A Noise Estimator with Rapid Adaptation in Variable-Level Noisy Environments

Bing-Fei Wu, Kun-Ching Wang*, Lung-Yi Kuo
Department of Electrical and Control Engineering
National Chiao-Tung University
Hsinchu, Taiwan, R.O.C.
Corresponding author mail: Kunching@cssp.cn.nctu.edu.tw

Abstract. In this paper, a noise estimator with rapid adaptation in a variable-level noisy environment is presented. To make noise estimation adapt quickly to highly non-stationary noise environments, a robust voice activity detector (VAD) is utilized in this paper and it depends on the variation of the spectral energy not on the amount of that. The noise power spectrum in subbands are estimated by averaging past spectral power values using a time and frequency dependent smoothing parameter, which is chosen as a sigmoid function changing with speech-present probability in subbands. The speech-present probability is determined by computing the ratio of the noisy speech power spectrum to its local minimum. Noise measurement, speech enhancement, spectral analysis, signal process.

1 Introduction

An accurate noise estimator used for speech enhancement in adverse environments is one of most essential parts. Inaccurate noise estimator will result in a perceptually annoying residual noise and speech distortion. In general, noise estimation is usually done by explicit detection of speech detection. This can be very difficult in the case of varying background noise. Furthermore, the background noise is assumed to be related stationary between speech pause. To overcome these problems, the noise spectrum needs to be estimated and updated continuously with a reliable speech detector. Among those algorithms, a recursive averaging is a commonly and easily used approach.

Martin [1] proposed a method which is based on minimum statistic (MS). The noise spectrum estimation is obtained by tracking the minimum of the noisy speech power spectrum over a specific window. To improve the computational complexity of estimating noise spectrum, Doblinger [2] proposed an efficient method. However, it fails to differentiate between a rise in noise power and a rise in speech power. Further, Cohen et al. [3] introduced a MCRA approach to estimate noise power spectrum using a smoothing parameter which is defined as the speech-present probability in subbands. The speech-present probability in subbands of a given frame can be determined by the ratio between the noisy speech power spectrum to its local minimum over a period of 0.5-1.5 sec. Finally, the ratio is compared to a specific threshold value to decide updating noise power or not. In recently, Lin et al. [4] proposed a simple and reliable noise estimation technique. To estimate subband noise adaptively and continuously, the smoothing parameter is adjusted by a sigmoid function. However, a variable-level of noise is not considered in this case. We summarize that the drawback of most methods are slow in adapting to suddenly increase level of noise.

In this paper, a noise estimator with rapid adaptation in a variable level noisy environment is presented. It depends only on the variation of the spectral energy but not on the amount of that. Based on the VAD, a noise estimator updates the noise spectrum fast and accurately even in suddenly increases of noise.

This paper is organized as follow. In order to make the estimator is robust against the time-varying level of noise. The utilized VAD in this algorithm is described in Section II. In Section III, the proposed noise estimation is presented in detail. In Section IV, the performance of the proposed method will be evaluated. Finally, we will discuss experimental results in Section V.

2 Voice Activity Detector

Shen et al. [5] first used an entropy-based parameter for speech detection under adverse conditions. Their experimental results revealed that the spectral entropy of a speech signal differs from that of a non-speech signal. The procedure for calculating a spectral entropy parameter is described as follows.
The short-time Fourier Transform (STFT) of a given time frame $s(n,l)$ is given by,

$$x(k,l)=\sum_{n=1}^{M} s(n,l)\cdot \exp(-j2kn\pi/M),\quad 1\le k\le M, \tag{1}$$

where $x(k,l)$ represents the spectral magnitude of the frequency component k in l^{th} frame index, and M is the total number of frequency components in FFT ($M=256$ in the proposed system). The spectral energy of each frame $x_{energy}(k,l)$ is described as follows.

$$x_{energy}(k,l)=\left|x(k,l)\right|^2,\quad 1\le k\le M/2, \tag{2}$$

Then, the probability associated with each spectral energy component $P_r(m,l)$ can be estimated by normalizing:

$$P_r(k,l)=\frac{x_{energy}(k,l)}{\sum_{m=1}^{M/2} x_{energy}(m,l)},\quad 1\le k\le M/2, \tag{3}$$

Following normalization, the corresponding spectral entropy H_l for a given frame is defined as follows.

$$H_l=\sum_{k=1}^{M/2} P_r(k,l)\cdot \log[1/P_r(k,l)], \tag{4}$$

The foregoing calculation of the spectral entropy parameter implies that the spectral entropy depends on the variation of the spectral energy not on the amount of that. Similarly, the spectral entropy parameter is robust against changing level of noise. Fig. 1 illustrates that the VAD can locate the speech-present regions, even in high level of background noise.

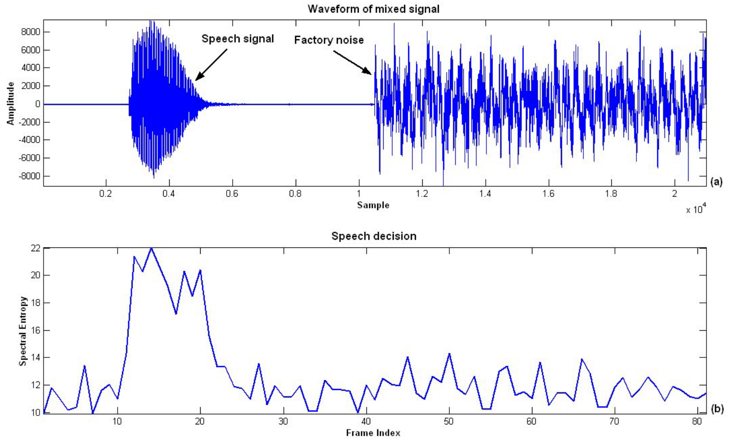


Fig. 1. Speech decision of a mixed signal (including a Factory noise)

3 Proposed Noise Estimator

Fig. 2 presents the flowchart of the proposed noise spectrum estimator. Let assume that noise $d(n)$ and speech $x(n)$ are uncorrelated. The smoothed power spectrum of noisy signal $P(k,l)$ is obtained by

$$P(k,l)=\eta P(k,l-1)+(1-\eta)\left|Y(k,l)\right|^2, \tag{5}$$

where $\left|Y(k,l)\right|^2$ is an estimate of the short-time power spectrum of $y(n)$, given by $y(n)=x(n)+d(n)$. η is a smoothing constant.
Since the noisy speech power spectrum in the speech-absent frames is equal to the noise power spectrum, the estimated noise power spectrum is updated by tracking the speech-absent. To make noise estimation track

speech-absent quickly in highly non-stationary noise environments, a robust voice activity detector (VAD), which depends on the variation of the spectral energy not on the amount of that, is utilized in this section.

First, computing the spectral entropy by Eqs. (1-4) during speech-present frame, a threshold σ is obtained as following:

$$\sigma = c \times E[H_l], \quad 1 \le l \le 5, \tag{6}$$

where c is constant by experiment.

If the spectral entropy value for a given frame is smaller than the threshold σ, then the current frame is regarded as a speech-absent frame. Moreover, the noise estimate is updated according to:

$$\bar{N}(k,l) = \lambda \cdot \bar{N}(k,l) + (1-\lambda) \cdot |Y(k,l)|^2, \tag{7}$$

where λ is a constant parameter.

Conversely, then the current frame is regarded as a speech-present frame. An algorithm that is suitable for estimating the noise spectrum during speech-present frame is used. First, finding the minimum of the noisy speech spectrum and using the minimum to determine signal-present probability in subbands. The signal-present probability is used to determine a time and frequency dependent smoothing parameter $\alpha(k,l)$, shown as following.

$$\bar{N}(k,l) = \alpha(k,l) \cdot \bar{N}(k,l-1) + (1-\alpha(k,l)) \cdot |Y(k,l)|^2. \tag{8}$$

To speed up the determination of local minimum of noisy speech spectrum, Doblinger's efficient method is used here [2], which is not constrained by any window length to update noise spectrum estimate.

$$\begin{aligned} &\text{If} \quad P_{\min}(k,l-1) < P(k,l), \\ &\text{then} \quad P_{\min}(k,l) = \gamma \cdot P_{\min}(k,l-1) + \frac{1-\gamma}{1-\beta}(P(k,l) - \beta \cdot P(k,l-1)), \\ &\text{else} \quad P(k,l) = P(k,l), \end{aligned} \tag{9}$$

where $P_{\min}(k,l)$ denote the local minimum of the noisy speech power spectrum and β and γ are constants determined experimentally.

Then, the local minimum is taken to determine speech-present probability $P_{sp}(k,l)$ in subbands, which is similar to that proposed in [3], and the ratio is shown as below:

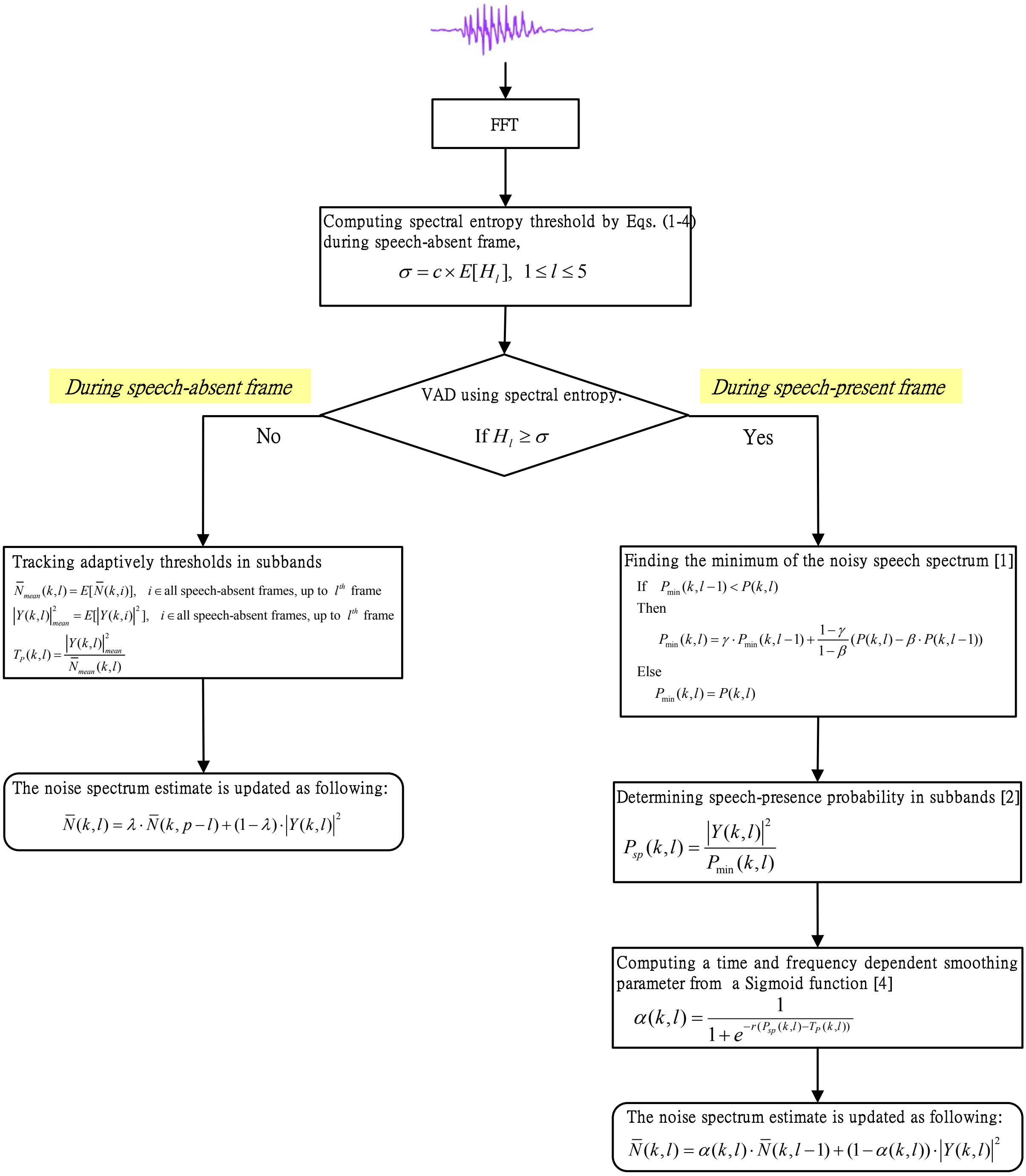
$$P_{sp}(k,l) = \frac{|Y(k,l)|^2}{P_{\min}(k,l)}. \tag{10}$$

To improve that a smoothing parameter is produced by comparing the ratio with a fixed threshold value [3], the smoothing parameter is chosen as a sigmoid function changing continuously with speech-present probability in subbands. The smoothing parameter is modified by

$$\alpha(k,l) = \frac{1}{1 + e^{-r(P_{sp}(k,l) - T_P(k,l))}}, \tag{11}$$

where $T_P(m,l)$ denotes a adaptive threshold in subbands and is determined during speech-absent frames and shown as following:

$$\begin{aligned} &\bar{N}_{mean}(k,l) = E[\bar{N}(k,i)], \quad i \in \text{all speech-absent frames, up to } l^{th} \text{ frame} \\ &|Y(k,l)|^2_{mean} = E[|Y(k,i)|^2], \quad i \in \text{all speech-absent frames, up to } l^{th} \text{ frame}, \\ &T_P(k,l) = \frac{|Y(k,l)|^2_{mean}}{\bar{N}_{mean}(k,l)} \end{aligned} \tag{12}$$



FFT
Computing spectral entropy threshold by Eqs. (1-4) during speech-absent frame,
$\sigma = c \times E[H_l],\ 1 \le l \le 5$
During speech-absent frame
VAD using spectral entropy:
If $H_l \ge \sigma$
During speech-present frame
No
Yes
Tracking adaptively thresholds in subbands
$\bar{N}_{mean}(k,l) = E[\bar{N}(k,i)],\ i \in$ all speech-absent frames, up to l^{th} frame
$|Y(k,l)|^2_{mean} = E[|Y(k,i)|^2],\ i \in$ all speech-absent frames, up to l^{th} frame
$T_p(k,l) = \frac{|Y(k,l)|^2_{mean}}{\bar{N}_{mean}(k,l)}$
The noise spectrum estimate is updated as following:
$\bar{N}(k,l) = \lambda \cdot \bar{N}(k,p-l) + (1-\lambda) \cdot |Y(k,l)|^2$
Finding the minimum of the noisy speech spectrum [1]
If $P_{\min}(k,l-1) < P(k,l)$
Then
$P_{\min}(k,l) = \gamma \cdot P_{\min}(k,l-1) + \frac{1-\gamma}{1-\beta}(P(k,l) - \beta \cdot P(k,l-1))$
Else
$P_{\min}(k,l) = P(k,l)$
Determining speech-presence probability in subbands [2]
$P_{sp}(k,l) = \frac{|Y(k,l)|^2}{P_{\min}(k,l)}$
Computing a time and frequency dependent smoothing parameter from a Sigmoid function [4]
$\alpha(k,l) = \frac{1}{1+e^{-r(P_{sp}(k,l)-T_P(k,l))}}$
The noise spectrum estimate is updated as following:
$\bar{N}(k,l) = \alpha(k,l) \cdot \bar{N}(k,l-1) + (1-\alpha(k,l)) \cdot |Y(k,l)|^2$

Fig. 2. The flowchart of the proposed noise spectrum estimator

4 Experimental Results

To evaluate the proposed noise estimator, the noisy speech is mixed with a suddenly increase level of Factory noise. Compare with Lin's estimator [4], the results are shown in Fig. 3 and Fig. 4. Fig.3 illustrates the comparison of the proposed noise estimator between Lin's one. Fig.3 (a) shows a phrase "May I Help You ?", which is pronounced from a man in English. It is observed that a background noise suddenly increase in 22000^{th} sample (or 2.75 sec under 8KHz sampling rate). Then, the noise power spectrum is estimated by Lin's noise estimator and a clean speech signal is produced by power spectral subtraction (PSS). The results are displayed in Fig.3 (b). It is observed that the noise is not removed completely later 22000^{th} sample (or 2.75 sec). Fig.3 (c) shows the estimated noise signal from Lin's noise estimator. It is found that the suddenly increase level of Factory noise is detected in 22000^{th} sample; however, the amplitude of estimated noise is enough large to meet idea noise later 22000^{th} sample. Fig.3 (d) shows the clean signal is generated by the proposed noise estimator and PSS. Compare with Fig.3 (b), the proposed noise can be performed well in suddenly changing level of noise. In Fig.3 (e), the estimated noise signal is produced by the proposed method. Fig.4 shows the spectrograms of a noisy speech signal, an enhanced speech signal of Lin's estimator and that of the proposed estimator, respectively. Similarly, due to the VAD is robust against a changing level of noise, the performance of speech enhancement in the proposed method is better than in Lin's method.

A noisy speech database is generated by applying various segmental SNRs in order to measure the segmental relative estimation error for various types and levels of noise. The segmental relative estimation error (SegErr) is defined by

$$SegErr = \frac{1}{M}\sum_{m=0}^{M-1}\frac{\sum_{\omega}[\bar{N}(\omega,m)-N(\omega,m)]^2}{\sum_{\omega}N^2(\omega,m)}. \tag{17}$$

Table I shows the outcomes of the SegErr measured by the proposed estimation method for four noise types with the SNRs range [-5 to 25dB]. The proposed approach is superior to the other methods.

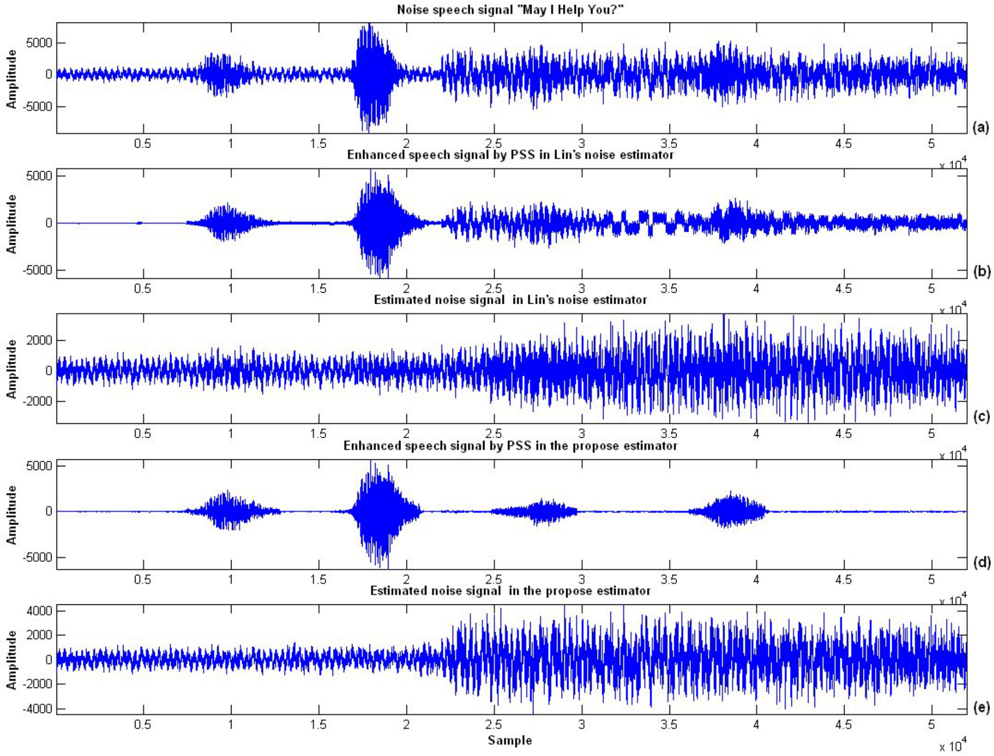
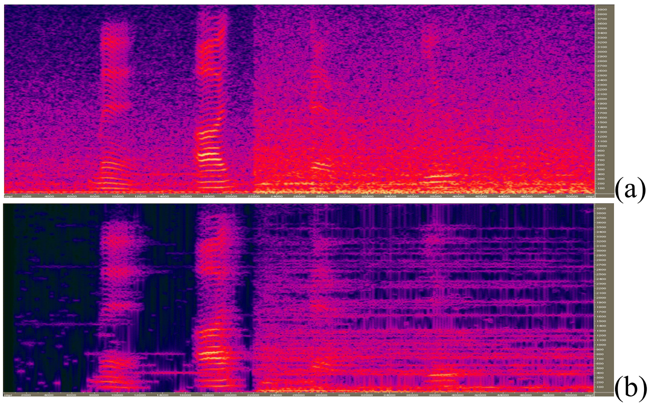


Fig. 3. Waveform of time signal



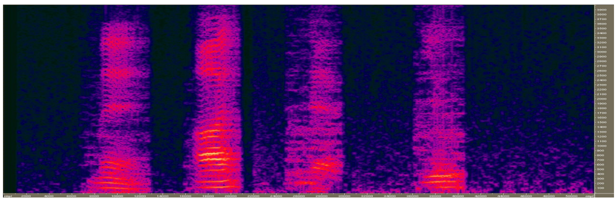

(c)

Fig. 4. Speech spectrogram (a) noisy speech signal (factory interior noise, suddenly raise in 2.75 sec) (b) speech enhanced with Lin's noise estimator (c) speech enhanced with the proposed noise estimator

Table 1. Example table

Input SegSNR [dB]	Car noise		Factory noise		Babble noise		White noise	
	Proposed	MCRA	Proposed	MCRA	Proposed	MCRA	Proposed	MCRA
-5	0.091	0.132	0.103	0.135	0.115	0.153	0.078	0.095
0	0.085	0.129	0.095	0.132	0.101	0.146	0.068	0.084
5	0.081	0.115	0.086	0.118	0.098	0.127	0.065	0.081
25	0.075	0.108	0.081	0.111	0.095	0.116	0.061	0.079

5 Conclusion

In this paper, a fast noise estimator, which is well suitable for suddenly varying level of noise, is presented. Based on the robust VAD, the speech decision can be determined accurately, and then the proposed algorithm can select the noise spectrum estimation which is suitable for the current frame. Unlike other method [1,3], the adaptation of this time and frequency dependent smoothing parameter does not depend on a specific time window and then updated continuously. Compare with Lin's estimator, the experimental results illustrate that the proposed estimator can remove the noise power spectrum by PSS.

6 Acknowledgement

The authors would like to thank the Promoting Academic Excellence of Universities for financially supporting this research under Contract No.91x104 Ex-91-E-FA06-4-4.

References

[1] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Trans. on Speech and Audio Processing*, vol. 9, no. 5, pp. 504-512, July 2001.

[2] G. Doblinger, "Computationally efficient speech enhancement by spectral minima tracking in subbands," *Proc. EUROSPEECH*, pp. 1513-1516, 1995.

[3] I. Cohen and B. Berdugo, "Noise estimation by minima controlled recursive averaging for robust speech enhancement," *IEEE Signal Processing Letters*, vol. 9, no. 1, Jan. 2002.

[4] L. Lin, W. H. Holmes, and E. Ambikairajah, "Adaptive noise estimation algorithm for speech enhancement," *Electronics Letters*, vol. 39, no. 9, pp. 754-755, May, 2003.

[5] J.L. Shen, J.W. Hung, and L.S. Lee, "Robust entropy-based endpoint detection for speech recognition in noisy environments,"*Proc. ICSLP-98*, 1998.