International Journal of the Physical Sciences Vol. 6(22), pp. 5361-5363, 2 October, 2011
Available online at http://www.academicjournals.org/IJPS
DOI: 10.5897/IJPS11.1277
ISSN 1992 - 1950 ©2011 Academic Journals

Short Communication

A comparison between minimum-statistics and soft-decision noise estimation methods

Roohollah Abdipour and Siamak Rasoulzadeh*

Department of Computer Engineering, Malayer Branch, Islamic Azad University, Malayer, Iran.

Accepted 2 September, 2011

Speech enhancement has been used in a variety of applications such as mobile phones, hearing aids and speech recognition systems. One of the most fundamental steps in many speech enhancement methods is noise estimation. Inexact noise estimation results in undesired effects in the enhanced signal. While over-estimating the noise leads to speech distortion, noise under-estimation leaves some annoying noise in the enhanced signal. Recently, minimum-statistics and soft-decision noise estimation methods have been paid high attention. In this paper, we evaluate these methods and compare their performance in speech+noise conditions.

Key words: Noise estimation, speech enhancement, minimum statistics, soft decision.

INTRODUCTION

Many speech enhancement methods assume that the noise power spectral density (PSD) is known apriori. Since this assumption is not valid in real conditions, the noise PSD should be estimated. So, the overall performance of the speech enhancement method will be dependent on that of noise estimation method.

The elementary voice activity detectors (VAD)-based noise estimation methods try to detect the silence intervals and update the noise estimation in these intervals. Because of low precision of VADs, these methods do not perform well. Among existing single-microphone noise estimation methods, the well-known minimum-statistics (MS) and soft-decision (SD) methods have been paid more attention. MS method is based on the idea that the minimum of the noise power in a sufficiently large time window can be considered as an estimation of the noise power (Martin, 1994, 2001; Martin and Lotter, 2001; Doblinger, 1995). The premise behind this idea is that in silence periods, the noisy signal power decreases to the noise power. Since there exist short silence periods between syllables and words, this method has the potential to track the noise power even in speech presence intervals.

In SD method, the speech presence probability in subbands is estimated and used to calculate a smoothing parameter needed in noise estimation step (Cohen, 2002, 2003). Although, this method performs well in stationary and semi-stationary noise conditions, its performance deteriorates in low signal-to-noise ratios (SNRs) and non-stationary noise conditions (Rangachari and Loizou, 2006). In this paper, we will review these methods and compare their performance with each other.

MINIMUM-STATISTICS METHOD

Assuming uncorrelated additive noise, the short-time Fourier transform (STFT) of input signal can be written as:

$$X(\lambda, k) = S(\lambda, k) + D(\lambda, k) \qquad (1)$$

where X, S and D show the STFT of noisy, clean and noise signals, and λ and k are the frame number and frequency bin index, respectively.

The PSD of the input signal ($P_X(\lambda, k)$) is obtained by recursively smoothing the input signal:

$$P_X(\lambda, k) = \alpha(\lambda, k).P_X(\lambda-1, k) + (1-\alpha(\lambda, k)).|X(\lambda, k)|^2 \qquad (2)$$

where $\alpha(\lambda, k)$ is the smoothing parameter. Noting the fact that in silence periods the noisy signal power decreases to the noise power, and noting that there exists short silence periods even during speech activity, the noise PSD can be estimated by finding

*Corresponding author. E-mail: siam.rasoolzade@gmail.com.

the minimum of the $P_X(\lambda,k)$ in D last samples ($P_{\min(\lambda,k)}$), and multiplying it to a bias compensation factor (B) (Martin, 2001, 2006; Martin and Lotter, 2001):

$$P_N(\lambda,k) = B(\lambda,k).P_{\min(\lambda,k)} \quad (3)$$

The smoothing parameter $\alpha(\lambda,k)$ should satisfy contradictory conditions. It should be near to one to lead to higher smoothing in silence periods. On the other hand, it should be small to let $P_X(\lambda,k)$ closely track the rapid changes of input spectrum in speech and periods. So, a constant value as in Martin (1994) is not appropriate. In Martin (2001), the following relation is proposed for $\alpha(\lambda,k)$:

$$\hat{\alpha}(\lambda,k) = \frac{\alpha_{\max}\ .\alpha_c(\lambda,k)}{1+\left(\frac{P_X(\lambda-1,k)}{\hat{P}_N^2(\lambda-1,k)}-1\right)^2} \quad (4)$$

where $\hat{P}_N^2$ and $\alpha_{\max}$ are estimated noise power and maximum of smoothing parameter, respectively, and $\alpha_c(\lambda,k)$ is a correcting factor which is calculated as:

$$?_?(?) = \frac{1}{1+(\frac{1}{?}\sum_{?=0}^{?-1} ?_?(?-1,?)\ /\ \frac{1}{?}\sum_{?=0}^{?-1}|?(?,?)|^2-1)^2} \quad (5)$$

The correcting factor is included to lessen the effects of incorrect smoothing parameters due to incorrect noise estimation in previous frames. The bias compensation factor is calculated as (Martin, 2001):

$$D(\lambda,k) \approx 1+(D-1)*\frac{2(1-M(D))}{\frac{2\hat{P}_N^4(\lambda-1,k)}{\widehat{var}\{P_X(\lambda,k)\}}-2M(D)} \quad (6)$$

where $\widehat{var}$ denotes the estimated variance, and D and *M(D)* are constant numbers which are respectively set to 15 and 0.668 in our experiments.

SOFT-DECISION (SD) METHOD

SD method tries to remedy the problems in VAD-based methods. In VAD-based methods, the silence periods are detected and then, the noise power is updated in silence periods. Formally, if H_0 and H_1 represent the silence and speech hypothesizes, the noise power is estimated as:

$$H_0(\lambda,k):\ \hat{P}_N^2(\lambda,k) - \alpha_d\hat{P}_N^2(\lambda-1,k)+(1-\alpha_d)|X(\lambda,k)|^2$$
$$H_1(\lambda,k):\ \hat{P}_N^2(\lambda,k) = \hat{P}_N^2(\lambda-1,k)$$

where $\alpha_d = 0.95$ is a constant smoothing parameter.

VAD-based methods face with two problems: performance fall due to low VAD's precision and lack of noise updating in speech periods. To overcome these problems, in SD method, the conditional speech presence probability in frequency bin *k* of frame λ (that is, $p(\lambda,k)$) is calculated and then, a soft transition between speech absence and presence hypotheses is performed:

$$\hat{P}_N^2(\lambda,k) = p(\lambda,k)\hat{P}_N^2(\lambda-1,k)+(1-p(\lambda,k))\left[\alpha_d\hat{P}_N^2(\lambda-1,k)+(1-\alpha_d)|X(\lambda,k)|^2\right] \quad (7)$$

The speech presence probability is calculated as (Cohen, 2003):

$$p(\lambda,k) = \left\{1+\frac{q(\lambda,k)}{(1-q(\lambda,k))(1+\xi(\lambda,k))\exp(-\upsilon(\lambda,k))}\right\}^{-1} \quad (8)$$

where $\upsilon = \frac{\gamma\xi}{1+\xi}$, ξ and γ are apriori and apostriori signal to noise ratios (Cohen, 2002, 2003), respectively, and q is the speech absence probability which is calculated as:

$$?(?,?) = \begin{cases} 1, & \text{if } \gamma_{\min}(\lambda,k) \le 1 \text{ and } \varsigma(\lambda,k) < \varsigma_0 \\ (\gamma_1-\gamma_{\min}(\lambda,k))/(\gamma_1-1) & \text{if } 1 < ?_{\min}(\lambda,k) \le \gamma_1 \text{ and } \varsigma(\lambda,k) < \varsigma_0 \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

$\gamma_{\min}$ and ς are instantaneous and smoothed aposteriori signal to noise ratios (Cohen, 2003), and $\varsigma_0 = 1.67$ and $\gamma_1 = 3$ are constant threshold values.

EVALUATION

Experiments have been conducted on a noisy file with white Gaussian noise at SNR = 15 db, containing a male person's speech sampled at 16 kHz. We used the hamming window of 256 point size. The speech file was selected from Noizeous database. The noisy file contained 3 s of silence in the beginning and the end, and 3 s of speech in between. The estimated noise power using MS and SD method is shown in Figure 1.

The benefit of SD method, rather than its simplicity and low computational cost, is its acceptable precision in speech absence intervals (Frames 1 to 80 in Figure 1). But, since SD method does not update the noise estimation in speech presence intervals (when $p(\lambda,k) = 1$), the estimated noise power will remain unchanged in these intervals (Frames 180 to 300). Therefore, SD method is not able to track the noise changes when speech is surely present.

Besides, the precision of SD method depends on the precision of speech presence probability estimation method. If speech presence periods are wrongly classified as silence, the noise will be over-estimated (Frames 310 to 340). On the other hand, if silence intervals are determined as speech ones, the estimated noise power will not be updated (Frames 360 to 380).

The main superiority of MS method is its ability to update noise estimation in speech presence intervals (Frames 180 to 310). The time and frequency dependent smoothing parameter in MS method enables it to track the rapid changes of speech signal, and so, detects the short durations of silence between syllables and words, which are used in noise estimation.

But MS method suffers from its high computational cost. Besides, finding the minimum in a relatively long window introduces some delays in noise estimation.

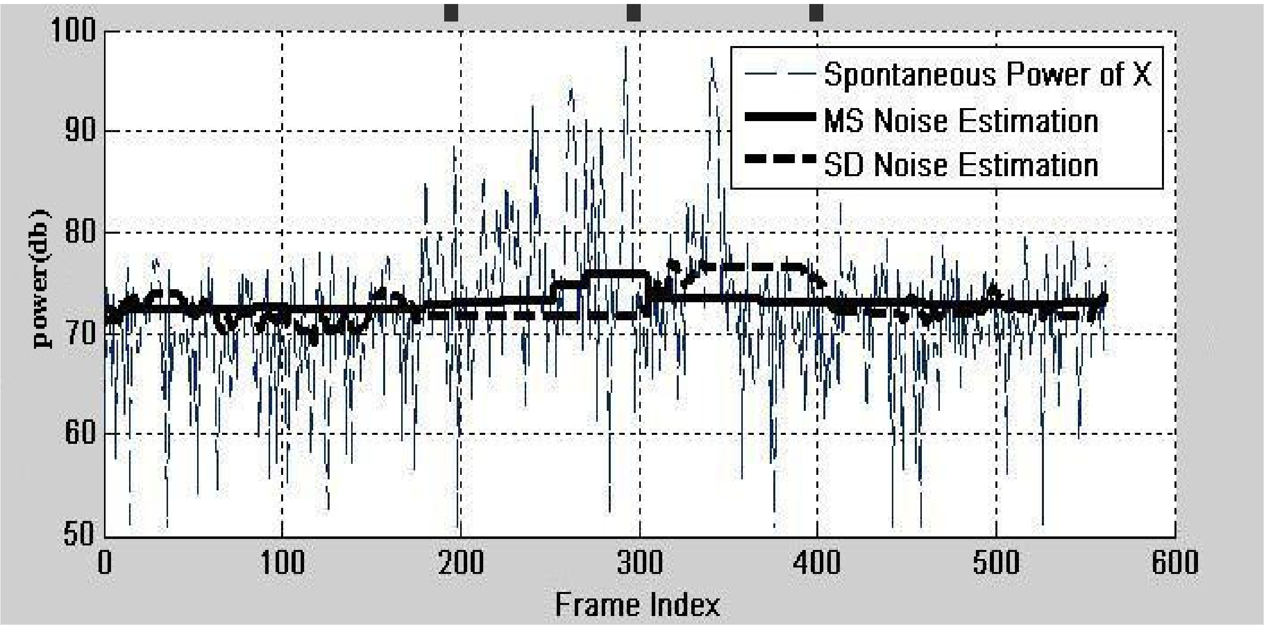

Figure 1. Comparison between MS and SD methods.

Rather, since the variance of the smoothed spectrum is twice that of an ordinary smoothed spectrum (Cohen, 2003; Rangachari and Loizou, 2006; Martin, 2006), the distance between minimum and mean increases and so, a larger bias compensation factor is needed.

Conclusion

The well-known MS and SD noise estimation methods were analyzed and compared with each other. While MS method is able to track noise changes in speech presence intervals, it suffers from high computational costs and noticeable noise tracking delays. SD method has the benefits of implementation simplicity, low computational cost and good precision in silence intervals, but it cannot track the noise changes when the speech is surely present.

REFERENCES

Doblinger G (1995). Computationally Efficient Speech Enhancement by Spectral Minima Tracking in Subbands. In Proc. EuroSpeech, 2: 1513–1516.

Cohen I (2002). Noise Estimation by Minima Controlled Recursive Averaging for Robust Speech Enhancement. IEEE Signal Proc. Lett., 9(1): 12-15.

Cohen I (2003). Noise Spectrum Estimation in Adverse Environments: Improved Minima Controlled Recursive Averaging. IEEE Trans. Speech Audio Proc., 11: 466-475.

Martin R (1994). Spectral Subtraction Based on Minimum Statistics. Proc. Eur. Signal Proc. Conf. (EUSIPCO), 3: 1182–1185.

Martin R (2001). Noise Power Spectral Density Estimation Based On Optimal Smoothing and Minimum Statistics, IEEE Trans. Speech and Audio Proc., 9(5): 504-512.

Martin R, Lotter T (2001). Optimal Recursive Smoothing of Non-stationary Periodograms, Proc. of IWAENC-2001, 1: 167-170.

Martin R (2006). Bias Compensation Methods for Minimum Statistics Noise Power Spectral Density Estimation. Signal Proc. 86: 1215-1229.

Rangachari S, Loizou PC (2006). A Noise-Estimation Algorithm for Highly Non-stationary Environments. Speech Commun., 48: 220-231.