SECOND GENERATION AND PERCEPTUAL WAVELET BASED
NOISE ESTIMATION
E. Jafer, A. E. Mahdi
Department Department of Electronic and Computer Engineering
University of Limerick
Limerick
Ireland

Abstract: - The implementation of three noise estimation algorithms using two different signal decomposition
methods: a second-generation wavelet transform and a perceptual wavelet packet transform are described in
this paper. The 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, robust noise estimation technique,
which combines a quantile-based algorithm with smoothing-based algorithm. 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.

Key-Words: - Speech processing, Wavelet-transform, Second-Generation wavelet transform, Noise estimation.

1 Introduction
In many speech enhancement systems and noise
compensation tasks, reliable noise estimation
remains a challenging problem. 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 based on tracking the minima of
the noise power [1], and quantile-based methods [2].
Although efficient, all these techniques involve
relatively high computational complexity.
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) [3];
(2) a one-pass quantile-based technique; and (3) a
 technique that is based on tracking the minimum
variance of the subband noisy signal [4], are
considered in this paper. First, The implementation
of these three algorithms will be described using two
signal representation schemes: the first is based on
the application of second generation wavelet
transform (SGWT) [5], and the second is based on
critical-band motivated perceptual wavelet packet
decomposition (PWPD) [6]. A new and robust
wavelet-based noise estimation technique, that is
based on combining the best features of algorithms
(1) and (2) is then proposed. This is followed by
performance evaluation of all the above noise
estimation techniques using a variety of speech
signals distorted by different types of noise. The
evaluation has been effected by using an objective
assessment measure based on the average relative
error in estimated noise.

2 WAVELET-BASED SPEECH SIGNALS DECOMPOSITION

2.1 Perceptual wavelet packet decomposition (PWPD)

A perceptually motivated wavelet packet
decomposition (PWPD) scheme designed to
approximate the critical-bands of the speech, similar
to that reported by Black and Zeytinoglu [6], has
been utilized in this work. The model is based on an
efficient 6-stage decomposition tree, which is
constructed by using 16-tap FIR filter bank derived
from the Daubechies wavelet function. The scheme
also provides for an exact invertible decomposition.
For speech signals sampled at 8 kHz, this
decomposition results in 18 critical bands.

2.2 Second-generation wavelet (SGWT)

The second-generation wavelet involves first splitting 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 [5]. In general, the lifting scheme speeds up the implementation as compared to the case of classical WT.

Fig.1: Representation of the forward and inverse SGWTs

3 Description of the noise estimation algorithms

A brief description of the three different noise estimation algorithms and their wavelet-based implementation is given in this section. 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, \(\hat{\sigma}^2_{wj} = E[w_i^2]\) will be used to denote the estimated noise power (or noise variance) at frame \(p\).

3.1 Adaptive smoothing-based noise estimation

In this technique, the noise and speech are assumed to be independent signals and that the noise power changes slowly. The adaptive noise estimation technique is based on the use of a smoothing parameter that is controlled by the estimated subband posterior SNR [3]. The subband noisy signal power (or variance), \(\hat{\sigma}^2_{y_i}(p) = E[y_i^2(n)]\), is estimated on a frame-by-frame basis using [3]:

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

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:

\[
\hat{\sigma}^2_{w_j}(p) = \alpha(p)\hat{\sigma}^2_{w_j}(p-1) + (1-\alpha(p))\hat{\sigma}^2_{y_i}(p) \tag{2}
\]

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

\[
\alpha(p) = 1 - \min \left\{ 1, \left( \frac{\hat{\sigma}^2_{y_i}(p)}{\hat{\sigma}^2_{y_i}(p-1)} \right)^{-1} \right\} \tag{3}
\]

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

\[
\hat{\sigma}^2_{w_j}(p-1) = 1/10 \sum_{k=0}^{10} \hat{\sigma}^2_{w_j}(p-k) \tag{4}
\]

3.2 Quantile-based noise estimation

In this technique, each subband noisy signal is framed into frames of length \(L_{frame}\) after the decomposition. 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 \(\{y_i^j(n), n = 0, \cdots, L_{win}-1\}\) in an ascending order of their values according to the requirement of the quantile-based approach [2]. The noise power in the \(i\)th subband of the \(p\)th frame, \(\hat{\sigma}^2_{w_j}\), is then estimated as:

\[
\hat{\sigma}^2_{w_j} = \beta \sum_{j=0}^{\text{int}(qL_{win})} (y_i^j(p))^2 / L_{win} \tag{5}
\]

where \(\beta\) 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

Both the noisy signal and the noise are considered to be stationary over a short period of time in this technique, such that the variance can be estimated on a frame-by-frame basis. The noisy signal variance, \(\hat{\sigma}^2_{y_i}\), for each band is calculated as [4]:

\[
\hat{\sigma}^2_{y_i}(p) = \alpha(p)\hat{\sigma}^2_{y_i}(p-1) + (1-\alpha(p))\hat{\sigma}^2_{y_i,\text{new}}(p) \tag{6}
\]
where \( \sigma_{y_i,\text{new}}^2(p) = \frac{1}{N} \sum_{k=0}^{N-1} y_i^2(pN + k) \) (7)

is the most recent approximation of the noisy signal variance using the new data at frame \( p \). The parameter \( \alpha \) is a smoothing factor chosen as \( 0.45 \leq \alpha \leq 0.95 \). The noise estimate \( \sigma_{w_i}^2(p) \) is updated such that

\[
\sigma_{w_i}^2(p) = \alpha \sigma_{w_i}^2(p - 1) + (1 - \alpha) \sigma_{w_i,\text{new}}^2(p)
\]

where \( \sigma_{w_i,\text{new}}^2 \) is the minimum value of \( \sigma_{y_i}^2(p) \) in the neighboring frames, i.e. if

\[
\sigma_{y_i}^2(p - 1) < \sigma_{y_i}^2(p) \quad \& \\
\sigma_{y_i}^2(p - 1) < \sigma_{y_i}^2(p - 2) \ldots \\
\& \sigma_{y_i}^2(p - 1) < 2\sigma_{w_i}^2(p - 1)
\]

then

\[
\sigma_{w_i,\text{new}}^2(p) = \sigma_{y_i}^2(p - 1)
\]

Otherwise

\[
\sigma_{w_i,\text{new}}^2(p) = \sigma_{w_i}^2(p - 1)
\]

3.4 A new noise estimation technique

Based on the modification of the quantile-based method presented in Section 3.2, a new noise estimation technique 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 \( i \)th subband of the \( p \)th frame, \( \hat{\sigma}_{w_i}^2(p) \), 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.4). Based on this, a smoothing factor, \( \alpha(p) \), is then introduced such that:

\[
\alpha(p) = 1 - \min \left\{ \frac{\sigma_{y_i}^2(p)}{\sigma_{w_i,\text{quantile}}^2(p)} \right\}^\beta
\]

(10)

where \( \sigma_{w_i,\text{quantile}}^2 \) is the quantile-based estimated noise power in the \( i \)th subband. As will be discussed in the next Section our experimental results have shown that in most cases setting \( \beta = 1 \) and \( \alpha = 0.5 \) result in the best performance of this new noise estimation technique.

4 Performance Evaluation

In the first part of the evaluation process, a large number of speech signals, sampled at 8kHz with an average length of 6 seconds each, have been employed. The signals were obtained from the TIMIT database. For the purpose of evaluation, these signals were distorted by additive noise of various types and levels and overlapped by 50%. For the SGWT case, signals were decomposed into 6 bands (details) using dB(7-9) wavelet filter [5]. In Figure.2, the real (solid) and estimated noise power (dashed) resulting from the implementation of each technique on band.3 (0.5-1 kHz) of speech signal decomposed by SGWT for the case of pink noise. Figure.3 shows the real and estimated noise for band.7 of the PWD case and band.7 for the PWPD decomposition, for the case of AWGN. In Figures 2 and 3, (a) corresponds to the adaptive smoothing-based technique, (b) the quantile-based technique, (c) the minimum variance tracking-based technique, and (d) the proposed technique.

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

\[
ARE = \frac{1}{N_{\text{frame}}} \sum_p \left| \frac{\hat{\sigma}_{w_i}^2(p) - \sigma_{w_i}^2(p)}{\sigma_{w_i}^2(p)} \right|
\]

(12)

Where \( N_{\text{frame}} \) 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 subband (band.2 for the SGWT case and band.7 for the PWD) 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.

<table>
<thead>
<tr>
<th>SNR (dB)</th>
<th>ARE – PINK NOISE</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>1.45 5.92 0.45 1.22</td>
</tr>
<tr>
<td>5</td>
<td>0.60 4.68 0.64 0.68</td>
</tr>
<tr>
<td>0</td>
<td>0.17 0.70 0.14 0.21</td>
</tr>
<tr>
<td>-5</td>
<td>0.078 0.16 0.13 0.076</td>
</tr>
<tr>
<td>-10</td>
<td>0.050 0.50 0.14 0.082</td>
</tr>
</tbody>
</table>

Table 1: Average relative error ARE in band-2 SGWT for the four-noise estimation methods.
In the second part of the evaluation process, the four noise estimation techniques have been used as part of speech enhancement algorithm, which is based on classical soft-thresholding using PWPD. Figure 4, for example shows the improvement in average segmental signal-to-noise ratio SegSNR of the enhanced speech signal resulting from the application of the four noise estimation techniques.

The presented results indicates that the minimum variance tracking-based method provides the best performance in tracking the average noise variations, while at the same time requires the minimum amount of memory storage. On the other hand, the adaptive smoothing-based method noise offers the highest precision in terms of tracking the instantaneous variation of the noise particularly for cases of speech signals of relatively low SNR. Presented results also demonstrate that the adaptability and trackibility of the quantile-based noise estimation method was enhanced when a smoothing factor based on the posteriori SNR is introduced as the proposed method illustrates.
References:


