A Speech denoising method Based on wavelet threshold and improved spectral subtraction

Ling Wei
College of Computer Science & Information
Guizhou University
Guiyang, China
Wling1987@163.com

Li Wang
College of Computer Science & Information
Guizhou University
Guiyang, China
Wangl_@tom.com

Abstract—In the speech de-noising method, wavelet threshold method and the spectral subtraction with simple algorithm and good performance has attracted the attention of many researchers. Based on the traditional spectral subtraction, the author found that the speech signal non-stationary noise cannot be effectively restrained, therefore, the improved algorithm based on wavelet threshold value and improved spectral subtraction is proposed The Experimental results show that this method in low signal-to-noise ratio, can effectively restrain "music noise", enhance the signal to noise ratio (SNR) and improve the quality of speech.

Keywords-Wavelet transform; Speech de-noising; Spectral subtraction

I. INTRODUCTION

Speech signal is a kind of human communication and transfer of information basic mode, but in actual speech communication environment, the speech signal is always will be affected by outside noise interference, so need to voice signal noise reduction processing, restore pure signal. How to remove the speech signal is not relevant noise, and improve the speech signal speech quality and signal-to-noise ratio, becomes the important topic of speech processing.

At present speech signal de-noising method has a lot of, such as wavelet transform, neural network, spectral subtraction and all. Wavelet transform[1] is a time-frequency analysis method, in time and frequency domain have good local characteristics, overcome the short time Fourier transform fixed resolution shortcomings, able to analysis signal details, has the higher frequency resolution, therefore in the signal de-noising has very good application. In 1992, P.Lockwood and J.Buody puts forward nonlinear spectral subtraction (NSS)[2], since then, spectral subtraction combining with other method produce a lot of effective speech enhancement methods, can remove spectral subtraction residual music noise in a certain extent. Although these algorithms processing effect is not obvious, but the spectral subtraction algorithm has outstanding advantages of simple algorithm and small amount of calculation, has been widely used in practice. The deficiency is after treatment can produce has certain rhythmic, sounds like music "music noise"[3]. By introducing two parameters α and β of the improved spectral subtraction algorithm, combined with wavelet threshold de-noising algorithm is put forward a new kind of de-noising method. The simulation results show that the improved method has better speech noise reduction effect.

II. WAVELET TRANSFORM AND DE-NOISING PRINCIPLE

Wavelet transform is a kind of time-frequency localization analysis method, time window and frequency window can be changed, in the high frequency part have high temporal resolution and low frequency resolution, low frequency part instead.

Continuous wavelet transform is defined as:

$$W_f(a,b) = |a|^{-1/2} \int_R f(t)\psi^*\left(\frac{t-b}{a}\right)dt \tag{1}$$

Where f (t) is any continuous function or signal, a is expansion factor, b is translation factor, $\psi(t)$ meet permissible conditions of wavelet base, namely mother wavelet or basic wavelet. The generating function $\psi(t)$ after expansion and translation is:

$$\psi_{a,b}(t) = |a|^{-1/2}\psi\left(\frac{t-b}{a}\right) \tag{2}$$

Usually, decretive scale parameter a and translation parameter b, then, $a = a_0^j, b = ka_0^j b_0, j \in Z$, get discrete wavelet transform defined as:

$$W_f(j,k) = |a_0|^{-j/2}\int_{-\infty}^{+\infty} f(t)\psi\left(a^{-j}t - kb_0\right)dt \tag{3}$$

Its reconstruction formula is:

$$f(t) = C\sum_{j=-\infty}^{+\infty}\sum_{k=-\infty}^{+\infty} C_{j,k}\psi_{j,k}(t) \tag{4}$$

Here, C is a signal independent constant.

In actual engineering, the useful signal average performance is low frequency signal or stationary signal and

noise is behaved for high frequency signal. So, wavelet threshold de-noising process generally according to the following three steps treatment:

1) The signal wavelet decomposition. To choose a suitable wavelet, determine the decomposition level N, then wavelet decomposition..

2) High frequency coefficient of Wavelet decomposition threshold quantification, get estimate wavelet coefficient $\hat{w}_{jk}$.

3) Wavelet reconstruction, get de-noising signal.

In the above three steps, threshold selection and threshold quantification is wavelet threshold de-noising core problem, directly affect the quality of the signal de-noising. Literature [4] adopting soft threshold method for noise reduction, the method of soft threshold function derivative discontinuous, direct impact on the reconstructed signal and the original signal difference, there have some limitations and needs further improvement.

III. SPECTRAL SUBTRACTION AND DE-NOISING PRINCIPLE

The spectral subtraction (SS) [5] is on the conditions of assumed additive noise and short time smooth speech signal independent of each other, from the power spectrum of signal with noise subtract the noise power spectrum, get more pure voice spectrum. Hypothesis background noise is the additive noise, and speech signal is not relevant. A frame of the speech signal with noise can be expressed as:

$$y(i) = s(i) + d(i), 0 \le i \le N-1 \quad (5)$$

Here, the N is a frame speech signal frames length, $s(i)$ is pure speech signal, $d(i)$ is noise signal.

Set Y_k is Fourier coefficient of $y(i)$, S_k is Fourier coefficient of $s(i)$, and D_k is Fourier coefficient of $d(i)$ .For the signal power spectrum can be expressed as:

$$|Y_k|^2 = |S_k|^2 + |D_k|^2 + S_k D_k^* + S_k^* D_k \quad (6)$$

Because $s(i)$ and $d(i)$ mutual independence, meet the Gaussian distribution and average is zero, so:

$$E\left(|Y_k|^2\right) = E\left(|S_k|^2\right) + E\left(|D_k|^2\right) \quad (7)$$

For a frame in the short time stationary process, there are: $|Y_k|^2 = |S_k|^2 + \lambda_n(k)$, $\lambda_n(k)$ is no sounds $|D_k|^2$ statistical average value, the original speech estimate:

$$\left|\hat{S}_k\right| = \left[|Y_k|^2 - E\left(|D_k|^2\right)\right]^{1/2} = \left[|Y_k|^2 - \lambda_n(k)\right]^{1/2} \quad (8)$$

Here, $\left|\hat{S}_k\right|$ is the enhanced speech amplitude.

IV. AN IMPROVED DE-NOISING METHOD

A. Improved spectral subtraction

Due to the spectral reduction process, is silent period statistical average noise variance instead of the current analysis frame the frequency point noise spectrum component; And noise spectrum with Gaussian distribution, namely its amplitude random variation range is very wide, so if the frame subtraction when a certain frequency point noise component is bigger, there will be a big part of noise residual, in the spectrum presents on the random peak, thus formed on the auditory rhythmical fluctuation of similar music of the residual noise, also known as "music noise" [6]. According to the basic spectrum subtraction of the residual noise too much and residual "music noise" strong, Berouti proposed important improvement scheme is introducing parameter α , and in the subtraction when noise spectrum takes on a greater than 1 parameter β [7], namely:

$$\left|\hat{S}_k\right| = \left[|Y_k|^\alpha - \beta \lambda_n^\alpha(k)\right]^{1/\alpha} \quad (9)$$

The improved form and ordinary spectral subtraction difference is introduced α and β parameters to improve the algorithm's flexibility. α is power correction coefficient, increase α , will improve the signal-to-noise ratio, but also will be more easy to cause the speech signal distortion. β is noise factor, it can be reduced to correction of noise power [8].

From the formula (9) known, when $\alpha = 2$, $\beta = 1$, is the basic spectrum subtraction. Figure 1 is basic principle diagram of the improved spectral subtraction.
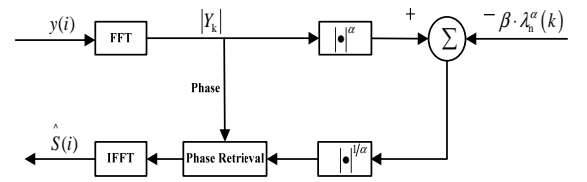

Figure 1. The basic principle diagram of improved spectral subtraction

B. The de-noising method base on Wavelet threshold and improved spectral subtraction

The speech signal after add window, frame, improved spectral subtraction treatment, get speech signal after the first

step noise reduction. Then, wavelet threshold de-noising, further inhibit the background noise. Specific steps are as follows:

1) Speech signal after add window ,frame, carries on the 256 point Fourier transform , get the spectrum of the speech signal with noise.

2) Each frame in Fourier transform, and calculation of frame of power spectral density, Set a zero threshold and the value of α, β. If the data is greater than this value, then according to the formula (9) get the frame of speech amplitude S(i) after denoising; If the data is less than the threshold value, then the frame S(i)= 0.

3) According to the phase of speech signal with noise, can get denoising signal phase, thus get denoising signal spectrum.

4) Fourier inverse transform.

5) After the above processing , the speech signal is decomposed by "Daubechies 6" Wavelet at 3-layers for each layer of Wavelet coefficients.

6) Wavelet decomposition coefficient threshold quantification.

$$\hat{w}_{j,k} = \begin{cases} \mathrm{sgn}(w_{j,k})(|w_{j,k}|-\lambda) & |w_{j,k}| \geq \lambda \\ 0 & |w_{j,k}| \geq \lambda \end{cases} \quad (10)$$

7) Wavelet reconstruction. According to wavelet decomposition of the Nth layer low frequency coefficient, and after the quantization process first to the Nth layer of high frequency coefficient. Wavelet reconstruction, get the speech signal after denoising.

V. SIMULATION RESULTS

Experimental speech signal were recorded by Cool Edit software in a laboratory quiet setting, speech content for the "according to industry experts analysis". Speech Signal with 8000Hz sample frequency, single channel, 16bit quantitative, PCM code. Noise is the 0.1*randn (1, n) of normal random noise, n is speech signal length value. The speech with noise through the window length is 256 hamming window, adjacent two frame shift is 80 points. Experimental graphs and results analysis as show follows:

Figure 2 is respectively with the spectral subtraction and wavelet threshold de-noising. From figure 2(c) can see, wavelet threshold function to noise clear enough clean, especially in no speech segment of noise removal is not complete, directly affect the audition effect. Also, the figure 2(d) in traditional spectral subtraction and cannot completely remove background noise, affects the speech sound effects; And from figure 2(e), the improved spectral subtraction introduce two parameters for the improved algorithm creates a lot of flexibility, by adjusting the two parameters can be obtained than the basic spectrum subtraction or wavelet threshold method better effect. According to the data in Table 1 will value α for 3, β for 5, relatively moderate, the

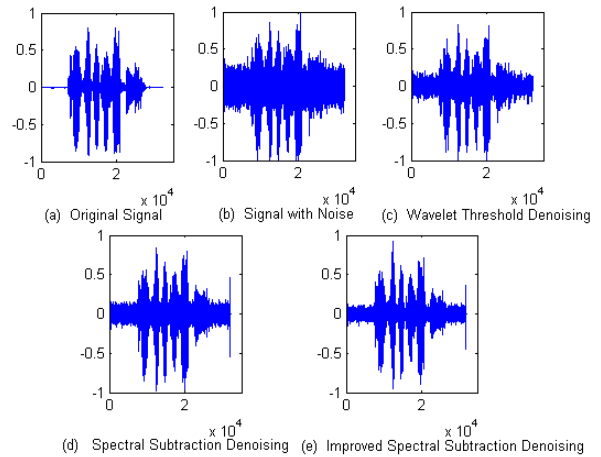SNR of the pronunciation of large and small relative distortion.

(a) Original Signal (b) Signal with Noise (c) Wavelet Threshold Denoising
(d) Spectral Subtraction Denoising (e) Improved Spectral Subtraction Denoising

Figure 2. Spectal subtraction and wavelet threshold de-noisingmethod

TABLE I. OUTPUT SNR MEAN VALUE AFTER THE IMPROVED ALGORITHM WHEN α AND β DIFFERENT VALUES

α	β	Output SNR mean value after the improved algorithm （dB）
2	3	7.601
2	5	8.341
2	7	7.850
3	3	9.023
3	5	9.097
3	7	7.858
4	3	9.081
4	5	8.803
4	7	8.301

Figure 3 and figure 4 respectively are in the SNR were 3 dB and 10 dB, using the wavelet threshold and improved spectral subtraction combined with the noise. It can be concluded from figure 3(c) that after such a process, the effect is better than the spectral subtraction or wavelet threshold de-noising method, speech amplitude slightly increase, in addition to the effective noise, enhance de-noising effect have been improved obviously. And from figure 4(c), the noise removal more clean, but to voice damage also will be very big, signal-to-noise ratio also declined; finally get the speech distortion is bigger. Thus it can be seen that this algorithm was suitable for the low SNR of speech signal.
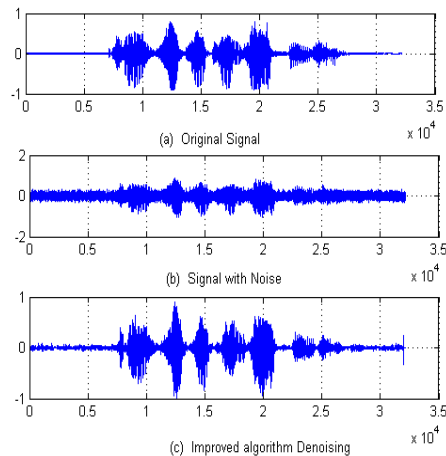
Figure 3.　signal after improved algorithm (SNR=3 dB)

Figure 4.　signal after improved algorithm (SNR=10 dB)

Speech de-noising effect evaluation based on input and output signal error size between, that is the output SNR of the judge. From the Table 2, it is known that the improvement compared to the spectral subtraction basic spectrum subtraction noise reduction effect improved, but improve amplitude and is not very big, Wavelet threshold method in a certain extent improve the de-noising effect. Wavelet threshold value and improved spectrum subtraction combining de-noising method, the output signal-to-noise ratio had the very big enhancement, "music noise" reduces, and to contain the white noise speech has good de-noising effect.

TABLE II.　　SPEECH DENOISING EFFECT EXPERIMENTAL RESULTS COMPARISON

Input SNR （dB）	Output SNR （dB）			
	wavelet threshold	*spectral subtraction*	*Improved spectral subtraction*	*wavelet threshold and Improved spectral subtraction*
-3	3.0339	-0.96043	0.61832	4.9505
0	4.7742	3.2796	4.7949	7.0603
3	7.9389	8.1102	8.7110	9.0968

VI.　CONCLUSION

Compared to the direct using spectrum subtraction or wavelet threshold method of speech de-noising, speech de-noising method not only introduced two parameters α and β, increase the algorithm flexibility, but also combines the wavelet threshold method, making "music noise" get obvious inhibition, signal-to-noise ratio has improved greatly, and improve the quality of speech, achieving very good speech enhancement effect based on the basic spectrum subtraction. The algorithm is suitable for low SNR environment of speech signal noise reduction, but slight speech distortion exists, continuous study is being made in view of the deficiency
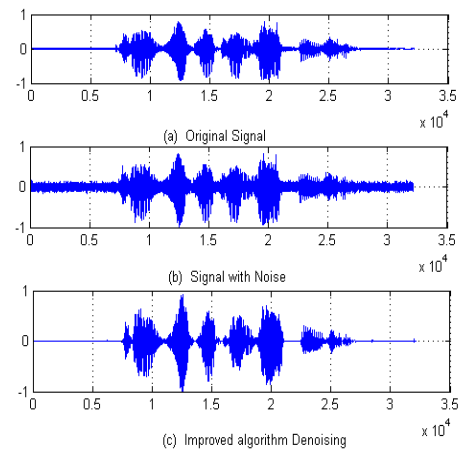
REFERENCES

[1] Jan P.H.van Santen, Richard W.Sproat, Joseph P.Olive, Julia Hirschberg, Progress in speech synthesis [M] ． New York:Springer，1997:25-30.

[2] P Lochwood, J Bandy. Experiments with a nonlinear special subtractiuon(NSS), hidden Markov models and projection for robust recognition in cars [J]. Speech Communication, 1992,11(6):215-228.

[3] Xiaoling Cao, Suli Zhang, Ping Wu, MATLAB -.based Spectral subtraction algorithm for speech enhancement [J]. Computer Simulation, 2006,23(3):278-283.

[4] Shuyu Li, Improved Wavelet threshold denoising method and its simulation using MATLAB [J]. Noise and Vibration Control, 2010,30(2):121-124..

[5] S.Boll. Suppression of acoustic noise in speech using spectral subtraction[J]. IEEE Trans on Acoustic Speech and Signal processing, 1979,27(2):113-120.

[6] Qin Lin, Daoxin Zhang, Xiaopei Wu, A new speech denoising algorithm based on improved spectral subtraction [J]. Computer Technology and Development, 2007,17(7):63-66..

[7] Berouti M. Enhancement of Speech Corrupted by Acoustic Noise[C]. IEEE International Conference on Acousties, Speech, and Signal Processing, Washington DC. 1979:208-211.

[8] Yong Xiao, Speech recognition algorithm is robust to noise research [D]. Central South University, 2009.