SPEECH ENHANCEMENT FOR NON-STATIONARY NOISE ENVIRONMENT BY ADAPTIVE WAVELET PACKET

Sungwook Chang, Y. Kwon, Sung-il Yang, and I-jae Kim*

Dept. of Electrical and Computer Engineering
Dept. of Physics*
Hanyang University, Korea
schang@ihanyang.ac.kr, yyhkwon@hanyang.ac.kr*, syang@hanyang.ac.kr, and junfrau@hotmail.com

ABSTRACT

We consider the non-stationary or colored noise estimation by wavelet thresholding method. First, we propose *node dependent thresholding* for adaptation in colored or non-stationary noise. Next, we suggest a noise estimation method based on *spectral entropy* using *histogram of intensity* instead of estimation method based on median absolute deviation (*MAD*). And we use a modified hard thresholding to alleviate time-frequency discontinuities. The proposed methods are evaluated on various noise conditions – white Gaussian noise, car interior noise, F-16 cockpit noise, pink noise, and speech babble noise. We compare our proposed methods with the conventional one with level dependent thresholding based on *MAD*.

1. INTRODUCTION

In many speech processing applications, speech has to be processed in the presence of undesirable background noise. During the last decades, various approaches to reduce the noise have been proposed. Among them, spectral subtraction [1] and wavelet thresholding [2] have been widely used. These methods require the estimation of statistical noise information. Hence, the accuracy of the estimated statistical noise information decides the performance of the enhancement system.

Recently, estimation method based on wavelet for statistical information of additive white Gaussian noise was introduced by Donoho and Johnstone. However, in spite of powerful performance for additive white noise, there are many problems to be resolved for a speech processing applications in real noisy environment.

White noise is useful as a conceptual entity, but it seldom occurs in real environment. Most of the noise captured by a microphone is colored, since its spectrum is not white. For example, pink noise is a particular type of colored noise that has a low-pass nature. The noise generated by a computer fan, an air conditioner, or an automobile engine can be approximated by pink noise. To solve this problem Johnstone and Silverman have proposed level dependent thresholding [3].

However, a great deal of real environmental noise is non-stationary, since its statistical properties change over time. Also, even the noise generated by a computer fan, an air conditioner, or an automobile engine are not perfectly stationary [4]. For the real environmental noise characteristics, we propose node dependent thresholding method as an extension of level dependent thresholding one in the entropy based adaptive wavelet packet tree.

Moreover, the conventional estimation method based on *MAD* has an assumption that the noise has Gaussian distribution. But, this assumption is not always valid in practice. For that reason, we propose a noise estimation method based on spectral entropy using histogram of intensity of wavelet coefficients for each node on adapted wavelet packet tree.

Finally, one of the problems to be resolved is time-frequency discontinuities in hard thresholding. For the problem, we use a modified hard thresholding function based on μ-law logarithm.

The proposed methods are tested on various noise conditions for comparison with the conventional wavelet thresholding method and spectral subtraction one.

2. DENOISING BY WAVELET THRESHOLDING

Wavelet transform has been intensively used as a powerful tool for noise reduction. The conventional wavelet based denoising algorithm can be summarized as follows; Let **s** clean speech with the finite length N and **x** the corrupted speech by white Gaussian noise **n** with variance σ^2,

$$\mathbf{x} = \mathbf{s} + \mathbf{n} \tag{1}$$

If **W** denotes a wavelet transform matrix, (1) in the wavelet domain can be given by;

0-7803-7402-9/02/$17.00 ©2002 IEEE

$$\mathbf{X} = \mathbf{S} + \mathbf{N}$$

where

$$\mathbf{X} = \mathbf{Wx},\ \ \mathbf{S} = \mathbf{Ws}, \mathbf{N} = \mathbf{Wn} \qquad (2)$$

Let $\hat{S}$ be an estimated speech of **S**, based on the noisy observation **X** in the wavelet domain. The speech $\hat{S}$ can be estimated by

$$\hat{S} = THR(\mathrm{X}, T) \qquad (3)$$

where $THR(\cdot,\cdot)$ denotes a thresholding function and T a threshold value.

Thresholding can be performed as Hard and Soft one defined as follows, respectively;

$$THR_h(X,T) = \begin{cases} X & ,|X| > T \\ 0 & ,|X| < T \end{cases} \qquad (4)$$

and

$$THR_s(X,T) = \begin{cases} \mathrm{sgn}(X)(|X| - T) & ,|X| > T \\ 0 & ,|X| < T \end{cases} \qquad (5)$$

The proper value of the threshold can be determined in many ways. A universal threshold T for the discrete wavelet transform is

$$T = \hat{\sigma}\sqrt{2\log(N)} \qquad (6)$$

where $\hat{\sigma} = MAD\ /\ 0.6745$ is the noise level. In the wavelet packet transform (WPT) case, the threshold becomes;

$$T = \hat{\sigma}\sqrt{2\log(N\log_2 N)} \qquad (7)$$

where N is the sample size.

For the correlated noise situation, a level dependent threshold was proposed [3] ;

$$T_j = \hat{\sigma}_j\sqrt{2\log(N)} \qquad (8)$$

with $\hat{\sigma}_j = MAD_j\ /\ 0.6745$, and MAD_j represents the absolute median estimated on the scale j.

3. NOISE ESTIMATION

As pointed out previously, wavelet thresholding method has many problems to be resolved for a speech processing applications in real noisy environment. To attack those problems, we propose a node dependent thresholding as the extension of level dependent thresholding and a new noise estimation method. And we use a modified hard thresholding based on μ-law logarithm to alleviate time-frequency discontinuities.

3.1. Node Dependent Thresholding

The main idea of the level dependent threshold is the estimation of threshold T_j for each wavelet packet scale in colored noise situation. This method shows better performance than universal threshold for colored noise.

But, noise level of colored or non-stationary noise shows different noise level to each scale and to each subband (i.e. to each node) on wavelet packet tree. For that reason, we define a threshold for each node;

$$T_{j,k} = \hat{\sigma}_{j,k}\sqrt{2\log(N)} \qquad (9)$$

with $\hat{\sigma}_{j,k} = MAD_{j,k}\ /\ 0.6745$ or proposed $\hat{\sigma}_{j,k}$, and $MAD_{j,k}$ represents the absolute median estimated at the scale j and subband k.

3.2. Noise Estimation based on Spectral Entropy using Histogram of Intensity

Generally, *MAD* based noise level estimation method is adopted for wavelet thresholding. *MAD* based noise level estimation method has an assumption that the noise has Gaussian distribution. But, this assumption is not always valid in practice. So we propose new noise estimation method based on spectral entropy using histogram of intensity. The proposed noise estimation method is summarized as follows:

Step 1. Estimate spectral pdf through histogram of wavelet packet coefficients for each node. Histogram is composed of B bins.

Step 2. Calculate normalized spectral entropy.

$$Entropy(n) = -\sum_{b=1}^{B} P \cdot \log_B(P) \qquad (10)$$

with

$n = 1, 2, \ldots,$ *No. of best nodes*

$$P = \frac{\textit{No. of Wavelet Packet Coefficients } c_{j,k} \textit{ in bin } b}{\textit{Node size in adapted wavelet packet tree}}$$

Step 3. Estimate *spectral magnitude intensity* by histogram and standard deviation of noise for node dependent wavelet thresholding. Pseudo code for estimation of spectral magnitude intensity by histogram is shown as follows;

```
for n = 1 : $node\_size$
  $b \approx \left| c_{j,k}(n) \right| / bin\_width$
    for i = $b$ : 1
      $histogram[i] = histogram[i] + 1$
    end
end
```

Next, we define an auxiliary threshold α.

$$\alpha(n) = Entropy(n) \cdot node\_size \cdot \beta \qquad (11)$$

where the range of β is from 0.7 to 0.9.

Finally, we estimate standard deviation of noise for node dependent wavelet thresholding.

$$\hat{\sigma}_{j,k} = [No.\, of\ bins\ bigger\ than\ \alpha(n)] \times bin\_width \qquad (12)$$

3.3. Modified Hard Thresholding

One of the major problems of hard/soft thresholding is time-frequency discontinuities. This leads to annoying artifacts and further degradation of output speech.

To resolve this problem, we adopt μ-law logarithm as a nonlinear function.

$$THR_h(X,T) = \begin{cases} X & ,|X| > T \\ T \cdot \left(\dfrac{1}{\mu} \cdot [(1+\mu)^{|X/T|} - 1] \cdot \mathrm{sgn}(X) \right) & ,|X| < T \end{cases} \qquad (13)$$

4. EVALUATION

In colored noise and non-stationary noise environment, the SNR cannot be used as faithful indication of speech quality. Thus we employ both objective and subjective tests for evaluation of proposed method. In objective tests, output SNR of the proposed methods was compared with the level dependent thresholding with *MAD* and spectral subtraction in various noise conditions. Spectral sub-traction test is performed by speech processing toolbox for Matlab [6]. In subjective tests, we employ an informal listening test and comparison of speech spectrum.

Various noise types, taken from Noisex-92 database, are used in our evaluation : white Gaussian noise, car interior noise, F-16 cockpit noise, pink noise, and speech babble noise. The performance results are averaged out using four different utterances from the TIMIT database. Half of the utterances are taken from male speakers, and others from female speakers. The parameters for proposed methods are selected as follows : $\beta = 0.8$, $\mu = 255$.

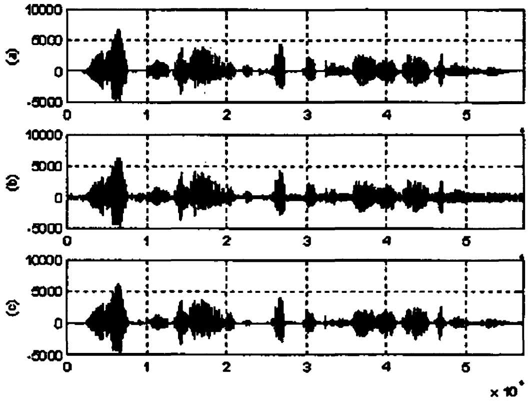

Figure 1. a) Original speech, b) Estimated by level dependent thresholding, c) Estimated by node dependent thresholding. (The speech degraded by car interior noise (SNR=10dB))

Fig.1 shows that node dependent thresholding has better performance than level dependent one. It is because the level dependent thresholding has adapted only to each scale. Table. 1 shows similar results. Moreover, the proposed methods are well suited to very strong noise and have a better performance than spectral subtraction and level dependent thresholding.

Although node dependent thresholding shows better denoising performance than level dependent one, subjective informal listening tests were in favor of the level dependent thresholding. It is because node dependent thresholding has a few artifacts.

However, another subjective tests show the opposite results. In Fig. 2, node dependent thresholding shows better performance in the spectral characteristic point of view, since we have used the spectral entropy to decide threshold. The spectral entropy gives an information for speech/non-speech region, approximately.

Table 1. Average SNR tests for pink noise corrupted speech

Corrupt Speech (dB)	Level Dependent with MAD	Node Dependent with MAD	Node Dependent with Proposed	Spectral Subtraction
-5	-3.70	3.53	3.31	0.10
0	1.11	5.43	5.91	1.77
5	5.79	7.44	8.30	2.35
10	10.15	9.49	10.47	2.83
15	14.15	11.39	12.15	4.08

Table 2. SNR(dB) tests for various noisy speech : "We like bleu cheese but Victor prefers swiss cheese." (SNR=10dB)

Noise Type	Level Dependent with MAD	Node Dependent with Proposed	Spectral Subtraction
White	10.29	10.35	2.39
Pink	9.47	10.49	2.42
F16	9.71	10.35	2.18
Car	9.65	13.50	1.95
Babble	9.59	10.18	2.23

Finally, in Table 2, we can see that the proposed method yields better performance than any conventional methods on various noise conditions.

5. CONCLUSION

The proposed methods were evaluated on various noise conditions – white Gaussian noise, car interior noise, F-16 cockpit noise, pink noise, and speech babble noise. We conclude that the proposed method gives two important results: 1) The method was well suited 'to the enhancement for very strong noise since it has better performance than spectral subtraction and level dependent thresholding; 2) Although subjective informal listening tests were in favor of the level dependent thresholding, the proposed method yields better spectral performance. It is very important characteristic for speech recognition or speaker verification.

6. REFERENCES

[1] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean square error short time spectral amplitude estimator," *IEEE Trans. on Acoust. Speech Signal Processing*, vol. 32, no. 6, pp. 1109-1121, 1984.

[2] D. L. Donoho, "Denoising by soft thresholding," *IEEE Trans. on Information Theory*, vol. 41, no. 3, pp. 613-627, 1995.

[3] I. M. Johnstone and B. W. Silverman, "Wavelet threshold estimators for data with correlated noise", *J. Roy. Statist. Soc. B*, vol. 59, pp. 319-351, 1997.

[4] X. Huang, A. Acero, H. Hon, "*Spoken Language Processing*," Prentice Hall, p. 474, 2001.

[5] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425-455, 1995.

[6] Imperial College of Science, Technology & Medicine, "VOICEBOX : Speech Processing Toolbox for Matlab", http: //www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html

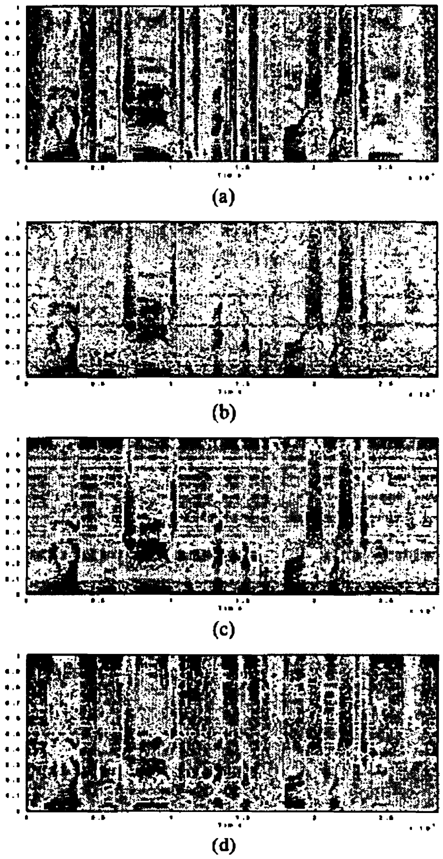

Figure 2. Speech Spectrograms. (a) Clean Speech : "We like bleu cheese but Victor prefers swiss cheese."; (b) Noisy Speech (additive F-16 cockpit noise at a SNR=10dB); (c) Speech enhanced with level dependent thresholding; (d) Speech enhanced with node dependent thresholding