International Journal of Electrical, Electronic and Communication Sciences
ISSN: 2517-9438
Vol:4, No:10, 2010

Noise Estimation for Speech Enhancement in Non-Stationary Environments-A New Method

Ch.V.Rama Rao, Gowthami., Harsha., Rajkumar., M.B.Rama Murthy, K.Srinivasa Rao and K.Anitha Sheela

***Abstract*—**This paper presents a new method for estimating the non-stationary noise power spectral density given a noisy signal. The method is based on averaging the noisy speech power spectrum using time and frequency dependent smoothing factors. These factors are adjusted based on signal-presence probability in individual frequency bins. Signal presence is determined by computing the ratio of the noisy speech power spectrum to its local minimum, which is updated continuously by averaging past values of the noisy speech power spectra with a look-ahead factor. This method adapts very quickly to highly non-stationary noise environments. The proposed method achieves significant improvements over a system that uses voice activity detector (VAD) in noise estimation.

***Keywords*—**Noise estimation, Non-stationary noise, Speech enhancement.

I. Introduction

In most speech enhancement algorithms, it is assumed that an estimate of the noise spectrum is available. But the performance of the speech enhancement algorithms, quality of the enhanced speech signal is depends upon the noise estimation. If noise estimate is too low, annoying residual noise will be audible, while if the noise estimate is too high, speech will be distorted resulting possibly in intelligibility loss. The simplest approach is to estimate and update the noise spectrum during the speech pauses of the signal using a VAD algorithm [1]. Although such an approach might work satisfactorily in stationary noise, it will not work well in more realistic environments where the spectral characteristics of the noise might be changing constantly. Hence there is a need to update the noise spectrum continuously over time and this can be done using noise-estimation algorithms. Several noise estimation algorithms have been proposed for speech enhancement applications [2-5]. These noise estimation methods do not adapt quickly to increasing noise levels. The noise estimate was updated by tracking the noise only regions of the noisy speech spectrum. The noise only regions are tracked based on VAD. In this paper, noise estimation is improved in the following aspects: update of the noise estimate without explicit VAD, estimate of speech presence probability exploiting the correlation of power spectral components in neighbouring frames. The proposed method updates the noise estimate in each frame using a time-frequency dependent smoothing factor computed based on the speech presence probability.

[1]Dept.of ECE, Gudlavalleru Engineering College, Gudlavalleru-521356, AP, India; * e-mail:chvramaraogec@gmail.com.
[2] Professor, Jayaprakash Narayan College of Engineering, Dharmapur, Mahabubnagar- 509001, AP, India.
[3] Principal, TRR College of Engineering, Pathancheru-502319, AP, India.
[4]Jawaharlal Nehru Technological University, Hyderabad- AP, India.

II. Proposed Noise Estimation Method

Let the noisy speech signal in the time domain be denoted as

$$y(n) = x(n) + d(n) \tag{1}$$

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

$$P(\lambda,k) = \eta P(\lambda-1,k) + (1-\eta)|Y(\lambda,k)|^2 \tag{2}$$

Where $P(\lambda,k)$ is the smoothed power spectrum, λ is the frame index, k is the frequency index, $|Y(\lambda,k)|^2$ is the short-time power spectrum of noisy speech and η is a smoothing constant.

A. Classifying Noisy Speech into Speech Present/Absent Frames

The power spectrum of the noisy speech is equal to the sum of the speech power spectrum and noise power spectrum since speech and the background noise are assumed to be independent. Also in any speech sentence there are pauses between words which do not contain any speech. Those frames will contain only background noise. The noise estimate can be updated by tracking those noise-only frames. To identify those frames, a simple procedure is used which calculates the ratio of noisy speech power spectrum to the noise power spectrum at three different frequency bands

$$\xi_L(\lambda) = \frac{\sum_{k=1}^{LF} P(\lambda,k)}{\sum_{k=1}^{LF} D(\lambda-1,k)}, \xi_M(\lambda) = \frac{\sum_{k=LF+1}^{MF} P(\lambda,k)}{\sum_{k=LF+1}^{MF} D(\lambda-1,k)}, \xi_H(\lambda) = \frac{\sum_{k=MF+1}^{\frac{Fs}{2}} P(\lambda,k)}{\sum_{k=MF+1}^{\frac{Fs}{2}} D(\lambda-1,k)} \tag{3}$$

Where $D(\lambda, k)$ is the estimate of noise power spectrum for the current frame and LF, MF and Fs correspond to the frequency bins of 1 KHz, 3 KHz and sampling frequency respectively. If all the three ratios mentioned in Eq. 3 are smaller than the threshold σ, that frame is concluded as a noise-only frame. Otherwise, if any one or all the ratios are greater than threshold that frame is considered as speech present frame.

B. Tracking the Minimum of Noisy Speech

For tracking the minimum of the noisy speech power spectrum over a fixed search window length, various methods were proposed. These methods were sensitive to outliers and also the noise update was dependent on the length of the minimum-search window. For tracking the minimum of the noisy speech by continuously averaging past spectral values [6], a different non-linear rule is used.

If $P_{\min}(\lambda-1,k) < P(\lambda,k)$ then

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

else

$P_{\min}(\lambda,k) = P(\lambda,k)$
end

where $P_{\min}(\lambda,k)$ is the local minimum of the noisy speech power spectrum and γ and β are constants. The look ahead factor β controls the adaptation time of the local minimum.

C. Speech Presence Detection

The approach taken to determine the speech presence in each frequency bin was similar to the method used in [4]. Let the ratio of noisy speech power spectrum and its local minimum be defined as

$$S(\lambda,k) = \frac{P(\lambda,k)}{P_{\min}(\lambda,k)} \quad (4)$$

This ratio is then compared with a frequency-dependent threshold, and if the ratio is greater than the threshold, it is taken as speech present frequency bin else it is taken as speech absent frequency bin. This is based on the principle that the power spectrum of noisy speech will be nearly equal to its local minimum when speech is absent. Hence smaller the ratio defined in Eq. 4 the higher the possibility that it will be a noise-only region or vice versa. This can be summarized as follows

If $S(\lambda,k) > \delta(k)$, then
$I(\lambda,k) = 1$ speech present
else $I(\lambda,k) = 0$ speech absent
end

Where δ(k) is the frequency dependent threshold whose optimal value is determined experimentally. $I(\lambda,k)$, is used to represent speech flag. Note that in [4], a fixed threshold was used in place of $\delta(k)$ for all frequencies. From the above rule, the speech-presence probability, $p(\lambda,k)$ is updated using the following first-order recursion

$$p(\lambda,k) = \alpha_p p(\lambda-1,k) + (1-\alpha_p) I(\lambda,k) \quad (5)$$

Where α_p is a smoothing constant. The above recursive implicitly exploits the correlation for speech presence in adjacent frames.

D. Calculating Frequency Dependent Smoothing Constant

Using the above speech-presence probability estimate, the time–frequency dependent smoothing factor is computed as follows [4]

$$\alpha_s(\lambda,k) == \alpha_d + (1-\alpha_d) p(\lambda,k) \quad (6)$$

Where α_d is a constant. Note that $\alpha_s(\lambda,k)$ takes values in the range of $\alpha_d \le \alpha_s(\lambda,k) \le 1$

E. Update of Noise Spectrum Estimate

Finally, after computing the frequency-dependent smoothing factor $\alpha_s(\lambda,k)$ using Eq. (6), the noise spectrum estimate is updated as

$$D(\lambda,k) = \alpha_s(\lambda,k) D(\lambda-1) + (1-\alpha_s(\lambda,k)) |Y(\lambda,k)|^2 \quad (7)$$

where $D(\lambda,k)$ is the estimate of the noise power spectrum. Hence, the overall algorithm can be summarized as follows. After classifying the frequency bins into speech present or absent, we update the speech-presence probability using Eq. (5) and then use this probability to update the time-frequency dependent smoothing factor in Eq. (6). Finally the noise spectrum is updated according to Eq. (7).

III. Experimental Results

The complete implementation and analysis of the proposed method were carried out in MATLAB. MATLAB allows offline implementation of signal processing based algorithms with relative ease in preliminary investigations without constraints of time and memory and use of computational power of the work station's processor. This is recommended before implementing any algorithm on a real-time system.

The performance of the proposed technique was compared with weighted average algorithm by combining with a Wiener-type speech-enhancement algorithm with the following spectral gain function

$$G(\lambda,k) = \frac{\hat{X}(\lambda,k)}{\hat{X}(\lambda,k) + \mu_k \hat{D}(\lambda,k)} \quad (8)$$

Where $C(\lambda,k)$ is the estimated clean speech spectrum computed from the noisy speech and noise estimates as follows

$$C(\lambda,k) = \max\{|Y(\lambda,k)|^2 - D(\lambda,k), \nu D(\lambda,k)\} \quad (9)$$

where $\nu = 0.001$ is a small positive number. The max operation is used to ensure positive values for the estimated clean speech spectra. The over subtraction factor μ_k is determined from the a posteriori segmental SNR. To evaluate the performance the following objective measures are considered.

A. Degree of Noise Reduction

There will be a trade-off between the noise suppression and speech quality. The performance of the proposed system is evaluated by the noise reduction, $NR(i)$, defined as

$$NR(i) = \frac{P_{in}(i)}{P_{out}(i)} \quad (10)$$

Where $P_{in}(i)$ is the background noise level in the corrupted speech signal and $P_{out}(i)$ is the noise level in the enhanced signal.

B. Time-domain SNR Measures

The time-domain segmental SNR (SNR_{seg}) measure was computed is given by

$$SNR_{seg} = \frac{10}{M}\sum_{m=0}^{\omega-1}\log\frac{\sum_{n=N_m}^{N_m+N-1}x^2(n)}{\sum_{n=N_m}^{N_m+N-1}\left(x(n)-\hat{x}(n)\right)^2} \quad (11)$$

Where $x(n)$ the input is (clean) signal, $\hat{x}(n)$ is the processed (enhanced) signal, N is the frame length and M is the number of frames in the signal.

To evaluate the performance, the proposed algorithm is employed on signals taken from NOIZEUS data corpus. Table 1 shows the noise reduction values in dB of proposed method and these values are compared with results obtained with weighted average algorithm. From this table it is observed that for 0 dB SNR values the background noise is greatly reduced using the proposed algorithm. Experiment values of SNR_{seg} are given in table 2 which shows higher values for the present algorithm**.** Performance of this work is also evaluated in terms of spectrograms and timing waveforms. Timing waveforms and spectrograms of noise corrupted speech signal by car noise at 5 dB and enhanced speech signals with weighted average and proposed methods are shown in Fig. 1 and Fig. 2 respectively. An examination of the result and listening tests indicates that there is improvement in speech quality and reduction in residual background noise by using the proposed noise estimation algorithm as compared with weighted average algorithm.

TABLE I
NOISE REDUCTION VALUES (DB)

Type of noise and SNR value		weighted average method (VAD is used)	proposed algorithm
Airport	0 dB	18.25	21.48
	5 dB	19.93	24.09
	10 dB	22.46	24.89
Car	0 dB	20.45	23.75
	5 dB	22.73	26.82
	10 dB	23.73	25.63
Train	0 dB	19.05	23.22
	5 dB	21.05	27.12
	10 dB	20.95	22.85

IV. CONCLUSION

This paper focussed on the issue of noise estimation for enhancement of noisy speech. The noise estimate was updated continuously in every frame using time–frequency smoothing factors calculated based on speech-presence probability in each frequency bin of the noisy speech spectrum. 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. This was confirmed by experimental results that indicated significantly higher preference for our proposed algorithm compared to the other existing noise-estimation algorithms.

REFERENCES

[1] Sohn. J, Kim. N, "Statistical model-based voice activity detection", IEEE Signal Process. Lett. 6(1), pp. 1-3, 1999.
[2] Malah.D, Cox.R, Accardi.A, "Tracking speech-presence uncertainty to improve speech enhancement in non-stationary environments", Proc. IEEE Internat. On Conf. Acoust. Speech Signal Process., pp. 789-792, 1999.
[3] Martin.R, "Noise power spectral density estimation based on optimal smoothing and minimum statistics", IEEE Tran. Speech Audio Process., 9(5), pp. 504-512,2001.
[4] Cohen.I, "Noise estimation by minima controlled recursive averaging for robust speech enhancement", IEEE Signal Process. Lett., 9(1), pp. 12-15, 2002.
[5] Cohen.I., "Noise spectrum estimation in adverse environments: improved minima controlled recursive averaging", IEEE Trans. Speech Audio Process., 11(5), pp. 466-475, 2003.
[6] Doblinger.G, "Computationally efficient speech enhancement by spectral minima tracking in subbands", Proc. Eurospeech, pp.1513-1516, 1995.

TABLE II
TIME-DOMAIN SSEGMENTAL SNR VALUES (DB)

Type of noise and SNR value		weighted average method (VAD is used)	proposed algorithm
Airport	0 dB	-9.01	-5.78
	5 dB	-5.77	-1.60
	10 dB	-0.41	2.02
Car	0 dB	-0.91	0.91
	5 dB	-0.40	1.13
	10 dB	-2.94	2.12
Train	0 dB	-8.35	-4.23
	5 dB	-4.61	-0.01
	10 dB	-3.18	-1.12

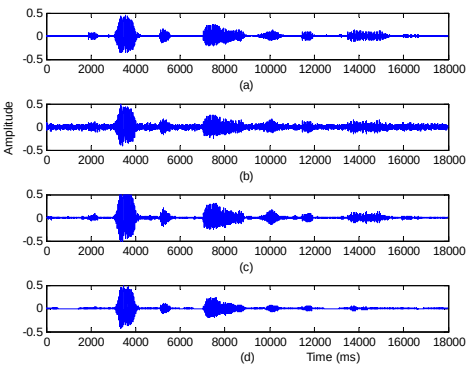

Fig. 1 Timing wave forms of (a) clean speech signal (b) noise corrupted speech signal and enhanced speech signals with (c) weighted average method (d) proposed method

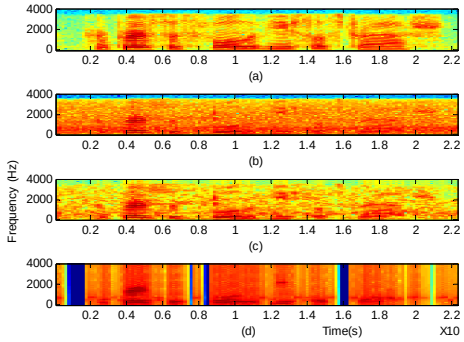

Fig. 2 Spectrograms of (a) clean speech signal (b) noise corrupted speech signal and enhanced speech signals with (c) weighted average method (d) proposed method