THE JOURNAL OF THE ACOUSTICAL SOCIETY OF KOREA Vol.21, No.4E 2002. 12. pp. 178~182

Adaptive Wavelet Denoising For Speech Rocognition in Car Interior Noise

E-jae Kim*, Sung-il Yang*, Y. Kwon**, Soon S. Jarng***
*School of Electrical and Computer Engineering, Hanyang University
**Department of Physics, Hanyang University
***Department of Control and Implementation Engineering, Chosun University
(Received 17 October 2002; accepted 23 December 2002)

Abstract

In this paper, we propose an adaptive wavelet method for car interior noise cancellation. For this purpose, we use a node dependent threshold which minimizes the Bayesian risk. We propose a noise estimation method based on spectral entropy using histogram of intensity and a candidate best basis instead of Donoho's best bases. And we modify the hard threshold function. Experimental results show that the proposed algorithm is more efficient, especially to heavy noisy signal than conventional one.

I. Introduction

A speech signal is often corrupted by noise in acquisition or transmission. The goal of denoising is to remove the noise while remaining the important features of original speech as much as possible. The classical solution to the noise removal problem is to use the Wiener filter, which utilizes the second-order statistics of the Fourier decomposition. And it can be achieved by a linear processing. Recently, signal denoising using a nonlinear processing has been studied. One of them is to use the threshold of wavelet coefficients pioneered by Donoho and Johnstone[5]. The main algorithm is as follows: The signal is discriminated from the noise by choosing an orthogonal basis, which efficiently approximates the signal (with few non-zero coefficients). Signal enhancement is achieved by discarding components below a predetermined threshold. The most well-known thresholding methods are *VisuShrink and SureShrink*[7]. These threshold choices enjoy an asymptotic minimax optimalities over function space.

Corresponding author: E-jae Kim (junfrau@ihanyang.ac.kr)
Department of Physics, Hanyang University, 17, Haengdang-dong, Seongdong-gu, Seoul, 133-791, South Korea

In this paper, we consider a method to remove car interior noise. For this, we propose a modified best basis searching method. The conventional best basis searching algorithm is based on entropy. But there are some possibilities that conventional best bases are not real best bases in the corrupted signal by colored noise, because car interior noise distribution is similar to speech's distribution. So, we need to find a candidate best basis different from conventional best bases. We select a candidate best basis which has smaller band width than conventional best basis. Then we apply node dependent thresholding which minimizes Bayesian risk of the signal.

To get suitable threshold value, we estimate the noise level using histogram of intensity[1,2]. Moreover, we apply some modified hard threshold function including an averaging filter concept.

The paper is organized as follows: In section 2, the wavelet thresholding is introduced and in section 3, we derive the node dependent threshold. Experimental results are explained in section 4, and conclusion is stated in section 5.

II. Wavelet Thresholding

Consider the signal $\{X_i, i = 1,\cdots,N\}$, where N is power of 2. If it is corrupted by an additive noise,

$$Y_i = X_i + \sigma\varepsilon_i, \qquad i=1,\cdots,N \tag{1}$$

where $\{\varepsilon_i\}$ is independent and identically distributed $(i,i,d.)$ as normal $N(0,1)$ and independent of $\{X_i\}$. The goal is to remove the noise and to obtain an estimate $\{\hat{X}_i\}$of $\{X_i\}$ which minimizes the mean squared error (MSE),

$$MSE(\hat{X}) = \frac{1}{N}\sum_{i=1}^{N}(\hat{X}_i - X_i)^2 \tag{2}$$

Here some additional requirements on the estimator can be imposed.

[SMOOTHNESS] *The estimator* $\{\hat{X}_i\}$ *should be, with high probability, as smooth as* $\{X_i\}$.

[ADAPTATION] *The estimator* $\{\hat{X}_i\}$ *achieves almost minimax risk over one of a wide range of smoothness classes, including the classes in which linear estimators do not achieve the minimax rate.*

So, how do we obtain an estimator $\{\hat{X}_i\}$ using wavelets? Donoho and Johnstone proposed a simple recipe based on thresholding in the wavelet domain. Their wavelet estimation procedure has three main steps[5-7].

STEP 1. Transform the observations $\{Y_i\}$ to the wavelet domain by applying a discrete wavelet packet trans-formation.

STEP 2. Estimate σ. Use this estimator to threshold (or shrink) the wavelet coefficients.

STEP 3. Invert the thresholded (shrunk) coefficients, recovering the estimator of function, $\{\hat{Y}_i\}$.

There are two thresholding methods widely used. The *soft-threshold* function takes the argument and shrinks it to zero by the threshold T.

$$S_T(x) = \mathrm{sgn}(x)\cdot\max(|x|-T,0) \tag{3}$$

The *hard-threshold* function keeps the input if it is larger than the threshold T; otherwise, it is set to zero.

$$H_T(x) = x\cdot 1\{|x|>T\} \tag{4}$$

Even though the wavelet thresholding is a simple method, it is not easy to find a good threshold. Donoho and Johnstone proposed the *VisuShrink* and *SureShrink* for threshold estimation. For a deterministic signal of length M, Donoho and Johnstone proposed, for *VisuShrink*, the universal threshold, $T=\sigma\sqrt{2\log N}$ (N is the node size), which is an asymptotically optimal estimate in the minimax sense. And *Sure-Shrink* is a method of level dependent selection of a threshold by minimizing Stein's unbiased estimator of risk.

III. Modified Wavelet Shrinkage

In this paper, we find the threshold to minimize error in sense of Bayesian risk[4,8]. We estimate the noise level based on spectral entropy. We propose some modified hard threshold function including an averaging filter concept.

3.1. Threshold for Minimizing the Bayesian risk

We want to find a threshold T which minimize the Baysian risk[3].

$$r(T) = E(\hat{X}-X)^2 = E_X E_{Y|X}(\hat{X}-X)^2 \tag{5}$$

where $\hat{X}=H_T(Y), Y\mid X\sim N(\theta,\sigma^2)$, and $X\sim G_{\sigma_x}$.

The optimal threshold is selected as follows:

$$T^*(\sigma_x) = \arg\min_T r(T) \tag{6}$$

For $X\sim N(0,\sigma_x^2)$, we get

$$E_X E_{Y|X}(\hat{X}-X)^2$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(H_T(y)-x)^2\,p(y\mid x)\,p(x)\,dydx \tag{7}$$

After some algebra, we find that the relation between threshold value (T) and deviation (σ_X) to minimize the Bayesian risk would be

$$T(\sigma_X) = \frac{1}{\sigma_X} \tag{8}$$

For general σ, from (7),

$$T(\sigma_X) = \frac{\sigma^2}{\sigma_X} \tag{9}$$

Adaptive Parameter Estimation

Generally, we estimate the signal's noise variance σ^2 by the median estimator,

$$\hat{\sigma} = \frac{median(|d^1 - median(d^1)|)}{0.6745} \tag{10}$$

where d^1 is the finest detail coefficients of the signal.

Let the signal be $Y = X + V$, with the X and V independent of each other,

$$\sigma_Y^2 = \sigma_X^2 + \sigma^2 \tag{11}$$

where σ_Y^2 is the variance of Y.

Thus, the estimated threshold value is as follows:

$$\hat{T}(\hat{\sigma}_X) = \frac{\hat{\sigma}^2}{\hat{\sigma}_X} \tag{12}$$

where

$$\hat{\sigma}_X = \sqrt{\max(\sigma_Y^2 - \hat{\sigma}^2, 0)} \tag{13}$$

In the case that $\sigma_Y^2 \leq \hat{\sigma}^2 + \hat{\sigma}_X$, is set to be zero and $\hat{T}(\hat{\sigma}_X)$is ∞. In practice, $\hat{T}(\hat{\sigma}_X) = \max(|d^1|)$.

3.2. Noise Estimation

To estimate noise variance (or level), MAD (equation (9)) is widely used. MAD shows a good performance to estimate noise level. But in the case of colored noise, MAD method shows some wrong estimation. To compensate this, we use spectral entropy by histogram of intensity.

Let's define an auxiliary threshold.

$$\alpha(n) = Entropy(n) \cdot node\_size. \beta \tag{14}$$

where n is node index and β is a value from 0.7 to 0.9.

$$\hat{\sigma}_{j,k} = [No.\ of\ bins\ bigger\ than\ \alpha(n)]\ bin\_width \tag{15}$$

where j is level index and k is node index Finally, node depended threshold is

$$\hat{T}_{j,k}(\hat{\sigma}_x) = \varepsilon \cdot \frac{\hat{\sigma}_{j,k}^2}{\hat{\sigma}_X}\sqrt{2 \cdot \log N} \tag{16}$$

where ε is a value from 0.01 to 0.1 w.r.t. signal's SNR.

3.3. Candidate Best Basis

The conventional best basis search algorithm is based on entropy. There are some possibilities that conventional best bases are not the real best bases in the corrupted signal by colored noise, because the colored noise distribution is similar to speech's distribution. And there are some trials to find another best basis search algorithm, so called 'near best basis search'. This method has a satisfactory performance, but it takes a lot of computation time. On the other hand, our proposed method to find candidate best basis is very simple. Our strategy is to take smaller bandwidth than conventional best basis;

$$BB(j,k) \begin{array}{l} \nearrow BB_c(j+1, 2\cdot k) \\ \searrow BB_c(j+1, 2\cdot k+1) \end{array}$$

where $BB_c(j,k)$ means candidate best basis at j level, k node.

3.4. Modified Hard Thresholding

There are many proposals on selecting thresholdings. These are based on hard or soft thresholdings. Soft threshold smoothes the signal and avoids spurious oscillations while hard thresholding achieves smaller mean square error. We use a modified hard thresholding.

$$\begin{aligned} H_T(x_n(t)) &= x_n(t), && \text{if } |x_n(t)| \geq T \\ &= 0, && \text{if } |x_n(t)| < T \\ &= 0, && \text{if } t = n \end{aligned} \tag{16}$$

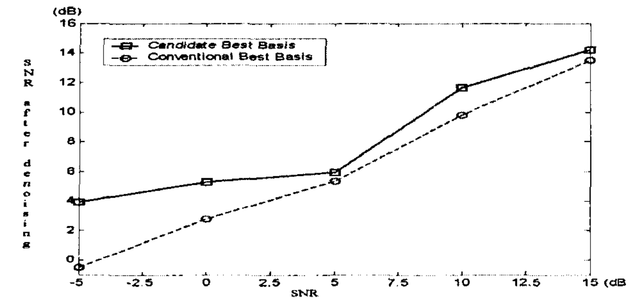

Figure 1. Comparison of SNR: Candidate best basis vs. Conventional best basis.

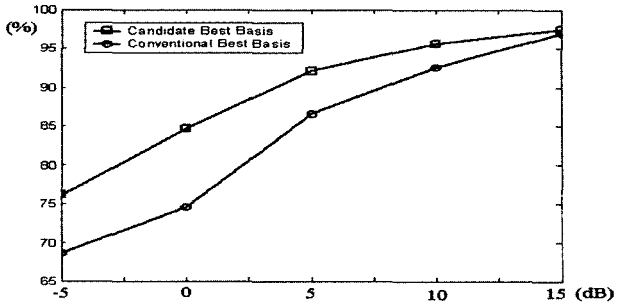

Figure 2. Recognition accuracy: Candidate best basis vs. Conventional best basis.

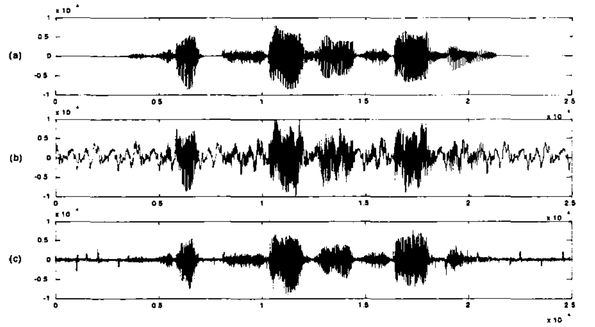

Figure 3. (a) Clean speech signal (six seven seven, English). (b) Noisy Speech Signal (SNR -5 dB). (c) Denoised Speech Signal.

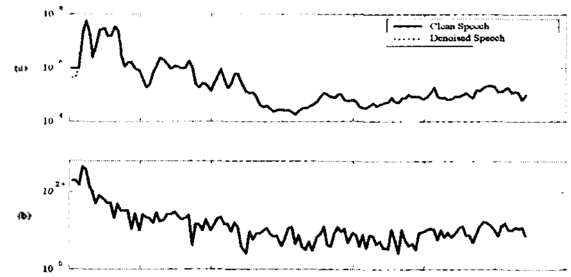

Figure 4. (a) Spectral density of clean speech and denoised speech from 15 dB noised speech. (b) Absolute difference between two speeches.

where n is ordered index of best basis.

The third term of equation (16) operates as a pseudo-function which averages the coefficient.

IV. Experimental Results

In experiment we used a TIDIGIT-64 database degraded by car interior noise taken from Noisex-92 database. And we use HTK toolkit for speech recognition accuracy test.

We found that conventional best bases are not real best bases, since entropy didn't reflect on signal's characteristics very well when the signals were corrupted by colored noise. And we selected candidate best basis at lower level. Fig. 1 and Fig. 2 indicate that our assumption is reasonable.

Fig. 3 shows denoised signal and we can see the experimental results in Tab. 1 and Tab. 2. Especially, it shows a good performance at low SNR signal. However at 15dB SNR signal, we cannot get efficient performance. It is because the signal loses its shape while noises are removed. Nevertheless, the signal's important features (formant, pitch, etc.) are kept well as shown in Fig.4.

Table 1. Average SNR for each signal.

[illegible]	[illegible]	[illegible]
15	14.06	14.19
10	11.41	11.64
5	5.51	5.95
0	4.46	5.3
-5	2.01	3.96

Table 2. Recognition accuracy for each speech signal.

[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	-5	50.8	72.4
	0	67.6	80.3
	5	80.9	86.7
	10	92.8	93.4
	15	95.9	95.9
[illegible]	-5	52.3	76.2
	0	73.3	84.7
	5	86.3	92.2
	10	94.7	95.6
	15	97.5	97.5

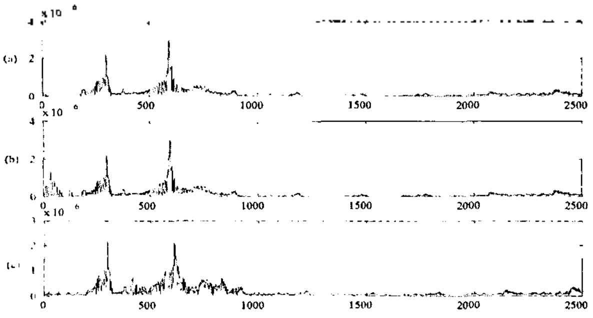

Figure 5. Spectrum of signal (SNR 15 dB): (a) Clean speech signal. (b) Noisy speech signal. (c) Denoised speech signal.

V. Conclusion

In this paper, we propose an adaptive wavelet shrinkage based on Bayesian modeling for a car interior noise reduction. To estimate noise level, we use a node dependent spectral entropy. We select a candidate best basis and apply modified hard thresholding for denoising car interior noise. We can confirm that our approach is reasonable, efficient denoising effect is obtained and SNR is improved. Even though we cannot nicely recover the noisy signal at high SNR signal but we can keep the signal's important features.

VI. Discussion

In this paper, we proposed a wavelet packet method for removing car interior noise, whose main distribution is at lower frequency domain than speech's. Even though we could not get satisfactory performance at high SNR speech signal, we can see that the low frequency components are removed in Fig. 5. Also there are some changes of spectral shape. However the recognition accuracy had no changes. That is, these changes have little effect on speech recognition accuracy.

References

1. E-jae Kim, Sung-il Yang, and Y. Kwon, "Adaptive wavelet denoising for a car interior noise," *SCI 2002*, IX, 344-347.
2. Sungwook Chang, Y. Kwon, Sung-il Yang, and E-jae Kim, "Speech enhancement for non-stationary noise environment by adaptive wavelet packet," *ICASSP* 2002, 1, 561-564.
3. S. Grace Chang, Bin Yu, and Martin Vetterli, "Adaptive Wavelet Thresholding for Image Denoising and Compression," *IEEE Trans. Image Processing*, 9, September 2000.
4. H. Chipman, E. Kolaczyk, and R. McCulloch, "Adaptive Bayesian Wavelet Shrinkage," *J. Amer. Statist. Assoc.*, 92 (440), 1413-1421, 1997.
5. D. L. Donoho, "Denoising by soft-thresholding," *IEEE Trans. Inform. Theory*, 41, 613-627, May 1995.
6. D. L. Donoho, "Wavelet Shrinkage: Asymptopia?," *J. R. Stat. Soc. B, ser. B*, 57 (2), 301-369, 1995.
7. S. Mallat, "A Wavelet Tour of Signal Processing," *Academic Press, 1998.*
8. Brani Vidakovic, "Statistical Modeling by Wavelet," John Wiley & Sons, INC, 1999.

[Profile]

• E-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 recognition.

• Sung-il Yang

He was born in Goesan, Chungbuk, Korea in 1956. He received his BS degree in Electronics Engineering with the greatest honors from Hanyang University, Seoul, Korea, 1984, and his MS and Ph.D. degrees in Electrical & Computer Engineering from the University of Texas, Austin, Texas, 1986 and 1989, respectively. Since 1990, he has been with Hanyang University and he is now a professor at the School of Electrical and Computer Engineering. His current research interest includes speech recognition, digital signal processing, and responsible technology. He is also a member of IEEE, Korea Institute of Telematics and Electronics, and the Acoustical Society of Korea.

• Y. Kwon

He was born in Seoul, South Korea in 1961. He received his BS degree in Mathematics and Physics with the greatest honors from Hanyang University, Seoul, Korea, 1984, and his MS and Ph.D. degrees in Physics from the University of Rochester, Rochester, NewYork, 1986 and 1987, respectively. Since 1995, he has been with Hanyang University and he is now an Associate Professor at Department of Physics. His current research interest includes Mathematical Physics, Theoretical Physics, Artificial Intelligence, Signal Processing and Quantum computing. He is also a fellow of the International Society for Complexity, Information, and Design , and a member of American Mathematical Society, Korean Mathematical Society and Korean Physical Society.

• Soon S. Jarng

He was born in Seoul, Korea in 1961. He received his BS degree in Electronics Engineering from Hanyang University, Seoul, Korea, 1984, and his MS and Ph.D. degrees in Electrical & Electronic Engineering from the University of Birmingham, U.K., 1988 and 1991, respectively. Since 1991, he has been with Chosun University and he is now associate professor at the School of Electronic Information Communication Engineering. His current research interest includes Cochlear Bio-mechanics, Underwater/Underground Acoustics, Piezoelectric Sensor Device, Finite Element Method (FEM), Boundary Element Method (BEM). He is also a member of IEEE, Korea Society of Noise and Vibration and Electronics, and the Acoustical Society of Korea.-jae Kim