International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol.8, No.12 (2015), pp.217-222
http://dx.doi.org/10.14257/ijsip.2015.8.12.21

A Weighted Recursive Averaging Approach for Noise Spectrum Estimation

Yan Feng

Modern Education and Technology Center, Tibet University, Lhasa, Tibet, China
fy4528@163.com

Abstract

In the paper, we present a new noise spectrum estimation algorithm which is simple and effective for non-stationary background noise environments. The new proposed algorithm continuously updates the estimated noise by weighted noisy speech with a constant smoothing factor, the weighting factor is adjusted by an estimated signal-to-noise ratio (SNR), and the SNR is controlled by the local energy which be obtained by frequency smoothing of the noisy power spectrum in each frame. Objective experimental results and a subjective comparison show that the improved noise estimation algorithm when integrated in speech enhancement is preferred over the competitive noise estimation algorithms.

Keywords: *Speech enhancement, Noise spectrum estimation, Weighted recursive averaging*

1. Introduction

Speech enhancement or noise reduction has been one of the main investigated problems in the speech community for a long time. Numerous speech processing devices such as telephony, speech recognition systems and digital hearing aids and so forth which are often used in the real world. Unfortunately, the speech intelligibility can be harmed due to background noise. Speech enhancement method such as minimum mean-square error short-time spectral amplitude estimator (MMSE) [1] can be used to increase the quality of these speech processing devices. Correct noise power spectrum estimation is essential to good quality of the enhanced speech [1][2]. In non-stationary noise environments, a useful noise power spectrum estimation approach, known as the minimum statistics (MS), is presented by Martin [3-4]. In this approach, minima values of a smoothed power spectrum estimate of the noisy speech are tracked, and multiplied by factor that compensates the estimate for possible bias. However, the variance of this noise estimate is about twice as large as the variance of a conventional noise estimator [3-5]. A computationally more efficient minimum tracking method is proposed in [6], its main drawbacks are the very slow update rate of the noise estimation in case of a abrupt rise in the noise energy level, and this tendency to cancel the signal [7-8].

We present here a recent algorithm, developed by Cohen and Berdugo [9], namely minima controlled recursive averaging (MCRA) that combines the robustness of the minimum tracking with the simplicity of the recursive averaging. The noise estimate obtained by averaging past power spectrum by a smoothing parameter that is controlled by the speech presence probability in subbands. The speech presence probability is adjusted by the minima values of a smoothed periodgram. In other words, the noise spectral estimation procedure comprises two iterations of smoothing and minimum tracking. The first iteration provides a rough voice activity in each frequency. Smoothing during the second iteration excludes relatively strong speech components, which makes the minimum tracking during speech activity more robust. The major drawback with the MCRA is the update of local minimum of noisy speech for increasing noise levels [8].

ISSN: 2005-4254 IJSIP

Even in the improved version of MCRA (IMCRA) [5] some variation of minimum statistics rules was used for minimum tracking. Even though the delay for this method was slightly less than that for minimum statistics approach, this method takes slightly less than 1.5s to update noise estimate for increasing noise levels.

Another popular algorithm, known as the weighted noise estimation (WN) [10], has a good speech quality for a wide range of SNRs and sufficient noise suppression simultaneously. The estimated noise is obtained as an average of the noisy speech weighted by an estimated SNR. The algorithm is a very simple and computationally efficient procedure. However, this method has a drawback that the weak speech segment often be regard as noise segment because of lower spectral amplitude especially in weak speech regions following high SNR speech segment.

In this work, we present an improved noise power spectrum estimation algorithm which can more accurately distinguish noise and speech. The presented improved algorithm does not wait for specific window time to update the noise estimate. Hence the tracking delay and the overestimates are all considerably reduced compared to a competitive noise tracking algorithms.

The paper is organized as follows. In Section 2, we present the improved noise estimator. The estimated noise is obtained as an average of the weighted noisy speech using a constant smoothing parameter, the weighting factor is adjusted by an estimated SNR, and the SNR is controlled by the local energy which be obtained by

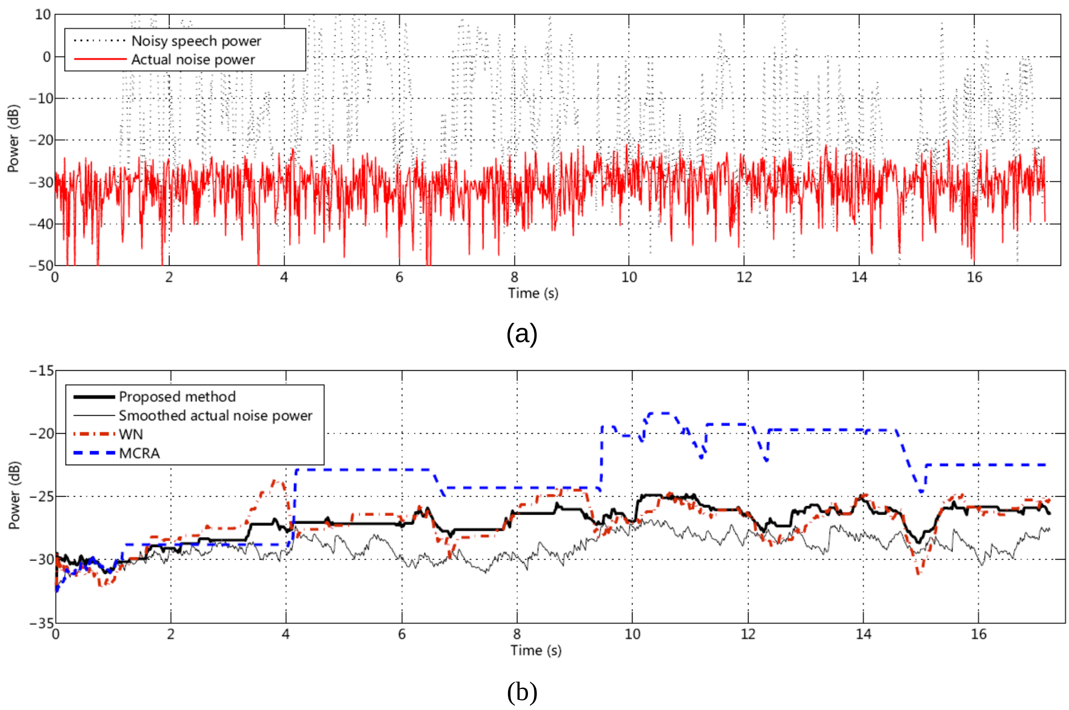


Figure 1. Example of Noise Power Estimation Obtained by the Proposed Method, MCRA and WN. Original Speech is Degraded by Car Noise Type at 5dB Input SNR and a Single Frequency Bin k = 13. (a) is Noisy Speech Power and Actual Noise Power; (b) is the Noise Power Estimation for a Speech Signal

frequency smoothing of the noisy power spectrum in each frame. In Section 3 we evaluate the proposed method, and discuss experimental results, which validate its effectiveness. Finally, in Section 4 we summarize the paper and draw conclusions.

2. Proposed Noise Power Spectrum Estimator

Let *x(t)* and *d(t)* denote speech and uncorrelated additive noise signals respectively, the observed noisy speech *y(t)* is given by

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

where *t* is the time index. Using a short-time Fourier transform, in the time-frequency domain we have

$$Y(n,k) = X(n,k) + D(n,k) \tag{2}$$

where *n* represents the frame index and *k* represents the frequency bin index.

In this section we present the noise power spectrum estimator and also analyze its performance. In contrast to the IMCRA and MS methods [3-5] which take into account the strong correlation speech presence in neighboring frequency bins of consecutive frames, the averaging of the noisy power spectrum is carried out in both time and frequency. In accordance with the IMCRA method [5], let *b* denote a normalized window function of length *2w+1*, the local energy $S(n,k)$ can be obtained by frequency smoothing of the noisy power spectrum in each frame,

$$S(n,k) = \sum_{i=-w}^{w} b(i)\,|Y(n,k-i)|^2 \tag{3}$$

Subsequently, we discuss the SNR calculation. First of all, because weighted noisy speech has a good speech quality for a wide range of SNRs and sufficient noise suppression simultaneously [10], and according to weighting factor $W(n,k)$ calculation of the weighted noise estimation [10], $W(n,k)$ is obtained by

$$W(n,k) = \begin{cases} 1, & \hat{\gamma}(n,k) \le \hat{\gamma}_1 \\ \dfrac{\hat{\gamma}_2 - \hat{\gamma}(n,k)}{\hat{\gamma}_2 - \hat{\gamma}_1}, & \hat{\gamma}_1 < \hat{\gamma}(n,k) < \hat{\gamma}_0 \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

where $\hat{\gamma}_0$, $\hat{\gamma}_1$ and $\hat{\gamma}_2$ are thresholds. $\hat{\gamma}(n,k)$ is an estimated SNR, in order to obtain a more accurate estimate of the SNR $\hat{\gamma}(n,k)$, we proposed a scheme as follows:

$$\hat{\gamma}(n,k) = \frac{S(n,k)}{\hat{\lambda}_D(n-1,k)} \tag{5}$$

In accordance with the MCRA and WN methods [5, 10], we proposed the new noise power spectrum estimator, the estimated noise is obtained as an average of the weighted noisy speech using a constant smoothing parameter during periods of speech absence, and holds the estimation during speech presence, specifically

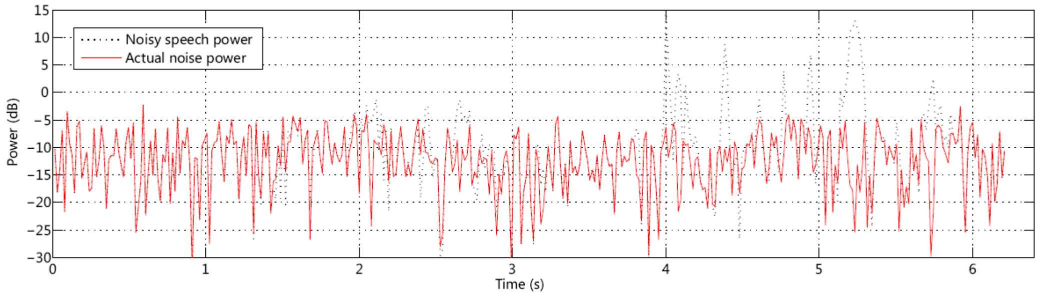
$$\hat{\lambda}_D(n,k) = \begin{cases} \hat{\lambda}_D(n-1,k), & W(n,k) = 0 \\ \alpha_d \hat{\lambda}_D(n-1,k) + (1-\alpha_d)[W(n,k)\,|Y(n,k)|^2], & \text{otherwise} \end{cases} \tag{6}$$

where $\alpha_d\,(0 < \alpha_d < 1)$ is the constant smoothing parameter, the speech is presence when the estimated SNR $\hat{\gamma}(n,k)$ is lower than a threshold $\hat{\gamma}_0$. The weighting factor

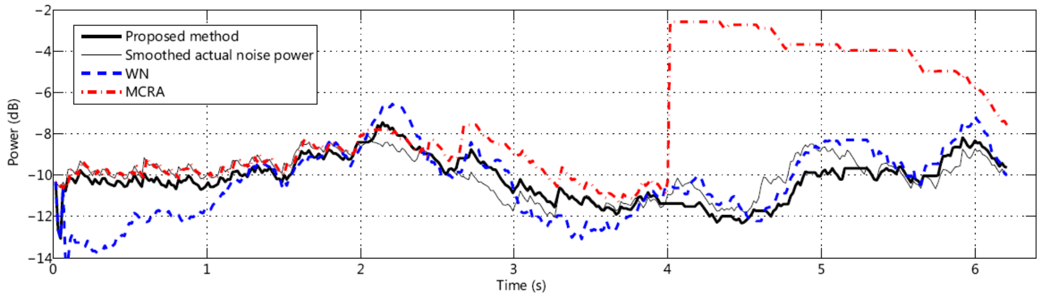
$W(n,k)$ is adjusted by the estimated SNR. The noise estimate $\hat{\lambda}_D(n,k)$ is initialized through average the first frames by

$$\hat{\lambda}_D(n,k) = \frac{1}{n}\sum_{i=1}^{n} |Y(i,k)|^2, n < T_{init} \tag{7}$$

where T_{init} is the first frames.



(a)



(b)

Figure 2. Example of Noise Power Estimation Obtained by the Proposed Method, MCRA and WN. Original Speech is Degraded by f16 Noise Type at 5dB Input SNR and a Single Frequency Bin k = 22. (a) is Noisy Speech Power and Actual Noise Power; (b) is the Noise Power Estimation for a Speech Signal

3. Evaluation

To evaluate the proposed noise estimation algorithm, we compare the algorithm with MCRA and WN methods. First, we measure the segmental relative estimation error for three noise types and four levels. Second, we integrate the noise estimation methods into a MMSE speech enhancement system, and determine the improvement in the segmental SNR. Three noise types in our evaluation are taken from the Noisex-92 database. They include Gaussian noise, car noise and F16 cockpit noise. The speech signal is constructed from six different utterances, half from male speakers and half from female speakers are taken from the NOIZEUS database. The speech signal is sampled at 8 kHz and degraded by the three noise types with segmental SNR in the range [0, 15] dB. The spectral analysis is implemented with hanning windows of 256 samples length and 128 samples overlap.

For the *a priori* SNR estimator [1], α =0.98, ξ_{min} = -25dB. For the MCRA method [9], α_d =0.95, α_s =0.8, α_p =0.2, ω =1, δ =5, L =125, b is hanning window. For the WN method [10], T_{init} =4, Lz =20, $\hat{\gamma}_1$ =0dB, $\hat{\gamma}_2$ =10dB, Θ_z =7dB. For the proposed method, T_{init} =4, α_d =0.95, $\hat{\gamma}_0$ =10dB, $\hat{\gamma}_1$ =3dB, $\hat{\gamma}_2$ =18dB, b is hanning window, ω =1.

Example of noise power estimation obtained by the proposed method, MCRA and WN, degraded by car noise type at 5dB input SNR and a single frequency bin k = 13, as illustrated in Figure 1, Figure 1 (a) is noisy speech power and actual noise power; Figure 1 (b) is the noise power estimation for a speech signal.

Another example of noise power estimation obtained by the proposed method, MCRA and WN, degraded by f16 noise type at 5dB input SNR and a single frequency bin k = 22, as illustrated in Figure 2, Figure 2 (a) is noisy speech power and actual noise power; Figure 2 (b) is the noise power estimation for a speech signal.

Table 1. Segmental Relative Estimation Error Obtained Using the Proposed Method, MCRA and WN Estimators

Input SNR [dB]	White Gaussian noise			Car noise			F16 cockpit noise		
	WN	MCRA	Proposed method	WN	MCRA	Proposed method	WN	MCRA	Proposed method
0	0.0913	0.1249	0.0622	0.2617	0.1024	0.0645	0.1240	0.0892	0.0581
5	0.1026	0.3124	0.0871	0.2592	0.1025	0.0645	0.1393	0.1410	0.0832
10	0.1242	1.7065	0.0827	0.2591	0.1022	0.0644	0.1776	0.4425	0.1129
15	0.1450	8.9268	0.1159	0.2597	0.1058	0.0642	0.1587	2.6243	0.0838

Table 2. Segmental SNR Improvement Obtained Using the Proposed Method, MCRA and WN Estimators

Input SNR [dB]	White Gaussian noise			Car noise			F16 cockpit noise		
	WN	MCRA	Proposed method	WN	MCRA	Proposed method	WN	MCRA	Proposed method
0	3.8655	3.9828	4.0633	5.8922	6.9141	7.2358	3.5742	3.6806	3.8209
5	3.2859	3.2412	3.4440	6.0393	6.8309	7.3266	2.9737	2.9309	3.1603
10	2.5612	2.2558	2.6697	5.7194	6.2169	6.8590	2.3319	2.0924	2.4948
15	1.6497	0.9588	1.7123	4.7385	4.8554	5.6937	1.5669	0.9983	1.6701

By contrast, the noise estimate obtained by the proposed method is much closer to the smoothed actual noise power. The proposed method can avoids noise overestimates, particularly in adverse noise environments, which involve weak speech components and low input SNR.

Table 1 presents the results of the segmental relative estimation error achieved by the proposed method, WN and MCRA estimation methods in various noise types and levels. It shows that the proposed method obtains significantly lower estimation error than the WN and the MCRA methods. Table 2 summarizes the results of the segmental SNR improvement for various types and levels. The proposed method yields a higher improvement in the segmental SNR than the WN and the MCRA methods.

This is confirmed by a subjective evaluation of speech spectrograms and informal listening tests.

4. Conclusions

A noise power spectrum estimation algorithm is proposed for non-stationary noise environments in the paper, resulting in more accurately distinguishing noise and speech, the tracking delay and the overestimates are all considerably reduced compared to the competitive noise tracking algorithms. Experimental results show that the proposed noise estimation algorithm when integrated in speech enhancement is preferred over other noise estimation algorithms.

References

[1] 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. ASSP-32, no. 6, **(1984)** December, pp. 1109-1121.
[2] "Speech enhancement using a minimum mean-square error logspectral amplitude estimator", IEEE Trans. Acoust., Speech, Signal Process., vol. ASSP-32, no. 2, **(1985)** April, pp. 443–445.
[3] R. Martin, "Spectral subtraction based on minimum statistics", in Proc. 7th Eur. Signal Processing Conf., ser. EUSIPCO',94, Edinburgh, **(1994)** September, pp. 1182-1185.
[4] "Noise power spectral density estimation based on optimal smoothing and minimum statistics", IEEE Trans. Speech and Audio Processing, vol. 9, no. 5, **(2001)** July, pp. 504-512.
[5] I. Cohen, "Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging", IEEE Trans. Speech and Audio Processing, vol. 11, no. 5, **(2003)**, pp. 466C475.
[6] G. Doblinger, "Computationally efficient speech enhancement by spectral minima tracking in subbands", in Proc. 4th Eur. Conf. Speech, Communication, and Technology, ser. EUROSPEECH',95, **(1995)**, pp. 1513-1516.
[7] K. U. S. J. Meyer and K. D. Kammeyer, "Comparison of one- and two-channel noise-estimation techniques", Proceedings 5th Int. Workshop on Acoustic Echo and Noise Control, ser. IWAENC', 97, **(1997)**, pp. 137-145.
[8] Y. Ephraim and I. Cohen, "Recent advancements in speech enhancement", in The Electrical Engineering Handbook, 3rd ed. Boca Raton, FL: CRC, Online, **(2004)**.
[9] I. Cohen and B. Berdugo, "Noise estimation by minima controlled recursive averaging for robust speech enhancement", IEEE Signal Process. Lett., vol. 9, no. 1, **(2002)**, p. 12C15.
[10] A. S. M. Kato and M. Serizawa, "Noise suppression with high speech quality based on weighted noise estimation and mmse stsa", Proceedings ser. IWAENC2001, **(2001)**, pp. 183C186.

Author

Yan Feng, He received the B. S. degree from Tibet University, China (2003) and M. S. degree from Southwest Jiaotong University, China (2010). Currently, he is an instructor at Tibet University, China. His research interests include signal and information processing and machine learning.