

Multi Band Spectral Subtraction Combined with Adaptive Noise Estimation



P. Sunitha, K. Satya Prasad

Abstract: *Speech enhancement is a prerequisite for many speech communication applications like speech coders, hearing aids, voice recognition systems, Man –Machine communication, Video conferencing etc . For all these practical systems input is a noisy speech , so it requires a pre-processing operation for further processing. This paper gives a method to improve the speech quality in presence of background noise, based on Multi Band Spectral Subtraction combined with adaptive noise estimation. Further the performance analysis of this method is done using objective quality measures like Segmental SNR(seg-SNR), Weighted SlopeSpectral Distance (WSSD), Log Likelihood Ratio(LLR), Perceptual Evaluation of Speech Quality (PESQ) and Composite Measures under highly non-stationary noise. Finally the results were compared against Spectral Subtraction(SS), Multi Band Spectral Subtraction(MBSS) methods and the results shows that the proposed method shows consistence performance against all the cases considered.*

Index Terms: *speech enhancement, Voice Activity Detection, NoiseEstimation, Multi Band Spectral Subtraction, adaptive noise estimation, Speech presence probability, Objective Performance Measures.*

I. INTRODUCTION

In speech enhancement techniques noise estimation plays a key role to enhance the speech quality in presence of highly non-stationary noise. Most of the speech enhancement techniques use Voice Activity Detection (VAD) to discriminate speech segments from non-speech segments. VAD algorithms extract features from noisy speech during speech absence segments and compared against a threshold value .If the measured value is greater than the threshold value then it is considered as a speech presence segment otherwise it is treated as a speech absence segment. Several VAD algorithms are available in literature .VAD algorithm based on statistical model given in[1], detects the speech presence and absence segments based on likelihood criterion. Statistical based algorithms are easy to implement and effective, but it fails when average of the likelihood ratio is greater than the threshold as it uses fixed threshold. One more drawback i.eit severely affectthe performance of noise estimation due to large delay.This can be bypassed using an adaptive threshold [2],by this we can detect many silence regions those cannot be detected when fixed threshold was used. Improvements can be found when VAD implemented with multiple observation likelihood ratio test[3].

Recursive averaging with adaptive parameter was used to implement VAD for non-stationary noise [4].These algorithms work well under stationary noise but fails in non-stationary noise. Speech enhancement applications requires accurate noise estimation at all times during speech presence segments also

In such cases VAD might not be sufficient to get precise noise power spectrum. When the noise power spectrum estimate is more accurate then enhanced signal is more nearer to the clean signal. To improve the speech quality and intelligibility in presence of highly non stationary noise,a speech enhancement algorithm requires noise estimation algorithms which update the noise spectrum continuously.Most of the speech enhancement applications requires accurate noise estimation at all times during speech presence segments also .In such cases VAD might not be sufficient to get precise noise power spectrum. When the noise power spectrum estimate is more accurate then enhanced signal is more nearer to the clean signal. To improve the speech quality and intelligibility in presence of highly non stationary noise,a speech enhancement algorithm requires noise estimation algorithms which update the noise spectrum continuously.Most of the speech enhancement applications in non-stationary scenarios use noise estimation methods algorithms which track the noise spectrum continuously [5]. Now, researchers focus their attention to improve the speech quality and intelligibility using efficient noise estimation algorithms. Minimum statistics is one among them,which updates the noise on a frame by frame basis by considering the minimum over a periodogramof input noisy speech signal over a small window[5].The main issue in this method is tracking of minimum leads to the over estimation of noise spectrum ,because of its inability to respond fast changes of the noise spectrum.Minima Controlled Recursive Algorithm (MCRA) proposed by Cohen estimates the noise power spectrum, by considering the average of past spectral values of speech against a time–frequency dependent smoothing factor.The key problem with this method is it requires twice the number of frames [6]. To overcome this Cohen proposed, Improved Minima Controlled Recursive Algorithm (IMCRA) by continuously monitoring the minimum of the noise estimate from noisy speech based on speech presence probability[7].Mostof the minimum statistics algorithms introduce too much residual noise because of inaccurate estimate of noise power. Time recursive averaging algorithms exploits the fact that, the noise spectrum has non-uniform effect on the speech spectrum, some regions will effect more adversely than others. For any type of noise spectrum estimate it is not possible to obtain estimation and updating of noise in each frequency bin of the entire noise spectrum.

Revised Manuscript Received on 30 July 2019.

* Correspondence Author

P.Sunitha, Research Scholar, Dept. of ECE, JNTUK,India,
Dr.K.Satya Prasad Rector, VFSTR,Guntur, India,

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

This observation makes the use of recursive type algorithms, whose noise spectrum is updated by considering the weighted average of past and present noisy speech spectrum estimates. The weight changes according to the speech presence probability [8]. This paper implements Multi Band Spectral Subtraction [13] with adaptive noise estimation and its performance is evaluated in terms of performance measures. The structure of the paper as follows, section II gives Multi Band Spectral Subtraction, Section III explores Noise Estimation by adaptive noise estimation Algorithm and Section IV gives objective quality measures finally results and conclusion are presented in section V.

II. MULTI BAND SPECTRAL SUBTRACTION

In realistic environments, noise is mostly colored, which affects the speech signal differently throughout the spectrum. Some of the frequencies are severely affected than others, based on the spectral characteristics of the noise. This section gives the details of it Here $y(n)$ represents noisy speech, $x(n)$ and $d(n)$ are clean and noise speech signals. Since Speech and audio signals are time-varying signals. If we compute the spectrum over the entire signal, then we get the average spectrum, but we cannot observe the individual phones or changes in fundamental frequencies. In real-time applications we also need to split up the signal in to segments such that we do not have to wait for the whole sentence to be finished before we can start processing. In this scenario the spectrum is computed using windowing and the Fast Fourier Transform Now FFT of the noisy signal is represented by

$$Y(K) = X(K) + D(K) \quad (2)$$

Multi Band Spectral Subtraction consists of different steps. Firstly, the spectrum of the windowed signal is computed using FFT. Secondly, the spectrums of the noise and noisy speech are divided into four different frequency bands and evaluate the over-subtraction factor in each band. In the third stage, subtraction is carried in individual frequency bands by readjusting the over subtraction factor. Finally, the modified frequency bands are merged and the time domain signal is reconstructed with the aid of the overlap-add method and taking the IFFT. In this algorithm over-subtraction factor is readjusted in each band. Therefore, the estimate of the clean speech spectrum in the j^{th} band is given by

$$|\widehat{X}_j(K)|^2 = \begin{cases} |Y_j(K)|^2 - \mu_j \cdot \delta_j \cdot |\widehat{D}_j(K)|^2, \\ \text{if } |\widehat{X}_j(K)|^2 > \phi |Y_j(K)|^2 \text{ else} \\ \phi \cdot |Y_j(K)|^2 \end{cases} \quad (3)$$

$b_j < K < e_j$

Where b_j and e_j are the starting and end frequency bins of the j^{th} frequency band. The band specific over subtraction factor is given by μ_j as a function of the segmental SNR in each band. The segmental SNR in j^{th} Band can be evaluated as

$$SNR_j(\text{db}) = 10 \log_{10} \left(\frac{\sum_{K=b_j}^{e_j} |Y_j(K)|^2}{\sum_{K=b_j}^{e_j} |\widehat{D}_j(K)|^2} \right) \quad (4)$$

The μ_j can be calculated as

$$\mu_j = \begin{cases} \mu_{\max}, & \text{if } SNR_j \leq SNR_{\min} \\ \mu_{\max} + (SNR_j - SNR_{\min}) \left(\frac{\mu_{\min} - \mu_{\max}}{SNR_{\max} - SNR_{\min}} \right), & \text{if } SNR_{\min} < SNR_j < SNR_{\max} \\ \mu_{\min}, & \text{if } SNR_j \geq SNR_{\max} \end{cases} \quad (5)$$

Here $\mu_{\min}=1$, $\mu_{\max}=5$, $SNR_{\min} = -10\text{dB}$, $SNR_{\max} = 10 \text{ dB}$. The δ_j provides an additional degree of control within each band which can be evaluated as

$$\delta_j = \begin{cases} 1 & f_j \leq 1\text{kHz} \\ 2.5 & 1\text{kHz} < f_j \leq \frac{F_s}{2} - 2\text{kHz} \\ 1.5 & f_j > \frac{F_s}{2} - 2\text{kHz} \end{cases} \quad (6)$$

III. ADAPTIVE NOISE ESTIMATION

Fundamental requirement of any speech enhancement algorithm is an accurate estimate of noise spectrum. Most of the Spectral subtractive type algorithms use Voice Activity Detection (VAD) for speech enhancement to separate the voiced and unvoiced segments which works well in case of stationary noise but fails in highly non-stationary noise. To overcome this, the proposed method uses adaptive noise estimation. Noise estimation algorithms works on the assumption that the analysis segment is, too long enough that it should contain both low energy segments and speech pauses. Noise is more stationary than speech in a analysis window. The spectrum of the noisy signal can be found by applying a window $w(n)$ with N samples and then compute the FFT of the windowed signal

$$Y(\lambda, K) = \sum_{n=0}^{N-1} y(\lambda N + n) w(n) e^{-j2\pi n k / N} \quad (7)$$

λ is the frame index and, 'K' represents the frequency bin index. Remembering the fact that both speech and noise are uncorrelated, the sum of the periodograms of speech and noise is approximately equals to the periodogram of noisy speech [7,8].

$$|Y(\lambda, K)|^2 = |X(\lambda, K)|^2 + |D(\lambda, K)|^2 \quad (8)$$

$|X(\lambda, K)|^2$, $|D(\lambda, K)|^2$ and $|Y(\lambda, K)|^2$ represents the periodograms of Clean speech, noise and noisy Speech respectively. Based on this assumption, an estimate of noise power spectrum can be obtained by tracking the minimum of the periodogram $|Y(\lambda, K)|^2$ of the noisy speech over a window. Here periodogram changes very rapidly over time, this led to the use of first order recursive version of periodogram, which results in smoothed noisy PSD [8], given as

$$P(\lambda, K) = g P(\lambda-1, K) + (1-g) |Y(\lambda, K)|^2 \quad (9)$$

Where 'g' is a smoothing constant whose value is in between '0' and '1'.

Adaptive Noise estimation algorithm involves the following steps

1. Separate the noisy speech into speech present and absent segments from the smoothed noisy PSD. Speech absent segments contains the background noise. The noise estimate is updated by tracking the noise only regions. To identify the noise only segments, obtain the ratio of noisy power spectrum ($P(\lambda, K)$

Where 'g' is a smoothing



speech presence and $S(\lambda, K)=0$ shows the speech absence regions. Where $\xi(K)$ is a frequency dependent threshold.

4. After computing speech presence probability, compute the new time-frequency dependent smoothing factor

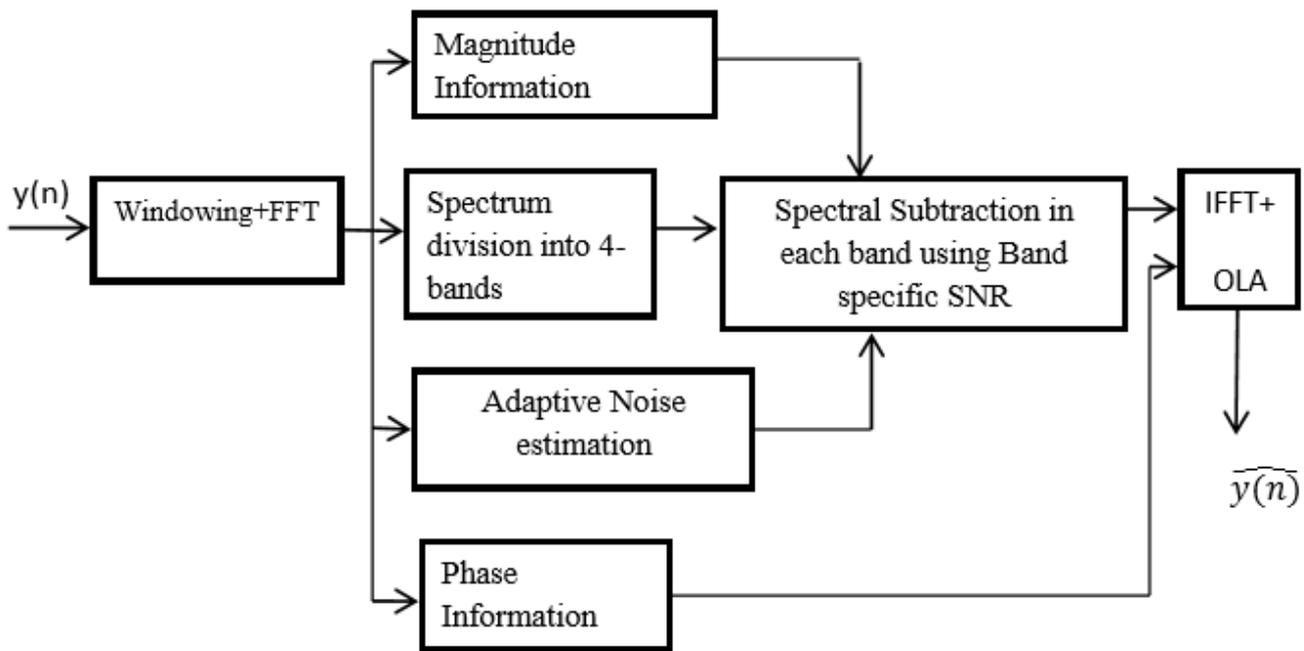


Fig.1. Block diagram of Multi Band Spectral Subtraction combined with adaptive noise estimation for speech enhancement.

IV. RESULTS AND CONCLUSIONS

Assessment of speech enhancement techniques can be done either by using objective quality or subjective listening tests. Comparative analysis of clean speech and enhanced speech signals by a group of listeners who are asked to identify the no of words correctly recognized on a pre-determined Scale is known as subjective listening test based on human auditory system.. Which involves a complex process and it is difficult to identify the persons with good listening skills. While objective evaluation is done on mathematical comparison of clean and enhanced signals .In order to calculate the objective measures, the speech signal is first divided into frames of duration of 10-30 msec. This result in a single measure which gives the average of distortion measures calculated for all the processed frames. This section gives the performance analysis of the proposed method by using four numbers of bands. Simulations were performed in the MATLAB environment. NOIZEUS is used as a speech corpus which is available at [9] and used by the most of the researchers, containing 30 sentences of six different speakers, three are male and other three are female speakers originally sampled at 25 KHz and down sampled to 8 KHz with 16 bits resolution quantization. Clean Speech is distorted by eight different real-world noises (babble, airport, station, street, exhibition, restaurant, car and train) at three distinct ranges of input SNR (0dB, 5dB, 10dB). In this algorithm speech sample is taken from a male speaker, English sentence is "we can find joy in the simplest things". This paper presents the performance evaluation based on different quality measures which are segmental-SNR, Weighted Slope Spectral Distance (WSSD) [10], Log

Likelihood Ratio, Perceptual Evaluation of Speech Quality (PESQ) [11] and three different composite measures[10].

A. Segmental SNR (seg-SNR): To improve the correlation between clean and processed speech signals summation can be performed over each frame of the signal [10] this results in segmental SNR .The segmental Signal-to-Noise Ratio (seg-SNR) in the time domain can be expressed as

$$SNR_{seg} = \frac{10}{M} \sum_{M=0}^{M-1} \log_{10} \frac{\sum_{n=NM}^{NM+N-1} x^2(n)}{\sum_{n=NM}^{NM+N-1} (x(n) - \hat{x}(n))^2} \quad (18)$$

Here $x(n)$ shows the original speech signal. \hat{x} is the processed speech signal, frame length is given by N and the number of frames is given by M . The geometric mean of all frames of the speech signal is seg-SNR [10], whose value was limited in the range of [-10, 35dB]

B. Log Likelihood Ratio(LLR): This measure was based on LPC analysis of speech signal.

$$LLR(\vec{a}_x, \vec{a}_{\hat{x}}) = \log \left(\frac{\vec{a}_{\hat{x}} R_x \vec{a}_{\hat{x}}^T}{\vec{a}_x R_x \vec{a}_x^T} \right) \quad (19)$$

$\vec{a}_x, \vec{a}_{\hat{x}}^T$ are the LPC coefficients of the original and processed signals. R_x is the autocorrelation matrix of the original signal .In LLR denominator term is always lower than numerator therefore LLR is always positive [10] and the LLR values are in the range of (0-2).

C. Weighted Slope Spectral Distance(WSSD): This measure can be evaluated as



Multi Band Spectral Subtraction Combined with Adaptive Noise Estimation

the weighted difference between the spectral slopes in each band can be computed using first order difference operation[10].Spectral slopes in each band of original and processed signals are given by

$$WSSD = \frac{1}{M} \sum_{m=0}^{M-1} \frac{\sum_{j=1}^K W(j, m) (X_x(j, m) - X_{\hat{x}}(j, m))^2}{\sum_{j=1}^K W(j, m)} \quad (20)$$

Table 4: Objective quality measures Segmental SNR(seg_SNR),Log Likelihood Ratio(LLR),Weighted Slope Spectral Distance (WSSD),Perceptual Evaluation of Speech Quality(PESQ),Signal Distortion(C sig),Background intrusiveness(C bak),Overall Quality(C ovl) for Spectral Subtraction(SS),Multi Band Spectral Subtraction(MBSS) and Proposed Method(PM)

Type	IP	Parameter																					
		Seg-SNR			LLR			WSSD			PESQ			C sig			C bak			C ovl			
		SS	MB SS	P M	SS	MB SS	P M	SS	MB SS	P M	SS	MB SS	P M	SS	MB SS	PM	SS	MB SS	PM	SS	MB SS	PM	
Airport Noise	0d	-	-	0.8	1.7	1.1	94.8	57.0	52.5	1.6	2.5	1.88	2.5	2.80	1.67	1.9	2.53	1.55	2.1	2.58	2.1	2.58	
	B	1.9	2.18	2	1.3	1.5	4	4	5	0	1.86	2.5	1	88	2	1.9	2.53	1.55	2.1	2.58	2.1	2.58	
	3	0.6	-	1.8	1.2	0.89	0.7	75.1	52.7	48.3	2.2	2.44	3.0	2.87	3.3	2.3	2.61	2.5	2.84	2.42	2.8	3.29	
Babble Noise	0d	-	-	1.9	1.33	1.1	97.1	60.9	56.5	1.2	2.4	1.49	2.4	2.71	1.44	1.9	2.48	1.15	2.0	2.51	2.0	2.51	
	B	1.9	2.60	0.4	1.33	1.1	3	8	5	4	1.82	2.4	1.49	2.4	2.71	1.44	1.9	2.48	1.15	2.0	2.51	2.0	2.51
	3	0.1	-	2.4	1.5	1.13	0.9	79.7	46.2	41.4	1.9	2.23	2.8	2.29	3.0	2.3	2.29	2.03	2.1	2.5	2.6	3.04	
Car Noise	0d	-	-	0.5	1.9	1.2	84.6	49.7	45.5	1.5	1.75	2.4	2.4	2.73	1.83	1.9	2.53	1.64	2.0	2.51	2.0	2.51	
	B	1.3	2.45	5	1.36	1.2	7	4	2	5	1.75	2.4	2.03	2.73	1.83	1.9	2.53	1.64	2.0	2.51	2.0	2.51	
	3	1.4	-	2.9	1.6	1.17	1.0	72.6	40.5	36.4	2.2	2.30	2.8	2.57	3.0	2.4	2.28	2.36	2.0	2.5	2.6	3.04	
Exhibition Noise	0d	-	-	0.8	2.0	1.43	1.2	95.4	50.0	45.5	1.4	1.65	2.3	1.78	2.1	1.9	2.48	1.45	2.1	2.39	2.1	2.39	
	B	1.5	2.16	4	1.43	1.2	8	3	5	8	1.65	2.3	1.78	2.1	1.9	2.48	1.45	2.1	2.39	2.1	2.39		
	3	1.0	-	2.7	1.9	1.16	1.0	82.2	42.8	38.1	1.9	2.21	2.8	2.20	3.0	2.3	2.09	2.7	2.1	2.5	2.6	3.03	
Restaurant Noise	0d	-	-	0.2	1.1	0.9	102.0	67.0	62.2	1.2	2.01	2.6	1.75	2.7	1.00	1.9	2.53	1.42	2.1	2.2	2.2	2.73	
	B	2.5	3.29	9	1.1	0.9	5	0	5	2	2.01	2.6	1.75	2.7	1.00	1.9	2.53	1.42	2.1	2.2	2.2	2.73	
	3	0.1	-	0.8	0.9	1.08	0.9	82.2	59.1	54.6	2.0	2.18	2.8	2.47	2.9	2.2	2.04	2.2	2.10	2.6	2.94	2.94	
Station Noise	0d	-	-	0.5	1.9	1.27	1.1	94.7	77.9	73.7	1.4	1.75	2.4	1.61	2.3	1.57	1.7	2.33	1.33	1.8	2.36	2.36	
	B	1.1	3.51	1	1.27	1.1	2	9	5	7	1.75	2.4	1.61	2.3	1.57	1.7	2.33	1.33	1.8	2.36	2.36		
	3	0.6	-	2.4	1.8	1.19	1.0	77.7	44.2	39.8	2.1	2.19	2.8	2.71	2.9	2.3	2.20	2.33	2.2	2.4	2.9	2.97	
Street Noise	0d	-	-	0.2	2.1	1.37	1.2	94.7	55.3	55.4	1.3	1.76	2.4	1.54	2.4	1.19	2.1	2.46	1.25	2.0	2.44	2.44	
	B	2.0	2.75	5	1.37	1.2	2	9	7	4	1.76	2.4	1.54	2.4	1.19	2.1	2.46	1.25	2.0	2.44	2.44		
	3	0.0	-	1.0	1.1	1.08	0.9	68.4	49.2	49.3	2.0	2.14	2.7	2.49	2.9	2.2	2.18	2.2	2.13	2.5	2.94	2.94	
Train Noise	0d	-	-	1.0	1.9	1.32	1.1	81.5	53.7	53.2	1.0	2.09	2.7	1.60	2.6	1.1	2.1	2.50	1.3	1.1	2.3	2.75	
	B	1.5	1.96	4	1.32	1.1	7	6	4	7	2.09	2.7	1.60	2.6	1.1	2.1	2.50	1.3	1.1	2.3	2.75		
	3	0.9	-	3.0	1.3	1.16	1.0	70.4	41.9	41.8	1.9	2.13	2.7	2.34	2.9	2.2	2.13	2.6	2.2	2.5	2.94	2.94	
Train Noise	0d	-	-	4.1	1.2	0.87	0.7	55.2	37.3	37.2	2.2	2.51	3.1	2.96	3.5	2.6	2.57	2.2	2.2	2.9	3.43	3.43	
	B	3.8	1.16	6	0.87	0.7	2	0	6	4	3	2.51	3.1	2.96	3.5	2.6	2.57	2.2	2.2	2.9	3.43	3.43	
	3	3.8	1.16	6	0.87	0.7	2	0	6	4	3	2.51	3.1	2.96	3.5	2.6	2.57	2.2	2.2	2.9	3.43	3.43	

D. Perceptual Evaluation Of Speech Quality (PESQ):

One among the objective quality measures which provides an accurate speech quality recommended by ITU_T [11] which involves more complexity in computation. A linear combination of average asymmetrical disturbance A_{ind} and average disturbance D_{ind} is given by PESQ.
 $PESQ = 4.754 - 0.186 D_{ind} - 0.008 A_{ind}$ (21)

E. Composite Measures: Linear combination of existing objective quality measures results in a new measure[10].This can be evaluated by using linear regression analysis.This paper uses the multiple linear regression analysis to obtain the following new composite

measures[10].These composite measures were measured on a five-point scale.

(i)Signal Distortion(C_{sig}):The linear combination of PESQ,LLR and WSSD measures results in anew composite measure named as Signal Distortion[10].This is evaluated using the following equation

$$C_{sig} = 3.093 - 1.029 * LLR + 0.603 * PESQ - 0.009 * WSSD \quad (22)$$

Table 4: Objective quality measures Segmental SNR(seg_SNR), Log Likelihood Ratio(LLR), Weighted Slope

Slope Spectral Distance (WSSD), Perceptual Evaluation of Speech Quality(PESQ), Signal Distortion(C_{sig}), Background intrusiveness(C_{bak}), Overall Quality(C_{ovl}) for Spectral Subtraction(SS), Multi Band Spectral Subtraction(MBSS) and Proposed Method(PM)

(ii) Noise intrusiveness(C_{bak}): The linear combination of PESQ, seg-SNR and WSSD measures results in a new composite measure named as noise Distortion[10]. This is evaluated using the following equation.

$$C_{bak} = 1.634 + 0.478 * PESQ + 0.007 * WSSD + 0.063 * seg_SNR \quad (22)$$

(iii) Overall Quality(C_{ovl}): Overall Quality is formed by Linear combination of LLR

PESQ and WSSD measures and is given by

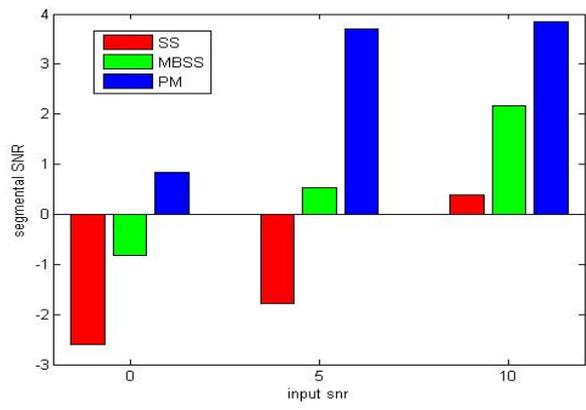
$$C_{ovl} = 1.594 + 0.805 * PESQ - 0.512 * LLR - 0.007 * WSSD \quad (23)$$

Table 1. Scale of Signal Distortion	
5	No degradation, very natural
4	little degradation, fairly natural
3	Somewhat degraded, somewhat natural
2	fairly degraded, fairly unnatural
1	very degraded, very natural

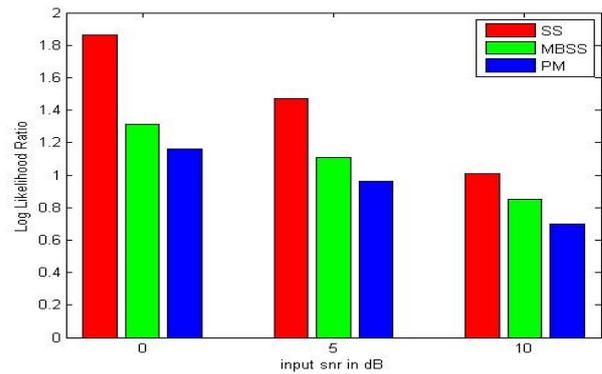
Table 2. Scale of Background Intrusiveness	
5	not observable
4	somewhat observable
3	observable but not intrusive
2	Somewhat intrusive
1	Very intrusive

Table 3. Scale of Overall quality	
5	excellent
4	good
3	fair
2	poor
1	bad

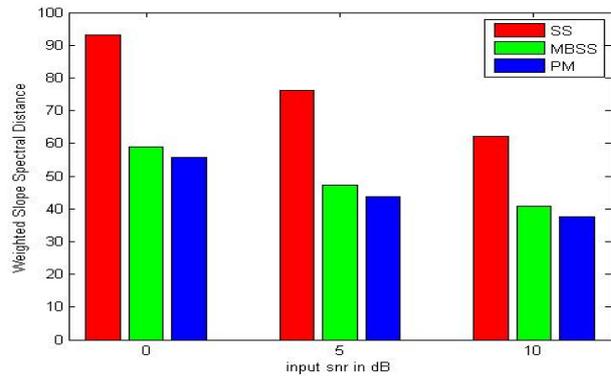
To obtain objective quality measures for the proposed method first the signal is windowed into 20 ms duration with 50% overlapping. Then the spectrum is divided into 4-bands with non-overlapping frequencies with over subtraction factor μ_j and spectral floor parameter $\phi = 0.002$. We can select the number of bands from 1 to 8 as suggested in [] but we can observe the markable improvements in terms of objective quality measures when the bands are increased from 1 to 4. Then the noise power spectrum is obtained by adaptive noise estimation algorithm discussed in section III by choosing smoothing constants $\beta = 0.96$ and $\gamma = 0.998$.



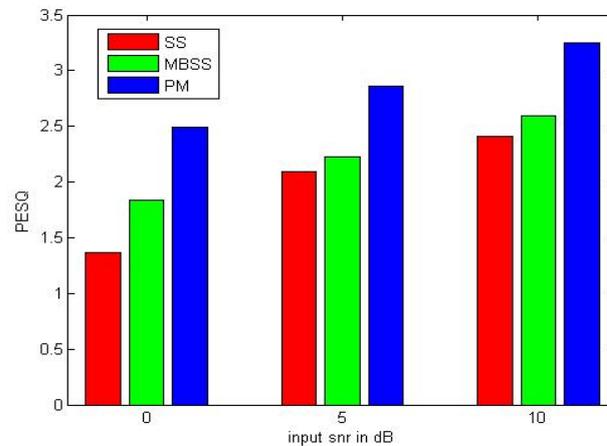
(a)



(b)



(c)



(d)



Multi Band Spectral Subtraction Combined with Adaptive Noise Estimation

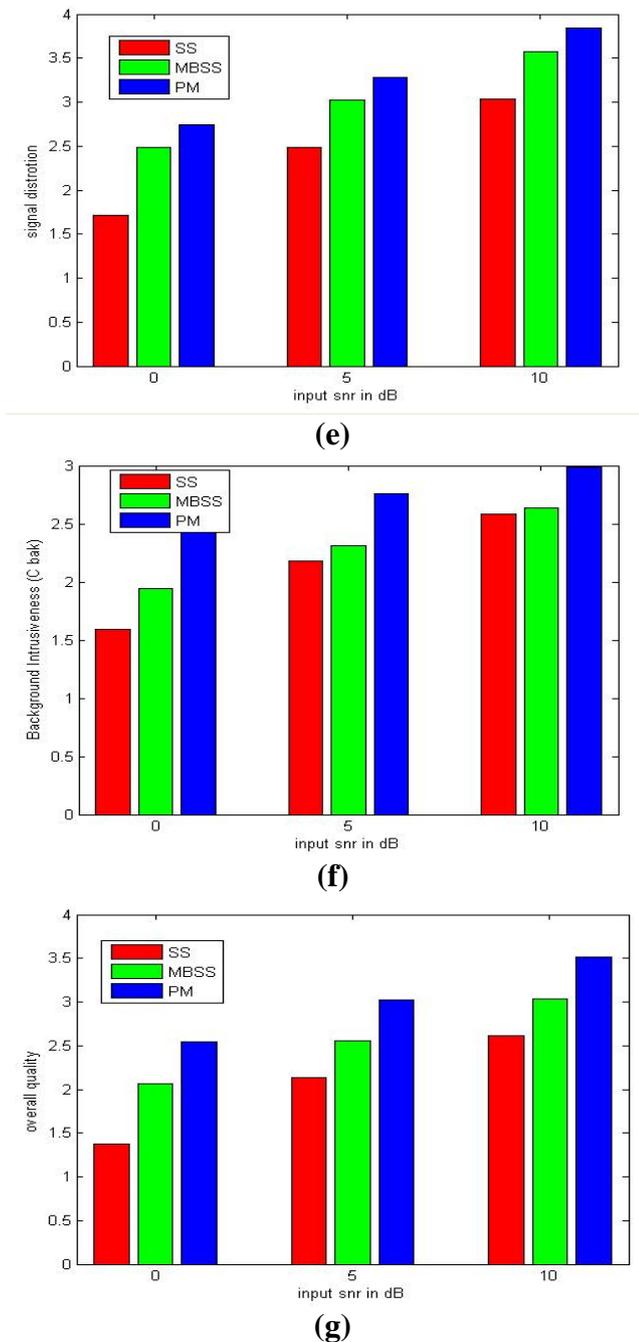


Figure 2: a)Segmental SNR b)Log Likelihood Ratio c)Weighted spectral slope distance d)PESQ e)Signal Distortion (C_{sig}) f)Background intrusiveness (C_{bak}) g) Overall quality (C_{ovl}) measures against input SNR.

Table 1. shows the objective quality measures for spectral subtraction [12], Multi Band Spectral Subtraction[13] and proposed method for three different levels of input SNR. The same can be observed in the form of bar graphs by taking the average over eight different types of noises in Fig.2.

From figure 2. It can be observed that the proposed method gives higher values of seg_SNR, PESQ, composite measures and lower values of Log Likelihood Ratio and Weighted Slope Spectral Distance.

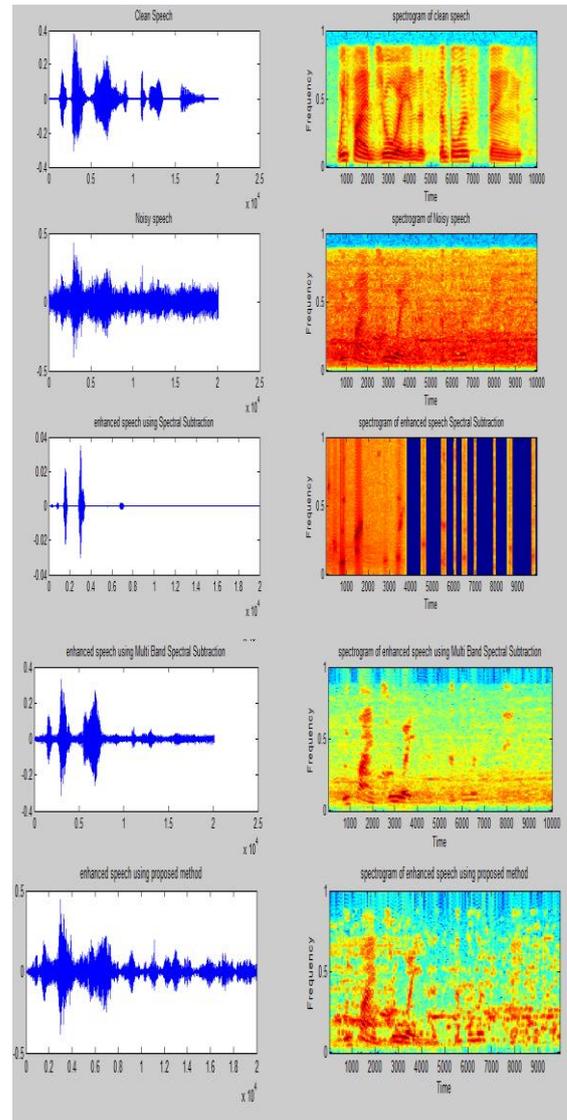


Fig.3. Time domain and spectrogram representation of Clean Speech noisy speech and enhanced speech signals by three different methods.

Spectrograms are widely used in speech processing to plot the spectrum of frequencies as it varies with time. The spectrogram can be evaluated as a sequence of FFTs computed over a windowed signal of duration of 20ms. From the above figure it can be observed that spectrogram of noisy speech appears in red colour. Clean speech consisting of both red, blue and yellow colour represents the voiced and unvoiced sounds. The spectrogram of proposed method is nearer to the original clean speech when compared to others. From the spectrogram of spectral subtraction method it can be observed that random spectral peaks introduces musical noise which can be reduced by Multi Band Spectral Subtraction method and further improvement can be achieved by proposed method.

REFERENCES:

1. Sohn, NS, Kim, and W. Sung, "A Statistical model-based Voice Activity Detection", IEEE Signal Processing Letters, 6(1), Pg.No:1-3, 1999.
2. YD. Cho, A. Kondo, "Analysis and improvement of a Statistical

- model-based Voice Activity Detector”, IEEE Signal Processing Letters,8(10), Pg.No:276-278,2001
3. J.Ramirez,J.C.Segura,C.Benitez,”Statistical Voice Activity detection using a multiple observation likelihood ratio test”, IEEE Signal Processing Letters,12(10), Pg.No :689-692,2005.
 4. S.Graf,T.Hrrbig,M.Buck,”Features for Voice Activity Detection:A Comparative Analysis”, EURASIP Journal on Advances in Signal Processing,2015(1),1-15,2015.
 5. R.Martin, “Noise Power Spectral Density Estimation based on Optimal Smoothing and Minimum statistics”, IEEE Transactions on Audio, Speech Processing 9 (5), pp.504–512, 2001.
 6. I.Cohen, “Noise Estimation by Minima controlled recursive averaging for robust speech enhancement ”, IEEE Signal Processing. Letter 9 (1), pp.12–15,2002
 7. Loizou, R.Sundarajan,Y. Hu,”Noise estimation Algorithm with rapid Adaption for highly non-stationary environments “,Proceedings on IEEE International Conference on Acoustic Speech Signal Processing,2004.
 8. .P.C.Loizou, R.Sundarajan,”A Noise estimation Algorithm for highly non-stationary Environments”. Speech Communication,48 ,Science Direct ,pp.220-231,2006.
 9. A Noisy Speech Corpus for Assessment of Speech Enhancement Algorithms. <https://ecs.utdallas.edu/Loizou/speech/noizeous>
 10. Yi.Hu ,P.C .Loizou,,”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.
 11. .ITU_T Rec ,”Perceptual evaluation of speech quality(PESQ),An objective method for end to end speech quality assessment of narrowband telephone networks and speech codecs”.,International Telecommunications Union ,Geneva Switzerland, February 2001.
 12. .Boll,S.F,”Suppression of acoustic noise in speech using spectral subtraction”. IEEE Transactions on Acoustics Speech and Signal Processing, 1979,27(2), 113–120.
 13. 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.

AUTHORS PROFILE



P. Sunitha received B.Tech Degree in E.C.E from JNTU, Kakinada, India in 2002 and M.Tech degree in Digital Electronics and Communication Systems from JNTU, Kakinada, India. She is currently a research scholar with the Department of E.C.E, JNTU, Kakinada, India. Her research interest is Speech Processing.



Dr. K. Satya Prasad received his B Tech. degree in Electronics and Communication Engineering from JNTU college of Engineering, Anantapur, India in 1977 and M. E. degree in Communication Systems from Madras University, Chennai , India in 1979 and Ph D from Indian Institute of Technology, Madras in 1989. He is currently working as Rector,

VFSTR,Guntur,India. His areas of Research include Communications Signal Processing, Image Processing, Speech Processing, Neural Networks & Ad-hoc wireless networks etc .He joined JNT University, Hyderabad as Lecturer in 1980 and served in different constituent colleges viz., Kakinada, Hyderabad and Anantapur and at different capacities viz., Associate Professor, Professor, and Head of the Department, Vice Principal and Principal. He has published more than 50 technical papers in different National & International conferences and Journals and Authored one Text book.