Adaptive Noise Estimation Using Second Generation and Perceptual Wavelet Transforms

Essa Jafer and Abdulhussain E. Mahdi

Department of Electronic and Computer Engineering
University of Limerick, Ireland
essa.jafer@ul.ie hussain.mahdi@ul.ie

Abstract

This paper describes the implementation and performance evaluation of three noise estimation algorithms using two different signal decomposition methods: a second-generation wavelet transform and a perceptual wavelet packet transform. These algorithms, which do not require the use of a speech activity detector or signal statistics learning histograms, are: a smoothing-based adaptive technique, a minimum variance tracking-based technique and a quantile-based technique. The paper also proposes a new and robust noise estimation technique, which utilises a combination of the quantile-based and smoothing-based algorithms. The performance of the latter technique is then evaluated and compared to those of the above three noise estimation methods under various noise conditions. Reported results demonstrate that all four algorithms are capable of tracking both stationary and non-stationary noise adequately but with varying degree of accuracy.

1. Introduction

Reliable noise estimation remains a challenging problem in many speech enhancement and noise compensation tasks. Accurate instantaneous noise power estimation is crucial for the success and robustness of any single-channel speech enhancement system. Over the last few years, various noise estimation techniques have been proposed and their performance evaluated. These include techniques that are based on tracking the minima of the noise power [1,2], and those which utilise a quantile computation algorithm [3-4]. Although efficient, all these techniques involve relatively high computational complexity.

In this paper, three different and recently-reported noise estimation algorithms: (a) an adaptive technique with a smoothing parameter that depends on the estimated subband signal-to-noise ratio (SNR) [5]; (2) a one-pass quantile-based technique [6]; and (3) a technique that is based on tracking the minimum variance of the subband noisy signal [7], are considered. First, we describe the implementation of these three algorithms using two signal representation schemes that provide different resolutions: the first is based on the application of second generation wavelet transform (SGWT) [8,9], and the second is based on critical-band motivated perceptual wavelet packet decomposition (PWPD) [10]. We then propose a new and robust wavelet-based noise estimation technique that is based on combining the best features of algorithms (1) and (2). This is then followed by performance evaluation of all the above four noise estimation techniques using a variety of speech signals distorted by different types of noise. The evaluation has been affected by using an objective assessment measure based on the average relative error in estimated noise.

2. Wavelet-based speech signals decomposition

2.1. Second generation wavelet (SGWT)

Classical wavelet transform (WT) is generated by using translation and dilation of a single mother wavelet function. This construction method requires a regular mesh and unbounded domain. Classical WT, therefore, works well for infinite or periodic signals, but special adaptations of the basis functions near the boundaries are required in order to handle non-periodic boundary conditions which are often encountered in natural speech. The second generation wavelet transform (SGWT) have been introduced to provide such adaptations as well as maintaining other powerful properties offered by classical WT such as time-frequency localization, multi-resolution and fast implementation [8]. The basic idea behind the second-generation wavelet is to first split a signal, $x(n)$, into an even set, $\{x(n): n \text{ even}\}$, and an odd set, $\{x(n): n \text{ odd}\}$, by predicting the odd signal from the even part. What is missed by the prediction is called the detail. The even samples are then adjusted to serve the coarse version of the original signal. The adjustment is needed to maintain the same average for the fine and coarse versions of the same signal. The above process can be summarized as follows (see Figure 1):

a) Split data: even and odd.
b) Predict odd using even: detail = odd – P (even).
c) Update even using detail: Coarse=even + U (detail).

The inverse transform can be easily constructed by "rewiring" the forward transform, as illustrated in Figure 1. The process of computing a prediction and recording the detail is called a lifting step. In general, the lifting scheme speeds up the implementation as compared to the case of classical WT. All operations within one lifting step can be done in parallel while the only the sequential part is the order of the lifting operations, resulting in an adaptive wavelet transform [9].

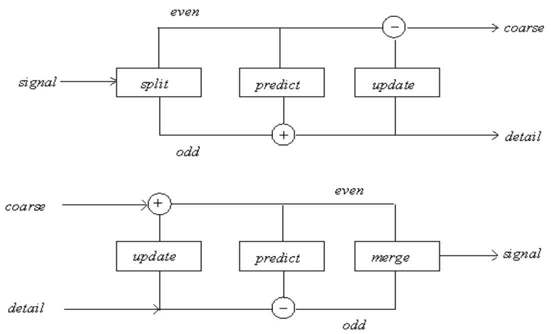

Figure 1. Representation of the forward and inverse SGWTs

2.2. Perceptual wavelet transform

For the implementation of the perceptual WT, we utilise a wavelet packet decomposition (PWPD) scheme designed to represent the critical bands of a given speech signal as widely used in perceptual auditory modeling. The scheme, which was first proposed by Black and Zeytinoglu [10], is based on an efficient 6-stage tree structure decomposition using 16-tap FIR filters derived from the Daubechies wavelet function, and provides for an exact invertible decomposition. For speech signals sampled at 8 kHz, this decomposition results in 18 critical bands.

3. Description of the noise estimation algorithms

In this Section, brief descriptions of the three different noise estimation algorithms and their wavelet-based implementation are given. In what follows we assume that $y(n)$ represents a band limited and sampled noisy speech signal, consisting of a clean speech signal $s(n)$ and a noise signal $w(n)$, such that $y(n) = s(n) + w(n)$. The noisy speech is first decomposed into a appropriate number of bandpass signals, $y_i(n)$, where i denotes the subband index, using either the SGWT or the PWPD, then framed using an appropriate sliding window. Also, $\hat{\sigma}^2_{w_i} = E\{w_i^2\}$ will be used to denote the estimated noise power (or noise variance) at frame p.

3.1. Adaptive smoothing-based noise estimation

This adaptive noise estimation technique is based on the use of a smoothing parameter that is controlled by the estimated subband *posteriori* SNR [5]. In this technique, the noise and speech are assumed to be independent signals and that the noise power changes slowly. The subband noisy signal power (or variance), $\sigma^2_{y_i}(p) = E\{y_i^2(n)\}$, is estimated on a frame-by-frame basis using [5]:

$$\hat{\sigma}^2_{y_i}(p) = \frac{1}{N}\sum_{n=0}^{N-1} y_i(pN+n) \qquad (1)$$

where $\hat{\sigma}^2_{y_i}(p)$ is the estimated noise power calculated at frame p, and N is the size of the frame. Similarly, the subband noise power is estimated using the smoothing filter:

$$\hat{\sigma}^2_{w_i}(p) = \alpha_i(p)\hat{\sigma}^2_{w_i}(p-1) + (1-\alpha_i(p))\hat{\sigma}^2_{y_i}(p) \qquad (2)$$

where $\hat{\sigma}^2_{w_i}(p)$ is the estimate of subband noise power at frame p. The smoothing parameter $\alpha_i(p)$ at frame p is chosen as:

$$\alpha_i(p) = 1 - \min\left\{1, \left(\frac{\hat{\sigma}^2_{y_i}(p)}{\bar{\sigma}^2_{w_i}(p-1)}\right)^{-Q}\right\} \qquad (3)$$

where Q is an integer and $\bar{\sigma}^2_{w_i}(p-1)$ is the average of the noise estimates of the previous 5 to 10 frames, such that

$$\bar{\sigma}^2_{w_i}(p-1) = 1/10\sum_{k=1}^{10}\hat{\sigma}^2_{w_i}(p-k)$$

(4)

3.2. Quantile-based noise estimation

The quantile noise estimation method considered here is based on a one-pass algorithm. After decomposition, each subband noisy signal is framed into frames of length L_{frame}. Let $L_{win} > L_{frame}$ be the length of a finite window observation of $y_i(n)$, ranging from 200ms to 2000ms. The method involves first sorting the previous set of data over the last M frames $\{y_i^p(n), n = 0, \cdots, L_{win}-1\}$ in an ascending order of their values according to the requirement of the quantile-based approach [6]. The noise power in the ith subband of the pth frame, $\hat{\sigma}^2_{w_i}$, is then estimated as:

$$\hat{\sigma}_{w_i}{}^2 = \beta \sum_{j=0}^{\mathrm{int}(q.L_{win})} (y_i^p(j))^2 \Big/ L_{win} \qquad (5)$$

where β is an appropriate scaling factor and $q = 0.2$. Here, L_{frame} and L_{win} are chosen to be equal to 64 ms and 512 ms, respectively, with the frames overlapped by 50 %.

3.3. Minimum variance tracking-based noise estimation

In this algorithm, both the noisy signal and the noise are considered to be stationary over a short period of time, such that the variance can be estimated on a frame-by-frame basis. The noisy signal variance, $\sigma^2_{y_i}$, for each band is calculated as:

$$\sigma^2_{y_i}(p) = \alpha_i\,\sigma^2_{y_i}(p-1) + (1-\alpha_i)\sigma^2_{y_i,new}(p) \qquad (6)$$

where

$$\sigma^2_{y_i,new}(p) = \frac{1}{N}\sum_{k=0}^{N-1} y_i{}^2(pN+k) \qquad (7)$$

is the most recent approximation of the noisy signal variance using the new data at frame p. The parameter α_i is a smoothing factor chosen as $0.45 \le \alpha_i \le 0.95$.

The noise estimate $\sigma^2_{w_i}(p)$ is updated such that

$$\sigma^2_{w_i}(p) = \alpha_i\,\sigma^2_{w_i}(p-1) + (1-\alpha_i)\sigma^2_{w_i,new}(p) \qquad (8)$$

where $\sigma^2_{w_i,new}$ is the minimum value of $\sigma^2_{y_i}(p)$ in the neighboring frames, i.e. if $\sigma^2_{y_i}(p-1) < \sigma^2_{y_i}(p)$ &

$\sigma^2_{y_i}(p-1) < \sigma^2_{y_i}(p-2)\ldots \ \& \sigma^2_{y_i}(p-1) < 2\sigma^2_{w_i}(p-1)$,
then

$$\sigma^2_{w_i,new}(p) = \sigma^2_{y_i}(p-1) \tag{9}$$

Otherwise $\sigma^2_{w_i,new}(p) = \sigma^2_{w_i}(p-1)$.

2.4. A new noise estimation technique

A new noise estimation technique based modification of the quantile-based method presented in Section 3.2 is proposed here. The modification is based on the addition of a smoothing parameter that depends on the estimated subband SNR, similar to that used in the smoothing-based technique presented in Section 3.1, such that a new quantile-based noise estimate that can be updated adaptively is obtained.

The new technique proceeds as follows. The noise power in the ith subband of the pth frame, $\hat{\sigma}^2_{w_i}$, is estimated as in the standard quantile-based method (eq. 5). This estimate of the noise power is considered here to be equivalent to the average of the noise estimates used in eq.3. Based on this, a smoothing factor, $\alpha_i(p)$, is then introduced such that:

$$\alpha_i(p) = 1 - \min\left\{1, \left(\frac{\hat{\sigma}^2_{y_i}(p)}{\sigma^2_{w_i,quantile}(p-1)}\right)^{-Q}\right\} \tag{10}$$

where $\sigma^2_{w,quantile}$ is the noise power in the ith subband of the pth frame as calculated by using eq. 5. As will be discussed in Section 4, our experimental results have shown that in most cases setting β=1 and α=0.5 result in the best performance of this new noise estimation technique.

4. Performance evaluation.

In the first part of the evaluation process, 60 seconds of speech divided into 10 signals, each of an average length of 6 seconds and sampled at 8 kHz, as acquired from the TIMIT database. The distorted speech frames are overlapped by 50% and different types of noise have been used to test the four noise estimation techniques using the SGWT. The noisy speech signals were decomposed into 6 bands (details) using the dB(7-9) wavelet filter [9], and each of the four techniques was then used to estimate the added noise. The real (solid line) and estimated noise for band 3 of the decomposed signal (0.5-1 kHz) resulting from each technique are shown in Figures 2 and 3 for the cases of additive white noise (AWGN) and pink noise, respectively. The second part of the evaluation process deals with the perceptual wavelet decomposition. Figure 4 shows the real and estimated noise for band 15 of the decomposition, for the case of AWGN. In Figures 2-4, (a) corresponds to the adaptive smoothing-based method, (b) quantile-based method, and (c) the minimum variance tracking-based method. Also, in (a), (b) and (c), a dashed line is used to mark the estimated noise, while in (d) a dashed line is used to mark the noise estimate obtained by a quantile-based method and a dotted line for that obtained by the proposed method. In Figure 5, the up-dating with time of the parameter $\alpha_i(p)$ used in eq.3 is shown for the case of one of SGWT subbands (band 3) of the noisy speech.

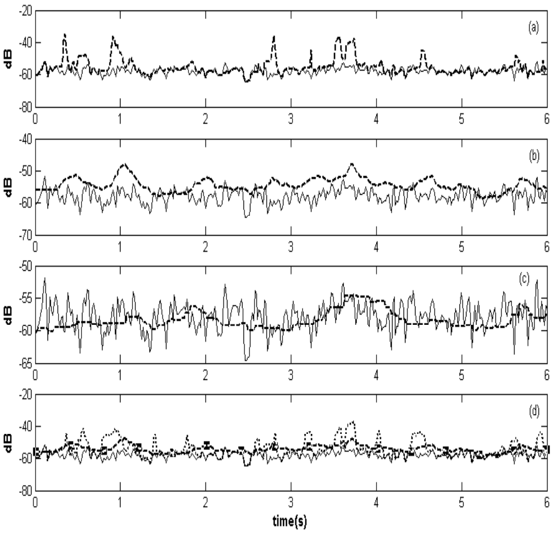

Figure 2: Real and estimated noise using SGWT-based noise estimation with AWGN at 5dB

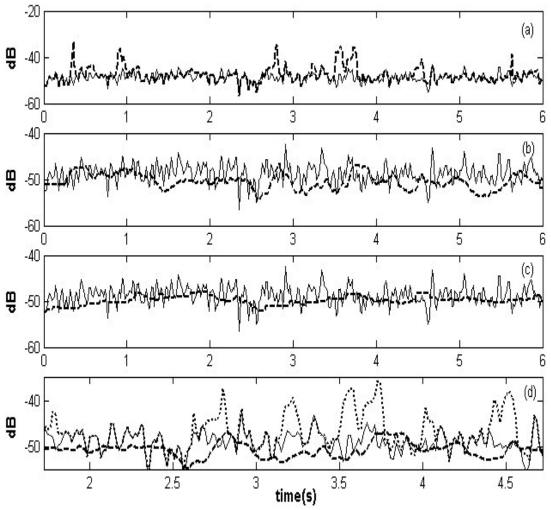

Figure 3: Real and estimated noise using SGWT-based noise estimation with pink noise at 0dB

To provide an objective performance measure, we also calculated the average relative error factor in the estimated noise defined as:

$$ARE = \frac{1}{N_{frame}} \sum_p \frac{\left|\hat{\sigma}^2_{w_i}(p) - \sigma^2_{w_i}(p)\right|}{\sigma^2_{w_i}} \tag{12}$$

where N_{frame} represents the number of frames in the test signal. Using this factor, tables 1 and 2 illustrate the performance of the four presented noise estimation techniques for one SGWT subband (band 1 for the SGWT case and band 7 for the PWPD) over different SNRs. Here, T1, T2, T3 and T4 refer to the first, second, third and the proposed noise estimation techniques in the sequence presented in Section 3.

It is obvious from this evaluation that all the four techniques considered here demonstrate capability in tracking various types of noise, but their performance accuracy varies depending on the rate of change of the noise under test. The minimum variance tracking-based method seems to offer the best performance in tracking the average noise variation. On the other hand, the adaptive smoothing-based method noise can track rapid changes of stationary and non-stationary noise depending on the value of smoothing parameter. Presented results also demonstrate that the performance of the quantile-based noise estimation method was improved when combined with the adaptive noise estimation method, as proposed in our new noise estimation technique. In particular, significant improvement was achieved by the proposed method for cases of speech signals of relatively low SNRs.

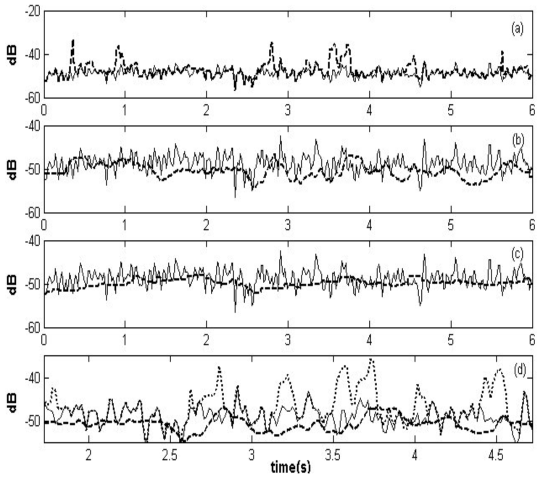

Figure 4: Real and estimated noise using PWPD-based noise estimation with F16 Cockpit noise at 0dB

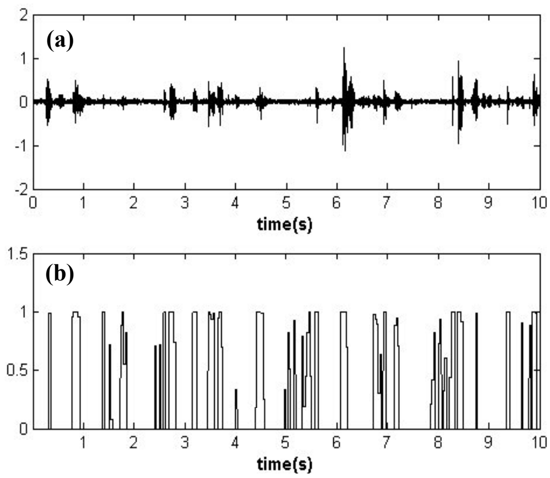

Figure 5: (a) A subband noisy signal at 10dB with, and (b) corresponding updating of $\alpha_i(p)$.

5. Conclusions

The implementation of three-noise estimation algorithms using both the SGWT and the PWPD have been investigated, and their performance evaluated and compared under different noisy conditions. Our results demonstrate that all three algorithms are capable of tracking both stationary and non-stationary noise, but with varying degree of accuracy depending on the level and rate of change of the noise under consideration. Reported results also show that by modifying the standard quantile-based algorithm, a new adaptive and robust noise estimation method with relatively superior performance to the above three techniques for cases of high additive noise, can be achieved.

Table 1: Average relative error *ARE* in band-1 SGWT for the four noise estimation methods.

SNR (dB)	*ARE* – WHITE NOISE			
	T1	T2	T3	T4
10	0.84	1.5	0.16	0.65
5	0.31	0.28	0.11	0.63
0	0.11	0.33	0.11	0.21
-5	0.061	0.64	0.12	0.07

Table 2: Average relative error *ARE* in band-7 PWPD for the four noise estimation methods.

SNR dB	ARE-PINK NOISE			
	T1	T2	T3	T4
10	4.09	2.97	0.48	3.00
5	1.41	1.01	0.48	1.22
0	0.56	0.50	0.48	0.47
-5	0.30	0.37	0.48	0.22

6. References

[1] Rainer, M., "An efficient algorithm to estimate the instantaneous SNR of speech signal ", *Proc. Eurospeech'93*, 1093-1096, 1993.

[2] Rainer, M, "Noise power spectral density estimation based on optimal smoothing and minimum statistics ", *IEEE Trans. Speech and Audio Proc*., 9(5): 504-512, 2001.

[3] Stahl, V., A. Fischer, A, and Bippus, R., " Quantile based noise estimation for spectral subtraction and weiner filtering*", ICASSP'2000*, Istanbul, 40:1875-1878, 2000.

[4] Evans, N., and Mason, J., "Time-Frequency quantile–based noise estimation", *Proc.EUSIPCO 2002.*

[5] Lin, L., Holmes, W., and Ambikairajah, E.,"An adaptive noise estimation algorithm from speech enhancement", *Proc. Australian Conf. on Speech Science and Techno.* Melbourne, 2002.

[6] Qiang, F. and Eric, W., " A novel speech enhancement system based on wavelet denoising", *Technical Report, OGI, School of Science and Engineering*, Oregon, 2003.

[7] Lin, L., Ambikairajah, E., and Holmes, W.," Speech enhancement for nonstationary noise environment", Proc. *Asia-Pacific Conference APCCAS '02*, 1:177-180, 2002.

[8] Sweldens, W.," The lifting scheme: a custom-design construction of biorthogonal wavelets", *Appl. Computer Harmon Analy.*, 3(2): 186-200, 1996.

[9] Daubechies, I. and Sweldons, W.," Factoring wavelet transforms into lifting steps", *J. Fourier Analy. Appl.;* 4(3): 247-269, 1998.

[10] Black, M., and Zeytinoglu, M.,"Computationally efficient wavelet packet coding of wide-band stereo audio signals", *Proc. ICASSP'95*, 3075-3078, 1995.