

# Wiener Filtering in Frequency Domain to Enhance Speech Corrupted by Colored Noise

V.Sailaja, P.Sunitha, B.Vasanth Lakshmi , V.Prasanth

**Abstract:** This paper presents a method for speech enhancement to predict speech quality in presence of highly non-stationary scenarios using basic wiener filtering in frequency domain with an adaptive gain function under eight different noises at three different ranges of input SNR. Its performance is evaluated in terms of objective quality measures like LPC based spectral distortion measures are Cepstrum Distance, Itakura Saito and Log Likelihood Ratio. This method was tested using Noizeous database, its performance measures were compared against spectral subtractive type algorithms and it shows its improvements in terms of objective quality measures.

**Key words:** Speech Enhancement, Wiener Filtering, Spectral Subtraction, Noise, Objective Quality Measures

## I. INTRODUCTION

An effective mode of human communication is speech to convey information. This is good when speaker and listener are nearer to each other and in a quite environment but it degrades due to environmental noise which is unavoidable in nature. Effect of this noise on speech varies depending on the type of noise as it is having different characteristics. Noise is classified into two different ways based on the spectral characteristics. White Gaussian noise is one in which has uniform effect on the entire speech spectrum which can be easily suppressed using Spectral Subtraction methods. Another type of noise is colored noise whose effect is non-uniform on the spectrum which can be eliminated by using Multi Band spectral subtraction suggested by S.D Kamath [2] by using band specific subtraction factor. Spectral subtraction method is one among the historical methods in which subtraction is carried in between noisy speech spectrum and estimate of noise spectrum. This method gains more importance as it involves Fourier transform to represent a speech signal in temporal domain to spectral domain and vice versa but it results in over estimation of noise further it introduces speech distortion[1]. Many improvements were suggested to this method to palliate the effect of noise with minimum speech distortion.

One among them is wiener filter and its primary aim is to estimate the noise spectrum based on apriori information. First this Wiener filter was suggested by P.Scalart to perform speech enhancement in frequency domain for single channel which makes a clear distinction among speech distortion and noise reduction[3]. Later different variants of Wiener filtering techniques are available in literature for noise reduction. An adaptive filter in which filter function adaption change according to its statistics like mean and variance of speech signal in frequency domain and this was tested in presence of additive white noise and colored noise shows its improvement[4]. This method uses noise estimation during speech absence regions only, updating of spectrum in continuous manner is an essential requirement of speech enhancement algorithms in real world applications. Such a method with time recursive noise estimation combined with Wiener filter to suppress the noise and its performance is evaluated in terms of objective quality measures like segmental Signal to Noise Ratio and Perceptual Evaluation of Speech quality for various values of smoothing parameter [5] gives improvements over basic wiener filtering. Wiener filter combined with wavelet thresholding the multitaper spectrum for improvement of speech quality in terms of quality measures was given in [6], in which first the spectrum was estimated by multitaper method. Further the spectrum was refined by wavelet thresholding and noise estimation was done by apriori SNR estimation.

Organization of the paper is as follows; starting with the historical information of Wiener filtering methods for noisy speech enhancement. In section II, presents basic operation of spectral subtraction, Wiener filtering for speech enhancement was presented in section III, quality measures for performance evaluation discussed in section IV. Finally, the results and conclusion has been provided in section V.

## II. BASIC SPECTRAL SUBTRACTION

Speech signal degraded by means of additive noise is named as noisy speech represented by  $y(n)$ . Basically the noise and speech are uncorrelated and related as

$$y(n) = c(n) + d(n), n \in (0, N - 1) \quad (1)$$

Here 'n' represents the time index.  $c(n)$  represents the clean speech and  $d(n)$  represents the noise only. Since speech is not a stationary signal to make it as a stationary windowing is used to break down the noisy speech signal into frames of small duration generally of 20-30 ms duration. In this paper noisy speech signal is divided into frames of duration 30 ms with 50% overlapping and then its transform is computed by means of Fast Fourier Transform.

**Revised Version Manuscript Received on 02 September, 2019.**

**Dr.V.Sailaja**, Professor, Pragati Engineering College, Surampalem, Andhra Pradesh

**P.Sunitha**, Assoc.Professor, Pragati Engineering College, Surampalem, Andhra Pradesh

**B.Vasanth Lakshmi**, Assoc.Professor, Pragati Engineering College, Surampalem, Andhra Pradesh

**V.Prasanth**, Assoc.Professor, Pragati Engineering College, Surampalem, Andhra Pradesh

$$Y_W(K) = C_W(K) + D_W(K) \quad (2)$$

'K' represents the discrete-frequency index in a frame and  $W$  is the window throughout this paper, the signal is divided into frames first and then windowed, hence for simplicity, we drop the use of subscript  $W$  from windowed signals. The spectrum of the windowed signals can be computed using the Fourier Transform equations which consists of both magnitude and phase information. Which are given as

$$Y(K) = |Y(K)| \exp(j\phi_y(K)) \quad (3)$$

$$C(K) = |C(K)| \exp(j\phi_c(K)) \quad (4)$$

$$D(K) = |D(K)| \exp(j\phi_d(K)) \quad (5)$$

$\phi(K)$  represents phase component of speech. The multiplication between  $Y(K)$  is multiplied by its complex conjugate  $Y^*(K)$  yields the short-time power spectrum of noisy speech.

$$|Y(K)|^2 = |C(K)|^2 + |D(K)|^2 + 2 \operatorname{Re}\{C(K) \cdot D(K)\} \quad (6)$$

$$|\hat{C}(K)|^2 = |Y(K)|^2 - |\hat{D}(K)|^2 \quad (7)$$

Where  $|\hat{C}(K)|^2$ ,  $|\hat{D}(K)|^2$  are the estimated spectrums of clean and noise signal and  $|Y(K)|^2$  is the spectrum of the noisy speech, respectively.

Clean signal estimate ( $\hat{c}(n)$ ) can be obtained using synthesis techniques like Filter Bank Summation or overlap-add method. This paper used overlap add technique to combine spectrums of individual frames and then Inverse Fourier Transform is used for reconstruction. During enhancement stage phase information is not used because phase information does not perceived by the human perception but in reconstruction process it has to be considered.

$$\hat{c}(n) = \text{IFFT}\left\{|\hat{C}(K)| \cdot \exp(j\phi_y(K))\right\} \quad (8)$$

### III. SPEECH ENHANCEMENT USING WIENER FILTERING

The spectral subtraction method considered as a filtering operation. Filter is used to remove the unwanted frequency components which are related to noise signal [10]. The Spectral subtraction filter can be written as

$$|\hat{C}(\omega)|^b = |Y(\omega)|^b - |\hat{D}(\omega)|^b = H(\omega) \cdot |Y(\omega)|^b \quad (11)$$

where 'b' represents the power if it is equivalent to 1 then it is Magnitude Spectral Subtraction and when  $b=2$  indicates the Power Spectral Subtraction. From this the filter function ( $H(K)$ ) can be written as

$$H(K) = \left[1 - \frac{|\hat{D}(K)|^b}{|Y(K)|^b}\right] \quad (9)$$

In general spectral subtraction methods estimate *a posteriori* SNR which are suitable for stationary white noise only. This problem can be elevated using smoothing time varying filter called Wiener filter [8]. Wiener filtering can be implemented in time and frequency domains. In time domain Wiener filter method can be implemented with Wiener Hopf equations. The spectral

subtraction method can also be treated as a filter in frequency domain by considering the filter function as a ratio of clean and noisy signal magnitudes [1]. Attenuation of this spectral variation accommodates colored noise, and updating takes place at any desired frame in order to handle non-stationary noise.

$$H_{\text{wiener}}(K) = \frac{P_c(K)}{P_y(K)} = \frac{P_c(K)}{P_c(K) + P_d(K)} = \frac{P_y(K) - P_d(K)}{P_y(K)} \quad (10)$$

The main limitation of this method is its frequency response is uniform at all frequencies and the requirement to estimate the power spectral density of the clean signal and noise prior to filtering. Therefore, non-causal WF cannot be applied directly to estimate the clean speech since speech cannot be assumed to be stationary. Therefore, an adaptive WF implementation can be approximate as

$$H_{\text{wiener}}(K) = \frac{|Y(K)|^2 - |\hat{D}(K)|^2}{|\hat{D}(K)|^2} \quad (11)$$

$$|\hat{C}(K)|^2 = H_{A.\text{wiener}}(K) \cdot |Y(K)|^2 \quad (12)$$

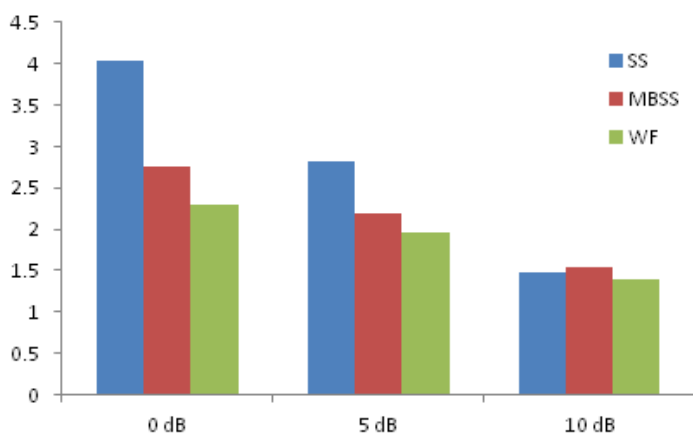
By comparing  $H(K)$  and  $H_{A.\text{wiener}}(K)$  from the equations 9 and 11 it can be noticed that the Wiener Filter which uses adapts the changes in filter function gives more accurate results than basic Wiener filter.

### IV. RESULTS AND CONCLUSION

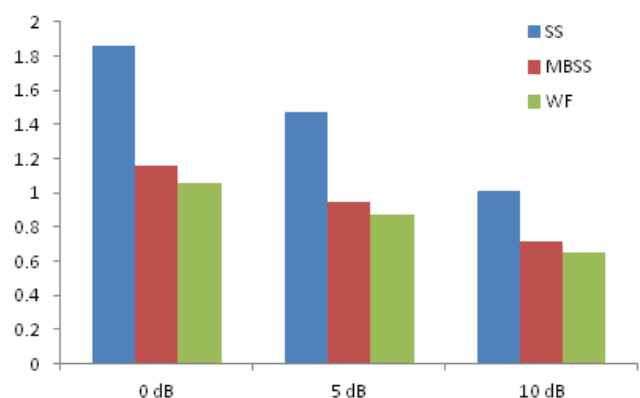
To predict speech quality of speech enhancement techniques in presence of noise several objective quality measures are available in the literature. Comparison of original clean speech and enhanced speech signals by a group of listeners on a predetermined scale is known as subjective quality evaluation. One can use objective evaluation to quantify the quality between the original clean and enhanced speech signals using mathematical equations. This paper presents the performance evaluation based on three different quality measures which are LPC based spectral distance measures (LLR, IS, CEP) [8]. To evaluate the objective quality measures the speech signal is taken from a Noisy database [9] consists of different speech sentences in American English Language under clean conditions and noisy speech at various levels of input SNR. The sentence is "We can find joy in the simplest things" taken from a male speaker. Then this signal is divided into frames of duration of 30 ms using Hamming Window with 50% overlapping. Then posteriori SNR was estimated in each frame based on a priori information and filter function was updated adaptively. The results were shown in table 1. shows the improvement of Wiener filter when compared to Spectral Subtraction and Multi Band Spectral Subtraction methods. The same results can be shown in the form of bar graphs by taking the average of eight different noises at three ranges of input SNR which shows the superiority of Wiener filtering for all the measures considered.

**Table 1:LPC based distortion measures for Spectral Subtraction(SS),Multi Band Spectral Subtraction(MBSS)and Wiener Filtering(WF)**

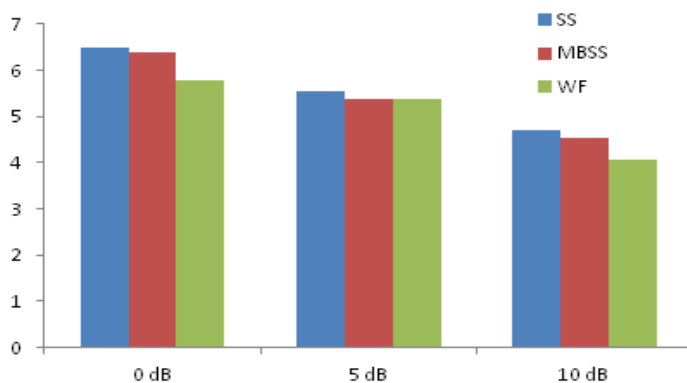
Type of Noise	Input SNR	Itakura Saito			Cepstrum Distance			LLR		
		SS	MBSS	WF	SS	MBSS	WF	SS	MBSS	WF
Airport	0dB	3.64	2.72	2.07	6.43	6.65	6.02	1.78	1.15	0.96
	5dB	1.91	1.46	1.25	5.12	5.01	4.77	1.24	0.74	0.67
	10dB	1.35	1.12	1.08	4.87	4.75	4.10	0.98	0.69	0.54
Babble	0dB	4.85	2.34	2.09	6.73	6.47	6.15	1.93	1.18	1.09
	5dB	2.07	1.93	1.92	5.56	5.77	5.65	1.57	0.98	0.92
	10dB	1.64	1.19	1.09	4.54	4.59	4.38	0.95	0.66	0.61
Car	0dB	3.51	3.26	2.48	6.33	6.45	6.04	1.98	1.21	1.04
	5dB	2.91	2.39	1.73	5.31	5.76	5.14	1.63	1.02	0.83
	10dB	1.66	1.67	1.12	4.58	4.89	4.19	1.09	0.76	0.58
Exhibition	0dB	3.75	3	2.62	6.35	6.49	5.98	2.01	1.28	1.04
	5dB	5.03	2.46	2.08	5.85	5.73	5.57	1.96	1.01	0.88
	10dB	0.87	1.54	1.41	5.03	4.98	5.01	1.12	0.78	0.73
Restaurant	0dB	3.77	1.84	1.81	6.06	5.82	5.72	1.15	0.95	0.93
	5dB	1.23	1.62	1.59	5.91	6.05	5.73	0.99	0.93	0.85
	10dB	1.16	1.16	1.11	4.45	4.64	4.12	0.79	0.63	0.52
Station	0dB	3.18	2.78	2.06	6.54	6.39	6.20	1.96	1.12	1.08
	5dB	2.28	2.46	1.98	5.65	5.81	5.14	1.87	1.04	0.81
	10dB	1.52	2.05	1.84	4.94	5.17	5.70	1.21	0.78	0.91
Street	0dB	4.96	3.25	2.65	6.20	6.52	6.48	2.10	1.22	1.17
	5dB	4.07	2.36	2.24	5.70	5.69	5.36	1.13	0.93	0.89
	10dB	1.91	1.95	1.89	4.08	4.01	3.90	0.68	0.59	0.56
Train	0dB	4.66	2.79	2.54	6.42	6.32	3.72	1.96	1.17	1.15
	5dB	3.09	2.94	2.91	5.35	5.66	5.50	1.34	1.01	1.09
	10dB	1.67	1.61	1.61	5.10	4.90	4.97	1.24	0.92	0.72



(a) Itakura Saito measure



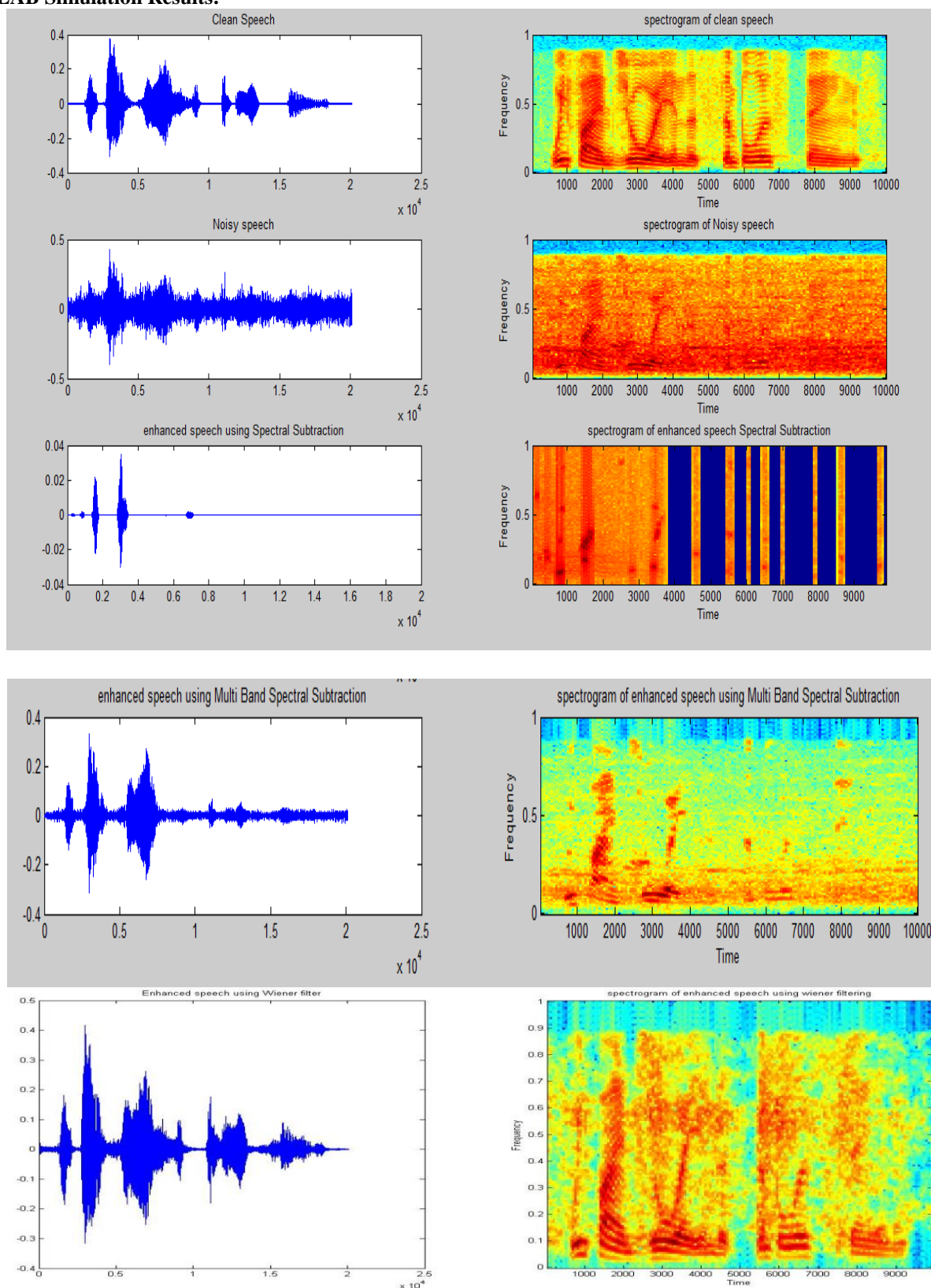
(c) Log Likelihood Ratio measure



(b) Cepstrum Distance measure

**Fig.2.Performance measures for Spectral Subtraction(SS),Multi Band Spectral Subtraction(MBSS) and Wiener Filtering(WF) methods.**

## MATLAB Simulation Results:



**Fig.3:** Left column represents the Time domain representations of clean, noisy and enhanced speech signal by spectral subtraction, Multi Band Spectral Subtraction and Wiener filtering methods. Right column represents the spectrograms of the clean , noisy and enhanced signals



## References:

1. Boll,S.F.(1979).”Suppression of acoustic noise in speech using spectral subtraction”. IEEE Transactions on Acoustics Speech and Signal Processing, 27(2), 113–120.
2. Kamath S., Loizou P., “A multiband spectral subtraction method for enhancing speech corrupted by colored noise”, Proc .IEEE Intl. Conf. Acoustics, Speech, Signal Processing, 2002.
3. P Scalart,J Viera –Filho,Speech enhancement Based on a Priori Signal to Noise estimation , proceedings of 21 st IEEE International Conference on Acoustics,Speech ,Signal Processing (ICASSP-96),pp.629-632.
4. Marwa A.Abd EI –Fattah et al,”Speech Enhancement with an adaptive Wiener Filter “,Int.Journal of Speech Technology,March 2014,Volume 17,Issue 1,pp 53-64.
5. Navneet Upadhyay and Abhijit Karmakar,” Single Channel Speech Enhancement using Wiener Filtering with Recursive Noise Estimation”, International conference on Intelligent Human Computer Interaction, IHCI 2015, Procedia Computer Science 84 , 2016 .
6. Yanna Ma and Akinori Nishara,” A modified Wiener filtering method combined with wavelet thresholding multitaper spectrum for speech enhancement”, EURASIP Journal on Audio, Speech, and Music Processing,volume 32 ,2014.
7. J Chen,J benesty,YHaung,New Insights into the noise reduction Wiener Filter,IEEE Trans.on Audio Speech Lang.Process,14(4),1218-1234,1996
8. A Noisy Speech Corpus for Assessment of Speech Enhancement Algorithms .<https://ecs.utdallas.edu/Loizou/speech/noizeous>
9. Yi.,Hu and Loizou P.C., "Evaluation of Objective Quality Measures for Speech Enhancement," IEEE Transactions on Audio, Speech and Language Processing ,vol.16,no.1,pp.229-238,Jan.2008.
10. P.Sunitha and K.Satya Prasad ,”Multi band Spectral Subtraction for speech enhancement with three different frequency spacing methods and their effect on quality measures”, International Journal of Image ,Graphics and Signal Processing,volume 11 No.5 ,May 2019.