ISSN 2321 3361 © 2019 IJESC

Research Article Volume 9 Issue No.11

Speech Enhancement of a Mobile Car-Noisy Speech Using Spectral Subtraction Algorithms

Afolabi, Lateef Olashile[1], Olaiya, Olayinka Oluwaseun[2]
Department of Electrical and Electronic Engineering[1], Department of Computer Engineering[2]
Federal Polytechnic Offa, Kwara State, Nigeria[1]
The Federal Polytechnic Ilaro, Ogun State, Nigeria[2]

Abstract:
Problem with speech are background noise from the environment and internal generated noise. Speech enhancements involve reduction of all these noise to a minimum level so has to increase the signal to noise ratio (SNR) without losing the quality and the content of the signal. Many algorithms are used to enhance speech signals but most are not as effective as Spectral subtraction algorithms. Hence, this paper is on the use of Spectral subtraction algorithms to reduce noise generated by mobile car, very close to the speech source. Different sample of polluted noisy speech signal were sample at -5dB, 5dB, 10dB, and so on, Spectral Subtraction noise reduction technique has a limit -5.0000dB to 18.0670dB SNR value for Car Noisy Speech Signal.

Keyword: Speech-enhancement, Spectral-subtraction, DFT, Noise signal, VAD.

I. INTRODUCTION

Speech, in general, is subject to noise and distortions by transducer such asmicrophone, loud speaker, and so on; transmission channelsuch astelephone wire, optical fiber, radio waves, and so on;and ambience orbackground noise such as industrial noise, environmental noise, terrestrial noise, and so on.

These degradations lower the intelligibility of the speech message thereby decreasing the quality and signal to noise ratio. Hence, to overcome, all these effect, there is need for speech processing before the usage. Speech processing entails measuring, filtering, producing and compressing analog speech signals in order to achieve discrete representation and digital processing of the analog speech signals. All these are to achieve a more quality and speech free from distortion with least noise level. Speech enhancement is the term used to describe this process. It is a process of increasing signal to noise ratio by reducing the noise level in the wanted signal. The noise can be either thermal noise, white, shot noise, and so on, which is called the internal generated noise, and industrial noise and terrestrial noise are called external generated noise. In literature, different methods have been used to improve the speech quality in different medium based on the fact that different media and their associated environmental noise affect speech differently [1, 2&3]. Hence, different signal processing approaches have been used to reduce the effect of noise on speech such that it will not affect the speech in term of quantity or quality with little success. The proposed approach tends to solve the issue by monitoring both quantity and quality of speech in the speech enhancement algorithm such that at the end the result can be preempted. Presently, many systems use speech processing applications as a result of the increased demand for automatic speech recognition, speaker recognition, verification and authentication packages and the recent technological advancement in the field of mobile communications and security threat [2]. In many speech processing applications, rapid degradation in performance of speech signal quality is often observed in noisy environments due to the resultant low SNR [3]. In order to address this challenge, noise reduction algorithms were proffered to improve or enhance the intelligibility of the noisy speech signal. There exist many noise reduction techniques but prominent among them is Wiener and Kalman filtering approach. However, different noise has different effects on speech signals [4]. Hence there is a need for an enhanced noise reduction algorithm that will be effective for most noisy environments at specified signal to noise ratio. This paper proposes an improved spectral subtraction algorithm deployed for a mobile car noise on a speech signal at different distance and noise level.

II. METHODS

In spectral subtraction, input signal is segmented into frames and multiplied with hamming window. We obtain the Discrete Fourier Transform (DFT) of these frames and separate magnitude and phase from speech. The noise power estimation and computation of spectral weighting takes place on magnitude. Once the noise estimate is subtracted from speech spectrum magnitude, it is recombined with the original phase of the noisy signal. Inverse Fourier Transform (IDFT) is taken and outputs are obtained by overlap and add method [3, 5]. The following assumptions were used in developing the analysis.

The background noise is acoustically or digitally added to the speech. The background noise environment remains locally stationary to the degree that its spectral magnitude expected value just prior to speech activity equals its expected value during speech activity. If the environment changes to a new stationary state, there exists enough time (about 300 ms) to estimate a new background noise spectral magnitude expected value before speech activity commences. For the slowly varying non-stationary noise environment, the algorithm requires a speech activity detector to signal the program that speech has ceased and a new noise bias can be estimated. Finally, it is assumed that significant noise reduction is possible by removing the effect of noise from the magnitude spectrum.

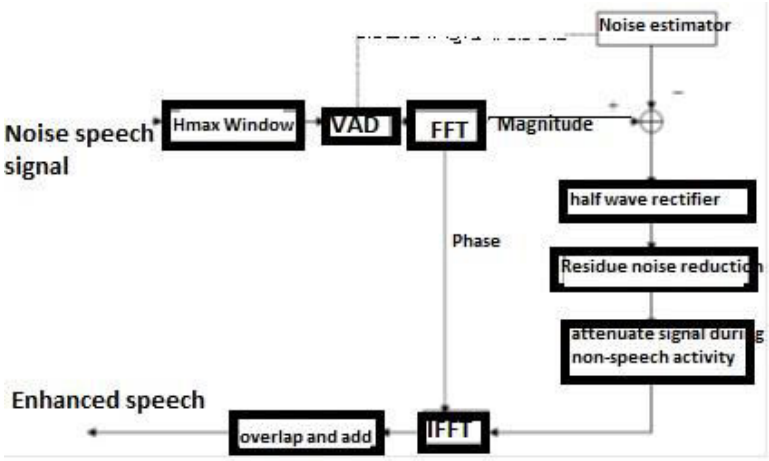


Figure.1.0. Block Diagram of the General representation of Spectral Subtraction [5]

From Fig.1.0 above, the noisy speech signal $G(n)$ which comprises of both the clean and noise signal is framed and windowed, each frame which is 20ms in length is simultaneously processed by both the VAD and DFT blocks, the voice activity detection (VAD) block is a binary classifier, that is the output can either be 1 or 0, the output is 1 if the frame contains speech samples and 0 otherwise. The DFT block converts each frame to its frequency domain to obtain the spectral magnitudes and phase. The non-speech segments or frames from the VAD block are used to determine the noise estimate, the noise estimate or magnitude from the noise estimation block is subtracted from the spectral magnitudes. The difference is then combined with the phase obtained from the DFT block to generate the enhanced speech spectral magnitudes, the output magnitude are then processed by the IDFT to obtain the enhanced speech signal in time domain. Assume that a windowed sequence $t(n)$has been added to a windowed speech signal $y(n)$, with their sum denoted by $g(n)$.

Then,

$g(n) = y(n) + t(n)$ (1)

Taking the Fourier transform gives

$G(e^{j\omega}) = Y(e^{j\omega}) + T(e^{j\omega})$ (2)

Where, $g(n) \leftrightarrow G(e^{j\omega})$

$G(e^{i\omega}) = \sum_{n=o}^{k-1} g(n)e^{-i\omega n}$ (3)

Where, k is the window length

$g(n) = \frac{1}{2\pi}\int_{-\pi}^{+\pi} G(e^{i\omega})e^{-i\omega n}d\omega$ (4)

Speech which is "contaminated" by noise can be expressed as

What spectral subtraction attempts to do is to estimate $Y(n)$ from $G(n)$.

If the noise process is represented by its power spectrum estimate $|\widehat{T}(w)|^2$, that of the noisy speech is $|G(w)|^2$, the power spectrum of the clean speech estimate $|\hat{Y}(w)|^2$ can be written as

$|\hat{Y}(w)|^2 = |G(w)|^2 - |\widehat{T}(w)|^2$ (5)

Since the power spectrum of two uncorrelated signals is additive. The clean speech phase $\Theta Y(w)$ is estimated directly from the noisy speech signal phase $\Theta G(w)$.

$\Theta Y(w) = \Theta G(w)$. (6)

Thus a general form of the estimated speech in frequency domain can be written as

$\hat{Y}(w) = \left(\max(|G(w)|^2 - k|\widehat{T}(w)^2 , 0)^{\frac{1}{2}}\right) . e^{j\Theta G(w)}$ (7)

Where k > 1 is used to overestimate the noise to account for the variance in the noise estimate. The inner term $(|G(w)|^2 - k|\widehat{T}(w)|^2$ is limited to positive values, since it is possible for the overestimated noise to be greater than the current signal. The spectral subtraction filter $H(e^{j\omega})$ is calculated by replacing the noise spectrum $N(e^{j\omega})$ with spectra which can be readily measured. The Magnitude $|T(e^{j\omega})|$ of $T(e^{j\omega})$ is replaced by its average value $\mu(e^{j\omega})$ taken during non-speech activity, and the phase $\theta_N(e^{j\omega})$ on $N(e^{j\omega})$ is replaced by the phase $\theta_x(e^{j\omega})$ of $G(e^{j\omega})$. These substitutions result in the spectral subtraction estimator$Y(e^{i\omega})$

$\bar{Y}(e^{i\omega}) = [|G(e^{i\omega})| - \mu(e^{i\omega})]e^{i\theta_x(e^{j\omega})}$ (8)

$\bar{Y}(e^{i\omega}) = H(e^{i\omega}) \times G(e^{i\omega})$ (9)

Within

$H(e^{i\omega}) = 1 - \frac{\mu(e^{i\omega})}{|G(e^{i\omega})|}$ (10)

$\mu(e^{i\omega}) = E\{|T(e^{i\omega})|\}$ (11)

The spectral error $\varepsilon(e^{i\omega})$resulting from this estimator is given by

$\varepsilon(e^{i\omega}) = \bar{Y}(e^{i\omega}) - Y(e^{i\omega}) = T(e^{i\omega}) - \mu(e^{i\omega})e^{i\theta_x(e^{j\omega})}$ (12)

A number of simple modifications are available to reduce the auditory effects of the spectral error. These include: a) magnitude averaging; b) half-wave rectification; c) residual noise reduction; and d) additional signal attenuation during non-speech activity.

Since the spectral error equals the difference between the noise spectrum N and its mean μ, local averaging of spectral magnitudes can be used to reduce the error. Replacing $|X(e^{j\omega})|$ with $\overline{G(e^{i\omega})}$where

$G(e^{i\omega}) = \frac{1}{M}\sum_{i=0}^{M-1}|G_i(e^{i\omega})|$ (13)

M is the number of frames over which averaging is done $G_i(e^{i\omega})$is the transform of the ith window of $g(n)$ $\overline{G(e^{i\omega})}$=*ith time-windowed transform of g(n)*

Gives

$Y_A(e^{i\omega}) = \left[|G(e^{i\omega})| - \mu(e^{i\omega})\right]e^{i\theta_x(e^{j\omega})}$ (14)

The rationale behind averaging is that the spectral error becomes approximately

$\varepsilon(e^{i\omega}) = Y_A(e^{i\omega}) - Y(e^{i\omega}) \cong |\bar{T}| - \mu$ (15)

Where,

$\left|\overline{T(e^{i\omega})}\right| = \frac{1}{M}\sum_{i=0}^{M-1}|T_i(e^{i\omega})|$ (16)

Thus, the sample mean of $|T(e^{j\omega})|$ will converge to $\mu(e^{j\omega})$ as a longer average is taken. Wherever the signal spectrum magnitude $|G(e^{j\omega})|$ is less than the average noise spectrum

magnitude $\mu(e^{j\omega})$, the output is set to zero. This modification can be simply implemented by half-wave rectifying $H(e^{j\omega})$. The estimator then becomes

$$Y(e^{i\omega}) = H_R(e^{i\omega}) \times G(e^{i\omega}) \qquad (17)$$

$$H_R(e^{i\omega}) = \frac{H(e^{i\omega}) + |H(e^{i\omega})|}{2} \qquad (18)$$

The input-output relationship between $G(e^{j\omega})$ and is shown in Fig

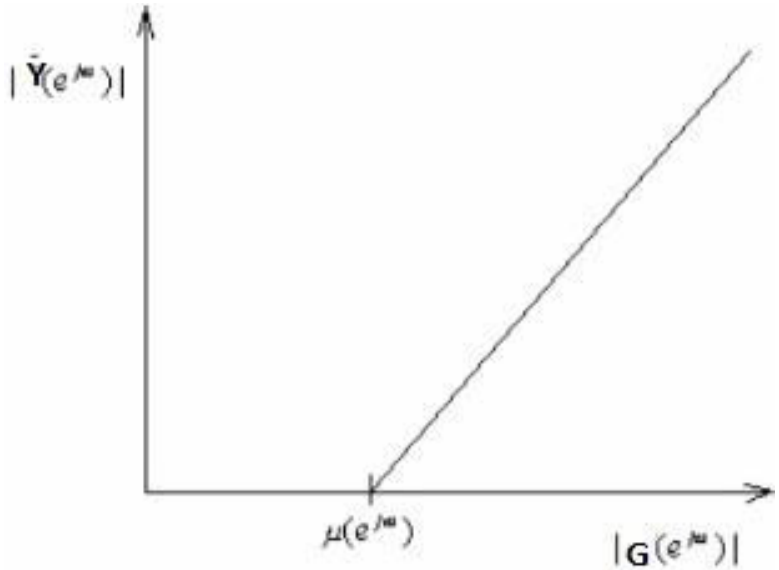
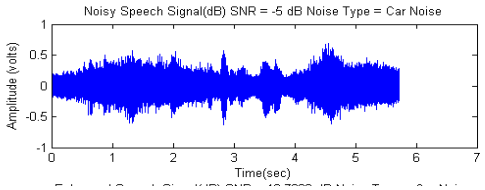
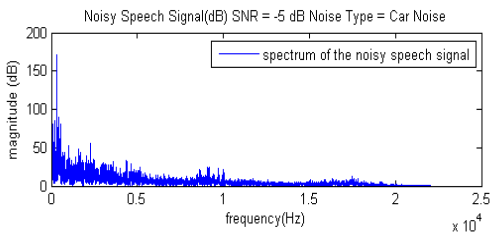
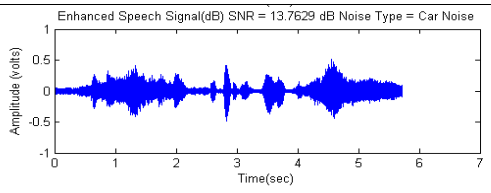
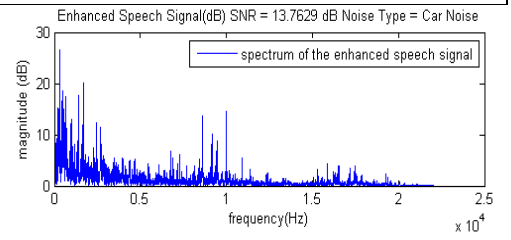
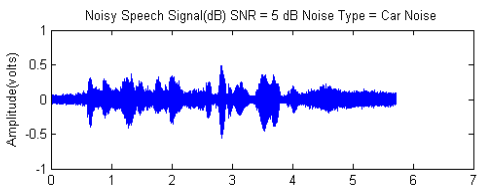
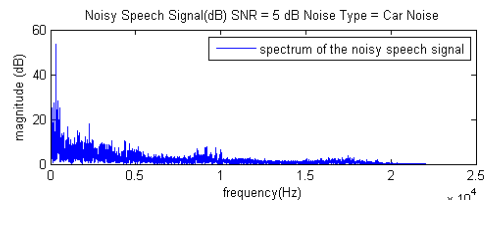


Figure.2.0. Input- Output relation between $Y(e^{i\omega})$ and $G(e^{i\omega})$

Thus, the effect of half-wave rectification is to bias down the magnitude spectrum at each frequency ω by the noise bias determined at that frequency. The bias value can, of course, change from frequency to frequency as from analysis time window to time window. The advantage of half-wave rectification is that the noise floor is reduced by $\mu(e^{j\omega})$. Any low variance coherent noise tones are also essentially eliminated. The disadvantage of half-wave rectification can exhibit itself in the situation where the sum of the noise plus speech at a frequency ω is less than $\mu(e^{j\omega})$. Then the speech information at that frequency is incorrectly removed, implying a possible decrease in intelligibility.

III. RESULTS AND DISCUSSION

Table.1.0. Speech signals of a noisy speech and enhanced speech in both time domain and frequency domain at different Signal to Noise Ratio (SNR)

SNR	Time domain	Frequency domain
-5dB noisy speech signal		
-5dB enhanced speech signal		
5dB noisy speech signal		

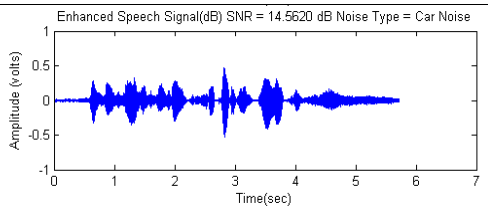
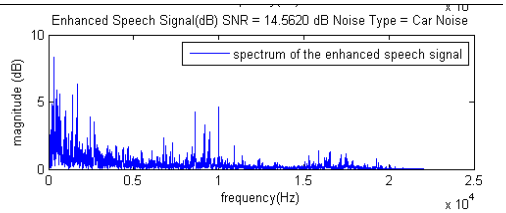
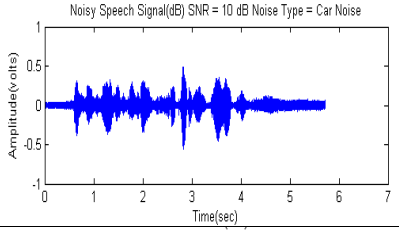
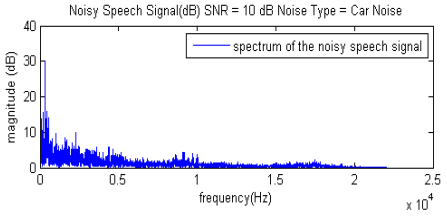
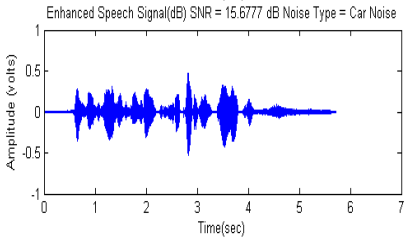
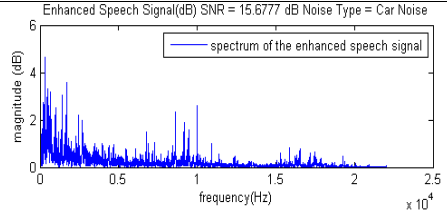
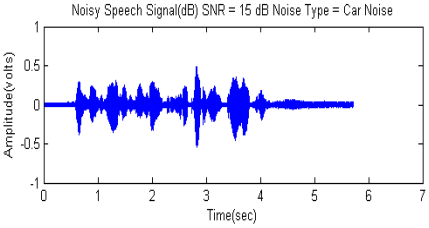
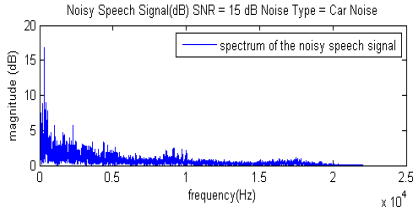
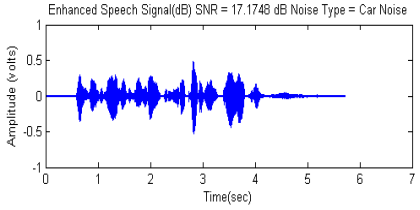
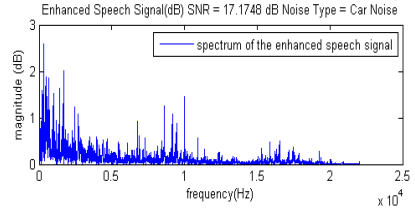
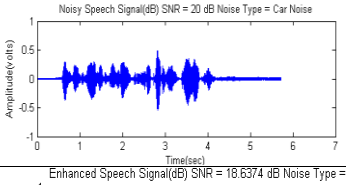
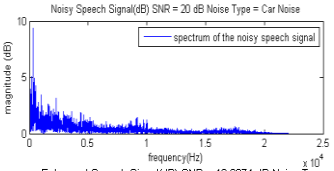
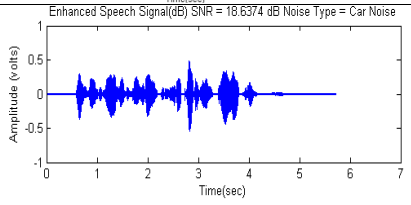
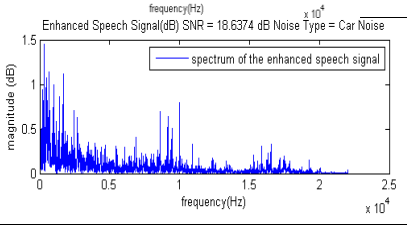
5dB enhanced speech signal		
10dB noisy speech signal		
10dB enhanced speech signal		
15dB noisy speech signal		
15dB enhanced speech signal		
20dB enhanced speech signal		
20dB noisy speech signal		

Table.2.0. Shows the Signal to Noise Ratio (SNR) Values of both the Noisy and Enhanced Speech Signals for Crowd

Signal to Noise Ratio of Car Noisy Speech Signal (dB)	Signal to Noise Ratio of the Enhanced Car Speech Signal (dB)
-5.0000	13.6729
5.0000	14.5620
10.0000	15.6777
15.0000	17.1748
20.0000	18.6374

At -5dB, 5dB, 10dB, and 15dB there exist a tremendous improvement in the signal to noise ratio, and the result are 13.7629dB, 14.5620dB, 15.6777dB, and 17.1748dB respectively. It was noticed that at 20dB and above there was deterioration in the signal to noise ratio i.e. at 20dB the SNR result value gives 18.6374dB.

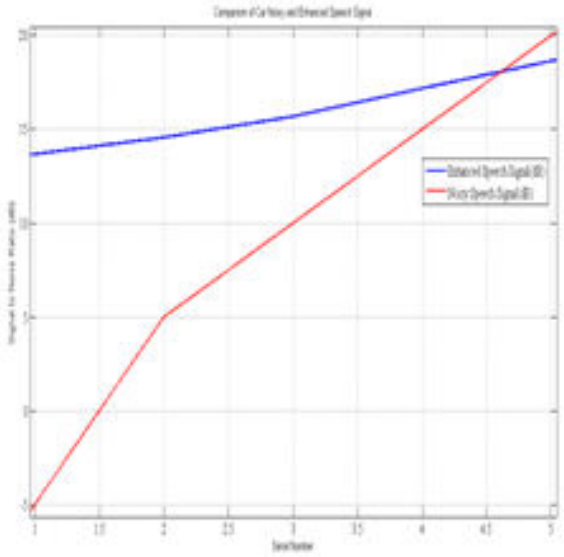


Figure.3.0. Comparison of both the Car Noisy and Enhanced Speech Signals

Fig 3.0 shows a graph that reveals the signal to noise ratios for Car Noisy and Enhanced Speech Signals. The upper limit of SNR for the noise signals can be deduced from the graph which is given by the points of intersection between the noisy speech signals and the enhanced speech signal plots. Beyond the upper limit point there can be no further enhancement to the noisy speech signals instead it deteriorates. For Car Noisy Speech Signals, the upper limit point is 18.0670dB, Therefore any SNR value above 18.0670dB will result in the deterioration of the Car noisy speech signals

IV. CONCLUSION

This paper has presented a noisy speech enhancement technique based on spectral subtraction filtering algorithm. It can be concluded that the Spectral Subtraction algorithm is efficient, fast and reliable for Car Noisy Speech Signals. But Spectral Subtraction noise reduction technique has a limit -5.0000dB to 18.0670dB SNR value for Car Noisy Speech Signal.

V. REFERENCES

[1].D.A.Enotes,(www.daenotes.com/electronics/communication -system/noise), accessed date: 15th April, 2015. 3.16pm

[2]. Y. Ephraim and D. Malah "Speech enhancement using a minimum mean-square error log spectral amplitude estimator *IEEE Trans. Acoustic Speech and Signal Processing*, vol. ASSP-33, pp. 443-445, Apr. 1985.

[3]. Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator*IEEE Trans. Acoustic Speech and Signal Processing,* vol. 32, pp. 1109-1121, Dec. 1984.

[4]. J. Sohn, N.S Kim and W. Sung "A statistical model-based voice activity detector," *IEEE Signal Process. Letters*, vol. 6, no. 1, Jan. 1–3 1999.

[5]. M. T. Sadiq, N. Shabbir and W. J. Kulesza "Spectral Subtraction for Speech Enhancement in Modulation Domain" JCSI International Journal of Computer Science Issues, Vol. 10, Issue 4, No 2, July 2013.

[6]. M Chugani and C Schuler, "Digital Signal Processing, a hand on approach" the Mcgraw Hills companies, International Editions, 2000.

[7]. A.V. Oppenheim, R. W. Schaffer, J. R. Buck, "Discrete-Time Signal Processing", Prentice Hall, 2nd edition. Upper Saddle River, NJ: *Prentice Hall*, 1999

[8]. N.S Kim and J.H Chang "Spectral enhancement based on global soft decision," *IEEE Signal Processing Letters*, vol. 7, pp. 108-110, May 2000.

[9]. K. R,Siddhart, B.Tech Thesis on Speech Enhancement, 2002

[10]. S. F. Boll, "Suppression of Acoustic Noise in Speech Using Spectral Subtraction", *IEEE Trans. on Acoustics, Speech, and Signal Processing*, Vol. ASSP-27,No.2 , April 1979, pp.113-120.

[11]. G. S. Kang, L. J. Fransen, "Quality Improvement of LPC- Processed Noisy Speech by Using Spectral Subtraction", *IEEE Trans. on Acoustics, Speech, and Signal Processing*, Vol. 37, No. 6, June 1989.

[12]. S.C. Venkateswarlu, K.S Prasad, A.S. Reddy, "Improve Speech Enhancement Using Weiner Filtering" In Global Journal of Computer Science and Technology Volume 11 Issue 7 Version 1.0 May 2011.

[13]. N. S. Banale, S. K. Sudhansu, S. M. Jagde, "Noise Free Speech Enhancement based on Fast Adaptive Kalman Filtering Algorithm" MBES College of Engineering

[14]. B. Chabane, B.D lamel, "On the use of Kalman Filter for Enhancing Speech Corrupted by Colored Noise", Faculte des Sciences de l'Ingenieur, Universite de Jijel., BP. 98, OuledAissa, 18000, Jijel, Algerie.

[15]. K.K. Paliwal and AnjanBasu, "A speech enhancement method based on Kalman filtering". In ICASSP, number 6.3, pages 1-4, 1987.