Adaptive De-noising of Speech Signals Using Wavelet Packet Transform

*Hirofumi Yogo, **Tadashi Kitamura and *Naoki Inagaki

* Dept. of Electrical and Computer Engineering, **Dept. of Computer Science
Nagoya Institute of Technology, Gokiso-cho, Showa-ku, Nagoya 466-8555, Japan

E-mail: yogo@elcom.nitech.ac.jp

Abstract: A de-noising using a Wavelet Packet transform is described. An optimum threshold for de-noising was obtained. A new selecting method of a transform level and an order of the wavelet function is also proposed. For the threshold of an adaptive de-noising, the noise estimated from the speech signals was used. The estimated noise was proportional to the optimum threshold. The other two de-noising methods were compared. The proposed method has the most prominent property among the three methods.

1. Introduction

The Wavelet Packet transform [1] is utilized for signal processing in many fields.[2] The method has a flexible and a detail processing properties. This paper describes the de-noising of speech signals using the Wavelet Packet transform. In this case, we need the selecting of the threshold, the transform level and the order of the wavelet function because the selecting of these parameters influences the de-noising results. Authors have already reported the selecting method of the transform level [3]. The threshold with constant value, which is coincident with the data length, has been utilized[4]. In order to perform the effective de-noising which corresponds to the real condition of the speech signals, the constant threshold is not appropriate. This paper describes about the new selecting method of the optimum threshold, the transform level and the optimum order of the wavelet function. The proposed method uses the estimated noise of the speech signals as the

optimum threshold. The de-noising of continuous speech signals was carried out using the three methods described next. They are (1) the Wavelet Packet transform, (2) a median filtration and (3) a moving -averaging. The most excellent method among them was the Wavelet Packet transform.

2. De-noising procedure by Wavelet Packet transform

The Wavelet Packet transform decomposes signal $S(t)$ into a coefficient vector (Detail) with higher frequency components and a coefficient vector (Approximation) with lower frequency components, as shown in Fig. 1. The coefficients ($w_{10}, w_{20}, \dots$) which are located on the left side correspond to "Approximation", and the coefficients ($w_{11}, w_{23}, \dots$) which are located on the right side correspond to "Detail". First of all, the noise corrupted speech signal is decomposed into the "detail" and the "approximation" by using the Wavelet Packet transform, as shown in Fig.2 (b) and (c). The threshold (th) is used

for the “detail” to fix the smaller coefficients than the threshold into zero, as shown in Fig. 2 (b).

Fig.1 Wavelet Packet decomposition tree

Fig.2 Schematic diagram of the de-noising procedure

And then, reconstruction of the waveform is performed by using these coefficients vector and the coefficients vector “approximation” shown in Fig.2(c), and the de-noised waveform shown in Fig.2 (d) is obtained.

3. Threshold

As the speech data, the NTT-AT’s continuous speech database was utilized. Noise corrupted speech signals were made by adding the white noise, which is generated in a computer, to the database. The threshold

with constant value defined in Eq. (1), which corresponds to the length of the data, has been utilized. [4]

$$th = \sqrt{2 \log(N \log N / \log 2)} \cdot \dots \cdot (1)$$

Where N means a length of the signal.

De-noising using the threshold was carried out for the signals with SNR=20 dB~0 dB. The order n was changed from 1 to 10 at each transform level L, and the errors between the original and the de-noised speech signals were calculated using the mean squared error (MSE) criterion defined by Eq. (2).

$$MSE = 1/N \sum_{i=1}^N (R_i - D_i)^2 \cdot \dots \cdot (2)$$

Where N means a length of the signal, R_i shows the original speech signal, and D_i means the de-noised one. The de-nosing properties are shown in Fig. 3(a) and (b). Fig.3 (a) shows the results at SNR=20 dB. We see that the smallest and stable MSE is obtained when the transform level L is 1 and the order n is 8.

(a)

(b)

Fig.3 De-noising properties (constant threshold)

It is also found that the MSE is stable when the level L is 2 and the order n is 8, as shown in Fig.3 (b). In the next, the threshold was changed for the optimum de-noising. The results are shown in Fig.4 (a) and (b). The order n was changed at each level L . Moreover, the threshold was changed in each case, and the MSE was calculated. We presumed that the optimum threshold was obtained when the MSE reached to the smallest value. We defined it as the optimum threshold $th(opt)$. Fig.4 (a) shows the results at the SNR=20 dB. We presumed that the optimum level $L(opt)$ was 3. Moreover, we decided that the optimum order $n(opt)$ was 6, because the MSE was stable at $n=6$. Fig.4 (b) shows the result at SNR=0 dB. We see that the optimum level $L(opt)$ is 3 and the optimum order $n(opt)$ is 6. By comparing Fig.3 (a) and Fig.4 (a), the MSE of Fig.4 (a) is about 5.3 dB smaller than that of Fig.3(a). The MSE of Fig.4 (b) is about 1 dB smaller than that of Fig.3 (b). It is found that the use of

(a)

(b)

Fig.4 De-noising properties (Optimum threshold)

the optimum threshold is effective.

4. SNR and Optimum threshold $th(opt)$

Fig.5 shows the relation between the SNR and the optimum threshold $th(opt)$ at the level $L=3$. The optimum threshold shown in Fig. 5 is fitted by the polynomial defined by Eq. (3).

$$\begin{aligned} \text{th}(\text{opt}) = & 0.0032S^3 - 0.0126S^2 + 0.0251S \\ & - 0.0068 \dots \dots \dots (3) \end{aligned}$$

where S means the SNR of the speech signal .
The result is shown in Fig.6.

5. Adaptive De-noising

Fig.5 Relation between SNR and th(opt)
(L=3).

Fig.6 Approximation by polynomial

Speech signals are influenced by many kinds of disturbance from the circumstance. The disturbance by the noise is always changed. Therefore, the information for de-noising is required to obtain from the speech signals themselves. So we estimated the noise component in the speech signals.

5.1 Estimated noise and the optimum threshold

By passing the signal Sn shown in Fig. 7(b) into the median filter, we obtained a filtered signal Sf shown in Fig.7(c). By subtracting the signal Sf from the signal Sn, We obtained the estimated noise Nd shown in Fig. 7(d). Then we estimated the noise component from the duration Dn in Fig.7(d). The estimated noise is shown in Fig. 8. The noise is proportional to the change of the SNR. We see that the excellent estimation of the noise is carried out. A logarithmic converted th(opt) is shown in Fig. 9. By comparing Fig. 8 and Fig. 9, it is found that two figures have similar properties. So we convinced that the estimated noise was able to use for the optimum threshold.

Fig.7 Estimating noise from the duration Dn

Fig.8 SNR vs. Estimated noise

Fig.9 Logarithmic converted th(opt).

5.2 De-noised results by three methods

The de-noised results based on the proposed method and the other two methods are shown in Fig. 10. We used 3 points median filter for de-noising. As the moving-average filter, we used a combined filter based on 3 points and 5-points moving-averaging. It is found that the de-noised result based on the Wavelet Packet transform has the most prominent property because the MSE is smallest. Then we show the original speech signal and the de-noised waveforms using the three methods. Fig.11

(a) shows the original speech waveform, (b) the noise corrupted speech waveform, (c) the result obtained by the moving-average,

Fig.10 De-noising properties by three methods

(d) the result obtained by the median filter and (e) the results processed by the Wavelet Packet transform, respectively.

Fig.11 Comparison of the de-noised results by the three methods

6. Discussions and conclusions.

The optimum threshold, the transform level and the order of the wavelet function were shown. We estimated the noise of the speech signal using the median filter. The magnitude of the estimated noise was proportional to the change of the SNR. Therefore the optimum threshold was selected based on the estimated noise. The other two methods were utilized for comparison. The method based on the proposed method showed the most prominent de-noising properties among the three methods. We intend to de-noise the speech signals in shorter duration by selecting the threshold in each duration.

7. References

- [1] R.R.Coifman, Y. Meyer, S.Quake and M.V.Wickerhauser, "Signal Processing and Compression with wavelet packet", in "Wavelets and Their Applications", edited by J.S.Byrnes,L.Byrnes, K.A.Hargreaves and K.Berry, Kluwer Academic Publications (1994).
- [2] Paul S. Addison, "The Illustrated Wavelet Transform Handbook, Institute of Physics Publishing (2002).
- [3] H.Yogo; T.Kitamura and N. Inagaki, " Restoration of Environmental Noise Corrupted Speech Using Wavelet Transform", Proc. of 31th ISCIE (SSS99), Nov. pp.141-146 (1999).
- [4] David.L,Donoh, "De-noising by Soft -thresholding", IEEE Trans. on Information Theory, Vol.41, No.3, May, pp.613-627 (1995).