech signal. Adaptive implementation of the wiener filter benefits from the local statistics (mean & variance) of the speech signal.



Assuming that the additive noise $v(n)$ is of zero mean and has white nature with variance of σ_v^2. Then the power spectral density is approximated by:

$$P_v(\omega) = \sigma_v^2$$

Consider a small segment of the speech signal, in which the signal $x(n)$ is assumed to be stationary. The signal $x(n)$ can be modelled by:

$$x(n) = m_x + \sigma_x w(n)$$

where m_x and σ_x are local mean and standard deviation of $x(n)$ and $w(n)$ is a unit variance noise

Within this **small segment of speech**, the Wiener filter transfer function can be approximated by:

$$H(\omega) = \frac{P_s(\omega)}{P_s(\omega) + P_v(\omega)} = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_v^2}$$

Because $H(\omega)$ is constant over this small segment of speech, the impulse response of the Wiener filter can be obtained by:

$$h(n) = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_v^2} \delta(n)$$

The enhanced speech signal $\hat{s}(n)$ in this local segment can be expressed as:

$$\hat{s}(n) = m_x + (x(n) - m_x) * \mathrm{h(n)}$$

If m_x and σ_s are updated at each sample, we can say:

$$\hat{s}(n) = m_x(n) + \frac{\sigma_s^2(n)}{\sigma_s^2(n) + \sigma_v^2} (x(n) - m_x(n))$$

the local mean $m_x(n)$ and $(x(n) - m_x(n))$ are modified separately from segment to segment and then the results are combined.

If σ_s^2 is much larger than σ_v^2 the output signal $\hat{s}(n)$ will be primarily due to $x(n)$ and the input signal $x(n)$ is not attenuated. If σ_s^2 is smaller than σ_v^2, the filtering effect is performed.

Notice that m_x is identical to m_s when m_v is zero. So, we can estimate $m_x(n)$ as:

$$\widehat{m}_s(n) = \widehat{m}_x(n) = \frac{1}{2M+1}\sum_{k=n-M}^{n+M} x(k)$$

where (2M+1) is the no: of samples in the short segment used in the estimation.

To measure the local statistics of the speech signal, we need to estimate the signal variance σ_s^2. Since $\sigma_x^2 = \sigma_s^2 + \sigma_v^2$, then

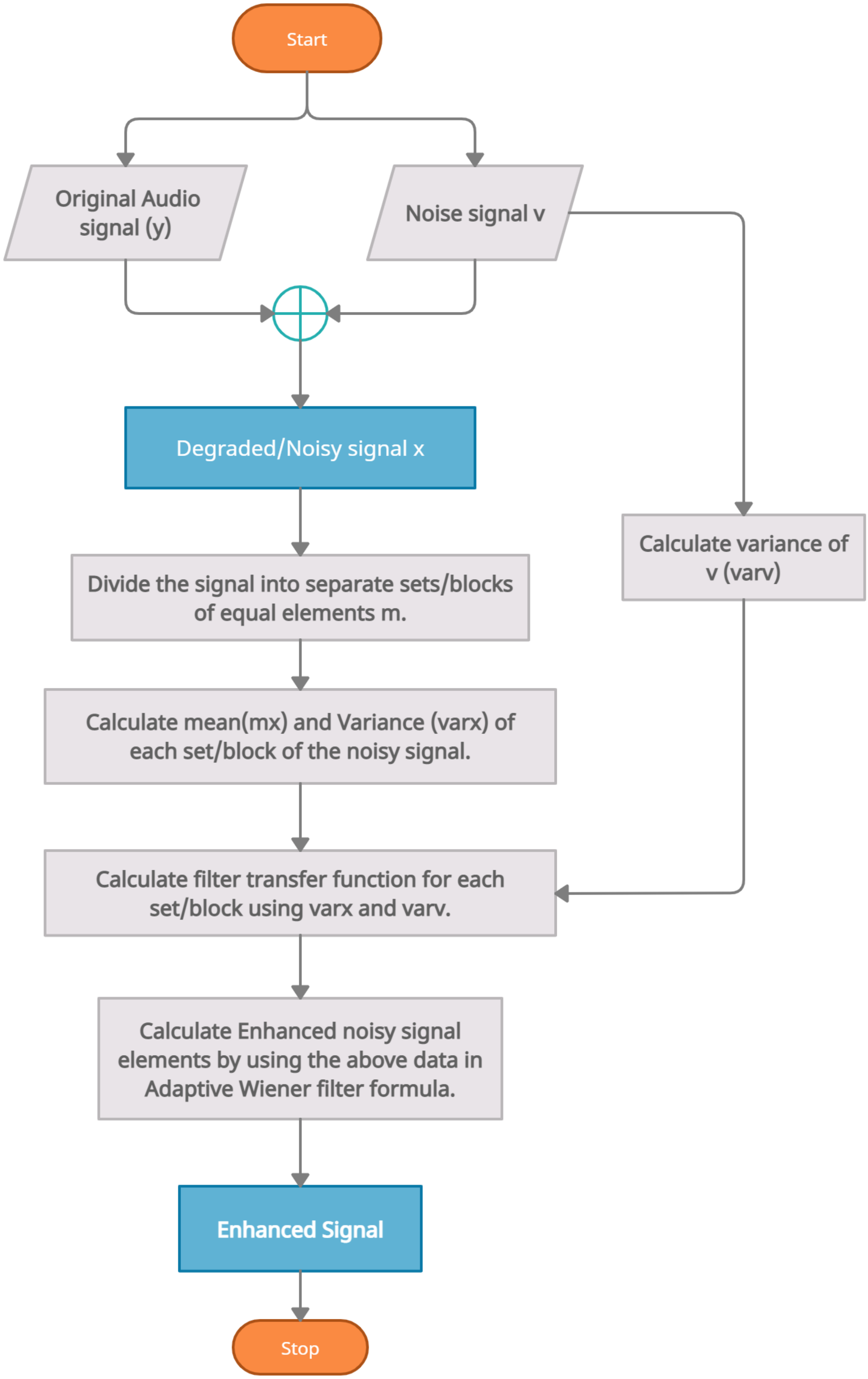
$\sigma_s^2(n)$ may be estimated from x(n) as follows:

$$\hat{\sigma}_s^2(n) = \begin{cases} \hat{\sigma}_x^2(n) - \hat{\sigma}_v^2, & \text{if } \hat{\sigma}_x^2(n) > \hat{\sigma}_v^2 \\ 0, & \text{otherwise} \end{cases}$$

where

$$\hat{\sigma}_x^2(n) = \frac{1}{(2M+1)}\sum_{k=n-M}^{n+M} (x(k) - \hat{m}_x(n))^2$$

Flowchart of Implementation:

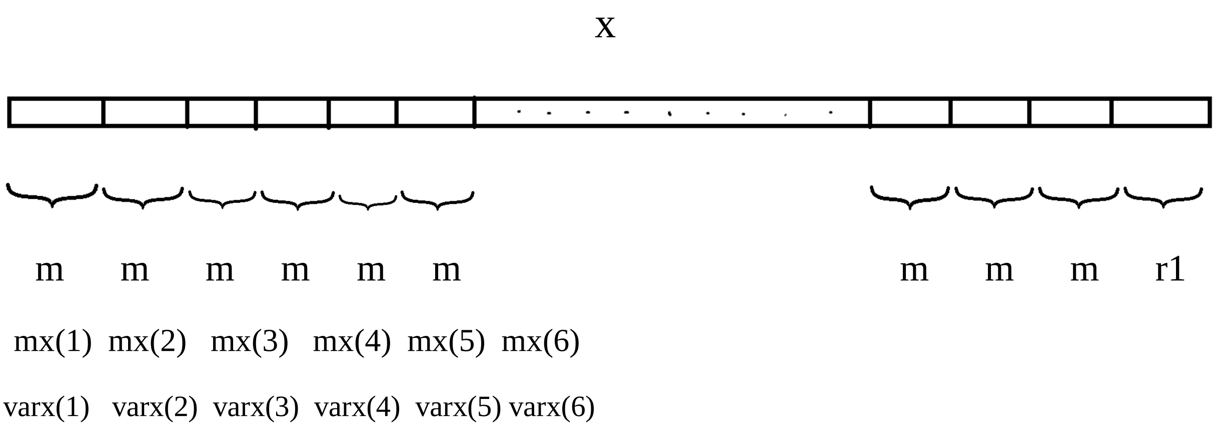


Details of algorithm and implementation

1)Read original clean sound into ‘r’.

2)Add a noise (White Gaussian noise ‘v’) of the same length as ‘r’, to ‘r’ for noisy signal x. Calculate variance of noise(v) ‘varv’.

3)Select the value of ‘m’ so that the noisy signal is divided into sets/blocks of m elements each as shown below…



the extra r1 elements are neglected. (Here x divided into sets/blocks using reshape).

4)Mean and variance for each set/block are calculated and stored in arrays ‘mx’ and ‘varx’ respectively. Each element in arrays ‘mx’ and ‘varx’ are repeated m times to make further calculations easy (repeated elements are stored in arrays Mx and Varx respectively).

5)Calculate filter transfer function.

- For i=1 to length(mx)
    - if (varx(i) > varv) then, hn(i) = (varx(i) - varv)/(varx(i));
    - else hn(i)=0;
    - i=i+1;

repeat each element in hn ‘m’ times and store it in Hn.

6)Calculate enhanced signal ‘s’. Every element in $s(i)$ is equal to

$$s(j) = Mx(j) + Hn(j) * ((x(j) - Mx(j))$$

- For j=1 to length(x-r1)
    - S(j)= Mx(j)+Hn(j).*(x(j)-Mx(j));
    - i=i+1;

7)Calculate SNR and plot the graph of the enhanced signal and compare it to the original clean signal.

Regarding Implementation:

We used MATLAB to implement this adaptive wiener filtering technique. Since Matlab is one of the best used in signal analysis and has adequate signal processing tools. Also it has the best tools to analyse the plots and graphs.

Also we are implementing for the additive noise case. There are many different kinds of noise like white noise, pink noise and other colour noises. We are considering **white gaussian noise** for our experimentation. The term *white* in white gaussian noise refers to the idea that it has uniform power density across entire frequency band. And the term *Gaussian* tells that it has a normal distribution in the time domain with an average time domain value of zero.

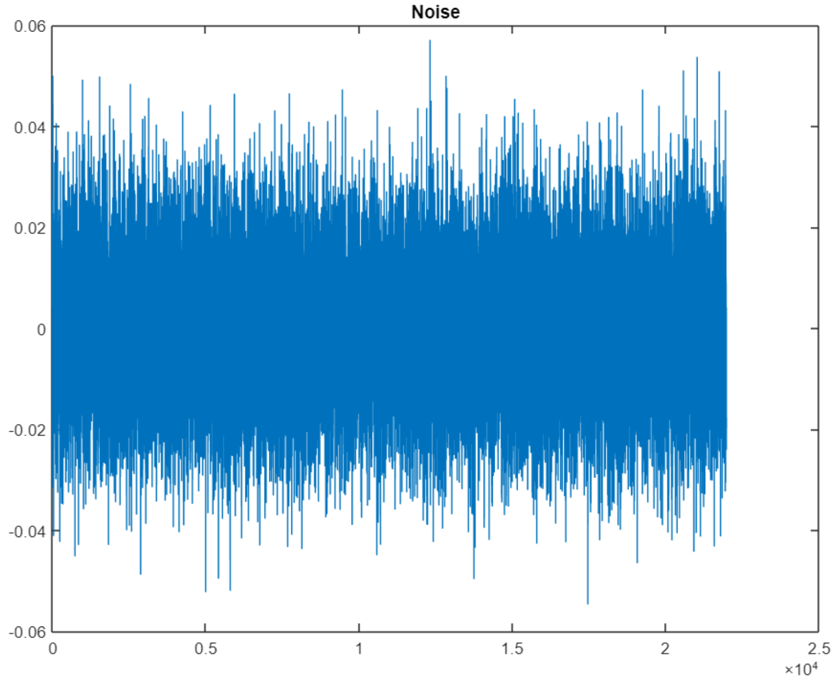
Matlab program:

```
close all;
clear all;
% r - reference signal
r = 'sp12.wav';
[y,Fs] = audioread(r);

%x - noisy signal
%Additive White Gaussian noise added to clean signal
x=awgn(y,10,'measured');

%Noise
v=x-y;

%n: number of segments with m samples each
m = 10;
n = round(length(x)/m);
r1=rem(length(x),m);
% variance of noise
varv = var(v);

%Mean of x
mx = mean(reshape(x(1:length(x)-r1),m,[]));
Mx = repelem(transpose(mx),m);
%replelem(mx,m) repeats each array element in mx 'm' times
varx = var(reshape(x(1:length(x)-r1),m,[]));
Varx = repelem(transpose(varx),m);
hn = zeros(length(mx),1);      %Filter transfer function
for i = 1:1:length(mx)
   if varx(i) > varv
   hn(i) =  (varx(i) - varv)/(varx(i));
   else if varv > varx(i)
   hn(i) = 0;
   end
   end
end
Hn = repelem(hn,m);
```

```
%Enhanced signal s
s = zeros(length(x),1);
for j =1:1:length(x)-r1
    s(j) = (Mx(j))+(Hn(j).*(x(j)-Mx(j)));
end

% Mean of noise
mv = mean(v)
% SNR of enhanced signal
SNR_f = snr(s,v)
figure(1);
plot(v);
title('Noise');
figure(2);
subplot(2,1,1)
plot(x,'b')
hold;
plot(y,'r');
title('Before Enhancement');
legend('Noisy signal','Clean Signal');
subplot(2,1,2);
plot(s,'k');
hold;
plot(y,'m');
title('After Enhancement');
legend('Noisy signal','Clean Signal');
%sound of enhanced signal
sound(s);
```

Results:

Here all the trails are done with only Additive White Gaussian Noise. 'mv' is the mean of noise signal. Example plot for White Gaussian Noise looks like this:

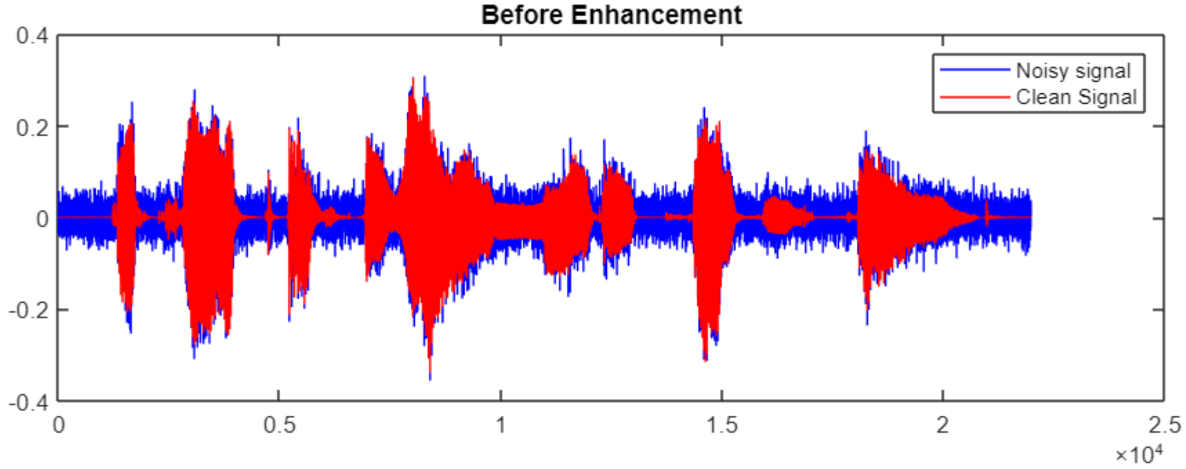


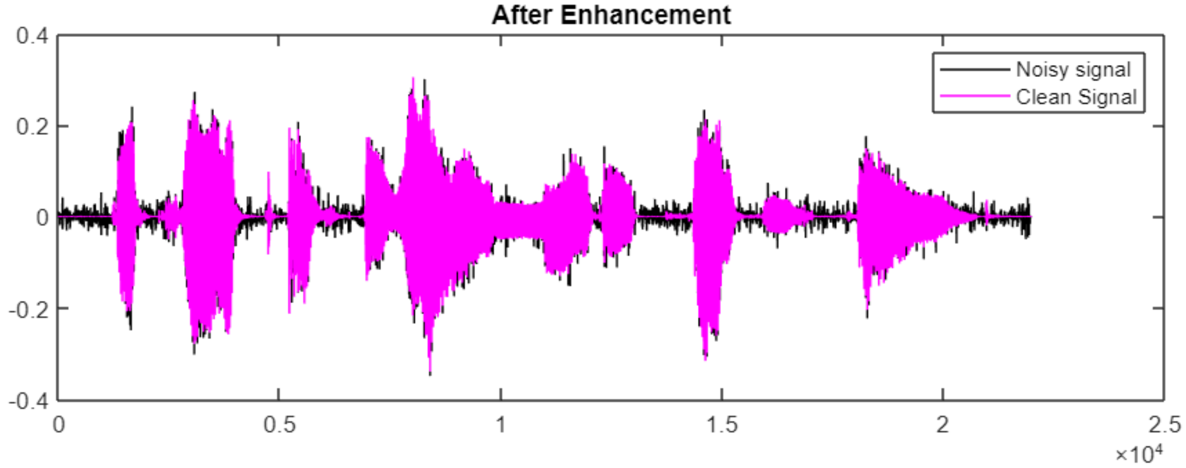
Audio samples collected:

3 Different audio samples of 2 seconds each are taken and then they are added with Additive White Gaussian Noise of different SNRs 5dB and 10dB, -5 dB, 15 dB etc. The resulting noisy signal is enhanced using Adaptive Wiener filtering technique and their plots are displayed and compared with the original signal.

Audio sample 1: ‘sp12.wav’- “The drift of the wind made a fuzzing sound” (2 sec).

- **When m=10 and SNR of noisy signal is 5dB**

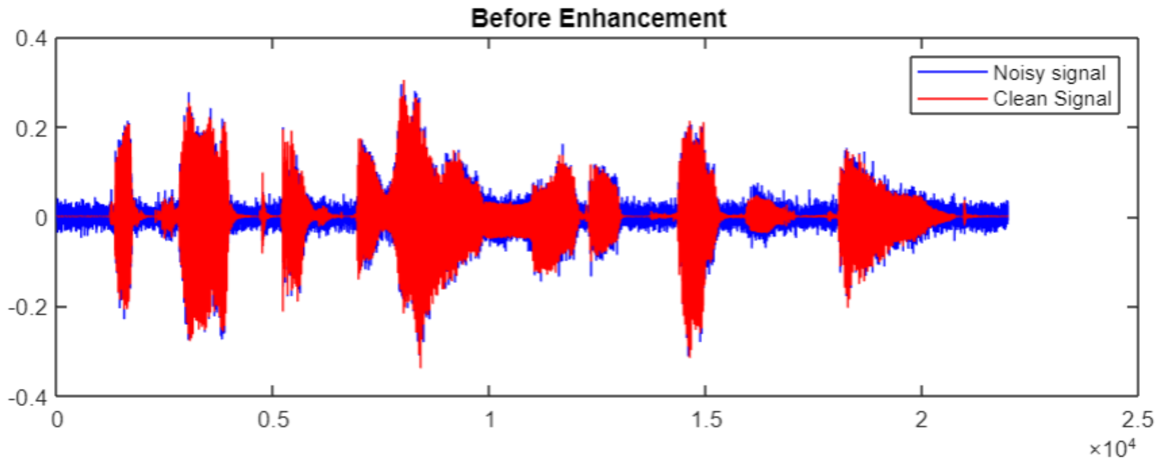


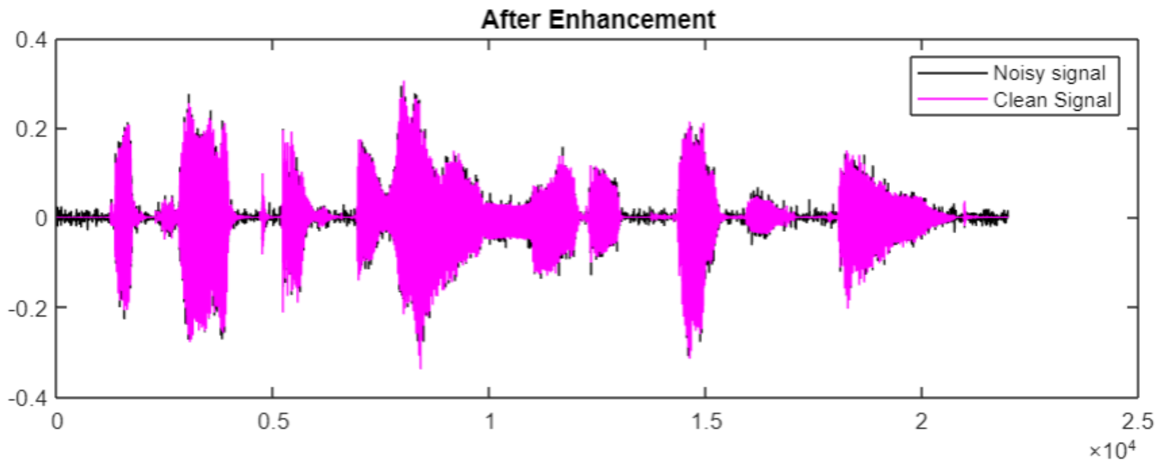



```
SNR_f =

      5.3158
```

SNR of the enhanced signal is 5.3158 dB.

- **When ‘m’=10 and SNR of noisy signal is 10 dB**

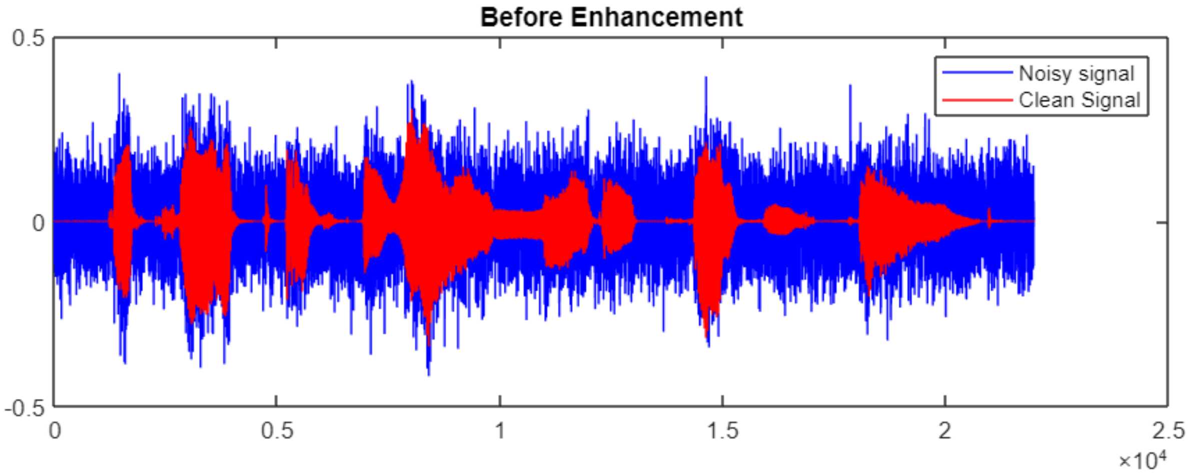


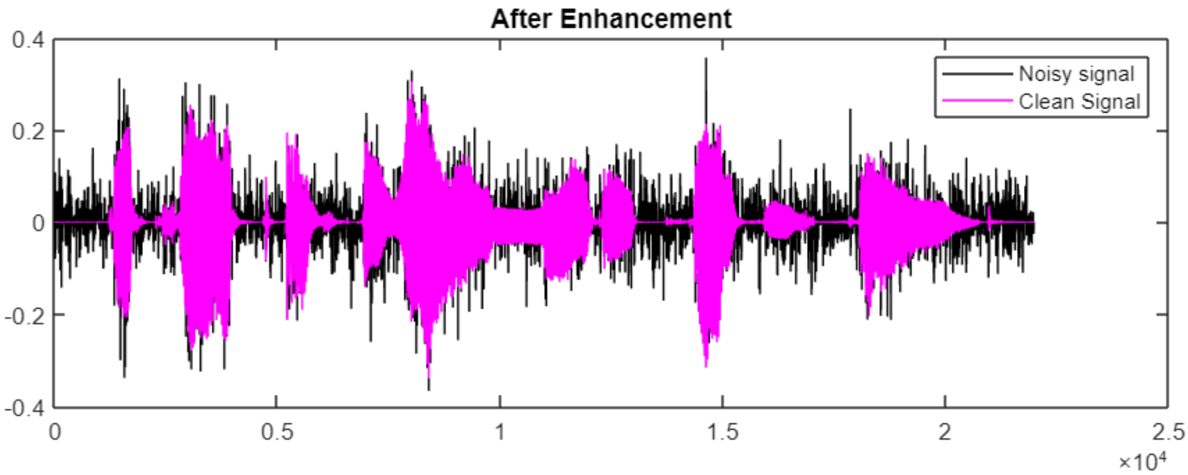



```
SNR_f =

    10.4837
```

SNR of the enhanced noisy signal is 10.4837 dB.

- **When ‘m’=10 and SNR of noisy signal is -5 dB**



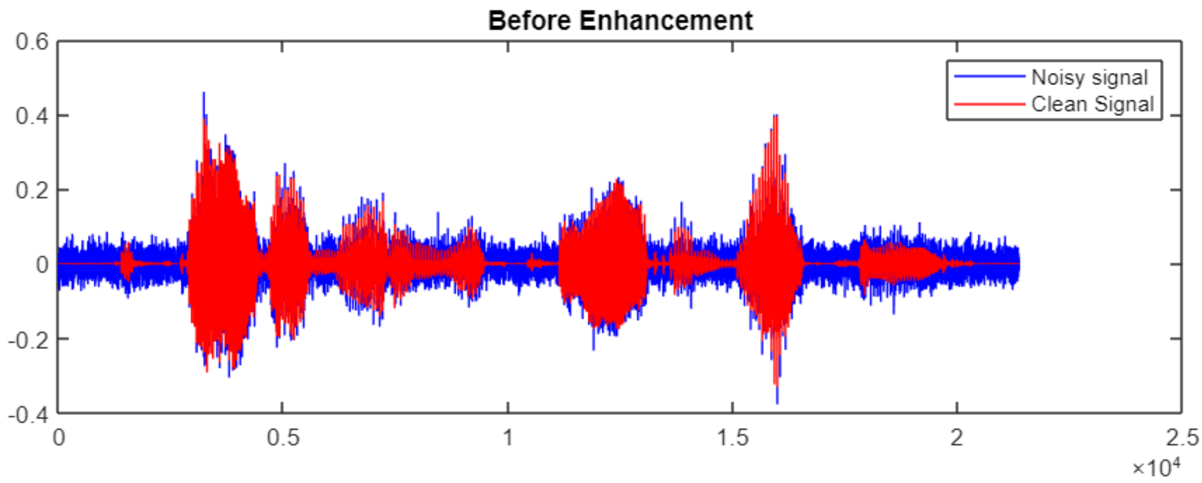


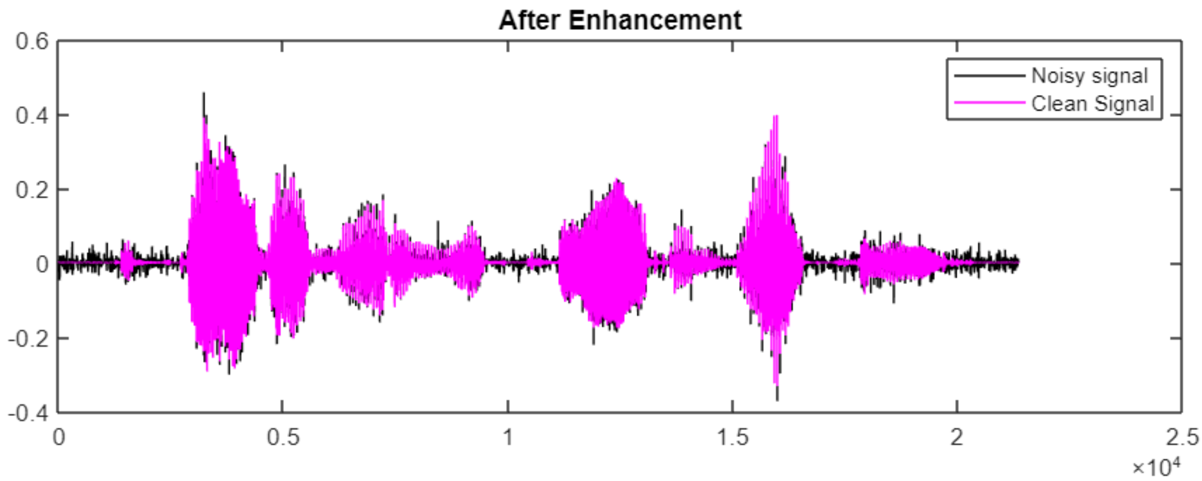

```
SNR_f =

   -3.7943
```

SNR of the enhanced noisy signal is -3.7943 dB. Here we can see that this particular can even enhance negative SNR.

Audio sample 2: ‘sp10.wav’- “The sky that morning was clear and bright blue” (2 sec).

- **When ‘m’=7 and SNR of noisy signal is 5 dB**

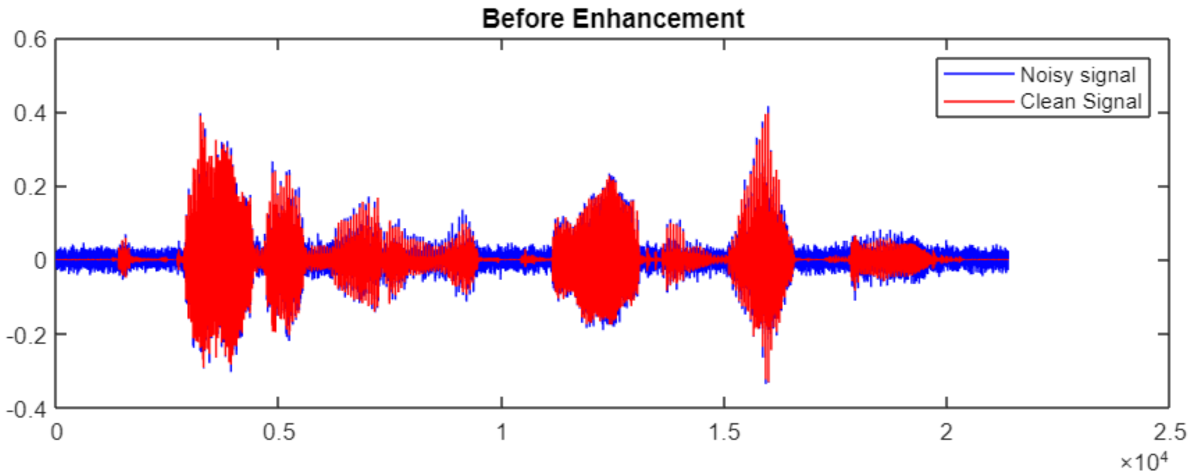


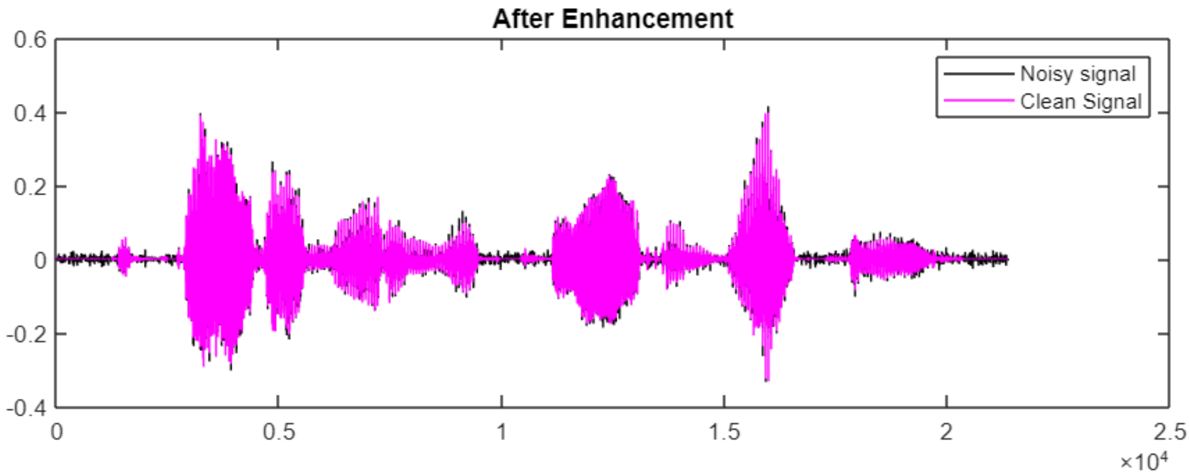



```
SNR_f =

    5.7661
```

SNR of the enhanced noisy signal is 5.7661 dB.

- When ‘m’=10 and SNR of noisy signal is 10 dB



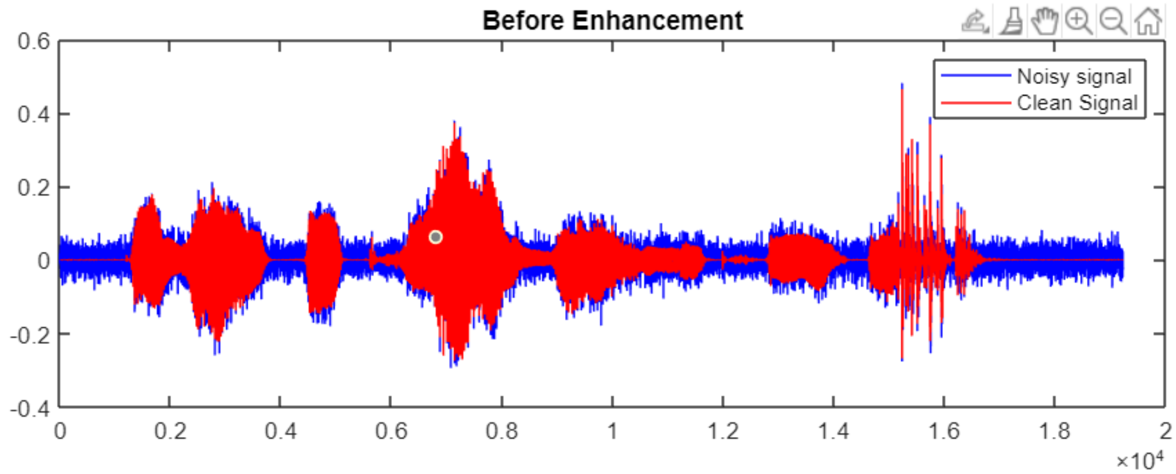


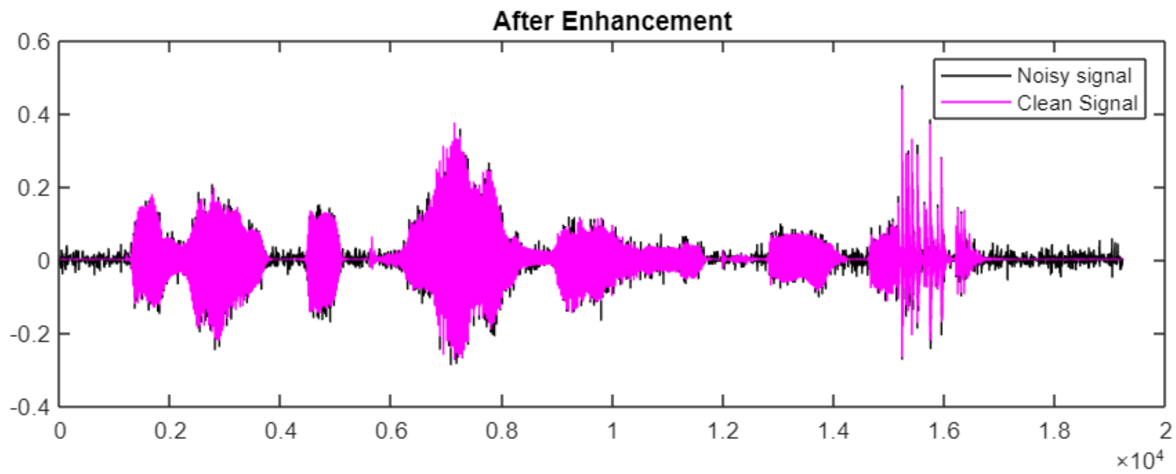

```
SNR_f =

    10.2704
```

SNR of the enhanced noisy signal is 10.2704 dB.

Audio sample 3: 'sp17.wav'- "The lazy cow lay in the cool grass" (2 sec).

- When 'm'=7 and SNR of noisy signal is 5 dB

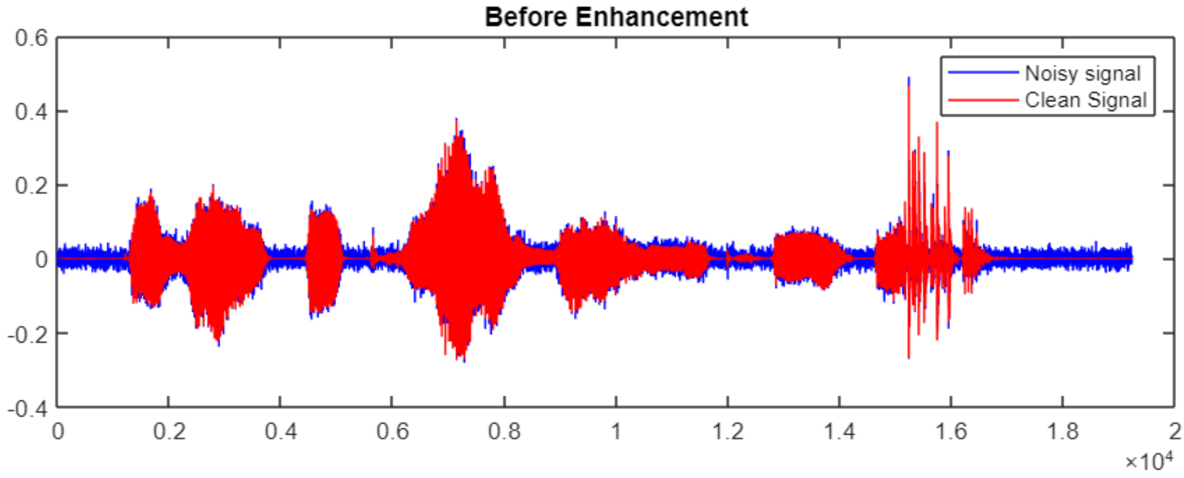


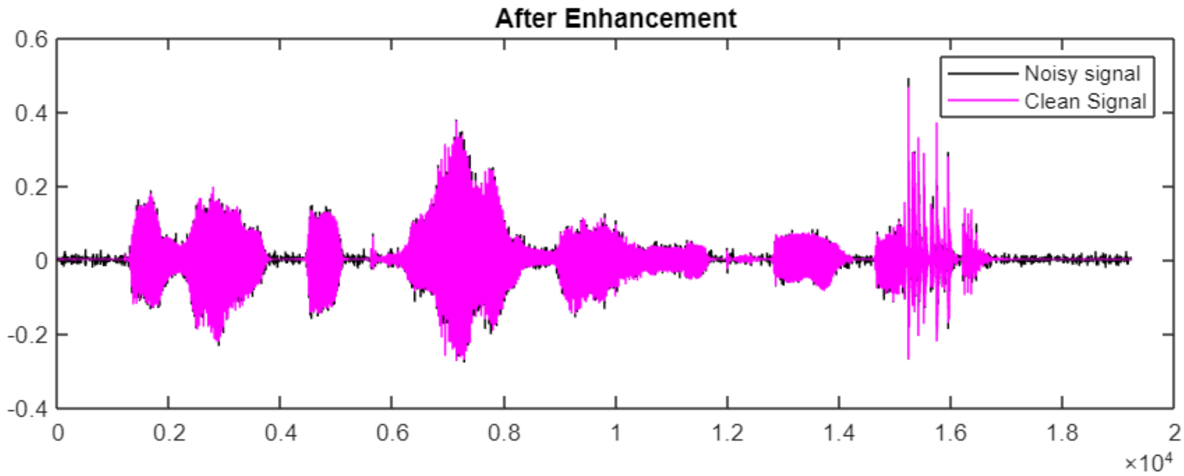



```
SNR_f =

    5.8528
```

SNR of the enhanced noisy signal is 5.8528 dB.

- When ‘m’=10 and SNR of noisy signal is 10 dB






```
SNR_f =

    10.5695
```

SNR of the enhanced noisy signal is 10.5695 dB.

Observation table

Samples	SNR before Enhancement	SNR after Enhancement
sp12.wav; m=10	5 dB	5.3158 dB
sp12.wav; m=10	10 dB	10.4837 dB
sp12.wav; m=10	-5 dB	-3.7943 dB
sp10.wav; m=7	5 dB	5.7661 dB
sp10.wav; m=10	10 dB	10.2704 dB
sp17.wav; m=7	5 dB	5.8528 dB
sp17.wav; m=10	10 dB	10.5695 dB

Conclusion

Speech signal enhancement has been achieved for different audio signals using Adaptive Wiener Filter technique. The enhancement is done in time domain. On observing the results from above experiments it very clear that the signals with white noise are all very enhanced now as the SNRs of the signals have been increased and the sound quality is also good when it is observed by hearing and looking at the plot. It is clear that the **best performance** is that of the **adaptive Wiener filter** because it considers the variation of the local statistics of the noisy speech signal compared to traditional Wiener filtering technique. One more advantage is Adaptive wiener filter can not only enhance the signals with positive SNR but it can also enhance the signals with negative SNR.

References

1) International Journal of Computer Applications (0975 - 8887) Volume 179 - No.17, February 2018 A Survey on Techniques for Enhancing Speech

2) Speech enhancement with an adaptive Wiener filter 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-Received: 12 February 2013 / Accepted: 1 August 2013 / Published online: 28 August 2013 © Springer Media New York 2013

3) ICASSP 86- TOKYO,CH2243-4/86/0000-3135 $1.00 0 1986 IEEE SPEECH ENHANCEMENT* by Jae S. Lim Massachusetts Institute of Technology Department of Electrical Engineering and Computer Science Cambridge, Massachusetts, USA

4) IJARCCE ISSN (Online) 2278-1021 ISSN (Print) 2319 5940 International Journal of Advanced Research in Computer and Communication Engineering ISO 3297:2007 Certified Vol. 5, Issue 8, August 2016 Copyright to IJARCCE DOI 10.17148/IJARCCE.2016.5859 296 A Review on Various Speech Enhancement Techniques Alugonda Rajani 1 , Soundarya .S.V.S2

5) IEEE, International Conference on Information, Communication, Instrumentation and Control (ICICIC-2017) Paper Id: 291 1 | Page ADAPTIVE WIENER FILTER FOR SPEECH ENHANCEMENT

6)A. Kawamura, Y. Iiguni, and Y. Itoh, "A noise reduction method based on linear prediction with variable step-size," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E88-A, no. 4, pp. 855–861, 2005.

7)S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 27, no. 2, pp. 113–120, 1979.

8)Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 32, no. 6, pp. 1109–1121, 1984.

9)B. Widrow, J. G. R. Glover Jr., J. M. Mccool et al., "Adaptive noise cancelling: principles and applications," *Proceedings of The IEEE*, vol. 63, no. 12, pp. 1692–1716, 1975.

10)P. J. Wolfe and S. J. Godsill, "Efficient alternatives to the Ephraim and Malah suppression rule for audio signal enhancement," *Eurasip Journal on Applied Signal Processing*, vol. 2003, no. 10, pp. 1043–1051, 2003.