Analysis of the SNR Estimator for Speech Enhancement Using a Cascaded Linear Model

Harjeet Kaur
Ph.D Research Scholar
I.K.Gujral Punjab Technical University
Jalandhar, Punjab, India

Rajneesh Talwar
Principal,Professor
CGC Technical Campus, Jhanjeri, Mohali
Punjab, India

Abstract—**Elimination of tainted noise and improving the overall quality of a speech signal is speech enhancement. To gain the advantage of individual algorithms we propose a new linear model and that is in the form of cascade adaptive filters for suppression of non-stationary noise. We have successfully deployed NLMS (Normalized Least Mean Square) algorithm, Sign LMS (Least Mean Square) and RLS (Recursive Least Square) as the main de-noising algorithms. Moreover, we are successful in demonstrating that the prior information about the noise is not required otherwise it would have been difficult to estimate for fast-varying noise in non-stationary environment. This approach estimates clean speech by recognizing the long segments of the clean speech as one whole unit. During experiment/implementation we used in-house database (includes various types of non stationary noise) for speech enhancement and proposed model results have shown improvement over conventional algorithms not only in objective but in subjective evaluations as well. Simulations present good results with a new linear model that are compared with individual algorithm results.**

Keywords—Least Mean Square (LMS); Normalized Least Mean Square (NLMS); Recursive Least Square(RLS); Speech Enhancement; Non- stationary

I. Introduction

The goal of speech enhancement is to improve the quality and intelligibility of speech that has been degraded by noise [11]. The speech varies according to the needs of specific applications, such as to increase the overall speech quality or intelligibility. In real-life contexts, there are wide variety of situations in which we need to enhance speech signals. Speech enhancement techniques have been successfully applied to problems as diverse as correction of disrupted speech due to pathological problems of the speaker, pitch and rate modification, restoration of hyperbaric speech, and correction of reverberation, but, noise reduction is probably the most important and most frequently studied Issues [19]. Prior to designing algorithms that cope with adverse conditions, it is crucial to understand the noise characteristics and the differences between the noise sources in terms of temporal and spectral characteristics. Noise can be impulsive, continuous, and periodic and its amplitude may vary in frequency range [19]. In previous literature various speech enhancement techniques are given for noise reduction [13]. Some are given as spectral subtraction, modified spectral subtraction, wiener and gain based method like MMSE STSA, log-MMSE, P-MMSE [15] etc. All methods give improvements in speech quality but do not improve intelligibility up to satisfactory levels [20]. In the real environment speech may be distorted by more than one noise source. Most of the time it is not possible to consider only single noise in communicated speech or number of noises in speech signal [20].Most noise estimation algorithms work well for stationary or slowly varying noise, but very few are working for non stationary noise. This is because of the weak predictability of fast-varying noise. Because of the non stationary nature of the speech signal, most current enhancement algorithms operate on a frame-by-frame basis [16]. Many algorithms ignore the temporal constraints between adjacent speech frames. Without context, and without specific knowledge about the noise, it is difficult to separate the speech from noise in the duration of a frame (typically about 20 ms). This is especially true when the noise is a form of speech (e.g., a crosstalk sentence). Previous research has revealed the importance of imposing cross-time spectral constraints in improving speech enhancement quality [16].In many such cases only a single-channel speech signal is available. Of the available solutions to the single-channel speech enhancement problem, Short Time Fourier Transforms (STFT) based methods achieve relatively good performance and compromise the majority [15]. It is appropriate to further categorize this class of speech enhancement algorithms into the sub-categories of spectral subtraction, wiener filter and statistical approaches [15]. Wiener filter is used for linearity whereas spectral subtraction is used for simplified mathematical expressions [15] .Almost all the papers work on speech enhancement with added known amount of noise and then use their proposed algorithm to enhance the speech or reduce the noise level. In this case, mostly noise is assumed to be white, Gaussian noise and colored noise [11]. However, if one records speech on the road or in the market, there is no guarantee that the noise is Gaussian. For this enhancement algorithm to be really useful, it must improve the quality of speech that was originally noisy due to some environmental conditions like railway station, fan, vehicle, machine gun, tank, factory etc that create distortions in clean speech signal and not due to explicit addition of noise by the researcher [13]. However, we have to bear this constraint in mind that the enhancement must be prominent in both quantitative as well as qualitative manner and at the same time, we should not be overlooking the complexity and the ease with which these algorithms can be implemented on hardware platforms [10].In this paper, we aim to reduce the requirement of priori information about the noise as this can be difficult to estimate with fast varying noise [16]. We propose a class of two stage adaptive architecture to address some properties of non-stationary noise by calculating energy for original speech and then calculate energy for processed speech with SNR, MSE. In the speech enhancement process, the estimation of the a priori signal-to-noise ratio (SNR), Mean square error (MSE), energy, Power Spectral Density (PSD) is one of the most important parts, especially in non-stationary environments [5].

This paper is divided into 6 sections. Section 1 gives the overview of speech enhancement. Section 2 is all about the basic principle of suppression of non stationary noise. Section 3 contains details about the Least Mean Square Algorithm, Normalized Least Mean Square Algorithm and Recursive Least Square Algorithm. The methodology, set up for problem solution presented in section 4. All the details about the data set used for analysis and results presented in section 5. At the end discussion and conclusion are included in section 6, respectively. In this paper, we study the problem of retrieving speech from non stationary noise assuming minimal noise prior.

II. Suppressing Non Stationary Noise

Noise can generally be classified into three major categories based on its characteristics: Stationary noise, Pseudo or Non-stationary noise and transient noise [19]. The spectral and temporal characteristics of pseudo or non-stationary noise change constantly. The task of suppressing this type of noise is more difficult than that of suppressing stationary noise [19]. Another distinctive feature of noise is their spectrum shape, particularly the distribution of noise energy in the frequency domain. For instance, most of the energy of car noise is concentrated in the low frequencies, i.e., it is low-pass in nature. In most speech enhancement methods, the estimation of power of the noise is a requirement. Fortunately, the bursty nature of speech makes it possible to estimate the noise during speech pauses [14]. Moreover, it should be mentioned that it is easier to deal with additive noise than convolutive noise [14]. For practical reasons, the estimation of the noise is almost performed in the spectral domain. Actually, spectral components of speech and noise are partially uncorrelated. Besides this, perception/hearing and psycho-acoustic models are well understood (and adapted) in the spectral domain [19]. Realizing the limitations of traditional speech enhancement methods in the presence of non stationary noise, Research efforts have been directed over the past decades to devise the new solutions. The solutions find into two categories: improvements to the noise estimators and modifications of the suppression rule [14]. The former class of methods essentially targets the limitations of Voice activation detection (VAD) based noise estimation.VAD is based on some prior knowledge of the speech signal. As discussed earlier, accurate estimation of noise spectrum would make effective with single channel speech enhancement methods in any background noise conditions. More research has been focused on improving the noise estimation. Due to the limitations of VAD, a number of methods are available for noise spectrum estimation. These methods are based on tracking some statics of power spectral values for each frequency bin over several frames. The test file has a continuous stretch of speech frames by high and low pulses in the plot. Noise is non stationary with random bursts has been detected by high and low pulses [14]. A different approach to carry out the adaptation of noise during both speech presence and absence is by estimate of SNR. Any sudden increase in the background noise level is not easily distinguished from speech and results in high estimated SNR making the method less effective in highly non stationary noise.

III. Proposed Method

As mentioned earlier, we focus on a common form of non stationary noise characterized by randomly occurring noise bursts in a stationary background. These noise bursts lead to the partial or complete corruption of the spectrum of the speech. A traditional STSA algorithms based on a simple noise estimator can effectively suppress only the stationary background noise leaving mainly the speech together with residual noise bursts in the enhanced signal. The proposed processing algorithm involves identifying regions in the spectrogram of the enhanced speech that are dominated by the residual noise. These cascaded adaptive algorithms contribute to improve the overall quality of the enhanced speech

A. *Least Mean Square Algorithms*

An adaptive filter is a filter that self-adjusts its transfer function according to an optimizing algorithm. Because of the complexity, most adaptive filters are digital filters that perform digital signal processing and adapt their performance based on the input signal [9]. Least mean squares (LMS) algorithms is such an adaptive filter used to find the filter coefficients by equation (1) that relate to producing the least mean squares of the error signal (difference between the desired and the actual signal). In stochastic gradient descent method, the adaptive filter is only adapted based on the current error [9].

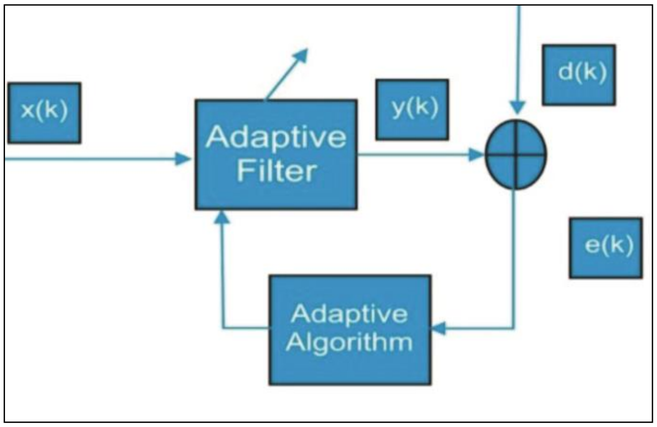

Fig. 1. Adaptive filter

$$W(k+1) = W(k) + \mu e(k) \quad (1)$$

$$e(k) = d(k) - W^H(k)X(k) \quad (2)$$

Equation (1) is used for calculating the updated weights and equation (2) is used for calculating error signal, where X(k) represents the input signal vector. The least mean square (LMS) algorithm uses the statistical properties of the signals. The main objective of this method is to minimize the mean square error. The LMS algorithm is widely used in the adaptive algorithm because of its simplicity in structure and its robustness for numerical analysis.

B. Normalized Least Mean Square (NLMS) Algorithm

NLMS is widely used algorithm because of its simplicity and robust performance. The stability of the basic NLMS is controlled by a step size. This parameter also governs the rate of convergence, speed, tracking ability and the amount of steady-state excess mean-square error (MSE) [9] aimed at solving conflicting objectives of fast convergence and low excess MSE. It achieves a certain degree of success that converges slowly with colored input signals. In the standard LMS algorithm if x (n) is large, it suffers from gradient noise amplification. But normalized LMS algorithm seeks to avoid gradient noise amplification. The step size is time varying and optimized to minimize error [9].

$$w(n+1) = w(n)1/2\mu(n)[-\nabla(n)] \qquad (3)$$

$$= w(n) + \mu(n)[p - Rw(n)] \qquad (4)$$

C. Recursive Least Square (RLS) Algorithm

Recursive algorithm is used for the design of adaptive transversal filters which provides the least square estimate of the tap weight vector of the filter at iteration (n-1) and also computes the updated estimate of the vector at the iteration n upon the arrival of new data [9]. RLS algorithms calculate J(n) by using the following equation:

$$J(n) = 1/N\sum \lambda^{i} e^{2}(n-1) \qquad (4)$$

In this Equation where N is the filter length and λ is the forgetting factor. This algorithm calculates not only the instantaneous value $e2(n)$ but also the past values, such as $e2(n-1), e2(n-2), ...e2(n-N+1)$. The range of the forgetting factor is (0, 1). When the forgetting factor is less than 1, it specifies that this algorithm places a larger weight on the current value and a smaller weight on the past values [9].

The resulting $E[e2(n)]$ of the RLS algorithms is more accurate than that of the LMS algorithms [10].The LMS algorithm require fewer computational requirements and memory as compare to RLS algorithm.

However, the Eigen value of the input correlation matrix, might affect the convergence speed of the resulting adaptive filters [9]. The convergence speed of the RLS algorithms is much faster than that of the LMS algorithms. However, the RLS algorithms require more computational complexity than the LMS algorithms.

$$E(i) = d(i) - y(i)$$

$$= d(i) - w^{H}(n)u(i) \qquad (5)$$

$$w(n) = w(n+1) + k(n)\xi^{*}(n)$$

IV. Cascade

The LMS algorithm provides only enhancement, NLMS has got the problem of musical tones, and in case of RLS algorithm rate of convergence is typically an order of magnitude faster than that of simple LMS filter. So RLS filter whitens the input data by using inverse correlation matrix of the data with zero mean, but this improvement in performance increase the computational complexity of the RLS filter [10]. After evaluating them, we come to a conclusion that NLMS-RLS is the optimum cascade for speech enhancement.

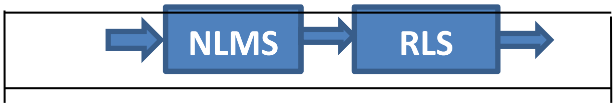

Fig. 2. Block Diagram

It has two stages: Normalized Least Mean Square algorithm and Recursive Least Square algorithm and the main idea of this approach are that of enhancing the original speech that is the combination of speech with non stationary noise.

V. Experimental Results

In this section, we present the experimental results of the proposed cascaded system that may use Output energy, SNR and MSE for the comparison. A comparison with the single algorithms will be used to study the merits and demerits. Due to limitations, we will replace single algorithm by cascaded algorithms.

A. Setup

For performance evaluation of the proposed method, English Language speech patterns have been recorded in different situations. The English speech patterns have been added with noise patterns. The listeners participated in different conditions. For each condition speech processed with circuit of adaptive filter algorithm. Ten sentences from the IIIT-H database are taken for clean speech that is produced by male and female speakers. Railway station and Restaurant noise were added to the clean speech files. Each male and female provided 2 to 6 minutes of "conversational speech" that is a story of anything. All recordings were originally digitized at a sampling rate of 16kHz.Then down sampled according to system requirement. Each story was cut to have different lengths of 6 sec and 3 sec.

We first show the performance of the traditional single channel filter which provides a benchmark for studying other noise reduction filters. Using a large ζ= 1 there is large variation in the value of SNR. SNR is increasing with the large value of leakage factor. But in case of RLS if we decrease forgetting factor delay is more .so best performance is achieved at moderate value of ζ .The output SNR reaches its peak when ζ= 1.0. The noise corrupted speech was processed by different circuits of adaptive filter algorithms that included cascaded version of NLMS algorithm and RLS algorithm based on a priori SNR and energy estimation. Different conditions are 1.Clean speech 2.speech with noise 3.Only noise. These listening experiments were conducted using headphones at a comfortable listening level. Each subject listened to a set of noise corrupted sentences to be formalized with the testing procedures.

The performance of the proposed method is compared in terms of parameters like Input energy, output energy, MSE, SNR. The values of these parameters are given in table for comparative analysis. In table 1 we calculated SNR with K=32, 64, 128 & 512 with new linear model and table 2 we calculated SNR with NLMS algorithm. It shows that SNR is improved in new linear model as compare to NLMS Circuit. In the fig spectrograms of speech signal voice 001.wav for 6 sec is shown and then that signal is processed by NLMS, RLS and combination of NLMS-RLS by setting the buffer values according to number of samples and each algorithm parameters are also adjusted according to system requirement.

The babble noise was recorded in the factory and cut that noise for 6 sec and 3 sec that are shown in fig. and that the signal was processed by NLMS, RLS and combination of both the algorithms. Such a NLMS filter with µ = 0.00095, N=512, w=0.1 and ζ= 0.85 .In this experiment we let the Output SNR is increasing and MSE is also increasing. But if we reduce the size of the input signal MSE should be same but SNR is still increasing. Consider leakage factor (ζ= 0.72, 0.85 and 1.0) we examine that performance varies with variation in the

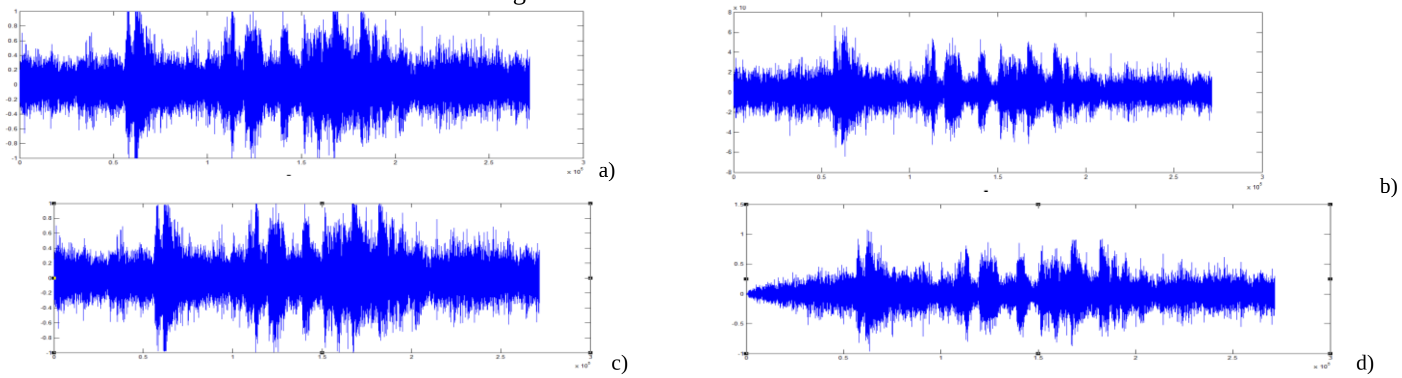

Fig. 3. (a) Speech corrupted by factory noise for 6 sec (b) Speech enhanced by NLMS (c) Speech enhanced by RLS (d) Speech enhanced by NLMS-RLS

B. Results

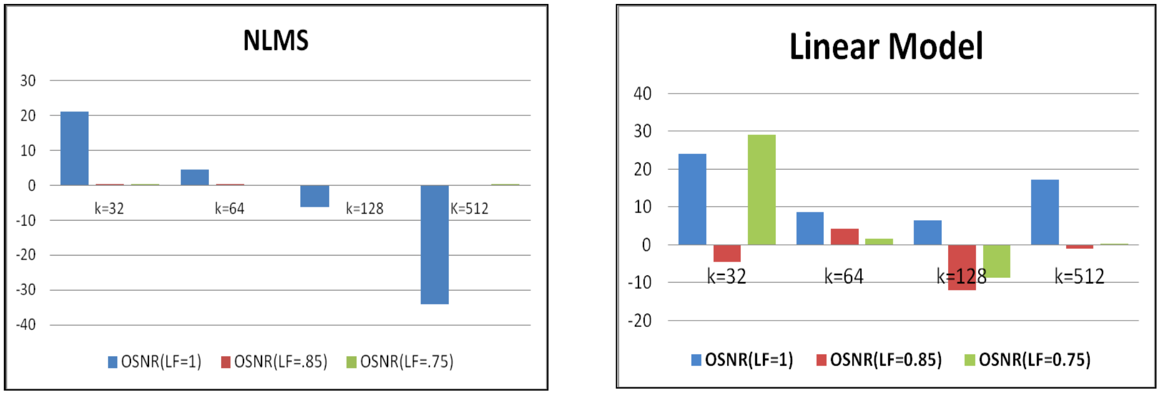

Fig. 4. Full band performance by forgetting factor 0.75, 0.85, 1 with window length 32, 64, 128, 512 by using a linear model and NLMS

TABLE I. ANALYSIS OF OSNR WITH DIFFERENT WINDOW SIZES WITH NEW LINEAR MODEL

Window size k	OSNR(LF=1)	OSNR(LF=0.85)	OSNR(LF=0.75)
k=32	24.13	-4.561	29.14
k=64	8.635	4.228	1.656
k=128	6.378	-12.09	-8.778
k=512	17.28	-0.9701	0.3721

TABLE II. ANALYSIS OF OSNR WITH DIFFERENT WINDOW SIZES WITH NLMS ALGORITHM

Window length k	OSNR(LF=1)	OSNR(LF=.85)	OSNR(LF=.75)
k=32	21.08	0.1903	0.03602
k=64	4.572	0.09006	-0.005574
k=128	-6.347	-0.2226	-0.08932
K=512	-34.14	-0.0125	0.0009499

TABLE III. COMPARISON TEST RESULTS FOR SIGNAL (6 SEC)

Parameters	NLMS	RLS	NLMS+RLS
In Energy	13970	560.7	13970
MSE	-0.04871	-0.004243	8749
SNR	-0.0125	20.47	17.28

TABLE IV. COMPARISON TEST RESULTS FOR SIGNAL (3 SEC)

Parameters	NLMS	RLS	NLMS+RLS
In Energy	5442	5442	5442
MSE	-0.01898	-0.04129	-0.01898
SNR	-0.00562	0.1718	0.2286

TABLE V. POST PROCESSING ALGORITHM VALUES FOR DIFFERENT TYPES OF SIGNALS

Parameters	Duration	No of samples	Energy	Power
Voice 001.wav	6 sec	286720	0.316694439	80.86 dB
Voice 001(1)	3 sec	132300	0.123402411	80.12 dB
Voice 006.wav	3 Sec	132300	0.238506168	82.98 dB
Voice 0012	3 Sec	132300	0.123402411	80.12 dB

The performance of the processing algorithm is evaluated for the real environmental noises like factory noise, canteen, digging noise. All these tables show that energy and SNR of the cascaded signal is higher than others individual Adaptive filters. All these noises are highly fluctuating and characterized by random energy bursts. Noise corrupted speech at selected SNR is generated by adding speech and noise digitally. For lower level of noise no need of using cascaded circuit. But as the noise ratio in the input increases, cascade of different filters are required.

VI. CONCLUSIONS

During the course of experiments, we have found that SNR tests alone can't reflect the effectiveness of a de-noising system and hence results are to be confirmed with listening tests. LMS algorithm can't be used alone as it provides trivial improvement. However it can be used as a preprocessing algorithm owing to its noise cancellation and channel equalization features.

Normalized Least Mean Square eliminates a good deal of noise however the Residual noise is heavy and undesirable as a good amount of vital speech information sometimes get subtracted. But this algorithm stops processing as soon as we achieve the target. From the SNR values of algorithms, we learn that SNR decreases drastically with increasing noise. By cascaded systems we can eliminate large amounts of noise whereas for lower levels of noise, any single algorithm (preferably Kalman Filter) will do. By pipelining NLMS with RLS, we slightly improve the efficiency of the system, resulting in providing stability with increasing noise proportions. Though cascade may show appreciable SNR improvements. NLMS-RLS is far better than any other algorithm. Moreover, the efficiency of this system varies only a little with increasing noise.

REFERENCES

[1] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," IEEE Trans. Acoust., Speech, Signal Process, vol 27, no. 2, pp.113–120,1979.

[2] D. Comminiello and M. Scarpinitie, "Functional link adaptive filters for Nonlinear Acoustic Echo Cancellation," IEEE Trans. Audio, Speech and Language Process, vol 21, no7, pp.1502-1512, 2013.

[3] Y. Ephraim and D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," IEEE Trans. Acoust., Speech, Signal Process, vol 32, no 6, pp. 1109–1121, 1984.

[4] Y. Ephraim and H. Van Trees, "A signal subspace approach for speech enhancement," IEEE Trans. Speech Audio Process, vol 3 no 4, pp.251–266, 1995.

[5] Y. Fang, G. LIU and J. GUO, "Speech enhancement based on modified a priori SNR estimation,"Front Electr. Electron. Eng. China.2011.

[6] Y. Hu and P.C. Loizou, "Subjective comparison and evaluation of speech enhancement algorithms," Speech Comm.,vol 49,no 7,pp.588–601, 2007.

[7] C. Kao , "Design of Echo Cancellation and Noise Elimination for Speech Enhancement," IEEE transactions. Vol 49,no 4, 2003.

[8] H. Kaur and R. Talwar , "Performance and Convergence Analysis of LMS Algorithm," IEEE International Conference on Computational Intelligence and Computing Research,2012.

[9] H. Kaur and R. Talwar, "Performance Comparison of Adaptive Filter Algorithms for Noise Cancellation,"IEEE International Conference on Emerging Trends in Communication, Control, Signal Processing and Computing Applications,2013.

[10] T.Kishore, "Optimum cascaded design for speech enhancement using Kalman filter," World Academy of Science, Engineering and Technology, pp. 355-359,2008.

[11] W. Jin and M.S. Scordilis, "Speech Enhancement by residual domain constrained optimization," Speech Communication, vol 48,pp. 1349-1364,2006.

[12] P.C. Loizou, "Speech Enhancement: Theory and Practice. Boca Raton," FL, USA: CRC, 2007.

[13] P.C.Loizou and G. Kim,"Reasons why current speech enhancement algorithms do not improve speech intelligibility and suggested solutions," IEEE Trans. Audio, Speech and Language Process, vol 19, no 1, pp. 47-56, 2011.

[14] K. Manohar and P. Rao, "Speech enhancement in non-stationary noise environments using noise properties," Elsevier, Speech Communication , vol 8, pp. 96-109, 2006.

[15] M. McCallum and B. Guillemin, "Stochastic-Deterministic MMSE STFT Speech Enhancement with General A Priori Information,"IEEE Trans. Audio, Speech and Language Process,vol 21, no 7, pp.1445-1457,2013.

[16] J. Ming, R. Srinivasan and D. Crookes , "A Corpus-Based Approach to Speech Enhancement From Non-stationary Noise," IEEE Trans. Audio, Speech and Language Process, vol 19, no 4, pp.822-836, 2011.

[17] U. Mittal and N. Phamdo, "Signal/ Noise KLT Based Approach for Enhancing Speech Degraded by Colored Noise," IEEE Trans. on Speech and Audio Processing, vol 8, no 2, pp.159- 16, 2000.

[18] K. Paliwal and A. Basu , "speech Enhancement Method Based on Kalman Filtering," EEE Int.Conf. Acoust. Speech, 1987.

[19] S.A. Selouani, "Speech Processing and Soft Computing," Springer Briefs in Electrical and Computer Engineering Springer Science + Business Media, LLC, 2011.

[20] S. Singh, M. Tripathy and R. S. Anand, " Single Channel Speech Enhancement for Mixed Non-Stationary Noise Environments," Advances in Signal Processing and Intelligent Recognition Systems (Springer) 264,pp. 545-555,2014..

[21] S.Singh, M. Tripathy, and R.S. Anand, " Noise removal in single channel Hindi speech patterns by using Binary Mask Thresholding function in various mother wavelets," IEEE International. Conference on Signal Processing, Computing and Control, 2013.