International Symposium on Chinese Spoken Language Processing, pp. 211-214, Oct. 13-15, 2000, Beijing

A NOISE CANCELLATION METHOD BASED ON WAVELET TRANSFORM

YANG Dali, XU Mingxing, WU Wenhu, ZHENG Fang

Center of Speech Technology, State Key Laboratory of Intelligent Technology and Systems,

Department of Computer Science & Technology, Tsinghua University, Beijing
ydl@sp.cs.tsinghua.edu.cn

ABSTRACT

In this paper, we present a frequency band threshold based on wavelet transform (FBT) noise cancellation method.

The noise cancellation is enable to improve on the articulation of the speech. Although the edge information of the speech is very important for recognition system to use, most traditional noise cancellation methods based on spectrum analysis smooth these edges of the original speech. We hope to get a noise cancellation method that keeps these edges information. We knew that the performance of edge detection based on wavelet transform is very high. So we use wavelet transform for noise cancellation.

Noise cancellation methods based on wavelet transform were referred to papers [1][2]. The method was given by paper [1] is not real-time. Hence this method is difficult to be used a practical system. Although the real-time property of the noise cancellation method was referred to paper [2] is perfect, the aural performance is defective. This method has a single threshold (ST). It ignored the difference of the frequency bands. FBT is presented by us in this paper possesses two characteristics as follow:
(1) These thresholds depend on frequency bands.
(2) These thresholds are self-adjusting.

Based on two judgement standards---signal noise rate (impersonal standard) and the articulation of the speech (subjective standard), we did comparison experiments between FBT and ST. Although FBT's signal noise rate inferior to the ST's, FBT's waveform distortion is less than ST's and FBT's articulation of the speech is remarkable superior to the ST's. We particularly analyzed the causes of the phenomena and did the comparison experiments of these two methods on the same speech recognition system. The conclusion is FBT is superior to ST.

Key words

Noise Cancellation Wavelet transform Frequency Band Threshold

1. INTRODUCTION

We only discuss speech signal in this paper. Speech recognition, transmission and processing systems operating in adverse environments, such as moving car, have to deal with a variety of ambient noises. Noise affects speech recognition system and working performance of practical system rapidly deteriorate with decreasing signal-to-noise ratio (SNR).

Suppose noise is additive. Let X(n) = S(n) + N(n),where the S(n) is a clean speech signal , the N(n) is a noise signal and X(n) is an input signal. Paper [3] presented a spectral subtraction noise cancellation method based on the case. There were numerous modifications and improvements of the method. Paper [4] showed these methods often result in excessive residual noise including unpleasant "musical tone artifacts". Our experiment showed the same result too. Another deplorable result is these methods smooth edges of the original speech signal. Because the edge information of the speech is very important for recognition system and the performance of edge detection based on wavelet transform is very high, we present a noise cancellation method based on wavelet transform.

2. TWO OLD METHODS AND A NOVEL METHOD

Paper [1] reviewed the mathematical characterization of singularities with Lipschitz exponents and theorems that estimate local Lipschitz exponents of functions from the evolution across scales of their wavelet transform. It is then proven that the local maxima of the wavelet transform modulus detect the locations of irregular structures and provide numerical procedures to compute their Lipschitz exponents. The wavelet transform of singularities with fast oscillations has a particular behavior that is studied separately. The local frequency of such oscillations is measured from the wavelet transform modulus maxima. It has been shown numerically that one- and two-dimensional signals can be reconstructed, with a good approximation, from the local maxima of their wavelet transform modulus. An algorithm is developed that removes white noises from signals by analyzing the evolution of the wavelet transform maxima across scales.

Although this method perhaps is not the best as its waveform distortion, its excellence is edges information is kept. But this method is not real-time. Hence this method is difficult to be used a practical system. We called the method *Mallat Method*

Donoho proposed a noise cancellation method in [2]. Although its real-time property is perfect, the aural performance is defective. It has a single threshold (ST) and ignored the difference characterization of the frequency bands. We called the method *ST Method.* In order to get over the defect of *ST Method*, we present a frequency band threshold (FBT) based on wavelet transform noise cancellation method in this paper. We called the method *FBT Method.* The method includes following steps:

(1) Noisy data X(n) is decomposed N frequency bands in wavelet domain with Mallat Tower method in paper [6]. Mother wavelet function was selected 4-order Daubeches wavelet in [7]. The wavelet coefficients of the frequency band i are noted W(i,j),i=0,…,6.

(2) The wavelet transform modulus maximums of the frequency bands on the first 100 milliseconds of the signal X(n) are searched. The maximums are noted M(i),i=0,…,N-1.

(3) The thresholds are calculated as following:

T(i) = k*M(i), i=0,…,N-1.

Where k is a constant and satisfies 0<k<1.

(4) The wavelet coefficients are changed. Let

W(i,j) – T(i), if W(i,j) > T(i)

W(i,j)=0, if W(i,j) <= T(i) and W(i,j) >= – T(i)

W(i,j) +T(i), if W(i,j) < – T(i)

(5) The signal is restructured with inverse wavelet transform to the W(i,j).

3. SUPPOSITION AND THRESHOLDS UPDATE

We suppose direct component of the signal X(n) is zero or else the direct component is subtracted from signal X(n). In addition, the noise N(n) is considered as a short-time stationary signal and X(n) only includes N(n) on the first 100 milliseconds. When the length of the signal X(n) is big enough, it is necessary that the thresholds is updated during the pause section of speech.

4. TASKS

To test our techniques, task was evaluated in the car environment. The test data was 300 sentences that were pronounced by 3 men. Each sentence includes one name and was calculated a SNR. Variance range of SNR is from 0 dB to 20 dB. Each name includes 2 or 3 syllables. Speech was sampled at 8k Hz and each sample point is digitized 16 bits. Giving the values of k and N, the 300 sentences were cleaned with the method was mentioned in section 2. Then 300 SNRs of the 300 sentences were computed. Finally, we calculated the mean of the first 300 SNRs (S0) and the mean of the later 300 SNRs(S1), respectively. Let S=S1-S0. S was filled in table 1 in corresponding position.

5. EXPERIMENTS

The experiment (Table 1) shows the effectiveness of the noise cancellation method *FBT Method*. Evaluating standard is an increase in SNRs. Those values are mean of the increase of 300 noisy sentences' SNRs. It is impersonal judgement standard. The subjective judgement standard is the articulation of the speech. Later, the relative result of the standard and recognition experiment will be mentioned.

Table 1 Evaluation of the cancellation method with *FBT Method*

	N=1	N=3	N=5	N=7
N=0.40	15	13	13	12
k=0.65	19	18	17	16
k=0.80	21	20	19	18
k=0.95	29	27	26	26

Figure 1 rude signal

In all the experiments the thresholds were calculated on the first 100 milliseconds and were not updated during the later pause section of speech.

When N=7 and k=0.65, the articulation of the speech is the best of them. Although the signal noise rate superior to other cases when k=0.95 and N=1, waveform distortion is the biggest too.

The articulation of the speech is unsatisfactory in the case.

Figure 2 Fourier Transform of rude signal

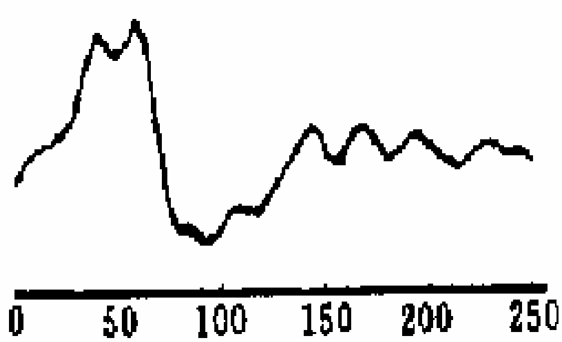


Figure 3 noisy signal

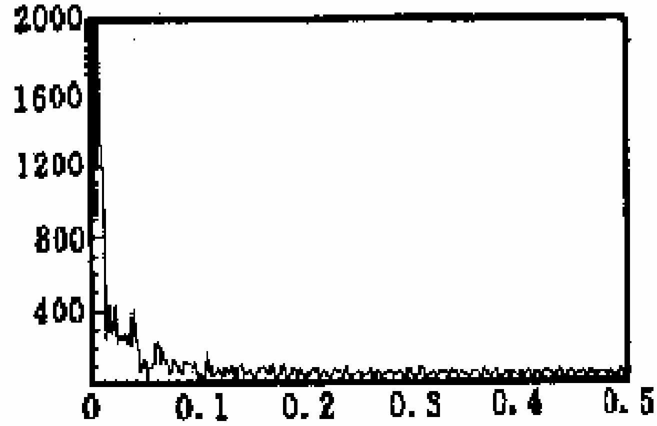
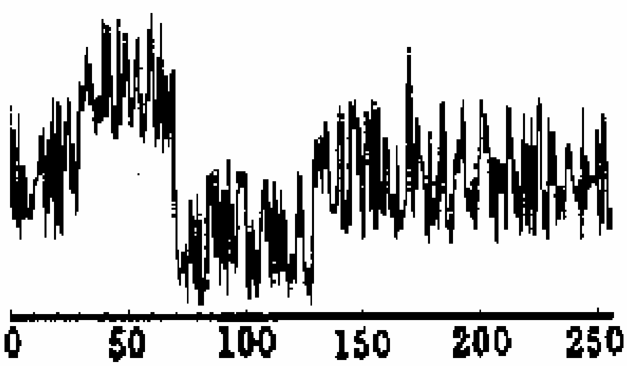


Figure 4 Fourier transform of noisy signal



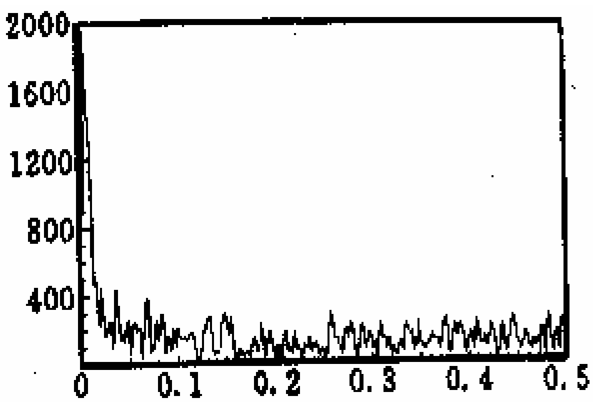


Figure 5 Ideal low-pass filtering

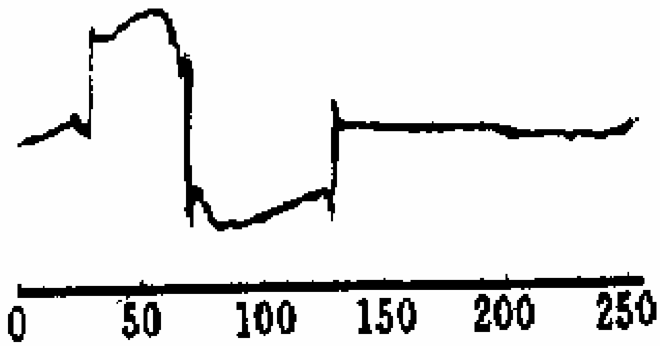
Figure 3 shows the waveform of the noisy signal and figure 4 shows its Fourier transform.

Figure 5 shows the waveform of ideal low-pass filtering the noisy signal.

Figure 6 shows the waveform of *Mallat method* filtering the noisy signal.

According to our experiment, if the speech was cleaned with *FBT Method*, the effectiveness of the recognition is worse than feature vector normalization 's in [5]. But the cleaning speech kept the edges information of the rude speech.

Figure 6 ***Mallat method*** **filtering**



6. CONCLUSION AND ANALYSIS

We presented a noise cancellation method called *FBT Method* in this paper. The method improved articulation of input speech and kept the edges information of rude speech. Although the effectiveness of the recognition is bad, the techniques can be use recognition system as a subprogram for detecting edges. It helps improving of recognition performance. We considered that the cleaned speech signal with *FBT Method* is unfit as a input signal for recognition system.

We make a comparison *FBT Method* and *ST Method*. *ST Method* possesses the simpleness and real-timeness. But it ignored the difference of the frequency bands. *FBT Method* considers the difference. The characterization of noise on each frequency band is different. For example, in the car environment energy of high frequency bands is the bigger. The thresholds of the *FBT Method* are more rational. Hence its waveform distortion is the less and the articulation is the better. *Mallat Method* enables to perfectly keep edges information. On the side, keeping edges information, *Mallat Method* is the best. We hope to combine merit *of Mallat Method* with *FBT Method's* for developing a new noise cancellation method based on wavelet transform in the further.

REFERENCES

[1] Mallat S and Hwang W L. *Singularity Detection and processing with wavelet*. IEEE Trans.on Information Theory. 1992, 38(2): 617~643

[2] Donoho and David L. De-noising by soft-thresholding. IEEE Trans. on Information Theory.1995, 41(3): 613~627

[3] Steven F.Boll, Suppression of Acoustic Noise in Speech Using Spectral Subtraction. IEEE Trans. on ASSP.1979,27(2): 113~120

[4] Serguei Koval, Mikhail Stolbov, Mikhail Khitrov, Broadband cancellation systems: new approach to working performance optimization.EuroSpeech99

[5] Viikki, Olli ,Bye, David, Laurila, Kari, Recursive feature vector normalization approach for robust speech recognition in noise. ICASSP99, 733~736

[6] Mallat S. A theory for multiresolution signal decomposition: the wavelet representation. IEEE Trans. on PAMI, 1989, 11(7): 674~693

[7] Daubeches I. The wavelet transform, time-frequency localization and signal analysis. IEEE Trans. on IT, 1990, 36(5): 961~1005