Research on Speech Signal Denoising Algorithm Based on Wavelet Analysis

Linglong Tan¹, Yehui Chen² and Fengzhi Wu³

¹,²Electronic communication engineering college, Anhui Xinhua University, Hefei, 230088, China
³Ashikaga University, Information Science and Manufacturing Engineering, Specialization in Renewable Energy and Environmental Engineering, Tochigi, Ashikaga-shi, 326-8558

*E-mail: tanlinglong@axhu.edu.cn

Abstract. In this paper, wavelet analysis is used to remove noise superimposed on the speech signal to achieve denoising of the speech signal. The prepared speech signal is superimposed with Gaussian white noise with different signal-to-noise ratios, and is removed by using a forced noise removal processing method and a given threshold denoising processing method; The wavelet soft and hard threshold denoising method is used to separately superimpose different Gaussian white noise speech signals, and the hard threshold and soft threshold denoising are respectively performed, and different wavelets are used for processing; Four threshold selection methods are used to denoise the noisy signal using different wavelets. The denoising effects of these methods are simulated in MATLAB software, and combined with the signal-to-noise ratio of the signal after denoising, the denoising effects of various methods are judged. Experimental results show that soft threshold denoising is slightly better than hard threshold denoising. Among the four threshold selection methods, Rigrsure rule denoising is the best. Although other methods are effective, they are not obvious.

1. Introduction
Voice is the medium through which people transmit information when they get along, and it is the most effective and convenient way for people to exchange information. However, in the living environment, the signal will be disturbed by various noises when it is propagated. Because of the presence of noise, it has many impacts on people's lives, but it cannot be completely eliminated, so only the noise can be reduced. The purpose of speech denoising is to remove the noise in the speech signal as much as possible, and filter out the background noise of the interference signal, so that people can get the original speech signal after removing the noise and improve the clarity of the speech [1]. Speech denoising is an effective means to improve speech signals, which is used to reduce noise pollution to speech signals, which can enhance the clarity of the signal, and plays an important role in noise filtering and improving the quality of speech signals. The wavelet denoising method uses the multi-resolution nature of wavelet when removing noise signals, so as to achieve the effect of removing noise. Compared with the traditional denoising method, the wavelet denoising algorithm can complete the denoising of various noisy signals.
2. Definition of wavelet transform

Suppose that \( \Psi \in L^1 \cap L^2 \), and the function value at the intersection is 0. According to the above assumption, the function formula (1) can be listed:

\[
\Psi_{a,b}(t) = |a|^{1/2}\psi\left(\frac{t-b}{a}\right)
\]

In the above function expression, \( \Psi \) is the mother wavelet, \( a \) is the scaling factor, and \( b \) is the translation factor.

From the above function expression, it can be obtained that \( \{\Psi_{a,b}\} \) is a continuous wavelet transform, and on the basis of \( f \in L^2 \), it can be further transformed into an expression of \( W_f(a,b) \):

\[
W_f(a,b) = \int_{\mathbb{R}} f(t)\psi\left(\frac{t-b}{a}\right)dt
\]

The requirements for the establishment of the above function expression are as follows:

\[
e^{-\pi \alpha^2} \int_{-\infty}^{\infty} \left|\Psi(\alpha\omega)\right|^2 d\alpha < \infty
\]

When the above requirements are met, the formula (3) is called an allowable condition.

Since the computer cannot directly perform the wavelet transform, the wavelet that the computer cannot handle is specially processed in the next step through the formula (2) to change, at the same time record the values of two different factors \( a \) and \( b \), so as to achieve the ideal Claim:

\[
a = a_0^m, (a_0 > 1), \quad b = nb_0a_0^m, (b_0 \in \mathbb{R}), (m, n \in \mathbb{Z})
\]

The following expression can be changed by formula (1):

\[
\Psi_{a,b}(t) = a^{-\frac{1}{2}}\psi(a^{-1}\omega t - b)
\]

The expression that can be obtained after wavelet transform is defined as:

\[
DW_{m,n}(f) = \int_{\mathbb{R}} f(t)\Psi_{m,n}(t)dt
\]

3. Wavelet threshold denoising

3.1. Basic principles of wavelet threshold

For signal noise reduction, smoothness and similarity must be observed. The threshold-based direct denoising algorithm proposed by Stanford University professor Donoho is a breakthrough in the field of modern wavelet applications [2]. The formula for eliminating noise signals is as follows:

\[
s(n) = f(n) + \sigma e(n), \quad 1 \leq n \leq N
\]

Among them, \( s(n) \) is the signal without desiccation, \( f(n) \) is the voice signal actually sent, \( e(n) \) is the noise of the surrounding environment, and \( \sigma \) is the noise level.

The basis of threshold denoising is mainly: in order to achieve the effect of noise reduction, the value after the signal processing of the information conversion is still large, but the energy concentration of the noise signal is relatively scattered, and the value is relatively small. Experiments can use this method to exclude values less than the given value. Not only that, the method will also control the wavelet coefficients within a certain range, and finally the signal will be changed to zero or retained and contracted.

The denoising of noise signals with high-frequency white noise can be divided into the following three aspects:

(1) Select a suitable wavelet base first, and then process and decompose the signal-to-noise \( s(n) \) to obtain the corresponding value;

(2) The corresponding similarity value \( \hat{W}_{f,a} \) is obtained through the processed wavelet coefficients;
(3) After the above two steps, after a series of changes and reorganizations, the noise-reduced speech signal can be obtained.

3.2. Common threshold functions

3.2.1. Hard threshold function. If the positive value of the wavelet coefficient is greater than or equal to the threshold, it is still the wavelet coefficient; otherwise, the part of the phase difference is changed to 0, as shown below:

$$w_j = \begin{cases} w_j & |w_j| \geq \lambda \\ 0 & |w_j| < \lambda \end{cases}$$

(8)

3.2.2. Soft threshold function. Assuming that the threshold is less than the positive value of the wavelet coefficient, a series of changes can be made to convert the extra part of the wavelet coefficient to an excess value; the less than part is converted to 0.

$$w_j = \begin{cases} \text{sign}(w_j)(|w_j| - \lambda) & |w_j| \geq \lambda \\ 0 & |w_j| < \lambda \end{cases}$$

(9)

3.3. Wavelet threshold noise reduction method

When the operation analysis of the wavelet coefficients is started, the threshold and function should be targeted for analysis, mainly for some different threshold functions. The usage of the specific threshold function is as follows [3]:

3.3.1. Hard threshold method.

$$W_{j,\lambda} = \begin{cases} W_{j,\lambda} & |W_{j,\lambda}| \geq \lambda \\ 0 & |W_{j,\lambda}| < \lambda \end{cases}$$

(10)

The main operation method of the hard threshold method is to compare the threshold value of the processed signal with the absolute value of the wavelet coefficient. According to the above formula, it can be known that the signal will be interfered from time to time at certain points, and there is a small probability of signal loss. These are some small defects of the hard threshold method. But the hard threshold method can more perfectly show the maximum value of the signal.

3.3.2. Soft threshold method. Compared with the hard threshold method, the soft threshold method has better continuity and stability. Due to the continuity of the soft threshold method, the soft threshold method makes the signal transmission more excellent and smooth. However, the shortcomings of soft threshold make the signal-to-noise results not accurate enough, and there may be deviations.

$$W_{j,\lambda} = \begin{cases} \text{sgn}(W_{j,\lambda})(|W_{j,\lambda}| - \lambda) & |W_{j,\lambda}| \geq \lambda \\ 0 & |W_{j,\lambda}| < \lambda \end{cases}$$

(11)

3.4. Wavelet threshold selection rules

Whether the denoising performance is good depends mainly on the threshold Th. If the threshold is too small, it will cause incomplete denoising, and a small amount of unremoved noise may still remain in the signal. If the threshold is greater than the true value, the accuracy of the signal may be affected to a certain extent, and information loss may occur. Therefore, the selection of the threshold value requires special attention, too large or too small will cause errors. In order to avoid the above errors, we can use the following methods [4,5]:

3.4.1. Sqtwolog rules. On the basis of the orthogonal wavelet variation, the expression of the threshold can be verified as

$$Th = a \sqrt{\text{sig}(n)}$$

The size of n in the above formula is the size of the signal during transmission. The formula for obtaining the noise variance is as follows:
\[ \hat{x} = \frac{M}{0.6745} \]  
(12)

\[ \sigma = \frac{M}{0.6745} \]  
(12)

\[ M \] is a piece of data that is relatively close to the true threshold value, and the noise signal is taken as the condition to be considered, and the signal length is \( n \), then the formula for the threshold value that can be calculated is:

\[ Th = \sigma \sqrt{2 \log(n)} \]  
(13)

However, the error of the fixed threshold value is still too large. In order to continue to reduce this error, the threshold value method of unbiased risk threshold rules can be used.

### 3.4.2. Rigrsure rules

This is a method of threshold selection with the smallest error. You can take out the two limits of each value range, and then bring each limit into the formula to calculate the minimum error value. This is the required threshold. The formula and calculation method are as follows:

In order to find the threshold, you need to create a vector array composed of the smallest to the largest, and add a vector \( r \) to represent the error. The formula is as follows:

\[ r_i = \left[ n-2i-(n-i) \omega_k + \sum_{k=1}^{i} \omega_k \right] / n, \quad i = 1, 2, \ldots, n \]  
(14)

Where \( i \) is the basic variable, and then find the minimum value in the range, and use this minimum value as the error value, and then find the \( \omega_{\text{min}} \) that is the minimum value, that is, the threshold is:

\[ Th = \sigma \sqrt{\omega_{\text{min}}} \]  
(15)

This and the SqtwoLOG rule above have their own characteristics. The Rigrsure rule has a great benefit in removing noise from a large range of noise, but in a small range of signal noise, a fixed threshold is a good choice.

### 3.4.3. Heursure rules

The Heursure rule is the optimization of the above two thresholding methods, and it is also the best choice for thresholding. The specific calculation formula is as follows:

Then use \( S \) as the sum of the squares of the signals, from which \( \eta = (S - \sigma^2) / n, \quad \mu = (\log_e n)^{3/2} / \sqrt{n} \) can be obtained. It can be obtained by combining the above two formulas:

\[ Th = \min(Th, Th_{0}), \quad \eta \leq \mu \]  
\[ Th = \sigma (0.3936 + 0.1829 \log_{x} x), \quad x \geq 32 \]  
\[ Th = 0, \quad x < 32 \]  
(16)

### 3.4.4. Minimaxi rules

This is a method to select a certain threshold, and the processing method with the smallest deviation can be realized in the Minimaxi rule. Since the experiment can treat the noise signal as an approximate value, the formula for selecting the threshold is as follows:

\[ Th = \sigma (0.3936 + 0.1829 \log_{x} x), \quad x \geq 32 \]  
\[ Th = 0, \quad x < 32 \]  
(17)

In the formula, \( x \) is the number of samples; \( \sigma \) is the variance of signal-to-noise.

### 4. Evaluation method

There are many factors that affect the effect of wavelet denoising. For example, the selection threshold is different, the wavelet basis function used is different, and the decomposition scale is different. Due to the influence of these series of factors, the effect of wavelet denoising will be changed. Therefore, this article needs to use some specific indicators to measure this. Mean square error (RMSE) and signal-to-noise ratio (SNR) are used as commonly used evaluation indicators.

The meaning of the mean square error refers to the square root of the variance between the original signal and the denoised estimated signal. Its definition is:

\[ \text{RMSE} = \left( \left[ f(n) - \hat{f}(n) \right]^2 / n \right)^{1/2} \]  
(18)
In this formula, the original signal is \( f(n) \), and the denoised signal is \( \hat{f}(n) \). The signal-to-noise ratio is a traditional method of measuring the noise measurement in a signal, and its definition is:

\[
SNR = 10 \log_{10} \left( \frac{p_s}{p_n} \right)
\]  

(19)

In the formula, the original signal power is \( p_s = \frac{1}{N} \sum_{n=1}^{N} f(n)^2 \) and the signal noise power is \( p_n = \text{RMSE}^2 \). When the signal-to-noise ratio is larger, the denoising effect will be better.

5. Wavelet threshold denoising simulation

Select a voice signal in WAV audio format as the original voice signal, as shown in Figure 1. Pre-process the original voice signal, remove the DC component, and normalize the amplitude. After processing, as shown in Figure 2, add Gaussian white noise to the processed voice signal.

Three methods are often used to denoise a noisy signal, forced denoising, given threshold denoising and general threshold denoising. This article discusses the use of forced denoising and given threshold denoising, using db8 wavelet the speech signals added with 0db and 5db Gaussian white noise are processed respectively. The signal after adding 0db Gaussian white noise is shown in Figure 3, and the signal after adding 5db Gaussian white noise is shown in Figure 4.

The signal-to-noise ratio of the 0db Gaussian white noise signal is -0.0083; the signal-to-noise ratio of the forced denoising and reconstructing signal is 11.7253; the signal-to-noise ratio of the denoising and reconstructing signal with a given threshold is 12.31171. The signal-to-noise ratio of the 0db Gaussian white noise signal is 5.051; the signal-to-noise ratio of the forced denoising and reconstructing signal is 13.3184; the signal-to-noise ratio of the denoising and reconstructing signal with a given threshold is 14.0669.
Through the forced denoising process and the given threshold denoising process respectively, the two noise-added signals are processed. The signal-to-noise ratio of the two methods can be compared. It is not difficult to find that as long as a suitable threshold can be found, use the given threshold denoising process. The signal-to-noise ratio is larger, which is better than the forced denoising process. Using sym4 wavelet and db8 wavelet, 7-layer decomposition is performed on the signal added with 5db Gaussian white noise, and hard threshold denoising and soft threshold denoising are applied to them respectively. Figures 5 and Figure 6 show that two wavelets perform noise on the noise signal. After denoising the signal with hard threshold and soft threshold.

The signal-to-noise ratio of the noise-added signal is 5.055; the signal-to-noise ratio after the hard threshold denoising is 6.1699; the signal-to-noise ratio after the soft threshold denoising is 6.1731. The signal-to-noise ratio of the noise-added signal is 4.9951; the signal-to-noise ratio after the hard threshold denoising is 6.1485; the signal-to-noise ratio after the soft threshold denoising is 6.1517. Compared with the results of soft and hard threshold denoising, the results of soft threshold denoising are better, but it is not particularly obvious. Next, using these four methods, db8 wavelet and sym4 wavelet are used to denoise the speech signal with 5db Gaussian noise respectively. The signal-to-noise ratio after denoising is 4.9983; the signal-to-noise ratio after the hard threshold denoising is 7.9343; the signal-to-noise ratio after Rigsure rule denoising is 7.5878; the signal-to-noise ratio after Sqtwolog rule denoising is 6.9705. The signal-to-noise ratio after Heursure rule denoising is 7.1212. The signal-to-noise ratio of the signal after adding noise is 5.0292. The signal-to-noise ratio after denoising using Heursure rule is 7.1834. The signal-to-noise ratio after denoising by Rigsure rule is 9.6179. After denoising with Sqtwolog rule, the signal-to-noise ratio is 6.1073. The signal-to-noise ratio after denoising with Minimaxi rule is 6.2925. Among the four threshold selection methods, the effect of Rigsure rule after denoising is the most obvious, followed by Heursure rule, however, the signal noise after using Sqtwolog rule denoising and Minimaxi rule denoising is not obvious compared with the first two methods.

Figure 5. Signal after sym4 wavelet denoising

Figure 6. db8 wavelet denoising signal

6. Summary
From the experimental results, we can see that compared with the forced denoising process and the given threshold denoising process, as long as a suitable threshold is found, the effect of the given threshold denoising process will be more satisfactory. Compared with soft and hard threshold denoising, using hard threshold processing can retain more peak features. The signal after soft threshold processing has better smoothing characteristics. Soft threshold processing is slightly better than hard threshold processing. In wavelet denoising, selecting different wavelets for denoising signals to denoise will also have different results. Through the above experiments, it is proved that the denoising method of speech signal based on wavelet analysis has a significant effect in removing the noise in the speech signal, and it is an effective wavelet denoising algorithm.
Acknowledgments
This work was supported by the Scientific Research Project of Anhui Natural Science Research Project No. KJ2017A628, the Quality Engineering Project No.2015zy073, the backbone Teacher Project No.2018xgg04 and Student Innovation Project No. 201812216022.

References
[1] Peng Lei, Meiling Chen, Jun Wang. Speech enhancement for in-vehicle voice control systems using wavelet analysis and blind source separation[J]. IET Intelligent Transport Systems, 2019, 13(4).
[2] Stepanov R, Bondar’ N I, Katsova M M, Sokoloff D, Frick P. Wavelet analysis of the long-term activity of V833 Tau[J]. Monthly Notices of the Royal Astronomical Society, 2020, 495(4).
[3] Sandar Oo. Study on Speech Compression and Decompression by using Discrete Wavelet Transform[J]. Journal of Trend in Scientific Research and Development, 2019, 3(3).
[4] Erol Seke, Kemal Özkan. A new speech signal denoising algorithm using common vector approach[J]. International Journal of Speech Technology, 2018, 21(3).
[5] Francesco Saltari, Daniele Dessi, Franco Mastroddi. Mechanical systems virtual sensing by proportional observer and multi-resolution analysis[J]. Mechanical Systems and Signal Processing, 2021, 146.