

# A New Speech Denoising Algorithm Based on Bilateral Filtering and Wavelet Transform

Caixia Liu, Yanyan Hou, Bin Yang

Department of information science and engineering, Zaozhuang University, Zaozhuang, China

cxsqz@126.com

**Keywords:** speech denoising; smoothing; bilateral filtering; wavelet transform

**Abstract.** A speech denoising algorithm based on bilateral filtering and wavelet transform is proposed in this paper in order to solve the noise problem in speech processing. After the discrete wavelet transform, the low-frequency part of the speech signal is filtered by one-dimensional bilateral filtering. Then the soft thresholding wavelet denoising method is performed on the high-frequency part. The experiments show that this method can effectively reduce the noise while fully reserving the signal detail.

## Introduction

Language is the most convenient and efficient means for human beings to communicate. The process of transmitting, storing, recognizing, synthesizing and enhancing the speech by digital method is one of the most important and basic components of the digital communication network. However, the speech signal will be subjected to interference from the surrounding environment, transmission media noise, internal electrical noise of communication equipment, and even other speakers in the coding and transmission process. These disturbances will eventually make the received speech become noise-contaminated signal. At the same time, the presence of noise can worsen the performance of speech processing system, which is unfavorable to the recognition of speech signal.

The purpose of speech enhancement is to remove the background noise, improve the anti-interference ability, and extract the effective speech signal from the noisy speech signal [1]. Speech denoising is an important research direction in the field of speech signal enhancement. It plays an important role in improving the quality of speech, improving the pronunciation and reducing noise pollution. The traditional adaptive filter [2] (such as Wiener Filter, Kalman Filter and so on) for speech denoising needs to know some characteristics or statistical properties of noise. Especially, when the noise is similar to the spectrum of speech, the traditional simple time domain or frequency domain processing often can not achieve good results. With the further research of speech denoising, a variety of techniques such as expert system, Fuzzy logic, neural network and wavelet transform are produced [3-4]. The paper [5] improves the variable step size by using the sigmoid function, improves the convergence speed and the tracking speed at the same time, and obtains the small steady-state error. But the error function changes too much in the bottom step of the curve. The threshold denoising method is another popular speech denoising method. For example, modular square method, soft-hard threshold tradeoff method [6], reduced threshold method [7] and so on. Threshold function can improve the denoising effect by modifying the threshold [8]. The hard threshold function is a discontinuous function, which leads to the oscillation of the reconstructed signal. The soft threshold function has a certain deviation between the denosing signal and the real value of the signal [9]. The realization of noise reduction method of EMD (Empirical Mode Decomposition) [10-11] is mainly use the adaptive generation of the IMF component to replace the original signal. It does not integrate and apply window function, the reduction effect is significant. But the effect of improving the uniformity is not good, that is, the error amplitude deviation is bigger than average value [12].

The bilateral filter is applied in image smoothing. Gaussian smoothing weights depend on the distance of a pixel's neighborhood and the function of generating weights is Gaussian function. So it

is called Gaussian smoothing or Gaussian filter. The effect is to blur the image while preserve edge. The improvement of bilateral filtering is to increase the effect of gray value, that is, if the gray value of the neighborhood is closer to the central pixel, the weights are bigger. Conversely, if the gray scale difference is very large, the weight is smaller, and so on, and the template coefficients are normalized. So the template's coefficients are no longer purely depended on the position. They are depended on the gray and the edges will be able to be effectively saved.

In this paper, the bilateral filtering is introduced into one-dimensional speech signal denoising, combined with the wavelet denoising technology, a speech denoising algorithm based on bilateral filtering speech and wavelet transform is proposed. The signal is divided into low frequency part and high frequency part by discrete wavelet transform. The bilateral filtering is performed on the low frequency part and soft threshold denoising on high frequency part. The method can reduce the noise while maintaining the details.

### The Speech Denoising Algorithm Based on Bilateral Filtering and Wavelet Transform

**Wavelet Transform.** The discrete wavelet transform of function  $f(t)$  in arbitrary mean square integrable space  $L^2(R)$  is:

$$W_x(j, k) = \int_R f(t)\overline{\phi_{j,k}(t)}dt \tag{1}$$

where, wavelet base  $\phi$  is:

$$\phi_{j,k}(t) = \frac{1}{\sqrt{2^j}}\phi\left(\frac{t}{2^j} - k\right) \tag{2}$$

Here, discretization is for continuous scaling parameters and continuous translation parameters. By the discrete wavelet transform of signal  $f(t)$ , the high frequency and low frequency components of the signal are obtained, as shown in Fig. 1.

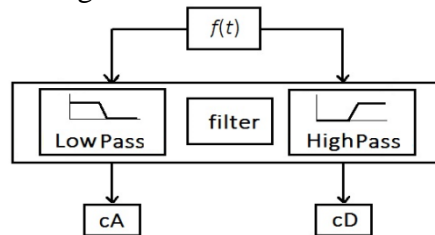


Figure 1. Discrete Wavelet Transform

The Low frequency component characterizes the signal itself and represents the profile of the signal. The High frequency component characterizes the subtle difference of the signal, while the noise is mainly concentrated in high frequency. In order to achieve the goal of smoothing and denoising on the premise of keeping the details, the low frequency component is smoothed by one-dimensional bilateral filtering, and the low frequency component is denoised with soft thresholding wavelet method.

**Bilateral Filtering on Low Frequency Component.** Bilateral filtering is an iterative edge-preserving filtering strategy. The basic idea of bilateral filtering is to calculate weighted average of adjacent points in space by spatial filtering. The weighted parameter reduces with the increasing of distance, and the aim of protecting edge is achieved. Weight average of similar gray value pixels is obtained by means of range filtering. The weighted parameter reduces with the increasing of difference value, and the aim of smoothing and denoising is achieved.

For an image  $f(x, y)$ , its bilateral filtering function is shown in the Eq. 3.

$$g(i, j) = \frac{\sum_{k,l} f(k,l)\omega(i, j, k, l)}{\sum_{k,l} \omega(i, j, k, l)} \tag{3}$$

where,  $\omega(i, j, k, l)$  is weight factor,  $f(k, l)$  is the pixel of Gaussian filter window.

For speech signal, we can improve the efficiency of bilateral filtering significantly by one-dimensional bilateral filtering. First, the two-dimensional bilateral filtering as shown in Eq. 3 is reduced dimension, and a one-dimensional bilateral filtering formula of signal  $f(t)$  is obtained, as shown in Eq. 4.

$$g(i) = \frac{\sum_l f(l)\omega(i,l)}{\sum_l \omega(i,l)} \tag{4}$$

where, the weight factor is:

$$\omega(i,l) = \exp\left(-\frac{(i-l)^2}{2\sigma_d^2} - \frac{\|f(i) - f(l)\|^2}{2\sigma_r^2}\right) \tag{5}$$

The value of  $f(i) - f(l)$  varies very little, and the spatial weight plays a major role and it achieves the aim of Gaussian smoothing. At the time point of large amplitude mutation,  $f(i) - f(l)$  will have a larger difference, and the range coefficient decreases, which resulting in the reduction of distribution of the entire kernel function and maintain the signal detail information.

The result of low-frequency filtering on Fig. 2 (a) is shown in Fig. 2 (b). It can be seen that the signal is smoothed while keeping the details.

**Wavelet denoising on High Frequency Component.** A model of one-dimensional signal containing noise can be expressed as follows:

$$s(t) = f(t) + \gamma \cdot e(t), \quad t = 0, 1, 2, \dots, n - 1 \tag{6}$$

where  $f(t)$  is the real signal,  $e(t)$  is noise,  $s(t)$  is a signal containing noise. The signal actually behaves as a low-frequency signal or some relatively stationary signal, and the noise signal is usually expressed as high frequency signal. There are three kinds of methods for eliminating noises in high frequency signal: forced denoising, default threshold denoising, soft or hard threshold denoising.

The high frequency part uses the soft threshold to denoise.

$$WT = \begin{cases} \text{sign}(WT)(|WT| - TH), & |WT| > TH \\ 0, & |WT| \leq TH \end{cases} \tag{7}$$

The results are shown in Fig.2 (c).

Combined with bilateral filtering of low-frequency and soft threshold denoising of high frequency, the result is shown in Fig. 2 (d).

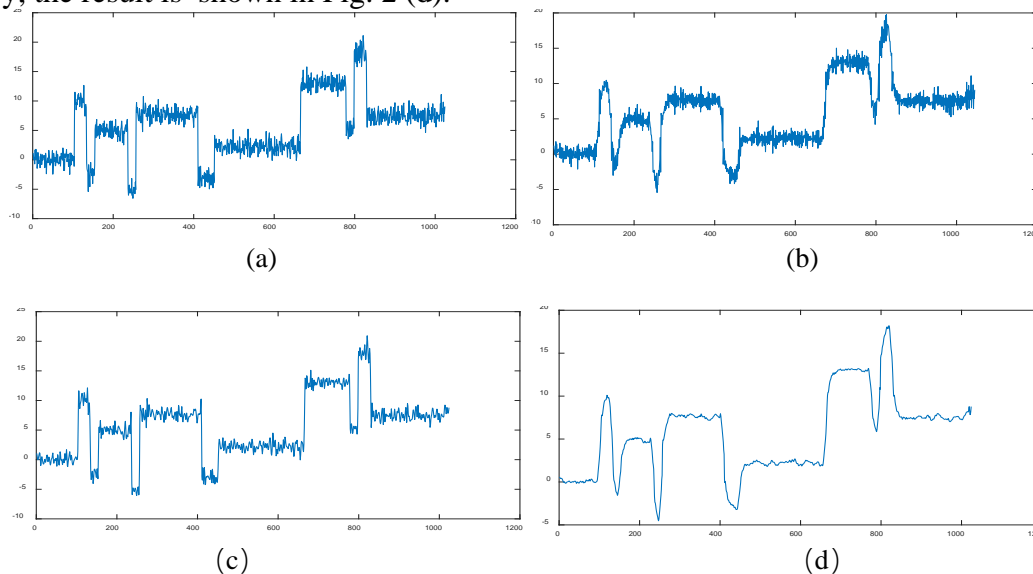


Figure 2 signal denoising. (a) is noisy signal; (b) is the result of bilateral filtering on low-frequency; (c) is the result of soft threshold denoising on high frequency; (d) is the final result.

### Results Analysis

The following are smooth and denoising results of the original signal in Figure 2 (a) with different methods. Fig. 3 (a) is the denoising result of wavelet soft thresholding. Fig. 3 (b) is the denoising result of wavelet hard threshold. Fig. 3 (c) is the result of FFT denoising. It is found that the denoising results of our algorithm are better.

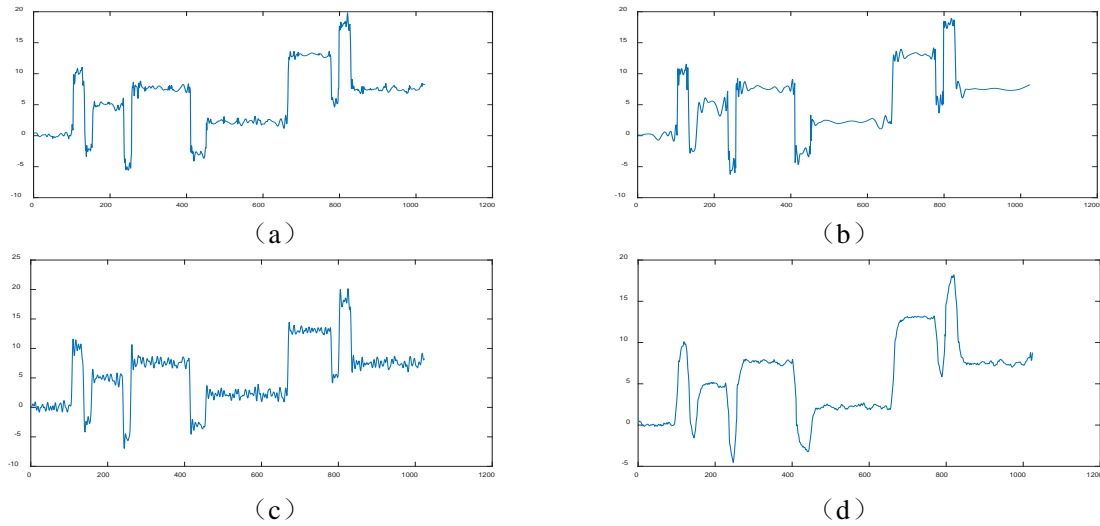


Figure 3. The denoising results with different method. (a) is with soft wavelet threshold; (b) is with hard wavelet threshold; (c) is FFT method; (d) is our method.

In the following, a speech signal is smoothed and denoised. Fig.4 (a) is the original voice signal, and Fig.4 (b) is noisy signal. Fig.4 (c) is the wavelet soft threshold denoising result. Fig.4 (d) is the wavelet hard threshold denoising result. Fig.4(e) is FFT denoising result and Fig.4 (f) is the result of our method. It is found that our denoising result is very obvious and the original signal is smoothed effectively at the same time.

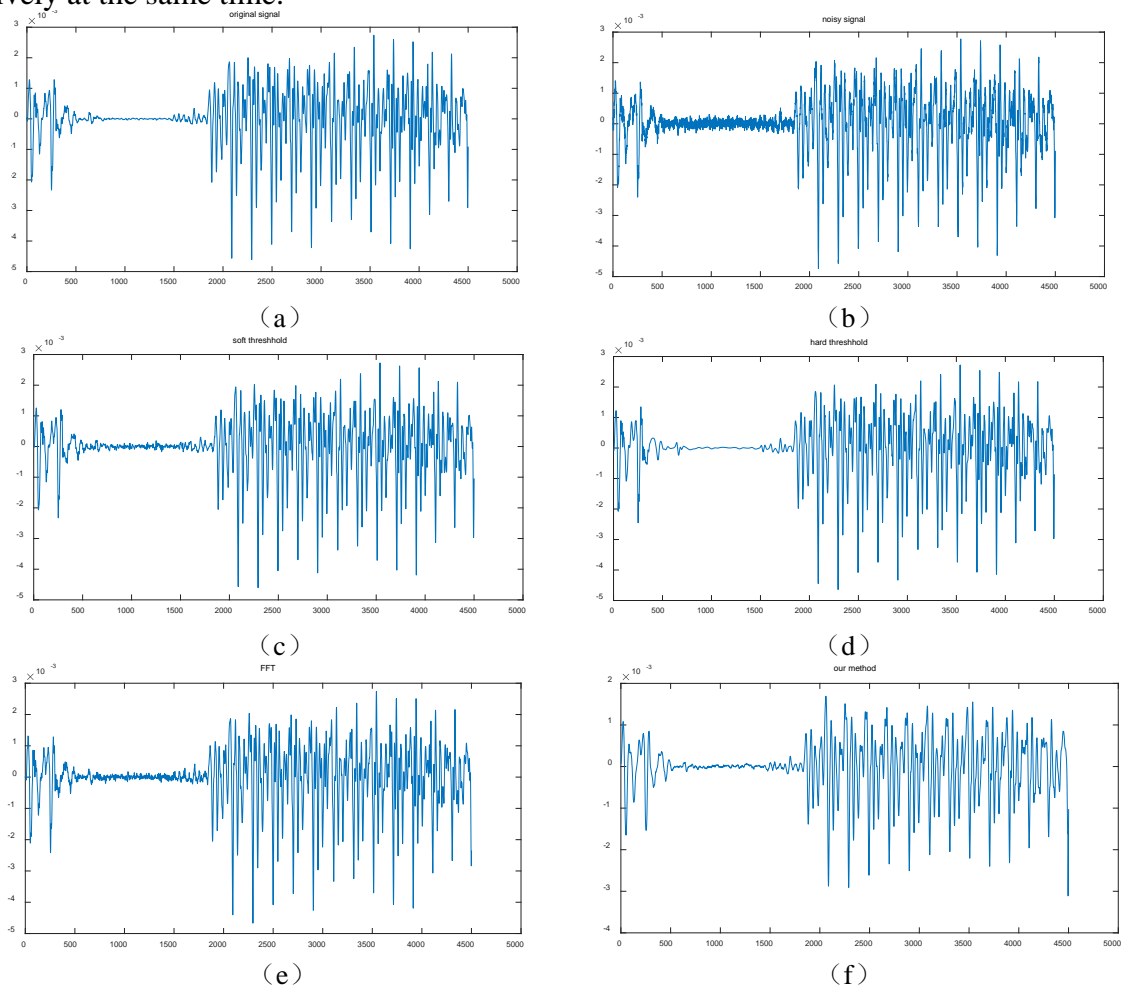


Figure 4. The speech signal denoising. (a) is the original voice signal; (b) is noisy signal; (c) is the wavelet soft threshold denoising result; (d) is the wavelet hard threshold denoising result; (e) is FFT denoising result and (f) is the result of our method

## Conclusion

A new algorithm of speech smoothing and denoising is presented in this paper. By using the discrete wavelet transform, the signal is divided into two parts. The bilateral filtering and soft thresholding method are used respectively in low frequency part and high frequency part. The result of smoothing and denoising is very good.

## References

- [1] Hang Hu. Modern Speech Signal Processing. Publishing House of Electronics Industry, Beijing, 2014, 15-20.
- [2] Xianda Zhang. Modern Signal Processing. Tsinghua University Press, Beijing, 2002.
- [3] PINNEGA RCR, MANSINNHA L. Time local Fourier analysis with a scalable, phase modulated analyzing function: the s-transform with a complex window. Signal Process, 2004, 84 (1) 1167-1176.
- [4] Zhang Ting, He Ling, Huang Hua. Speech endpoint detection based on critical band and energy entropy. Journal of Computer Applications. 2013, 33 (1) 175-178.
- [5] Qin J F, Yang J Z. A novel variable step size LMS adaptive filtering algorithm based on sigmoid function. Journal of Data Acquisition & Processing, 1997, 12 ( 3) 171-172.
- [6] Zhang Rui, Zhang Youchun. New threshold Function in Wavelet Threshold de-noising. Computer Engineering and Applications, 2013, 49 (15) 215-218.
- [7] Wang Biao. Research on an Improved Threshold Denoising Method. Computer and Digital Engineering, 2012, 40 (3) 30-31.
- [8] Zhu Yashu, Ni BinBin. Gu Xudong. A Study of Speech Denoising Threshold Functions Based on Wavelet Transformation. Journal of Hebei North University (Natural Science Edition). 2017, 33 (9) 29-33.
- [9] Chen Xiaojuan, Wang Wenting, Jia Mingchao, Song Na. New denoising method for speech signal based on wavelet entropy and adaptive threshold. Application and Research of Computers, 2014, 31 (3) 753-755.
- [10] KEOGH E , CHU S , HART D , PAZZANI M .An online algorithm for segmenting time series .In International Conference on Data Mining 2001.Los Alamitos :IEEE Computer Society , 2001.289-296 .
- [11] HAGIT S , STANLEY B .Approximate queries and representations for large data sequences .In Proceedings of the Twelfth International Conference on Data Engineering .Washington DC :IEEE Computer Society, 1996 .536 -545.
- [12] Wan Jian, Zhao Chunhui, Fu Jin. Study on speech signal denoising algorithm based on EMD. Journal of Harbin University of Commerce (Natural Sciences Edition) 2005, 21 (4) 479-485.