An Optimization Adaptive BWT Speech Enhancement Method

1Cao Bin-Fang, 2Li Jian-Qi, 3Qu Peixin and 1Peng Guang-Han
1College of Physics and Electronics Sciences,
2Communication and Electric Engineering College, Hunan University of Arts and Science, Chang de, China
3School of Information and Engineering, Henan Institute of Science and Technology,
Xinxiang, 453003, China

Abstract: Since it is difficult to choose the threshold function and value for noisy speech signal after wavelet transform, this study proposed a bionic wavelet method of hierarchical threshold based on PSO. It firstly used bionic wavelet transform for wavelet decomposition of noisy speech signal. Then, the proposed threshold function was used for threshold processing and PSO algorithm was introduced to complete the hierarchical optimization. Finally, high frequency noise component separated by bionical wavelet transform is used as the input of the adaptive filter, to ensure complete removal of the signal relevant noise. Experimental results show that the method has a prominent effect of speech enhancement under different SNR conditions and achieves the optimal estimate of valuable signal and noisy components of the same frequency.

Keywords: Speech enhancement, bionic wavelet transform, multilevel threshold, adaptive filter, PSO

INTRODUCTION

Speech enhancement technology is an important problem within the field of speech and signal processing with impact on much computer-based recognition, coding and communication application. The underlying goal of speech enhancement is to improve the quality and intelligibility of signals, as perceived by human listeners and to remove the noisy signals (You et al., 2007). Over the development in the past four decades, researchers have proposed sorts of speech enhancement algorithms which include the traditional Wiener filtering (Liu et al., 1979), spectral subtraction (Boll, 1979), method based on subspace (Jensen et al., 1995), method based on statistics model (Ephraim and Malah, 1984) and method based on wavelet (Donoho, 1995). These methods have advantages such as simple algorithm or apparent noise component removal.

Wavelet Transform (WT) is a tool used for analyzing the time-variant and non-stationary signal where optimum filtering cannot be achieved by Wiener filtering or Kalman filtering. WT can considerably check the noisy signals. Donoho proposed soft-hard threshold transform which enjoyed a wide application and comparatively has achieved better result. But, this method has inherent defects: firstly, noisy signals separated are not necessarily the optimum estimation when overlapping between noises and useful signals; secondly, amplitude and frequency of auditory effect is unable to be represented; thirdly, hard to choose the best Bionic Wavelet Transform (BWT) threshold function and threshold value.

Therefore, adaptive Bionic Wavelet Transform was proposed in reference (Cao and Li, 2010) to solve the first two problems above. Particle Swarm Optimization algorithm (PSO) was based on the conception: Firstly, to decompose noisy speech signal using BWT can guarantee the auditory effect of amplitude and frequency; secondly, to threshold the threshold function proposed in sub-band which has been separated and to optimize the threshold by introducing particle swarm algorithm; thirdly, to reconstruct the signal processed by threshold function. Compared with standard bionic filtering, adaptive bionic wavelet transform based on PSO can guarantee its better auditory effect and displays better filtering property, as considered to be a new approach to remove noisy speech signals.

BIonic WAVELET TRANSFORM (BWT) OF NOISy SPEECH SIGNALS

Human ear is a perfect device which collects and deals with audio signals (Bahoura and Rouat, 2001). The outer ear, middle ear and inner ear constitute the human peripheral auditory system. The main function of outer and middle ear is to enhance and compensate the incoming sound. Receipt of sound and decomposition are performed in membrane, the main component of the inner
ear which can adapt signal analysis according to the frequency. This analysis ability to low frequency is better in comparison with high frequency. So we must consider the auditory property in the process of speech enhancement. Many previous methods of speech enhancement introduce the band when decompressing noisy signal but it cannot embody the auditory property of amplitude and frequency. Therefore, bionic wavelet transform was proposed to analyze the speech signal in this study.

**Bionic wavelet transform (BWT):** Based on the modified Giguerre’s auditory circuit model (Jun and Zhang, 2001), Yao proposed a new time-frequency method named Bionic Wavelet Transform (BWT) which integrates cochlea mechanism into wavelet transform. “Bionic” means that it is rooted in an active biological mechanism. The BWT is distinguished from standard wavelet transform in that the resolution in the time-frequency domain achieved by the BWT can be adaptively adjusted not only by signal frequency changes but also by signal’s instantaneous amplitude and its first-order differential coefficient. In the process of standard wavelet transform Hu and Loizou (2004), mother bionic function h(t) must meet the admissible requirement. So, we use h(t) to represent the envelope function.

\[
h(t) = \frac{1}{\sqrt{a}} h(t) \exp(jw_0t) \tag{1}
\]

where, \(w_0 = 2\pi f_0\), \(f_0\) is the windowed central frequency of \(h(t)\).

BWT of signal \(f(t)\) can be defined as:

\[
\text{BWT}_T f(a, \tau) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) h\left(\frac{t - \tau}{a}\right) \exp(-jw_0\left(\frac{t - \tau}{a}\right)) dt \tag{2}
\]

where, \(\alpha, \tau\) represent time shift and scale variables, respectively. In order to simulate the active basilar membrane using BWT, Yao integrates a new adjustment parameter \(T\) into bionic mother function:

\[
h_T(t) = \frac{1}{T} h(t) \exp(jw_0t) \tag{3}
\]

where, \(T\) was introduced by following the active auditory model proposed by Giguerre. It can be defined as:

\[
T(\tau, \Delta) = 0 - \frac{\text{BWT}_T f(a, \tau)}{\text{BWT}_T f(a, \tau) + G_1 + G_2(\text{BWT}_T f(a, \tau))} \tag{4}
\]

Here, \(\text{BWT}_T f(a, \tau)\) is bionic wavelet coefficient with time shift \(\tau\) and scale variable \(a\); where, \(G_1\) and \(G_2\) represent the active gain factor, \(\text{BWT}_T f(a, \tau)\) is saturation factor representing non-linear saturation effect in cochlear model; \(\Delta\) is the calculation step. So we can define bionic wavelet transform as:

\[
\text{BWT}_T f(a, \tau) = \frac{1}{T} \int_{-\infty}^{\infty} f(t) h\left(\frac{t - \tau}{a}\right) \exp(-jw_0\left(\frac{t - \tau}{a}\right)) dt \tag{5}
\]

**Fast BWT algorithm:** Though BWT has sorts of advantages as compared with standard wavelet transform, in early times its application is limited because of its time-consuming calculation. The relationship are found existent among Morlet mother function, bionic wavelet function and standard bionic wavelet transform by making further improvement in BWT, as shown in Chen and Wang (2004):

\[
\text{BWT}_T f(a, \tau) = \frac{1.772\sqrt{\pi} T}{\sqrt{1 + T^2}} \cdot \text{WTf}(a, \tau) \tag{6}
\]

where, \(\text{BWT}_T f\) is BWT coefficient; \(\text{WTf}\) is WT coefficient and \(T\) the constant of Morlet mother function.

**Threshold denoising:** De-noise method using soft threshold and hard threshold was proposed by Donoho and has received a widespread popularity form scholars home and abroad. The threshold function can be defined as:

- **Hard threshold:**
  \[
  W^H = \begin{cases} 
  W, & |W| > \lambda \\
  0, & |W| \leq \lambda 
  \end{cases} \tag{7}
  \]

- **Soft threshold:**
  \[
  W^S = \begin{cases} 
  \text{sign}(W)(|W| - \lambda), & |W| > \lambda \\
  0, & |W| \leq \lambda 
  \end{cases} \tag{8}
  \]

where, \(\text{sign}(x)\) represents sign function. Threshold function proposed by Donoho has been widely applied in many areas but the defects still exist both in hard threshold and soft threshold. The wavelet coefficient obtained by hard threshold could possibly cause fluctuation in the process of reconstructing signals due to the fact that the threshold value chosen is discontinuous; despite its continuity of wavelet coefficient obtained by using soft threshold, in reality, the stable deviation remains which directly affects the signals that have been reconstructed.

At present, researchers have proposed many improved threshold methods which are inappropriate in that all the coefficients below a predefined value zero are set to zero. To guarantee the smooth transition of wavelet coefficient as well as good
quality of removing noisy signals, a new threshold function was put forward in this study.

The threshold function can be properly described as:

$$W_k = \begin{cases} \text{sign}(W_k)\sqrt{(W^2 - (\alpha A)^2)}, & W \geq \lambda \\ 0, & W < \lambda \end{cases}$$

(9)

where, $\alpha$ is the parameter between 0 and 1, namely, wavelet coefficient value steadily reduces to zero when $W_k$ is below the threshold value. Otherwise, square of coefficients subtracts square of wavelet threshold value, square roots for the results so that musical noise can be reduced. By selecting threshold function, level-dependent particle swarm optimization is introduced in the study.

**THRESHOLD ALGORITHM BASED ON PARTICLE SWARM OPTIMIZATION**

**Principle of particle swarm optimization (PSO):** Particle Swarm Optimization (PSO) is a very useful tool similar to other evolutionary algorithms. PSO can achieve optimum research in the complex space (Kennedy and Eberhart, 1995) through individual cooperation and competition. It first generates initial clustering, namely, to randomly initialize a group of particles with each particle which is a possible solution for the optimization problems. Objective function determines a fitness value for each particle. Each particle moves in the solution space with its moving direction and distance determined by one certain velocity. Usually, every particle motions followed by the optimal particle evaluated by fitness function to explore the optimal location through generations. During each generation, particles track two of these extreme points, one is the newly-found optimal solution gbest in particle itself, the other being the optimal solution gbest explored in the entire clustering group.

Suppose there exists a exploration $n$-dimensional space in which $m$ particles constitute the clustering group:

$$X = \{x_1, x_2, \ldots, x_m\}$$

where, $x_i = \{x_{i1}, x_{i2}, \ldots, x_{in}\}$ represents the location of particle $i$, $v_i = \{v_{i1}, v_{i2}, \ldots, v_{in}\}$ is the velocity; $p_i = \{p_{i1}, p_{i2}, \ldots, p_{in}\}$ is the individual extremum and $p_g = \{p_{g1}, p_{g2}, \ldots, p_{gn}\}$ is the clustering’s overall extremum. According to principles of current optimal particles, $x_i$ can update its velocity and location by following expression (Lei and Tung, 2005; Kennedy and Eberhart, 1995):

$$v_i^{(t+1)} = v_i^{(t)} + \xi_1(p_i^{(t)} - x_i^{(t)}) + \xi_2(p_g^{(t)} - x_i^{(t)})$$

(10)

$$x_i^{(t+1)} = x_i^{(t)} + v_i^{(t+1)}$$

(11)

**Optimal hierarchical threshold based on PSO:** Specific steps for exploring optimal hierarchical threshold using PSO is as follows:

**Step 1:** Initialization: To preset parameters such as cluster dimensions (wavelet decomposition levels), particle scale (membership), maximum iteration degrees and to randomly initialize the cluster position and velocity

**Step 2:** Calculating initial fitness value of the objective function: Taking the Signal-to-noise Ratio (SNR) of denoised signals as the fitness value function which indicates better denoising quality with higher fitness value of SNR.
This step associates the denoising method using wavelet threshold with PSO where location parameter X is the optimal wavelet threshold value and with denoising the noisy signals using wavelet threshold value of the threshold function. The estimated signals can be obtained and each particle's objective function value can be calculated according to the function.

**Step 3:** Evaluating particle's individual extremum: By comparing the previous objective function value with that of current one it can be calculated that current value can be regarded as the optimal location, labeled as pi, when the current function value is lower, otherwise, pi remains constant.

**Step 4:** Evaluating cluster global extremum: As for each particle, pg will be reset with better result when comparing particle’s fitness value, namely, global optimization pg

**Step 5:** Particle updating: To update each particle's velocity and location in term of Eq. 10 and 11

**Step 6:** Examining whether to conform to the ending conditions: Iteration terminates and the optimal output will be obtained if current iteration degrees reach to the maximum degrees that have been preset (or the results satisfy the precision required), otherwise, steps from 2 to 5 repeat.

Optimal level-dependent threshold value, $\lambda_{\text{opt}}$, can be found by using these steps mentioned above, then $\lambda_{\text{opt}}$ could be applied to the method of removing noisy signals using wavelet threshold. Denoised signals can be obtained by thresholding wavelet decomposition coefficient of the denoised signals using $\lambda_{\text{opt}}$ and one certain threshold function and then by reconstructing wavelet coefficients according to these that have been processed.

**RELATED NOISE REMOVAL**

**Principle of adaptive noise canceling:** Adaptive filter can be used to construct adaptive noise canceller (Cao and Li, 2010), as shown in Fig. 2. Original signal $x(n)$ includes pure signals and noises, $n(n)$ represents the reference noise input. This filter is actually a just a completion of estimate of noise for $x(n)$ and noise removal by estimate value $y(n)$ subtracting original signals.

Adaptive noise canceling is intensely affected by its algorithm. Two types of algorithms can be seen which are minimum mean-square (LMS) algorithm and Recursive Least Squares (RLS) algorithm.

**Adaptive noise canceling algorithm**

**Properties of the algorithms concerned:** Minimum mean-square (LMS): Detuning is proportionate to input vector. LMS filter meets the problem of amplified gradient noise with larger input but it can avoided using normalization processing for normalized least mean square algorithm (NLMS).

Different from LMS, NLMS converges with its convergence condition being non-related to eigenvalue. Convergence is faster and robustness is better with input being the speech signals. Moreover, NLMS has the same amount of calculation with LMS. So it’s a better choice to select NLMS in restricting noise.

Gain $g(n)$ in RLS is not constant and its value changes when there are different input functions. It is because of this improvement that makes RLS superior than LMS and more adaptable to non-stable signals. But it requires a higher demand of hardware due to more calculation.

**BIONIC WAVELET SPEECH ENHANCEMENT BASED ON PSO**

At present, Wavelet Transform (WT) has been a powerful tool for removing noises from speech and images. Wavelet-domain speech enhancement algorithm based on threshold has been widely applied and proves efficient. On the basis of standard threshold denoising, in this study we proposed a hierarchical threshold speech enhancement based on bionic wavelet modification of particle swarm optimization algorithm.

**Algorithm step:** Algorithm steps are given as follows:

**Step 1:** According to the bionic wavelet transform of noisy speech signals, low frequency coefficient $a$ and high frequency, $d_0, d_1, d_2, d_3 \ldots d_{j-1}$ can be obtained. Suppose

$$x(n) = f(n) + n(n), n = 0,1, \ldots, N-1$$

(12)
where, \( x \) is the noisy speech signal; \( f \) represents the clean speech signal; \( n \) is additive noisy signal and \( N \) is the length of signal \( f \). We can get the following expression by applying BWT to noisy speech signals:

\[
\text{BWT}_f, f = \text{BWT}_f(x)
\]

(13)

where, \( \text{BWT}_f \) is the coefficient of noisy speech signal after the bionic wavelet transform.

**Step 2: Thresholding wavelet coefficient:** It is critical to choose the right threshold function and threshold value when thresholding wavelet threshold coefficient. Because speech signal lies mostly in lower frequency subband, we can filter the noise by selecting appropriate threshold value and by thresholding the wavelet coefficient. We obtain the optimal level-dependent threshold by combining threshold function 13 proposed in this study with the application of PSO. To evaluate the effectiveness of this new BWT-based method for speech enhancement, we generally use the level-dependent threshold value, cited in references Bahoura and Rouat (2001) and Jun and Zhang (2001) which varies according the scales where signals and noisy signals present different transmission characteristics:

\[
\lambda(j) = \sigma \sqrt{2 \log(N)} / \log(j+1)
\]

(14)

where, \( \sigma \) represents noise variance, \( N \) is the signal length and \( j \) is the decomposition level. Threshold value reduces with the increase of \( j \) which just accords with the characteristic discrepancy exhibited in different scales of wavelet transform of the noisy signals.

**Step 3:** Obtaining the de-noise speech signals using bionic wavelet inverse transform

We can obtain the denoised speech signals by reconstructing original speech signals with using wavelet coefficient processed, in other words, performing inverse auditory perceived wavelet transform for these processed wavelet coefficients.

**Simulation and analysis:** In this study, we selects function Morlet as the mother wavelet function with its adaption scale set as 22 and window function Hamming, however, \( T_0, G_1 \) and \( G_2 \) are particularly important in the process of speech enhancement of BWT with \( T_0 = 0.000 \), \( 0.26 \), \( G_1 = 0.6 \) and \( G_2 = 72 \). We set cluster number at 25, spatial dimension (decomposition level) at 3, maximum iteration degrees at 100 and the learning factors \( C_1 \) and \( C_2 \) equally at 1.5012.

**Experimentation conditions:** In this experiment, we use speech signal with the sampling frequency at 8 kHz from speech corpus, additive white Gaussian noise \( N(0, \sigma^2_w) \), thus controlling the input signal-to-noise ratio of contaminated signals by updating noise variance \( \sigma^2_w \) which ranges from -10 to 10 dB. Simulation figure is presented below with SNR being at -10 dB, where \( a \) is clean speech signal and \( b \) is noisy speech signal. The filtering of signals is performed by using adaptive filtering based on RLS, BWT, adaptive BWT proposed by Donoho (1995) and the method proposed in this study, thus the results are presented in the following Fig. 3c-f.

Signal-to-noise ratio (SNR) has traditionally been the measurement of the distortion of speech enhancement algorithm but method based on SNR works only when we try to copy waveform coding of the input signal or the enhancement algorithm.

Suppose \( x(n) \) is noisy signal, \( s(n) \) is clean speech signal. \( S(n) \) is enhanced signal. We consider all signals to be energy signal.

The classical equation of SNR can be defined as:

\[
\text{SNR} = 10 \log_{10} \left( \frac{\sum_{n=1}^{N} s^2(n)}{\sum_{n=1}^{N} [s(n) - \hat{s}(n)]^2} \right)
\]

(15)

SNR can be calculated only if \( s(n) \) is known but in reality, this process is impossible to complete, as a result, the application of classical equation of SNR lies mostly in areas in which the algorithm such as clean speech signal and noisy signal is known. We use SNR to measure the quality of speech signal due to its simple mathematic expression.

The classical equation of SNR is hardly related with the subjective attribute of speech quality, thus resulting in a rough SNR and an unsatisfactory estimation. Because the energy of speech signal is time-variant while the energy of noisy is evenly distributed, the fuzzy signal with lesser energy is considerably affected by noise which makes it hard to human to perceive the signals. The shortcomings mentioned above can be overcome by using scale-dependent SNR, the measurement to evaluate how speech waveform in time domain is distorted. It can be defined as follows:
The objective and subjective listening test indicate that the proposed method in this study can obviously improve the enhancement result.

**CONCLUSION**

Speech enhancement algorithm is completed by using the adjustment threshold function and threshold optimization based on PSO, so the methods possess the following characteristics for considering the hierarchical optimization threshold of PSO:

- To ensure the auditory effect of signal frequency and amplitude
- To be applied to colored noise and non-stationary signal environment
- To ensure the high SNR of the enhancement in time domain, no residual noise and algorithm self-adaptation property
- To make speech signal better to be perceived using level-dependent threshold.

The shortcomings of the algorithm proposed also need to guarded, instance, time-consuming calculation, requiring highly effective bionic decomposition, reconstruction and adaptive algorithm with fast...
convergence and stable, low noise variance. These shortcomings affect its application in real-time systems and are the key research topic in the near future.

ACKNOWLEDGMENT

This work was supported by Science and Technology Planning Project of Hunan Province, China 2012SK3131 and 2013SK5069.

REFERENCES