SPEECH ENHANCEMENT USING WAVELET PACKET TRANSFORM

Sungwook Chang, Sungil Jung, Y. Kwon*, and Sung-il Yang

Department of Electrical and Computer Engineering
Department of Physics*
Hanyang University, Korea
schang@ihanyang.ac.kr, sijung@ihanyang.ac.kr, yyhkwon@hanyang.ac.kr*, and syang@hanyang.ac.kr

ABSTRACT

In this paper, we present denoising algorithm using wavelet packet transform based on VAD (Voice Activity Detection). For the purpose, the weighting parameter estimation method based on speech dominant indicator and the modified Berouti’s method are suggested. Subjective evaluation shows that the performance of the proposed algorithm with wavelet bases is better than that with Fourier basis.

1. INTRODUCTION

In many speech processing applications, speech has to be processed in the background of undesirable noise. During the last decades, various approaches to reduce the noise have been proposed. Among them, spectral subtraction has been widely used. This method requires the estimation of statistical noise information. Hence, the accuracy of the estimated statistical noise information decides the performance of the enhancement system. Thus, accurate estimation of spectral information for a speech signal is essential. However, Fourier basis in which is assumed that the input speech is periodically sinusoidal cannot shows good estimation of spectrum in noisy environment. To overcome the problem of Fourier basis, we use wavelet packet transform for robust spectrum estimation.

The proposed denoising algorithm can be divided into two main procedures. The first procedure is VAD with wavelet packet transform [1]. And the other is a modified spectral subtraction method [2] based on speech dominant indicator (SDI) [1].

2. ROBUST VOICE ACTIVITY DETECTION

VAD by wavelet packet transform is used for noise estimation in noise only area. In VAD procedure, differential power spectrum [3] using wavelet packet transform is used for estimations of noise level and noise magnitude intensity (NMI) [1] as follows:

\[
D(k) = \left[ \sum_s \omega_k \left\{ WPT_{coef}(k + h) \right\}^2 \right]
\]

(1)

where \( D(k) \) and \( WPT_{coef} \) are the differential power spectrum of speech signal and coefficient of wavelet packet transform respectively. \( S \) and \( E \) denote the orders of the differential equation and \( \omega_k \) means weighting coefficients. In this procedure, we first perform noise level estimation using conventional noise estimation technique [2] as follows:

\[
\hat{N}_j(i) = \begin{cases} 
\alpha \cdot \hat{N}_{j-1}(i) + (1 - \alpha) \cdot P_j(i), & P_j(i) \leq \beta \cdot \hat{N}_{j-1}(i) \\
\hat{N}_{j-1}(i), & \text{otherwise}
\end{cases}
\]

(2)
which \(\alpha = 0.98\), \(1.5 \leq \beta < 2.5\)

\(i\) : wavelet packet filterbank index
\(j\) : frame index
\(P_j(i)\) : average power spectrum of \(i^{th}\) wavelet packet filterbank at \(j^{th}\) frame
\(\hat{N}_j(i)\) : estimated noise level

And NMI estimation is performed using below algorithm.

**Algorithm**: Estimation of noise magnitude intensity NMI

1. **Step 1**: Initialize noise estimation intensity, NMI = 0.
   Initialize wavelet packet filterbank number, \(n = 0\).
2. **Step 2**: If \(n\) is larger than the highest wavelet packet filterbank index, then exit.
   \(n = n + 1\);
   go to Step 2.
3. **Step 3**: If \(1 \leq \frac{\hat{N}(n)}{P(n)}\), then NMI = NMI + 1.

For robust VAD and estimation of spectral subtraction weight, we use the speech dominant indicator \(\kappa\) as follows:

\[
\lambda = \log_{10} \left( \frac{N_{FB}}{NMI} \right) \quad (3)
\]

\[
\kappa = 10 \log_{10} \left( \left( \prod_{i=1}^{n} P(i) \right)^{\frac{\alpha}{\beta}} \right)^{\lambda} \quad (4)
\]

where \(N_{FB}\) : number of filterbank

\(\gamma\) : experimental exponent parameter

a value between 1 and 3 is used, experimentally.

SDI with NMI provides good discrimination between noise and unvoiced speech in noisy environment.

Next, we find a threshold for robust detection of a boundary between speech and silence in noisy environment. The procedure is explained in ref[1].

Then we perform VAD procedure using SDI and the above thresholds.

### 3. DENOISING ALGORITHM

In denoising algorithm based on spectral subtraction, noise estimation and estimation of spectral subtraction weight are crucial parts. If a faulty estimated noise is applied to denosing process, it induces an elimination of speech component or sediment of a noise component. That is, the accuracy of the estimated statistical noise information decides the performance of the enhancement system. Hence, reliable estimation of spectral information for a speech signal is essential. However, Fourier basis cannot show good estimation of spectrum in noisy environment. And instead of Fourier basis, we use wavelet packet transform for robust spectrum estimation. It can provide a reliable spectral estimation. For that reason, we use a slightly modified noise estimation method.[2]

#### 3.1. Noise Estimation

We estimate noise spectrum by modified noise estimation method in noise area taken from VAD process as follows:

\[
\hat{N}_j(i) = \begin{cases} 
\alpha \cdot \hat{N}_{j-1}(i) + (1-\alpha) \cdot P_j(i), & \text{if } \hat{N}_j(i) < \beta \cdot \hat{N}_{j-1}(i) \\
\hat{N}_{j-1}(i), & \text{otherwise}
\end{cases}
\]

which \(\beta = 2\), \(\alpha = 0.98\)

\(1 \leq i \leq \text{filterbank size}\)

\(j = \text{frame index}\)

#### 3.2. Spectral Subtraction Weight

Berouti et al. proposed a modified version of the power subtraction rule in which the amount of noise subtraction depended on the SNR of the particular frame as shown in Eq. 8 and Eq. 9 [4].

\[
G(\omega) = \begin{cases} 
1 - \alpha \left( \frac{\hat{N}(\omega)}{|X(\omega)|} \right)^{\gamma_2}, & \text{if } \frac{|\hat{N}(\omega)|}{|X(\omega)|} < \frac{1}{\alpha + \beta} \\
\beta \left( \frac{\hat{N}(\omega)}{|X(\omega)|} \right)^{\gamma_2}, & \text{otherwise}
\end{cases}
\]

\[
|\hat{S}(\omega)| = G(\omega) \cdot |X(\omega)|
\]

\(\hat{N}(\omega)\): Estimated Noise Spectrum

\(X(\omega)\): Input Spectrum

\(G(\omega)\): The filter for Power Spectral Subtraction

\(\hat{S}(\omega)\): Estimated Speech Spectrum

In the above method, estimation of spectral subtraction weight is crucial. Thus, we propose a new estimation method of spectral subtraction weight for reliable denoising algorithm. The proposed method which depends on the SDI is shown below.

\[
\alpha_j = \left( \frac{K_{\max} - K_{\min}}{(K_i - K_{\max}) \cdot \lambda + K_{\max} - K_{\min}} \right)^2
\]

where \(\alpha_j, j\) and \(\lambda\) are spectral subtraction weight, frame index and normalization level, \(0.4 \leq \lambda \leq 0.6\)

Finally, we define the filter for magnitude spectral subtraction with estimated noise level found from Eq. 7 as shown in Eq. 11 and perform the spectral subtraction by Eq. 12.

\[
G_j(i) = \begin{cases} 
1 - \alpha_j \left( \frac{|\hat{N}_j(i)|}{|P_j(i)|} \right)^{\gamma_3}, & \text{if } \alpha_j \cdot |\hat{N}_j(i)| < |P_j(i)| \\
\beta, & \text{otherwise}
\end{cases}
\]

\[
|\hat{S}_j(i)| = G_j(i) \cdot |WPT_j(i)|
\]
which \(0 \leq \beta \leq 0.02\), \(1 \leq i < \text{filterbank size}\).

\[ j \]: frame index

\[ |\hat{N}_i(j)| \]: estimated noise level

\[ |P_i(j)| \]: average power spectrum of \(i^{th}\) wavelet packet

filterbank at \(j^{th}\) frame

In the proposed spectral subtraction method, we use spectrum magnitude which is known as more effective than power spectrum [4].

4. EVALUATION

In noisy environment, the SNR cannot be used as faithful indication of speech quality. Thus, we employ subjective tests for evaluation of the proposed method. For subjective tests, we use an informal listening test and spectrum test. TIDIGIT 64 is used as speech database and various noise (babble, leopard, pink, and volvo noise), taken from Noisex-92 database, is added for our evaluation. Filterbank is composed of 64 uniform bands both in Fourier basis and wavelet bases.

Fig.2 shows speech spectra obtained by the proposed algorithm with Fourier basis and wavelet bases. The proposed algorithm with wavelet bases yields the better result than that with Fourier basis for the most noise conditions. Also, in informal listening test, the proposed algorithm with wavelet bases shows better performance than that with Fourier basis as shown in Table 1. And, as shown in Fig.2 (d) and (e), Coiflet and Daubechies’ based result shows similar spectrogram.

However, as shown in Table 2, preference performance depends deeply on the characteristics of additive noise. That is, Daubechies’ basis is suitable for speech babble noise and Leopard (military vehicle) noise, and Coiflet basis for the others. This result is matched with the characteristics of the wavelet bases.

Table 3 shows the preference percentage between the outputs of the proposed denoising algorithm using wavelet packet transform based on Daubechies’ basis and noisy speech signal. Although the performance of informal listening test depends on the characteristic of background noise, most listeners preferred the output of the proposed algorithm to the non-processed noisy signal. But, there were cases that some listeners preferred the noisy signal due to the induced distortions and artifacts.

<table>
<thead>
<tr>
<th>Daubechies’ Basis</th>
<th>Coiflet Basis</th>
<th>No Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Babble 50 %</td>
<td>14 %</td>
<td>36 %</td>
</tr>
<tr>
<td>Leopard 79 %</td>
<td>21 %</td>
<td>0 %</td>
</tr>
<tr>
<td>Pink 7 %</td>
<td>53 %</td>
<td>40 %</td>
</tr>
<tr>
<td>Volvo 7 %</td>
<td>53 %</td>
<td>40 %</td>
</tr>
</tbody>
</table>

Table 2: Preference percentage between enhanced speech using Daubechies’ basis and Coiflet basis. (average of –5 dB, 0 dB, 5 dB, 10 dB and 15 dB)

<table>
<thead>
<tr>
<th>Proposed Algorithm</th>
<th>Noisy Signal</th>
<th>No Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Babble 68 %</td>
<td>32 %</td>
<td>0 %</td>
</tr>
<tr>
<td>Leopard 68 %</td>
<td>32 %</td>
<td>0 %</td>
</tr>
<tr>
<td>Pink 72 %</td>
<td>28 %</td>
<td>0 %</td>
</tr>
<tr>
<td>Volvo 76 %</td>
<td>16 %</td>
<td>8 %</td>
</tr>
</tbody>
</table>

Table 3: Preference percentage between the outputs of proposed denoising algorithm using wavelet packet transform based on Daubechies’ basis and noisy speech signal. (average of –5 dB, 0 dB, 5 dB, 10 dB and 15 dB)
5. CONCLUSIONS

We proposed the denoising algorithm based on wavelet packet transform for various environments. We can see that the proposed denoising algorithm shows good performance even though it is sensitive to choosing the wavelet basis.

6. REFERENCES