BIOENGINEERED
2016, VOL. 7, NO. 5, 352–356
http://dx.doi.org/10.1080/21655979.2016.1197617

RESEARCH PAPER

# Wavelet speech enhancement algorithm using exponential semi-soft mask filtering

Gihyoun Lee[a], Sung Dae Na[a], KiWoong Seong[b], Jin-Ho Cho[c], and Myoung Nam Kim[d]

[a]Department of Medical & Biological Engineering, Graduate School, Kyungpook National University, Daegu, Korea; [b]Department of Biomedical Engineering, Kyungpook National University Hospital, Daegu, Korea; [c]School of Electronics Engineering, College of IT Engineering, Kyungpook National University, Daegu, Korea; [d]Department of Biomedical Engineering, School of Medicine, Kyungpook National University, Daegu, Korea

**ABSTRACT**

In this paper, we propose a new speech enhancement algorithm based on wavelet packet decomposition and mask filtering. In the traditional mask filtering such as ideal binary mask (IBM), the basic idea is to classify speech components as target signal and non-speech components as background noises. However, speech and non-speech components cannot be well separated in target signal and background noise. Therefore, the IBM has residual noise and signal loss. To overcome this problem, the proposed algorithm used semi-soft mask filter to exponentially increase. The semi-soft mask minimizes signal loss and the exponential filter removes residual noise. We performed experiments using various types of speech and noise signals, and experimental results show that the proposed algorithm achieves better performances than the traditional other speech enhancement algorithms.

**ARTICLE HISTORY**
Received 21 January 2016
Accepted 15 April 2016

**KEYWORDS**
binary mask filtering; semi-soft filtering; speech enhancement; wavelet shrinkage; wavelet transform

## Introduction

As the recent development of the communication device, speech enhancement is required in many speech signal processing applications. Traditionally, speech enhancement algorithms are based on linear processing techniques such as Wiener filtering, spectral subtraction, signal subspace approach, and linear prediction. Although these techniques can be used to suppress the background noises, they distort the speech signal and also tend to introduce a perceptually annoying residual noise, often referred to as musical noise.[1] Recently, Wavelet transform have been applied to a lot of research such as signal and image de-noising, compression, detection, and pattern recognition. Donoho[2] proposed wavelet shrinkage, which processes by thresholding wavelet coefficients, as a powerful tool in de-noising signals. Unfortunately, it is not always possible to separate the components corresponding to the target signal from those of noise by a simple thresholding.[3,4] More recently, improved algorithms that used wavelet shrinkage threshold have been proposed by Zhu.[5] However, the algorithm has problems hindering the application of the algorithm to speech signals. It is the inability to maintain signal continuity and the signal loss of speech information.

On the other hands, ideal binary mask (IBM) method[6] is proposed for its efficiency to increase speech intelligibility. It is established as a goal of binary time-frequency (T-F) masking approach. A noisy speech signal is gridded and respectively analyzed in each T-F decomposed unit.[7] A target signal is estimated and separated from the residual called as interference noise. However, such voiced and unvoiced components still cannot be well separated in target signal and interference noise.[7-9]

In this paper, the proposed algorithm has a new mask filtering with matrix of mask feature for each wavelet bands and the exponential semi-soft mask. The proposed speech enhancement algorithm maintains the signal continuity and has good performance of speech enhancement.

## Theory and method

### *Modified wavelet packet decomposition for speech signals*

Wavelet packet decomposition is widely used in speech signal processing because of very simple and powerful. And it is possible to resolve high

**CONTACT** Myoung Nam Kim kimmn@knu.ac.kr Department of Biomedical Engineering, School of Medicine, Kyungpook National University, 680, Gukchaebosang-ro, Jung-gu, Daegu 41944, Korea.

frequency components with in a small time window of a speech signal. A noisy speech signal $y(n)$ can be described as:

$$y(n) = s(n) + v(n) \tag{1}$$

where $s(n)$ is clean speech and $v(n)$ is background noise in $n$th frame. Generally, wavelet packet decomposition decomposes the noisy signal using wavelet packet transform into time-frequency wavelet coefficients of multiple sub-bands. The decomposition of multiple sub-bands is designed to mimic the critical bands as widely used in perceptual auditory modeling.[10]

In this paper, wavelet packet decomposition was modified to enhance speech bands based on Daubechies6 wavelet basis. The structure of the critical bands in modified wavelet packet decomposition (MWPD) is optimized to departmentalize speech bands and it evenly distributes energy of noise bands. The speech signal is decomposed to 20 sub-bands of the wavelet coefficient $w_{j,m}(k)$ using MWPD. In other words, $w_{j,m}(k)$ is the $j$th level, $k$th wavelet coefficient of the $m$th sub-band in MWPD, where $j$ = 3, 4, 5, $m$ = 1, … , 20, and $k$ = 1, … , $N$/2$j$. The structure and frequency band of MWPD is shown Fig. 1. Fig. 1 shows that the structure of MWPD more finely decompose a signal in main speech frequency (2 ~ 3.5 kHz) than other frequency bands. And then, $w_{j,m}(k)$ can be modified to the time ($t$) and critical band ($m$) axis. The modified $w_{j,m}(k)$ expressed in matrix form by (Eq. 2).

$$\Psi_m(t) = \begin{bmatrix} \psi_1(t) \\ \psi_2(t) \\ \vdots \\ \psi_{20}(t) \end{bmatrix} = \begin{bmatrix} \psi_1(1) & \psi_1(2) & \dots & \psi_1(t) \\ \psi_2(1) & \psi_2(2) & & \vdots \\ \vdots & \dots & \ddots & \vdots \\ \psi_{20}(1) & \psi_{20}(2) & \dots & \psi_{20}(t) \end{bmatrix} \tag{2}$$

where $\Psi_m(t)$ is the signal composed of the $m$th sub-band at specific time ($t$), and $\Psi_m(t)$ is a matrix of the wavelet coefficient information in time and wavelet bands.

## Wavelet shrinkage and ideal binary mask

Wavelet shrinkage algorithm is very simple and powerful tool for de-noising from noisy signal. It can separate white noise from noisy signal using universal threshold ($\lambda$).

$$\lambda = \sigma\sqrt{2\log(N\log_2 N)} \tag{3}$$

$$\sigma = MAD/0.6745 \tag{4}$$

$$\hat{\Psi}_m^O(t) = \begin{cases} \mathrm{sign}(\Psi_m(t))(|\Psi_m(t)| - \lambda), & |\Psi_m(t)| \geq \lambda \\ 0 \qquad\qquad\qquad\qquad\qquad , & |\Psi_m(t)| < \lambda \end{cases} \tag{5}$$

where MAD is the absolute median estimation of wavelet coefficients. Although the wavelet shrinkage is satisfactory for removing white Gaussian noise, most wavelet shrinkage has hard or soft threshold

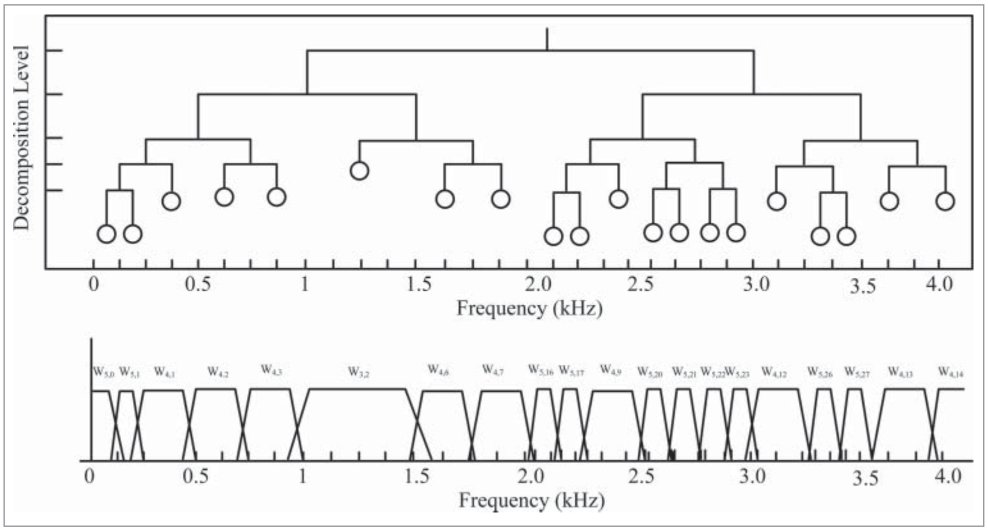


Figure 1. Frequency band and structure of MWPD. (A) Clean speech (B) Noisy speech (C) IBM algorithm (D) Proposed algorithm.

function. The hard threshold function is easy to generate the Pseudo-Gibbs phenomenon when reconstructing the signal.[5] The soft threshold function can also make a variance of the de-noised signal too greatly. Recently, to compensate for these disadvantages, a new wavelet shrinkage algorithm is proposed by Zhu.[5]

Zhu's wavelet shrinkage algorithm:[5]

$$\hat{\Psi}_m^Z(t) = \begin{cases} \mathrm{sign}(|\Psi_m(t)|)\left(|\Psi_m(t)| - \dfrac{a\lambda}{a+\exp(|\Psi_m(t)|-\lambda)}\right), & |\Psi_m(t)| \geq \lambda \\ 0 & , \ |\Psi_m(t)| < \lambda \end{cases} \tag{6}$$

where, $a$ is positive. Zhu's algorithm solved the problem of constant deviation between the estimated wavelet coefficients and noise signal.

In the ideal binary mask (IBM) method, $s(n)$ and $v(n)$ in (Eq. 1) are decomposed by shot-time Fourier transform. Given $Y_n$, $S_n$, and $V_n$ as the corresponding time-frequency representations of $y(n)$, $s(n)$, and $v(n)$, the IBM matrix $M$ for the target signal $S$ is defined as:[7]

$$M_{c,n} = \begin{cases} 1, & \text{if} |S_{c,n}| \geq |V_{c,n}| \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

where $S_{c,n}$ and $V_{c,n}$ are the spectral values of $S_n$ and $V_n$ in units indexed by the frequency point $c$, and $S_{c,n}$ and $V_{c,n}$ are pre-estimated with the instantaneous SNR, which are experimentally obtained by the Wiener filter in the assumption of no correlation to each other.[6]

In Sections 3, the performance of the proposed algorithm will be presented compare with Zhu's wavelet shrinkage algorithm and IBM method[10] with Wiener wavelet threshold (WT) filter.[11]

## *Proposed wavelet speech enhancement algorithm using exponential semi-soft mask filtering*

The proposed wavelet speech enhancement algorithm has a new mask filtering that applies a mask feature matrix ($\Pi_m$). It was made in consideration of variances in wavelet coefficients of each critical band, and it is used for extracting the mask. The mask feature matrix ($\Pi_m$) is as follows:

$$\varepsilon_m = \sqrt{\frac{1}{l}\cdot\sum_{i=1}^{l}\Psi_m(i)^2 - \left(\frac{1}{l}\cdot\sum_{i=1}^{l}\Psi_m(i)\right)^2} \tag{8}$$

$$\Pi_m = \varepsilon_m\sqrt{1/2\ \cdot\log N} \tag{9}$$

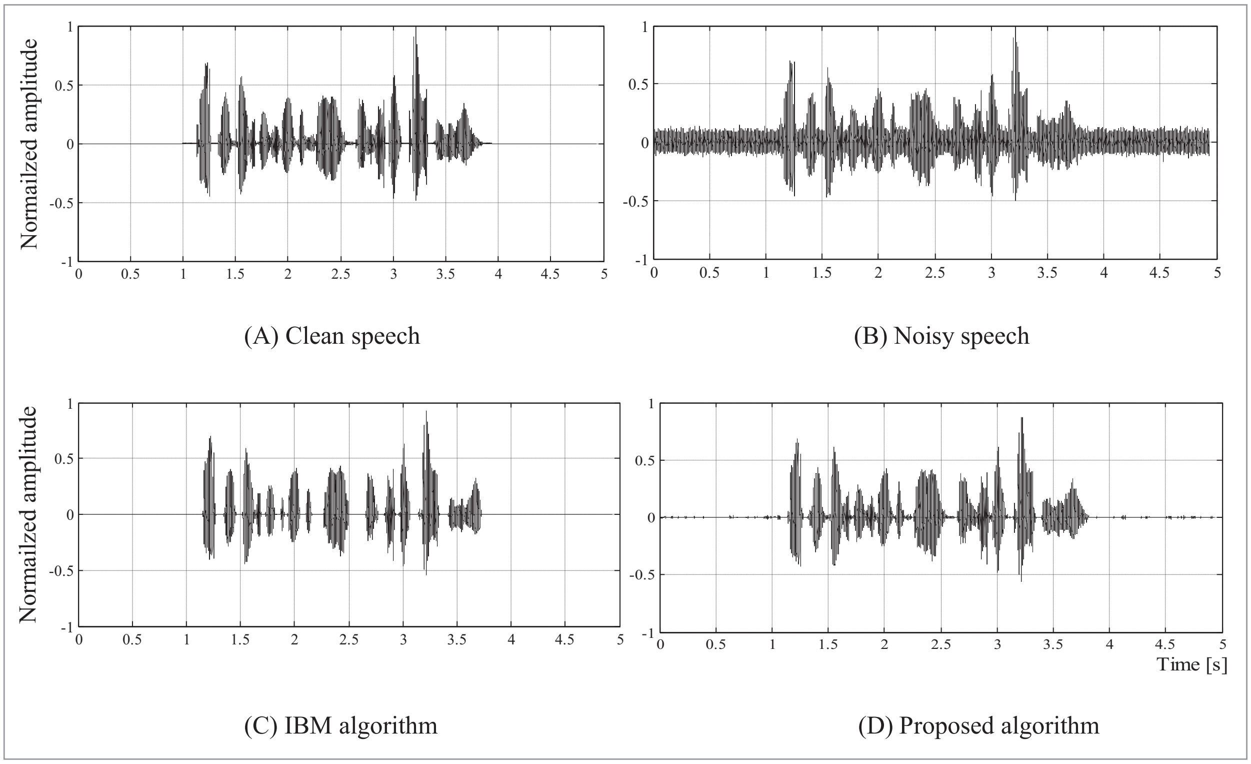
where $N$ is sample number of each frame and $l$ is sample number of 25 ms. Then the proposed semi-soft filtering mask is made based on the mask feature matrix ($\Pi_m$).

$$M_m^{\mathrm{prop}}(t) = \begin{cases} 1, & \text{if } \Pi_m \leq |\Psi_m(t)| \\ 0.5\cdot e^{B\cdot\varepsilon_m/|(I-m)+1}, & \text{else if } \overline{\Psi}_m(t-N:t) > \varepsilon_m \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

where $B$ is number of wavelet critical band ($B$ = 20), $I$ is fundamental wavelet band, and $\overline{\Psi}_m$ is average of $\Psi_m(t)$. The bands of speech are determined by the mask feature matrix ($\Pi_m$), and the speech signal is passed by the mask filtering. It maintains the speech signal between speech bands using the variances of wavelet coefficient ($\varepsilon_m$). The semi-soft mask exponentially increases and emphasizes the near fundamental wavelet bands, moreover, it adjusts balance of wavelet coefficients power. Therefore it has a feature which maintains signal continuity and minimizing signal loss. When the wavelet coefficients have low value than the variances of wavelet coefficient ($\varepsilon_m$), the semi-soft mask determined to be meaningless signal. Finally, we can be obtained enhanced speech signal after filtering process using the mask.

## Experiment and results

To perform experiment of the proposed algorithm, speech samples from the TIMIT database[12] and a noise samples from NOISEX-92[13] are used. The data samples have a sampling rate of 16 kHz and a bit rate of 16 bps. A variety of noises (white, pink, and engine of tank) and SNR environments (0 dB, 5 dB, 10 dB, and 15 dB) are used. The graphical results of speech enhancement are shown in Fig. 2. (Fig. 2A) is a clean speech signal and (Fig. 2B) is a noisy speech signal with white noise SNR of 5dB, (Fig. 2C) and (Fig. 2D) is the enhanced signal obtained using IBM and the



(A) Clean speech

(B) Noisy speech

(C) IBM algorithm

(D) Proposed algorithm

**Figure 2.** Results of the speech enhancement.

proposed algorithm. IBM has good performance of noise reduction. However, it shows speech signal loss on between syllables for example in 1.8 ~ 2.2 seconds and also has information loss on end point of speech for example in 3.8 second. On the other hand, (Fig. 2D) shows that almost completely restore the clean speech. (Fig. 2D) is very natural and clean speech signal but we can confirm the very small residual noises in non-speech part. To determine effect of residual noise, an improvement of signal-to-noise ratio (SNR) is evaluated on various noise environments and many speech samples. A random extracted speech data from TIMIT database, which comprised male and female speech signals with a variety of accents, included more than samples of 50 that have length of 3 ~5 seconds. Table 1 shows detailed values of enhanced SNR for a variety of SNRs (0, 5, 10, and 15 dB) and 3 noise environments.

The proposed algorithm has high values of enhanced SNR than the other algorithms. This signifies that the proposed algorithm has a better speech enhancement performance. All algorithms have better score in white noise than other noises. Zhu's wavelet shrinkage algorithms have poor performances in all environments. Therefore, it is not suitable for the speech signal. IBM has very good performances in high SNR environments, but relatively bad performance in low SNR environments. The proposed algorithm not only has very good performances in nice environments but also has comparatively good performances in unkind environments. Moreover the proposed algorithm exhibits a significant increase in white noise environment. This is because of MWPD and the characteristics of proposed algorithm, which adjusts balance of wavelet coefficients power. Therefore, the proposed algorithm is relatively weak in colored noise. As a result, Table 1 shows that the proposed algorithm

**Table 1.** Speech Enhancement obtained using SNR.

| Environment | | Enhanced SNR (dB) | | |
|---|---|---|---|---|
| Noise | SNR (dB) | Zhu[5] | IBM[10] | Proposed |
| White | 0 | 1.19 | 6.97 | 9.53 |
| | 5 | 5.56 | 11.42 | 12.95 |
| | 10 | 9.38 | 15.88 | 16.61 |
| | 15 | 12.23 | 20.27 | 20.49 |
| Pink | 0 | 0.61 | 4.76 | 7.04 |
| | 5 | 4.95 | 8.79 | 11.11 |
| | 10 | 8.87 | 13.56 | 14.98 |
| | 15 | 12.05 | 18.45 | 19.19 |
| Engine | 0 | 0.57 | 6.47 | 7.09 |
| | 5 | 5.03 | 9.75 | 11.66 |
| | 10 | 9.29 | 14.09 | 15.56 |
| | 15 | 13.02 | 18.88 | 19.56 |

has the best performance compared with the other algorithms in the all noise environment.

## Discussion

In this paper, we propose a new speech enhancement algorithm using exponential semi-soft mask filtering. It has modified wavelet packet decomposition shrinkage to departmentalize speech bands. And it proposed exponential semi-soft mask filtering which effectively removes the noise and minimize the loss of speech. The proposed algorithm shows good performance in a variety of noisy environments. The performance of the speech enhancement was confirmed by performing 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.

## Disclosure of potential conflicts of interest

No potential conflicts of interest were disclosed.

## Funding

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 No. 2015R1A2A2A03006113).

## References

[1] Lei SF, Tung YK. Speech enhancement for nonstationary noises by wavelet packet transform and adaptive noise estimation. In Intelligent Signal Processing and Communication Systems (ISPACS 2005), Proceedings of 2005 International Symposium on IEEE 2005; 41–44.

[2] Donoho DL. Denoising by soft thresholding. IEEE Trans. on Information Theory 1995; 41(3):613–627; http://dx.doi.org/10.1109/18.382009

[3] Bahoura M, Rouat J. Wavelet speech enhancement based on the teager energy operator. Signal Processing Letters IEEE 2001; 8(1): 10–12; http://dx.doi.org/10.1109/97.889636

[4] Gao HY, Bruce AG. Waveshrinkage with semisoft shrinkage. StatSci divition of mathsoft Inc., 1995

[5] Zhu JF, Huang YD. Improved threshold function of wavelet domain signal de-noising. In: Proc. ICWAPR 2013; 14–17.

[6] Li N, Loizou PC. Factors influencing intelligibility of ideal binary-masked speech: Implications for noise reduction. J Acoust Soc Am 2008; 123(3):1673–1682; PMID:18345855; http://dx.doi.org/10.1121/1.2832617

[7] Sun J, Tang Y, Jiang A, Xu N, Zhou L. Speech enhancement via sparse coding with ideal binary mask. In Signal Processing (ICSP 2014) 12th International Conference on IEEE 2014; 537–540; http://dx.doi.org/10.1109/ICOSP.2014.7015062

[8] Han J, Yook S, Nam KW, Lee S, Kim D, Hong SH, Kim IY. Comparative evaluation of voice activity detectors in single microphone noise reduction algorithms. Biomed Engineering Lett 2012; 2(4):255–264; http://dx.doi.org/10.1007/s13534-012-0078-3

[9] Lee G, Na SD, Cho JH, Kim MN. Voice activity detection algorithm using perceptual wavelet entropy neighbor slope. Biomed Mater Eng 2014; 24(6):3295–3301; PMID:25227039

[10] Li Y, Wang D. On the optimality of ideal binary time–frequency masks. Speech Commun 2009; 51(3):230–239; http://dx.doi.org/10.1016/j.specom.2008.09.001

[11] Hu Y, Loizou P. Speech enhancement based on wavelet thresholding the multitaper spectrum. IEEE Trans. on Speech and Audio Processing 2004; 12(1):59–67; http://dx.doi.org/10.1109/TSA.2003.819949

[12] Zue V, Seneff S, Glass J. Speech database development at MIT: TIMIT and beyond. Speech Commun 1990; 9(4):351–356; http://dx.doi.org/10.1016/0167-6393(90)90010-7

[13] Varga A, Steeneken HJM. Assessment for automatic speech recognition: II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems, Speech Commun 1993; 12(3):247–251; http://dx.doi.org/10.1016/0167-6393(93)90095-3