International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol.7, No.2 (2014), pp.355-364
http://dx.doi.org/10.14257/ijsip.2014.7.2.33

Noise Estimation based on Entropy without using VAD for Speech Enhancement

B. Ravi Teja[1] and S. Bhavani[2]

[1]*Assistant professor, Gudlavalleru Engineering College, Krishna Dt. AP, India*
[2]*Assistant professor, Gudlavalleru Engineering College, Krishna Dt. AP, India*
[1]*braviteja.22@gmail.com*, [2]*ammav10@gmail.com*

Abstract

A practical speech enhancement system consists of two major components, the estimation of noise power spectrum, and the estimation of speech.In single channel speech enhancement systems, most algorithms require an estimation of average noise spectrum since a secondary channel is not available. This requires a reliable speech/silence detector. Thus the speech/silence detection can be a determining factor for the performance of the whole speech enhancement system. The speech/silence detection finds out the frames of the noisy speech that contain only noise. If the speech/silence detection is not accurate then speech echoes and residual noise tend to be present in the enhanced speech. The performance of noise estimation algorithm is usually a tradeoff between speech distortion and noise reduction. In existing methods, noise is estimated only during speech pauses and these pauses are identified using Voice Activity Detector (VAD). This paper describes novel noise estimation method to estimate noise in non-stationary environments. This approach uses an algorithm that classifies noisy speech signal into pure speech, quasi speech and non-speech frames based on adaptive thresholds without using of VAD.Speech presence is determined by computing the ratio of the noisy speech power spectrum to its local minimum, which is computed by averaging past values of the noisy speech power spectra with a look-ahead factor. To evaluate proposed method performance, segmental SNR as evaluation criteria and compared with weighted average noise estimation method. The simulation results of the proposed algorithm shows better performance than conventional methods.

Keywords: *Entropy, Noise Estimation, Quasi Speech, Smoothing Constant, Speech Enhancement, Voice Activity Detector (VAD)*

1. Introduction

Background noise is a well-known and well researched problem in real time speech enhancement applications particularly multimedia, wireless communications, communications between pilot and air traffic control tower, speech recognition, speech coding, *etc*. The presence of noise in speech signals can result in appreciable degradation in both the quality and intelligibility. In automatic speech recognition systems, the performance degrades badly in the case of adverse environments with very low SNR. In the case of mobile communication, the speech signal is degraded by different types of noise in the communication channel. Noise is an unwanted signal and there are many types such as background noise, vehicle noise, *etc*. Unless, the nature of noise is known, it is difficult to enhance the speech. Due to random nature and inherent complexities of various types of noises, it is therefore Noise Spectrum Estimation is an important aspect of speech enhancement. In most situations we have only the noisy speech signal available while the

ISSN: 2005-4254 IJSIP

noise may be non-stationary and its power is unknown. Noise information has to be extracted from the noisy speech signal alone. Noise power estimation is crucial to effective speech enhancement, If Noise Estimate is too low, annoying residual noise will be audible, while if the noise estimate is too high, and speech will be distorted resulting possibly in intelligibility loss.

In single channel speech enhancement systems there will be access only to noisy speech and hence the noise statistics have to be estimated from the noisy speech itself. The Main objectives of speech enhancement techniques are to improve quality, intelligibility, robustness and to increase the accuracy of the speech Recognition. Speech enhancement techniques are concerned with algorithms that mitigate these unwanted noise effects and thus improve signal quality. Many speech enhancement systems have been developed based on spectral subtraction and Wiener filtering principles. The common features of all these methods are to estimate the power spectrum of clean speech using the power spectrum of noisy speech. Speech enhancement is an extremely difficult problem if we don't make any assumptions about the nature of the noise signal we aim to remove, since it is difficult to extract the information from noisy speech signals. Usually the noise spectrum estimate is obtained from the first few milli-seconds of noisy speech which are silence regions. This assumption is valid for the case of stationary noise in which the noise spectrum does not vary much over time. Traditional VADs also track the noise only frames of the noisy speech to update the noise estimate. But the update of noise estimate in those methods is limited to speech absent frames. This is not enough for the case of non-stationary noise in which the power spectrum of noise varies even during speech activity. Hence there is a need to update the noise spectrum continuously over time. Since it is difficult to extract the information from noisy speech signals, many noise estimation algorithms were proposed.

2. Related Works

The author Martin [6] proposed minimum statistics algorithm based on the observation that the power level of noisy speech signal often decays to the power level of noise. Hence by tracking of minima of noisy speech spectrum, we can get an estimate of the noise spectrum. But this algorithm failed to predict rise in speech power and rise in noise power during voiced speech intervals and it requires large window length to encompass long segment of speech to work effectively. The authors Cohen I. and Berdugo [12] proposed minima controlled recursive average algorithm based on averaging the past spectral values of noisy speech which was controlled by a time and frequency dependent smoothing factors, but it requires twice the number of frames for updating the local minima of noise level. Hirsch and Ehrlicher [13] proposed weighted average algorithm based on smoothing the spectral values of noisy speech, Noise estimation will never be updated, if SNR (Signal to Noise Ratio) is at high level [5].To overcome this drawback, this paper addresses a reliable and fast noise estimation technique based on entropy for speech enhancement in the real time environment. The section 3 describes the proposed noise estimation algorithm, Implementations and results in section 4 and section 5 conclude the work.

3. Proposed Work

Let the noisy speech signal is denoted as

$$y(n) = x(n) + d(n) \quad \ldots\ldots(1)$$

Where x(n) is the original speech and d(n) is the noise. The smoothed power spectrum of noisy speech is computed using the following first - order recursive equation

$$\widehat{N}(l,k) = \alpha\widehat{N}(l-1,k) + (1-\alpha)|Y(l,k)|^2 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

Where $\widehat{N}(l,k)$ is the smoothed power spectrum, l is the frame index, k is the frequency index, $|Y(l,k)|^2$ is the short time power spectrum of noisy speech and α is a smoothing constant [7]. Smoothing constant is not fixed but varies with time and frequency. The above recursive equation provides a smoothed version of periodogram $|Y(l,k)|^2$[11].

3.1. Proposed Noise Estimation Algorithm

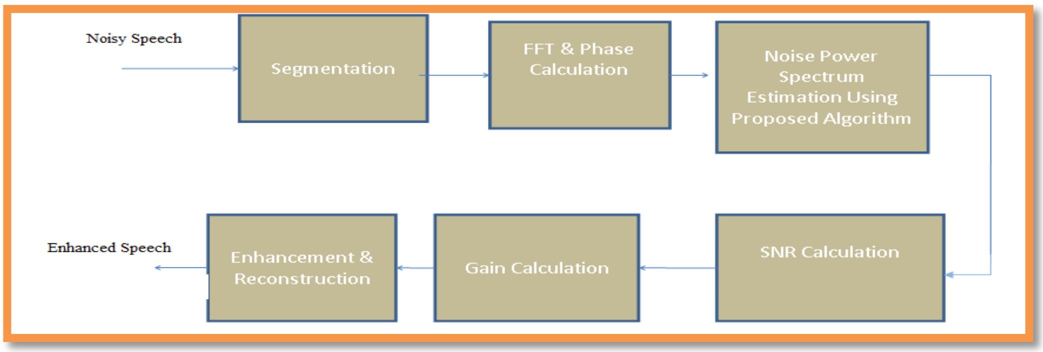


Figure 1. Proposed Noise Estimation Algorithm

Figure 1 is the algorithm of our work and in this the noise signal is segmented into number of frames and FFT is performed on those frames [8]. To discriminate various frames of noisy speech signal entropy is calculated. Based on threshold determination, classification of noisy speech signal and how the noise power is estimated and updated is shown in Figure 2.

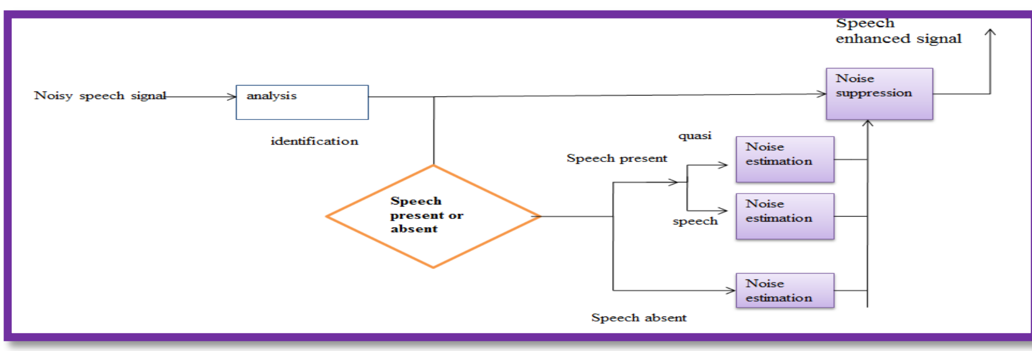


Figure 2. Classification of Noisy Speech Signal

3.2. Noise Power Spectrum EstimationMethod

The proposed noise estimation method classifies the noisy speech into three categories as pure speech, non-speech and quasi speech precisely. For this purpose, two thresholds are introduced for entropy H (*l*)

$$H(l,k) = -\sum_{k=1}^{2M} S(l,k)\log(S(l,k)) \ldots\ldots\ldots\ldots\ldots\ldots (3)$$

H (*l*) is called entropy of the noisy speech signal, which is a quantitative measure of how certain the outcome of a random noisy speech signal.

$$S(l,k) = \frac{Y_{energy}(l,k)+R(l)}{\sum_{k=1}^{2M} Y_{energy}(l,k)+R(l)} \quad (4)$$

$Y_{energy}(l, k)$ is the energy of noisy speech.

$R(l) = \max_k\{Y(l,k)\}$is a constant used to stabilize the S(l, k).

It is highly dependent on SNR of noisy speech and controlled by max (y (l, k)). The stabilization parameter R(l) is adjusted in each frame in order to rapid change in noise power spectrum.

$$\text{Let } T_1(l) = r_1E[H_{avg}(l)] \quad (5)$$

$$T_2(l) = r_2E[H_{avg}(l)] \quad (6)$$

Where $T_1(l)$, $T_2(l)$ are thresholds to classify noisy speech into Non speech, original speech & quasi speech. r1 r2 are 0.98, 0.95 respectively which are determined by experiment. E [Havg (l)] means an average over the recent number of initial silence frames including lth frame.If $H_{avg}(l) > T_1(l\text{-}1)$ then$T_1(l)$, $T_2(l)$ are updated by (5) and (6)

3.2.1. Noise Estimation for Non-Speech

If $H_{avg}(l) > T_1(l)$ then

$$\hat{N}(l,k) = \alpha\hat{N}(l-1,k) + (1-\alpha)|Y(l,k)|^2 \quad (7)$$

N(l, k) is the noise spectrum estimated in non-speech frame. α is known as forgetting factor (or) look – ahead factor (or) smoothing factor lies between 0.7 to 0.9 [9],[10].

3.2.2 Noise estimation for Quasi – speech

The purpose of introducing quasi – speech frame is to analyze noisy speech signal accurately.

If$T_2(l)<H_{avg}(l)<T_1(l)$then

$$\hat{N}(l,k) = P(l,k)\hat{N}(l-1,k) + \left(1-P(l,k)\right)|Y(l,k)|^2 \quad (8)$$

$$P(l,k) = a_d + (1-a_d)\, P_{sp}(l,k) \quad (9)$$

Where a_d is adaptive threshold and $P_{sp}(l, k)$ is a speech present probability is given by

$$P_{sp}(l,k) = \frac{|Y(l,k)|^2}{P_{min}(l,k)} \quad (10)$$

$P_{min}(l, k)$ is minimum noisy speech spectrum and it is updated by the following equation

$$P(l,k) = \xi P(l-1,k) + (1-\xi)|Y(l,k)|^2 \quad (11)$$

Where **ξ**is smoothing factor, P(l,k) is average noise spectrum.

If $P_{min}(l-1,k) \leq P(l,k)$ then

$$P_{min}(l,k) = \gamma P_{min}(l-1,k) + \frac{1-\gamma}{1-\beta}\left(P(l,k) - \beta P(l-1,k)\right) \quad (12)$$

If $P_{min}(l-1,k) > P(l,k)$ then

$$P_{min}(l,k) = P(l,k) \quad (13)$$

γ=0.998, β=0.96 &**ξ**=0.6 to 0.7 were determined experimentally.

Adaptive filter is used to produce estimated pure signal from a given noise speech signal. It

is a class of optimum linear filter, involves linear estimation of desired signal by minimizing minimum mean square error based on adaptive gain function is shown in Figure 3.

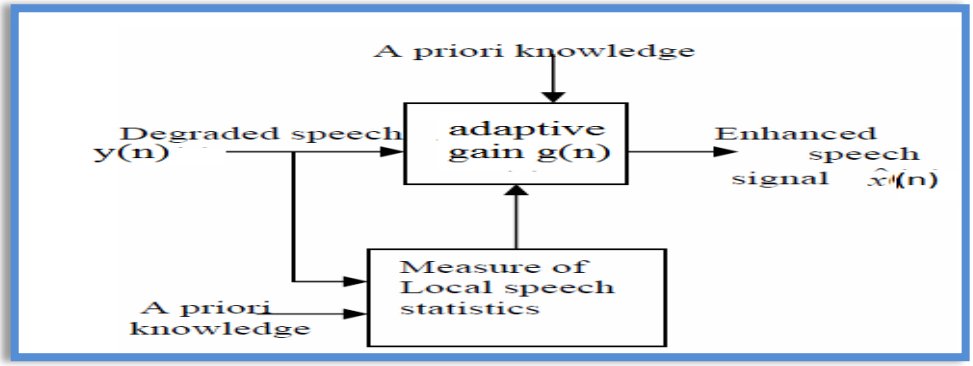


Figure 3. Typical Speech Enhancement System for Noise Reduction

Adaptivefilter implementation is useful for obtaining enhanced speech signal from varying statistics of noisy speech based on following adaptive gain function.

$$g(\mathrm{n}) = 1 - \min\{1, \left(\frac{\sigma^2{}_{\mathrm{yi}}(\mathrm{n})}{\sigma^2{}_{\mathrm{wi}}(\mathrm{n-1})}\right)^{-\mathrm{Q}} \qquad \text{.....,.....} \quad (14)$$

Where Q is aninteger, represents the average noise estimate

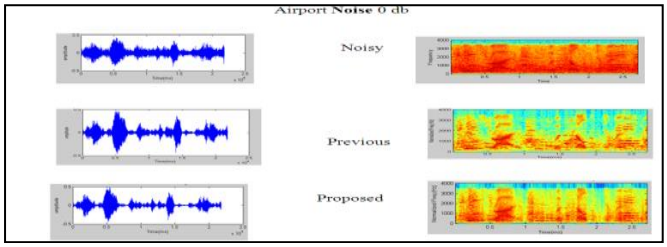
4. Implementation & Results

This section describes performance evaluation of the proposed algorithm. The complete implementation and analysis of the proposed method were carried out in MATLAB. This is recommended before implementing any algorithm on a real-time system. In order to evaluate if the processed speech signal is improved, well defined assessment techniques are needed, which can be both subjective and objective methods. The major goal of objective measures is to obtain high correlation with subjective measures, indicates the speech quality or improvement of the signal in order to evaluate performance of speech processing algorithms. Thus, having a good objective score gives an indication of whether or not the perception and/or quality have been improved. Noise reduction metrics are useful in determining e.g. the SNR-improvement before and after speech enhancement. This was confirmed by formal listening tests that indicated significantly higher preference for proposed algorithm compared to the other existing noise-estimation algorithms. Unlike other methods, the update of local minimum was continuous over time and did not depend on some fixed window length. Hence the update of noise estimate was faster for very rapidly varying non-stationary noise environments. We measure the segmental SNR over short frames and obtained the final result by averaging the values of each frame over all the segments. The Experimental results of proposed algorithm works effectively compared with weighted average algorithm are shown in Table 1. Spectrograms and Timing waveforms of speech corrupted by Airport noise and enhanced speech signal with weighted average technique & proposed algorithm are shown in the Figures 5, 6,

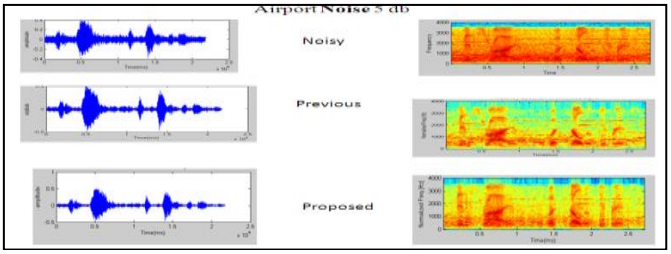
7. Experimental values of segmental SNR and LLR (Log Likelyhood Ratio) for airport noise, car noise are shown in Figure 7 and Figure 8 respectively which show higher values for the proposed algorithm by comparing with weighted averaging technique. LLR is used to compare the fit of two models one of which is nested within the other. Likelihood ratio test is a statistical test used to compare the fit of two models, one of which is a special case of the other. The test is based on the likelihood ratio, which expresses how many times more likely the data are under one model than the other.

Table 1. Comparisons

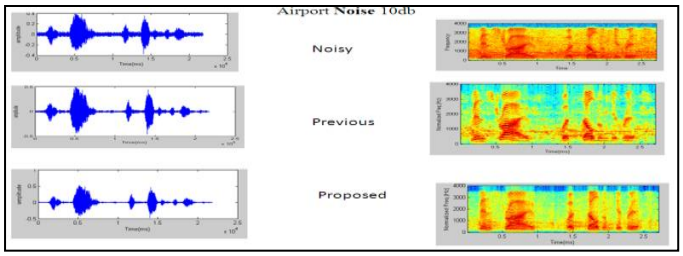
Type of Noise (db)	Segmental SNR		LLR	
	Weighted Average Technique	Proposed (SERA) Technique	Weighted Average Technique	Proposed (SERA) Technique
AIRPORT				
0	-3.802483	-3.440414	1.237398	1.057377
5	-2.781458	-2.526855	1.124488	0.934859
10	-0.731036	-0.083965	0.919158	0.736983
15	1.310788	3.080826	0.910468	0.549913
CAR				
0	-6.806391	-6.716270	1.687827	1.500914
5	-5.668619	5.485975	1.842711	1.596159
10	-4.866237	-3.861581	1.976017	1.602708
15	-4.335797	-3.537122	1.831509	1.580956
TRAIN				
0	-6.486321	-6.296185	2.091845	1.798190
5	-5.559169	-4.970945	2.322675	1.845213
10	-5.251629	-4.206358	2.036162	1.759774
15	-3.211548	4.284449	2.230337	1.827800



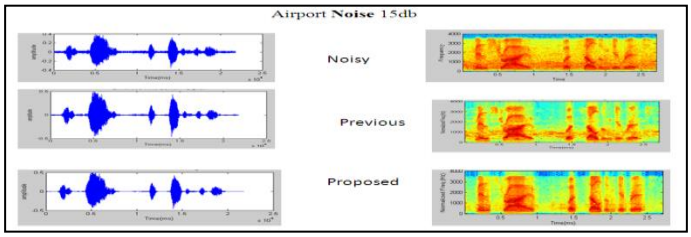
(a)



(b)

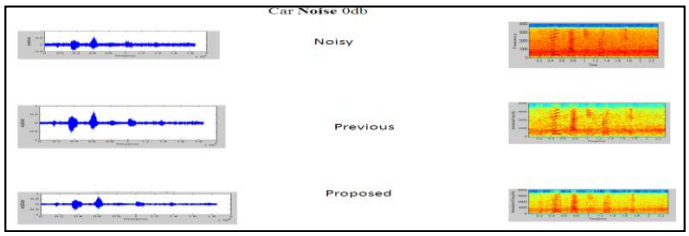


(c)

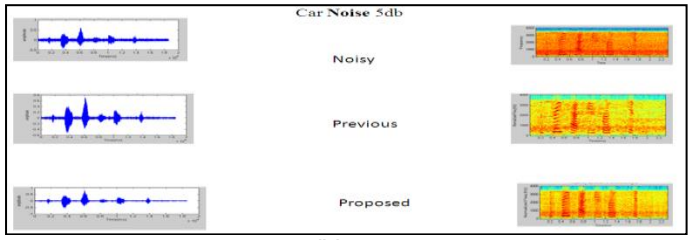


(d)

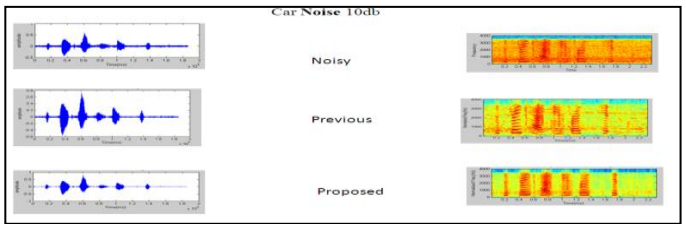
Figure 4. Timing Waveforms and Spectrograms of (a) Original Speech Signal (b) Noise Corrupted Speech Signal and Enhanced Signals with (c) Weighted Average Method (d) Proposed Method for Airport noise 0, 5, 10db Respectively



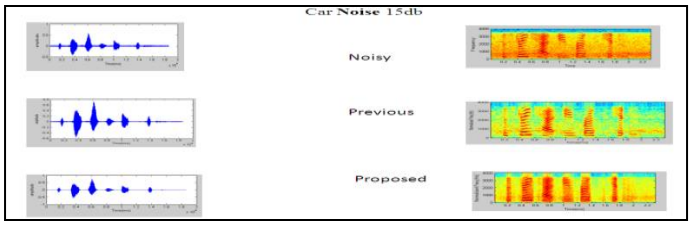
(a)



(b)

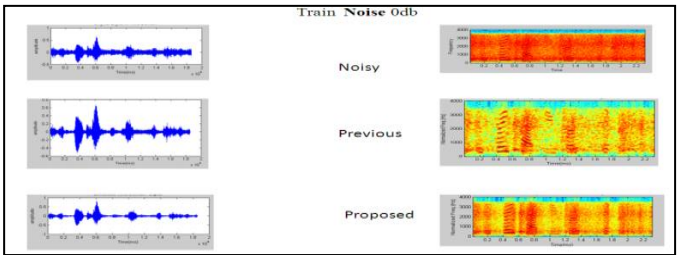


(c)

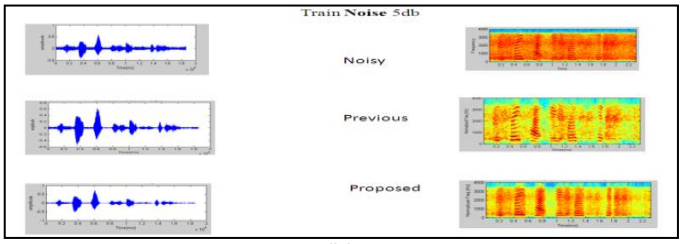


(d)

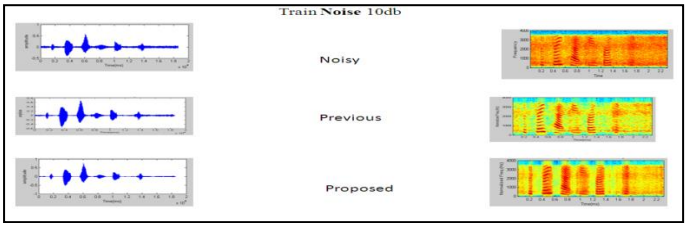
Figure 5. Timing Waveforms and Spectrograms of (a) original speech signal (b) Noise Corrupted Speech Signal and Enhanced Signals with (c) Weighted Average Method (d) Proposed Method for Car Noise 0, 5, 10,15db Respectively



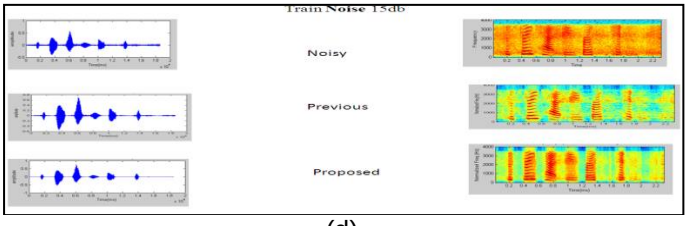
(a)



(b)



(c)



(d)

Figure 6. Timing Waveforms and Spectrograms of (a) Original Speech Signal (b) Noise Corrupted Speech Signal and Enhanced Signals with (c) Weighted Average Method (d) Proposed Method for Train Noise 0, 5, 10,15db Respectively

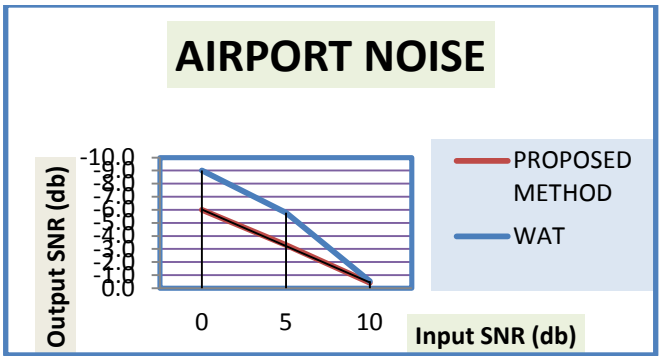


Figure 7. Segmental SNR for Airport Noise

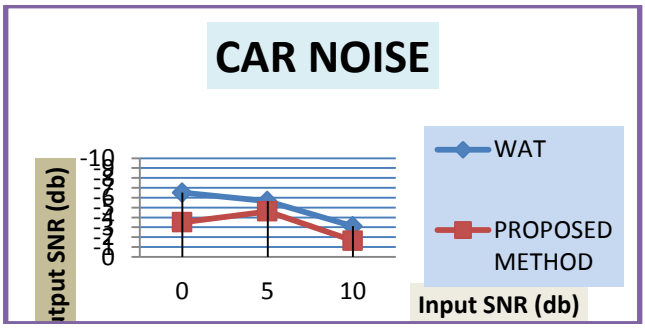


Figure 8. Segmental SNR for Car Noise

5. Conclusion

In this paper, we proposed a new approach to the enhancement of speech signals that have been corrupted by noise in real time environments. This approach uses a method that separates the speech presence and speech absence of noisy speech signal based on adaptive threshold to work effectively in non-stationary noise environments speech present signal is further divided into quasi-speech, & pure speech. According to increase or decrease in noise level, initial noise power spectrum is updated and estimated in each frame based on classification. The proposed method resulted in great reduction in noise while providing enhanced speech with lower residual noise. Performance of the proposed algorithm is better compared to existing algorithms. In future, we plan to evaluate its possible applications in preprocessing for new communication systems.

References

[1] R. SundarRajan and C. L. Philipos, "A Noise Estimation Algorithm for Highly Non-stationary environments", speech communication, vol. 48, **(2006)**, pp. 220-231.
[2] C. V. RamaRao, "Noise Estimation for Speech for Enhancement in Non-stationary environments – a new method", World Academy of Science, Engineering and Technology, vol. 70, **(2010)**, pp. 739-740.
[3] T. Lalith Kumar and R. SundarRajan, "Speech Enhancement using Adaptive Filters", VSRD-JJEEE, vol. 2, vol. 2, **(2012)**, pp. 92-99.
[4] C. GaneshBabu and P. T. Vanathi, "Performance Analysis of Voice Activity Detection Algorithm for Robust Speech Recognition System under Different Noisy Environment", Journal of Scientific & Industrial Research, vol. 69, **(2010)** July, pp. 515-522.
[5] G. Poblinger, "Computationally Efficient Speech Enhancement by Spectral Minima Tracking in Subbands", Proc. Euro Speech 2, **(1995)**, pp. 1513-1516.
[6] R. Martin, "Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum Statistics", IEEE Trans speech Audio Process, **(2001)**, pp. 504-512.

[7] P. Loizou, R. Sundarajan and Huy, "Noise Estimation Algorithm with Rapid Adaption for Highly Non-stationary Environments", prec. IEEE international conference on acoustic speech signal Proc, **(2004)**.
[8] J. Sohn and N. Kim, "Statistical Model Based Voice Activity Detection", IEEE signal ProcLett, vol. 6, no. 1, **(2001)**, pp. 1-3.
[9] S. Tanyer and H. Ozer, "Voice Activity Detection in Non – Stationary Noise", IEEE Speech Audio Proc., vol. 8, no. 4, **(2000)**, pp. 478-482.
[10] P. Loizou, "A Noise Estimation Algorithm with Rapid Adaption for Highly Non-stationary environments", Speech Communication Science direct, **(2006)**, pp. 220-231.
[11] A. Radha and R. Fuknu, "Noise Estimation Algorithms for Speech Enhancements in Highly Non-stationary Environments", IJCSI, vol. 8, **(2011)**, pp. 39-44.
[12] I. Cohen and B. Berdugo, "Noise Estimation by Minima Controlled Recursive Averaging for Robust Speech Enhancement", IEEE Signal Proc. Letters, vol. 9, no. 1, **(2002)** January, pp. 12-15.
[13] H. G. Hirsch and C. Ehrlicher, "Noise Estimation Techniques for Robust Speech Recognition", Proc. 20th IEEE Int. Conf. Acoustics, Speech, Signal Processing, Detroit, MI, **(1995)** May 8-12, pp. 153-156.

Authors

Ravi Teja Ballikura received the B.Tech and M.Tech degree in electronics and communication engineering in 2010 from bapatla engineering college, Digital Electronics and Communication Systems in 2012 from Gudlavalleru engineering college affiliated to JNTUK, Kakinada respectively. Working as a Assistant Professor in Gudlavalleru engineering college from 2012 to till date. Research Interests in speech processing and more especially in enhancement of speech signal.

Bhavani Samanthapudi received the B.Tech and M.Tech degree in electronics and communication engineering in 2010 from Rao and Naidu engineering college, Digital Electronics and Communication Systems in 2012 from Gudlavalleru engineering college affiliated to JNTUK, Kakinada respectively. Working as a Assistant Professor in Gudlavalleru engineering college from 2013 to till date. Research Interests in speech processing and more especially in enhancement of speech signal

