LETTER

Speech Enhancement Algorithm Using Recursive Wavelet Shrinkage

Gihyoun LEE†, **Sung Dae NA**†, **KiWoong SEONG**††, **Jin-Ho CHO**†††, *Nonmembers*, *and* **Myoung Nam KIM**††††a), *Member*

SUMMARY Because wavelet transforms have the characteristic of decomposing signals that are similar to the human acoustic system, speech enhancement algorithms that are based on wavelet shrinkage are widely used. In this paper, we propose a new speech enhancement algorithm of hearing aids based on wavelet shrinkage. The algorithm has multi-band threshold value and a new wavelet shrinkage function for recursive noise reduction. We performed experiments using various types of authorized speech and noise signals, and our results show that the proposed algorithm achieves significantly better performances compared with other recently proposed speech enhancement algorithms using wavelet shrinkage.
key words: *speech enhancement, wavelet shrinkage, recursive algorithm, hearing aids*

1. Introduction

Speech enhancement is required in many speech signal processing applications. Generally, the algorithms can be classified into two major categories, single-channel source and multi-channel source. Multi-channel algorithms have good performance but still many devices use only a single microphone. Moreover, the speech enhancement algorithm of most hearing aids must have simple structure. Therefore, there is a need for more studies on speech enhancement algorithms of hearing aids. This is a particularly challenging issue because of the widespread use of sources that are a combination of noise and speech. The algorithms based on wavelet transforms have been proposed for speech enhancement and de-noising of signals. Donoho [1] proposed wavelet shrinkage as a powerful tool in de-noising signals. Because wavelet transforms have the characteristic of low entropy, multi-resolution, correlation, and flexible selection of basis functions, wavelet shrinkage based on wavelet transforms have achieved successful application in the field of de-noising and speech enhancement. More recently, various improved algorithms that used wavelet shrinkage threshold have been proposed by Zhu [2], Xue [3], and Zhang [4]. However, all of these algorithms have problems hindering the application of the algorithm to speech signals. The problems involve the inability to maintain signal continuity and the signal loss of speech information.

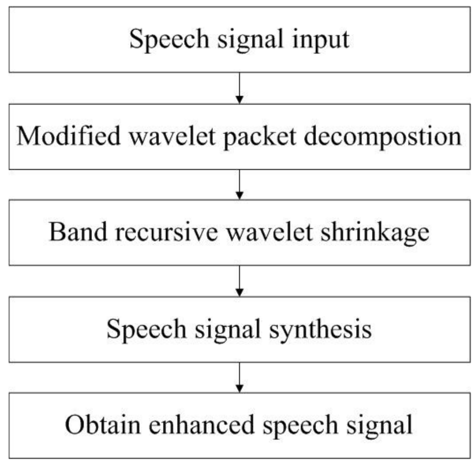

Fig. 1 Block diagram of the proposed algorithm.

To overcome these problems in this paper, the proposed algorithm has a new threshold function that needs a threshold matrix for each band and a new thresholding function using recursive noise estimation. The proposed threshold function based on noise variance maintains the signal continuity and has good speech enhancement performance. In order to express the overall concept of the proposed algorithm, a block diagram of the proposed algorithm is shown in Fig. 1.

The noisy speech signal is decomposed by Modified Wavelet Packet Decomposition (MWPD). Then, speech enhancement and noise reduction processes are performed using the proposed band recursive wavelet shrinkage (BRWS). After signal synthesis process, a enhanced speech signal is obtained. On the other hand, the proposed algorithm should be possible to real-time processing for use in hearing aids. The wavelet transform of MWPD had a lot of studies about real-time processing [5], [6]. BRWS only use logarithm operator to calculate threshold values, and the other calculating processes are consisted by addition and multiplication operator. Consequently, BRWS was performed with $5ms$ per a

Manuscript received December 8, 2015.
Manuscript revised March 7, 2016.
Manuscript publicized March 30, 2016.
†The authors are with the Department of Medical and Biological Engineering, Graduate School, Kyungpook National University, 680 Gukchaebosang-ro, Jung-gu, Daegu 41944, Korea.
††The author is with the Department of Biomedical Engineering, Kyungpook National University Hospital, 130 Dongdeok-ro, Jung-gu, Daegu, 41944, Korea.
†††The author is with the School of Electronics Engineering, College of IT Engineering, Kyungpook National University, 80 Daehakro, Bukgu, Daegu, 41566, Korea.
††††The author is with the Department of Biomedical Engineering, School of Medicine, Kyungpook National University, 680 Gukchaebosang-ro, Jung-gu, Daegu, 41944, Korea.
a) E-mail: kimmn@knu.ac.kr
DOI: 10.1587/transinf.2015EDL8251

frame (25ms) in Windows 7, Intel Core i5 750, and Matlab R2012b environment. And the process of speech signal synthesis has only addition operator. It is shown that real-time processing of the proposed algorithm is possible. Furthermore, the proposed algorithm can be used in hearing aids and embedded systems through system optimization.

2. Theory and Method

2.1 Modified Wavelet Packet Decomposition (MWPD)

While speech enhancement algorithms that use the wavelet transform have the ability to decompose signals that are similar to the human acoustic system, they also enable conciseness and practicality [6]. The structure of the critical bands in MWPD, which was modified from wavelet packet decomposition, is optimized to classify speech bands and distributing noise bands. For a given level j, the wavelet packet (WP) transform decomposes the speech signal $x(n)$ into 2^j sub-bands corresponding to wavelet coefficient sets $w_{j,m}(k)$.

$$w_{j,m}(k) = WP[x(n), j] \tag{1}$$

The speech signal is decomposed to 20 sub-bands of the wavelet coefficient $w_{j,m}(k)$ using MWPD. $w_{j,m}(k)$ is the jth level, kth wavelet coefficient of the mth sub-band in MWPD, where $j = 3, 4, 5, m = 1, \ldots, 20$, and $k = 1, \ldots, N/2$. $w_{j,m}(k)$ can be modified in the time and critical band. The modified $w_{j,m}(k)$ can be also expressed in matrix form by Eq. (2).

$$\Psi_m(t) = \begin{pmatrix} \varphi_1(1) & \varphi_1(2) & \ldots & \varphi_1(t) \\ \varphi_2(1) & \varphi_2(2) & \ldots & \varphi_2(t) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_{20}(1) & \varphi_{20}(2) & \ldots & \varphi_{20}(t) \end{pmatrix} \tag{2}$$

where $\Psi_m(t)$ is the signal composed of the mth sub-band at specific time t. We used signal entropy to extract the envelope of the signal.

Recent speech enhancement algorithms use wavelet shrinkage threshold. Wavelet shrinkage is the simplest, and a variety of threshold functions that apply wavelet coefficients can be employed for speech enhancement [7], [8]. Recently, several wavelet shrinkage algorithms have been proposed using a hard threshold, soft threshold, and semi-soft threshold. With the hard threshold function, it is easy to generate the Pseudo-Gibbs phenomenon when reconstructing the signal [2]. The soft threshold function has better continuity, but it makes the variance of the de-noised signal become too great. To compensate for these disadvantages, Zhu's algorithm [2], Xue's algorithm [3], and Zhang's algorithm [4] are proposed. These algorithms are as follows:

Zhu's algorithm [2]:

$$\lambda = \sigma\sqrt{2\log N} \tag{3}$$

$$\Pi^{f1} = |\Psi_m(t)| - \frac{a\lambda}{a + \exp(|\Psi_m(t)| - \lambda} \tag{4}$$

$$\hat{\Psi}_m^{f1}(t) = \begin{cases} \text{sign}(|\Psi_m(t)|)\,\Pi^{f1} & \text{if } |\Psi_m(t)| \geq \lambda \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

Xue's algorithm [3]:

$$\Pi^{f2} = |\Psi_m(t)| - \frac{\beta\lambda}{\beta + |\Psi_m(t)| - \lambda} \tag{6}$$

$$\hat{\Psi}_m^{f2}(t) = \begin{cases} \text{sign}(|\Psi_m(t)|)\,\Pi^{f2} & \text{if } |\Psi_m(t)| \geq \lambda \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

Zhang's algorithm [4]:

$$\Pi^{f3} = |\Psi_m(t)| - \frac{\lambda}{\exp\frac{|\Psi_m(t)| - \lambda}{A}} \tag{8}$$

$$\hat{\Psi}_m^{f3}(t) = \begin{cases} \text{sign}(|\Psi_m(t)|)\,\Pi^{f3} & \text{if } |\Psi_m(t)| \geq \lambda \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

where a is positive ($a = 0.125$), β is a positive regulatory factor ($\beta = 6$), σ is standard deviation of the window, N is number of samples in the window and A is an arbitrary positive constant. a and β were set for a best performance value from results of Zhu's paper [2]. Zhu's algorithm solves the problem of constant deviation between the estimated wavelet coefficients and noise signal. However, it has poor continuity [2]. Although Xue and Zhang's algorithms have good continuity, they exhibit poor performance with complex signals such as contaminated speech signals, and they remove too much information from the speech parts. Therefore, we propose a multi-band threshold function in order to maintain signal continuity. In Sect. 3, the performance of the proposed algorithm will be presented compared to other recent wavelet shrinkage methods.

2.2 Proposed Band Recursive Wavelet Shrinkage (BRWS)

In this section, a new wavelet shrinkage function is presented for speech enhancement. In order to use features of noise and speech wavelet band, BRWS applied wavelet band recursive shrinkage function and threshold matrix (Λ_m) for each wavelet band that is based on the wavelet coefficients from each critical band. The threshold matrix (Λ_m) is as follows:

$$\Upsilon_m = \sqrt{\frac{\sum_{i=t}^{t+N}(\Psi_m(i) - M_m)^2}{N}} = [\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{20}] \tag{10}$$

$$\Lambda_m = \sqrt{2\log N}\,\Upsilon_m = [\lambda_1^{prop}, \lambda_2^{prop}, \ldots, \lambda_{20}^{prop}] \tag{11}$$

Υ_m is standard deviation of mth wavelet band and M_m is mth mean value of wavelet coefficients. A new threshold metrix (Λ_m) of each critical band is calculated using Υ_m. The process of speech enhancement is as follows:

$$\Pi_m^{prop} = |\Psi_m(t)| - \sum_{i=1}^{D} a_i(\Psi_m(t-1)) \tag{12}$$

$$\hat{\Psi}_m^{prop}(t) = \begin{cases} \text{sign}(|\Psi_m(t)|)\,\Pi_m^{prop} & \text{if } |\Psi_m(t)| \geq \lambda_m^{prop} \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

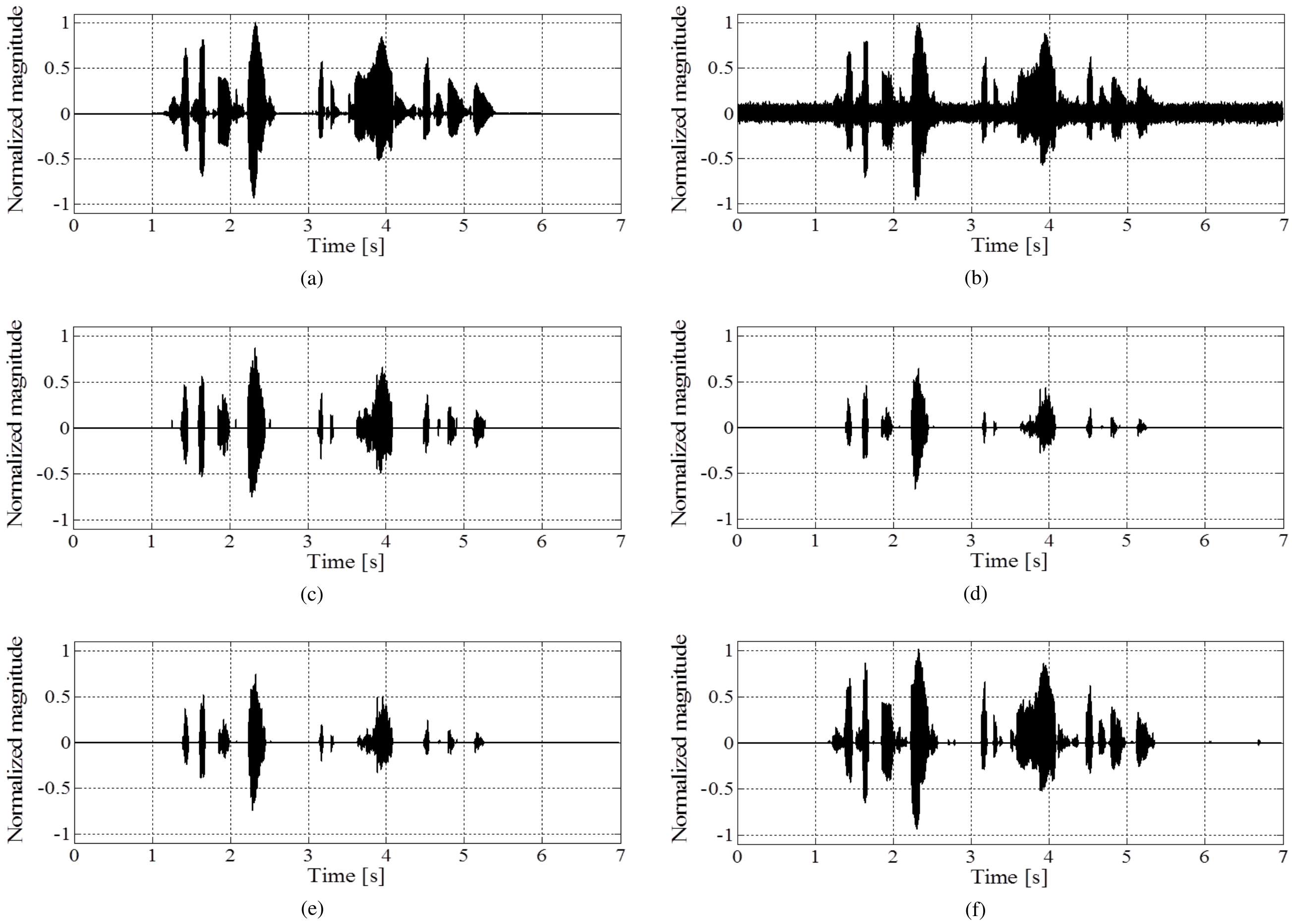

Fig. 2 Result of the noise reduction for (a) a clean speech signal and (b) a contaminated speech signal, (c) the speech enhancement result obtained using Zhu's algorithm, (d) the speech enhancement result obtained using Xue's algorithm, (e) the speech enhancement result obtained using Zhang's algorithm, and (f) the enhanced speech.

where Π^{prop} is recursive noise estimation factor, D is recursive order, which is set 32, and a_i is recursive coefficients, which are updated to have the least error value using past wavelet coefficients of noise bands. The recursive noise estimation factor (Π^{prop}) has function that remove noises and keep speech information in speech bands. The bands of noise are determined by the threshold matrix (Λ_m). When the input wavelet coefficients have a lower value than the threshold, BRWS determines the information to be meaningless and removes it. If the input wavelet coefficients have features that are much closer to speech, BRWS keeps the speech information. Moreover, because of recursive noise estimation, BRWS has a good noise reduction performance despite maintaining the signal information.

3. Experiment and Results

To test the performance of the proposed algorithm, a speech signal sample from the TIMIT database [9] and a noise signal sample from NOISEX-92 [10] are used. The data samples have a sampling rate of 16 kHz and a bit rate of 32 bps. We also experimented using a variety of noises (white, car, and babble) and SNR environments (0 dB, 5 dB, 10 dB, 15 dB, and 20 dB) to evaluate the performance of the proposed algorithm. The graphical results of speech enhancement and noise reduction are shown in Fig. 2.

Figure 2 (a) is a clean speech signal, and (b) is a contaminated speech signal with white noise SNR of 5dB, Fig. 2 (c), (d), and (e) are enhanced signals respectively obtained using reference algorithms, and Fig. 2 (f) is the enhanced signal obtained using the proposed algorithm. All results show good performance of noise reduction. However, because of too much noise reduction, Fig. 2 (d) and (e) show attenuation of the speech signal. (c) also shows attenuation of the speech signal at weak signal magnitude. While Fig. 2 (c) to (e) show that much of the speech signal is lost, Fig. 2 (f) shows a speech signal that is very natural and clean. (f) shows clean signal at noise parts and maintain information of speech parts at not only strong speech signals but also very weak speech signals. The sample speech data, which comprised male and female speech signals with a variety of accents, included more than 50 random samples with lengths of 3–5 seconds. An objective measure was utilized for evaluation purposes, namely, the percep-

Table 1 Speech quality obtained using PESQ.

Environments		PESQ			
Noise	SNR(dB)	Zhu [2]	Xue [3]	Zhang [4]	Proposed
White	0	−0.19	−0.15	−0.15	1.25
	5	0.21	0.20	0.20	1.75
	10	0.92	0.83	0.83	2.26
	15	1.57	1.42	1.42	2.80
	20	2.14	1.96	1.96	3.26
Babble	0	−0.21	−0.16	−0.16	1.34
	5	0.33	0.32	0.33	1.87
	10	1.02	0.96	0.96	2.37
	15	1.67	1.52	1.52	2.79
	20	2.24	2.05	2.05	3.21
Car	0	−0.21	−0.16	−0.17	2.69
	5	0.09	0.10	0.09	3.01
	10	0.78	0.71	0.70	3.22
	15	1.45	1.32	1.32	3.56
	20	2.05	1.87	1.87	3.87

tual evaluation of speech quality (PESQ), which is a widely used measure when attempting to obtain an objective evaluation of speech quality [11]. The PESQ measure is the most complex to compute, and it is the one recommended by the ITU-T P.862 for the speech quality assessment of narrowband speech codecs [12]. Therefore, PESQ is utilized for objective evaluation, and a higher value of PESQ indicates better speech quality. Detailed values of the speech quality obtained from PESQ are given in Table 1. The table gives detailed values for a variety of SNRs (0, 5, 10, 15, and 20 dB) and three noise environments. The proposed algorithm has higher values of PESQ than the other algorithms.

This signifies that the proposed algorithm has a better speech enhancement performance. Zhu's algorithm has a good performance in low noisy environments (SNR of 10, 15, and 20 dB). However, it has a poorer performance than Xue and Zhang's algorithms in very noisy environments (SNR of 0 and 5 dB). Xue and Zhang's algorithms have similar performances in almost all environments. The proposed algorithm not only has very good performance in low noisy environments but also has comparatively good performance. Moreover, the proposed algorithm exhibits a significant increase in the car noise environment. This is because of the characteristics of car noise, which has a frequency that is different from that of the human voice. The proposed algorithm is based on the band recursive noise reduction algorithm, so it has a better performance than the other algorithms and the other environments. Table 1 shows that the proposed algorithm has the best performance when compared with the algorithms in the all noise environments.

4. Conclusion

In this paper, we proposed a new single-channel speech enhancement algorithm of hearing aids. It has a very simple construction based on wavelet shrinkage. The proposed algorithm shows good performance in a variety of noisy environments. The performance of the speech enhancement was confirmed by experiments using many signal samples and in a variety of noisy environments. Currently, we are extending our research to enable us to successfully realize a usable system.

Acknowledgments

This work was supported by a grant from the National Research Foundation of Korea (NRF), which was funded by the Korean government (MSIP) (No. 2013R1A2A1A09015677 and 2015R1A2A2A03006113).

References

[1] D.L. Donoho, "De-noising by soft thresholding," IEEE Trans. Information Theory, vol.41, no.3, pp.613–627, 1995.

[2] J.-F. Zhu and Y. Huang, "Improved threshold function of wavelet domain signal de-noising," Proc. ICWAPR, pp.190–195, 2013.

[3] W. Xue, F.H. Guan, and L.W. Chen, "Radar signal de-noising based on a new wavelet thresholding function," Computer Simulation, vol.25, no.8, pp.319–322, 2008.

[4] W.Q. Zhang and G.X. Song, "Signal de-noising in wavelet domain based on a new kind of thresholding function," Journal of Xidian University, vol.31, no.2, pp.296–299, 2004.

[5] H.D.O. Mota, F.H. Vasconcelos, and R.M.D. Silva, "A real-time system for denoising of signals in continuous streams through the wavelet transform," In Signals, Circuits and Systems international Symposium on IEEE, vol.2, pp.429–432, 2005.

[6] H.D.O. Mota, F.H. Vasconcelos, and R.M.D. Silva, "Real-time wavelet transform algorithms for the processing of continuous streams of data," In Intelligent Signal Processing, IEEE International Workshop on. IEEE, pp.346–351, 2005.

[7] G.H. Lee, S.D. Na, J.H. Cho, and M.N. Kim, "Voice activity detection algorithm using perceptual wavelet entropy neighbor slope," Bio-medical Materials and Engineering, vol.24, no.6, pp.3295–3301, 2014.

[8] H.Y. Gao and A.G. Bruce, "Waveshrinkage with semisoft shrinkage," StatSci divition of mathsoft Inc., 1995.

[9] V. Zue, S. Seneff, and J. Glass, "Speech database development at MIT: TIMIT and beyond," Speech Communication, vol.9, no.4, pp.351–356, 1990.

[10] A. Varga and H.J.M. Steeneken, "Assessment for automatic speech recognition: II. NOISEX-92: A database and an exper-iment to Study the Effect of Additive Noise on Speech Recognition Systems," Speech Communication, vol.12, no.3, pp.247–251, 1993.

[11] A. Rix, J. Beerends, M. Hollier, and A. Hekstra, "Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs," ITU-T Recommendation, pp.862, 2001.

[12] M.A. Rashed, T.A. El-Garf, I.F. Tarrad, and H.A. Almotaafy, "The effect of weight factor on the performance of G.729A speech coder," IJESET, vol.6, no.1, pp.1–8, 2013.