THE JOURNAL OF THE ACOUSTICAL SOCIETY OF KOREA Vol.20, No.2E 2001. 6. pp. 25~29

A New Method for Selecting Thresholding on Wavelet Packet Denoising for Speech Enhancement

I-jae Kim*, Hyoung-soo Kim*, Kwang-hyun Koh*, Sung-il Yang*, Y. Kwon**

*Dept. of Control & Inst. Eng., Hanyang Univ.
**Dept. of Physics, Hanyang Univ.
(Received 15 Desember 2000 ; revised 2 April 2001 ; accepted 14 June 2001)

Abstract

In this paper, we propose a new method for selecting the threshold on wavelet packet denoising. In selecting threshold, the method using median is not efficient. Because this method can not recover unvoiced signal corrupted by noise. So we partition a speech signal corrupted by noise into the pure noise section and voiced section using autocorrelation and entropy. The autocorrelation and entropy can reflect disorder of noise.

The new method yields more improved denoising effect. Especially unvoiced signal is very nicely reconstructed, and SNR is improved.

Keywords: *Denoising, Wavelet packet, Wavelet shrinkage*

I. Introduction

There are several methods for noise cancelation, but we proposed a denoising method with wavelet packets[1]. Wavelet transform is an efficient tool in analysis and process of data and applied variously in signal processing recently. The wavelet packet transform is a generalization of the ordinary wavelet transform. Applications of wavelet packets that an entropy criterion allows to select the best basis are compression and denoising of signal.

In order to reduce the effect of noise, we propose preprocessor in speech signal processing using wavelet packet denoising[1]. In particular, we want to recover unvoiced signal corrupted by noise. In this paper, we propose a new wavelet packet denoising method for selecting threshold. We briefly describe wavelet packets, denoising with wavelet packets and the proposed new method in selecting threshold in section 2. In section 3, we show an experiment and the result. Section 4 gives our concluding remarks.

Corresponding author: Sung-il Yang (junfrau@hanmail.net)
Hanyang University 1217 Sa-doing, Ansan, Kyunggi-do
425-170, Korea

II. Denoising with Wavelet Packet and selecting threshold

2.1. Wavelet Packet Transform

The wavelet packet system was proposed by Ronald Coifman to allow a finer and adjustable resolution of frequencies at high frequencies. It gives a rich structure that allows adaptation to particular signals or signal classes.

Wavelet packets are particular linear-combinations or superpositions of wavelets. They form bases which retain the orthogonality, smoothness, and localization properties of their parent wavelets. In order to generate a basis system that would allow a higher resolution decomposition at high frequencies, we have to iterate (split and down-sample) the highpass branch of the Mallat algorithm tree as well as the low pass scaling function branch. If we choose to split, the function is replaced by two half-size sets of functions created by lowpass and highpass filtering[2].

2.2. The Best Basis Algorithm

From the wavelet packet tree as we have shown in Fig. 1, one can select a number of orthonormal bases for representing a given signal. Coifman and Wickerhauser proposed a rule to select the best orthonormal basis, which minimized the Shannon-Weaver (SW) entropy of the vectors in the decomposition, Since entropy represents a state of disorder, a property of noise, the minimization of entropy would improve the SNR of the decomposed signal[3].

The SW entropy of a signal $x=\{x_i\}$ is defined as

$$H(x) = -\sum p_i \log p_i \tag{1}$$

where, and we set plogp = 0 if p = 0. This is not an additive information cost function. We can construct a new additive cost function from $H(x)$ by defining

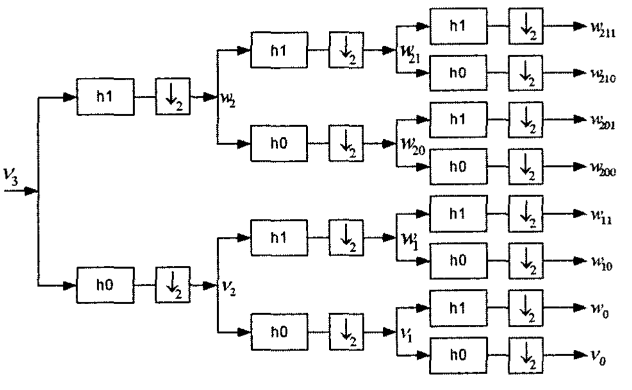

Figure 1. The binary tree for wavelet packet transform.

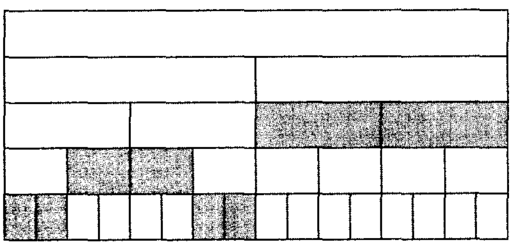

Figure 2. An example of Wavelet packet best basis.

$$H(x) = ||x||^2 \lambda(x) + \log ||x||^2 \tag{2}$$

$$\lambda(x) = -\sum_i |x_i|^2 \log |x_i|^2 \tag{3}$$

From Eqs. (2) and (3) we know that minimizing λ(x) automatically *minimizes H(x).*

After calculating entropy from each packet nodes, we may choose the best basis. Pairs of nodes are considered starting from the lowest level and the sum of the entropies of the nodes is compared with that of the parent node from where these originated. If the parent node has a larger entropy, it is replaced by the sum of entropy of children. This process is repeated until the original signal at the top is reached. Those nodes whose entropy values are unchanged in this procedure are marked. Starting from the root, the tree is examined and the nodes that come from the selected nodes are pruned away. The remaining marked nodes as shown shaded in Fig. 2 form the best bases.

2.3. Conventional Thresholding Methods [4] and A New Threshold Method

We have a finite length signal of observations x_i of a signal s_i corrupted by an i.i.d, zero mean, white Gaussian noise n_i with standard deviation σ, i.e.

$$x_i = s_i + \sigma n_i \tag{4}$$

where i = 1, 2, 3, ..., N[5]. Our aim is recovering s_i from x_i.

Let W and W^1 an $N \times N$ wavelet transform matrix and its inverse, respectively. In the wavelet domain, Eq. (4) becomes

$$X = S + N \tag{5}$$

where

$$X = W_x,\ S = W_s,\ N = W_n, \tag{6}$$

and x, s and n are column vectors containing x_i, s_i, and n_i, respectively, and W maps n to N, which is also zero mean white Gaussian with variance σ^2 whereas S is typically sparse containing only a few large wavelet coefficients. Therefore, to estimate signal s, one resorts to modification of X by suitable thresholding. A diagonal filtering operation, Δ, can represent this thresholding

$$\Delta = diag[l_1, l_2, \cdots, l_N] \tag{7}$$

where l_i 's are wavelet threshold filters.

With this, the signal estimate is

$$\hat{s} = W^{-1} \Delta Wx \tag{8}$$

There are two basic types of the wavelet thresholding. The policy for hard thresholding is 'keep or kill'. The absolute values of all transform coefficients are compared to the fixed threshold . If the magnitude of the coefficient is less than , the coefficient is replaced by zero.

$$l_i^{hard} = \begin{cases} 1 & \text{if } |x_i| \geq \tau \\ 0 & otherwise \end{cases} \tag{9}$$

Soft threshloding shrinkage all the coefficients towards the origin in the following way:

$$l_i^{soft} = \begin{cases} \dfrac{sign[x_i][|x_i| - \tau]}{x_i} & \text{if } |x_i| \geq \tau \\ 0 & otherwise \end{cases} \tag{10}$$

Soft thresholding smoothes the signal and avoids spurious oscillations while hard thresholding achieves smaller mean square error (MSE) [2].

Let (z_i) be i.i.d. N(0,1). Then

$$\pi_n \equiv \Pr ob\{ \|(z_i)\|_{l_n^\infty} \leq \sqrt{2 \log M_n} \} \rightarrow 1,\ n \rightarrow \infty. \tag{11}$$

where $M_n = n \cdot \log_2 n$ in case of wavelet packets[4]. From Eq. (11), we can get Stein's Unbiased Risk Estimate (SURE), that is, an adaptive threshold selection rule defined as

$$\rho = \sqrt{(2 \log_e (N \times \log_2(N)))} \tag{12}$$

where N is the number of samples in the signal vector [6]. With this approach, obtaining risks and minimizing them with respect to ρ values give a threshold selection. The method is adaptive through searching a threshold level for each wavelet decomposition level.

The above signal model assumes the noise is normally distributed with zero mean and variance of 1, which means that we have to rescale the threshold values when dealing with unscaled and non-white noise. Calculated thresholds is rescaled by the standard deviation of noise estimated from the finest level of the decomposition of each signal so that $\tau = \rho \cdot \sigma$. The noise level is estimated scale by scale to take account of the obviously strong high-frequency content. In wavelet packet case, σis estimated from the first node at each subdecomposition band which give the best statistics for the noise level estimation.

If $\{X_p\}_{0 \leq p < P}$ are P independent Gaussian random variable of zero-mean and variance σ^2 then

$$E(Med(|X_p|)_{0 \leq p < P}) \approx 0.6745\ \sigma \tag{13}$$

As a robust estimate of the standard deviation

$$\sigma = median(|d^1|)/0.6745 \tag{14}$$

can be used. From Eq. (13) we can define τ again[7].

$$\tau = \sigma \times median(|d^1|)/0.6745 \tag{15}$$

The noise in the wavelet transform is, at each resolution level, a Gaussian noise which is again approximately stationary. We estimate the variance of the noise by assuming that most of the empirical wavelet coefficients at each resolution level are noises, and hence that the median absolute deviation reflects the size of the typical noise[8].

But this method can not reconstruct unvoiced signal in noise. In this paper we propose an efficient method. The main idea is to partition the noisy signal into the pure noise signal and the voiced signal including the unvoiced signal in noise. And we apply several different threshold values to each partitioned signal.

We will use the absolute magnitude mean of noisy to a new threshold. Our intension is to preserve unvoiced signal in the corrupted signal. Though the σ is a good threshold for noise cancellation, it is not suitable to preserve unvoiced signal in noise. To preserve unvoiced signal in noise, we need to find smaller value than for a suitable threshold. We assume that the first one second of the signal is a pure noise. The absolute magnitude mean

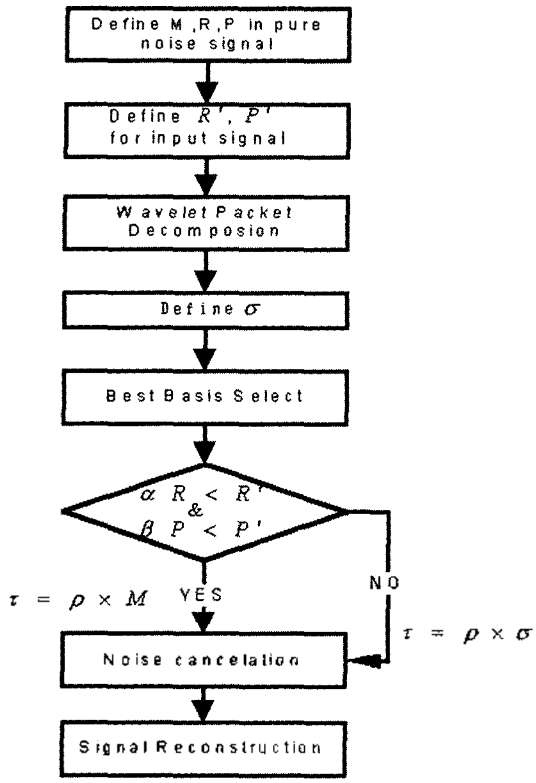

Figure 3. A block diagram of the proposed wavelet packet denoising.

M, autocorrelation R, and entropy P, of the signal in stationary section are defined as follows:

$$M=\frac{1}{L}\sum_{i=0}^{L}x(i) \tag{16}$$

$$R=\sum_{i=0}^{L}\sum_{j=0}^{L}x(i)x(i-j) \tag{17}$$

$$P=-\sum_{i=0}^{L}|x(i)|^2\log|x(i)|^2 \tag{18}$$

where L is the stationary section.

The selection of autocorrelation and entropy is reasonable, because they can well reflect disorder of noise. Accordingly, we use these two variables to partition the signal into the pure noise and the voiced signal including the unvoiced signal in noise. And to each section, we apply different thresholds.

$$\tau=\begin{cases}\rho\times M & \text{if}(\alpha\times R)<R' \text{ and}(\beta\times P)<P' \\ \rho\times\sigma & otherwise\end{cases} \tag{19}$$

where R' is signal's autocorrelation, P' is signal's entropy, α=1.5, β=1.7. α and β are decided by experiment.

III. Experiments and Results

In this experiment, we use a speech sample with 8kHz sampling rate, 16 bits/sample. And we made noisy speech signal by adding Additive White Gaussian Noise. A experiment is done as the following steps in Fig. 3.

As a result, we can get more improved denoising signal compared with speech signal which is denoised by selecting in all speech sections.

In particular, unvoiced signal is very nicely reconstructed as you can see in Fig. 6. And SNR is improved from 20dB to 25.61dB.

IV. Conclusion

In this paper, we proposed a new method, which partition the noisy signal into the pure noise section and

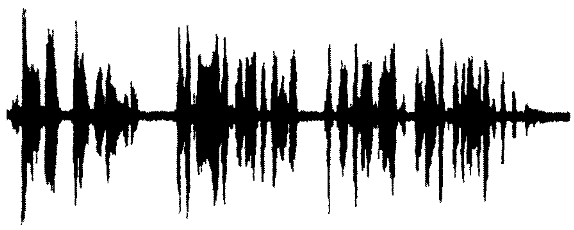

Figure 4. A noisy speech signal.

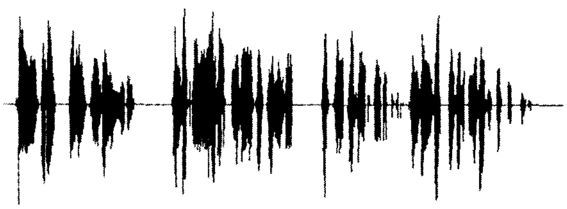

Figure 5. A denoised speech signal after wavelet packet denoising by selecting in all speech section.

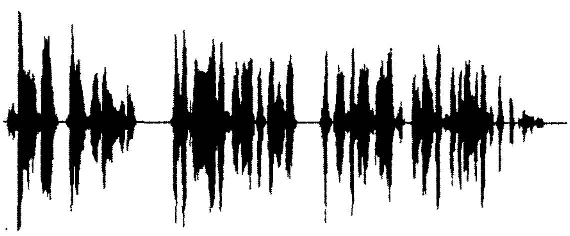

Figure 6. A denoised speech signal after wavelet packet denoising by the proposed method.

the voiced signal including the unvoiced signal in noise and apply different thresholds to each partitioned signal. After denoising by selecting in all speech sections, we lose much information of denoised speech signal and SNR is 4.9dB. But after applying the proposed method, we can get more improved denoising effect. Especially unvoiced signal is very nicely reconstructed, and SNR improves to 25.6dB for the noisy signal with 20dB.

Acknowledgements

This work was supported by the Hanyang University Research Foundation.

References

1. Hyung-soo Kim, Kwang-hyun Koh, "Speech Recognition of Corrupted Speech using Wavelet Packets," ISSA SCI2000, Vol. 6 Part 2, pp. 25-28.
2. C. S. Burrus, R. A. Gopinath and H. Guo, *Introduction to Wavelets and Wavelet Transforms*, Prentice Hall, 1998.
3. M. V. Wickerhauser, *Adapted Wavelet Analysis from Theory to Software*, IEEE Press, 1994.
4. D. Donoho, "Denoising by Soft-Thresholding," IEEE Trans. on Information Theory 46, pp. 612-627, 1995.
5. S. K. Nath et al. "Wavelet based compression and denoising of optical tomography data", Optics Communications, Vol. 167, pp. 37-46, 1999.
6. P. E. Tikkanen, "Nonlinear wavelet and wavelet packet denoising of electrocardiogram signal," Biological Cybernetics, Vol. 80, pp. 259-267, 1999.
7. B. Walczak and D. L. Massart, "Noise Suppression and Signal Comppresion using the Wavelet Transform," Chemometrics & Intelligent Laboratory Systems, Vol. 36, 1997.
8. D. Donoho, "Nonlinear Wavelet Methods for Recovery of Signals, Densities, and Spectra from Indirect and Noisy Data," Proceedings of Symposia in Applied Mathematics, Vol. 10, 1993.

[Profile]

• I-jae Kim

He was born in Inchon, Korea, in 1976. He received the B.S. and M.S. degrees in control and instrumentation engineering from Hanyang University, Ansan, Korea in 1999 and 2001, respectively. He is a graduate student for Ph.D currently in the course of Ph.D. degree at Hanyang University. His current research interests include digital speech and wavelets, DSP, speech recog-nition.

• Hyoung-soo Kim

He was born in Suwon, Korea, in 1972. He received the B.S. and M.S. degrees in control and instrumentation engine-ering from Hanyang Un-iversity, Ansan, Korea in 1997 and 1999, respec-tively. His current re-search interests include speech, audio and image signal processing using wavelets, speech recognition and synthesis. He is a Asistant Research Engineer in LG Electronics/In-formation & Communication.

• Kwang-hyun Koh

He was born in Seoul, Korea, in 1972. He re-ceived the B.S. and M.S. degrees in control and instrumentation engine-ering from Hanyang Un-iversity, Ansan, Korea in 1998 and 2000, respec-tively. His current re-search interests include speech proce-ssing ,wavelets, speech recognition. He is a Asistant Research Engineer in Samsung Electronics.

• Sung-il Yang

The Journal of the Acoustical Society of Korea Vol. 16, No. 2E, September 1997.

• Y. Kwon

The Journal of the Acoustical Society of Korea Vol. 16, No. 2E, September 1997.