論文 / 著書情報
Article / Book Information

Title	A modified Wiener filtering method combined with wavelet thresholding multitaper spectrum for speech enhancement
Authors	Yanna Ma, AKINORI NISHIHARA
Citation	EURASIP Journal on Audio Speech and Music Processing, Volume 2014, Number 1, Page 32
Pub. date	2014, 8
Creative Commons	See next page.

License

EURASIP Journal on
Audio, Speech, and Music Processing
a SpringerOpen Journal

RESEARCH Open Access

A modified Wiener filtering method combined with wavelet thresholding multitaper spectrum for speech enhancement

Yanna Ma* and Akinori Nishihara

Abstract

This paper proposes a new speech enhancement (SE) algorithm utilizing constraints to the Wiener gain function which is capable of working at 10 dB and lower signal-to-noise ratios (SNRs). The wavelet thresholded multitaper spectrum was taken as the clean spectrum for the constraints. The proposed algorithm was evaluated under eight types of noises and seven SNR levels in NOIZEUS database and was predicted by the composite measures and the SNR_{LOSS} measure to improve subjective quality and speech intelligibility in various noisy environments. Comparisons with two other algorithms (KLT and wavelet thresholding (WT)) demonstrate that in terms of signal distortion, overall quality, and the SNR_{LOSS} measure, our proposed constrained SE algorithm outperforms the KLT and WT schemes for most conditions considered.

Keywords: Speech enhancement; Multitaper spectrum; Wavelet thresholding; Constrained Wiener filtering

1 Introduction

The objective of speech enhancement (SE, also called noise reduction) algorithms is to improve one or more perceptual aspects of the noisy speech by decreasing the background noise without affecting the intelligibility of the speech [1]. Research on SE can be traced back to 40 years ago with two patents by Schroeder [2], where an analog implementation of the spectral magnitude subtraction method was described. Since then, the problem of enhancing speech degraded by uncorrelated additive noise, when only the noisy speech is available, has become an area of active research [3]. Researchers and engineers have approached this challenging problem by exploiting different properties of speech and noise signals to achieve better performance [4].

SE techniques have a broad range of applications, from hearing aids to mobile communication, voice-controlled systems, multiparty teleconferencing, and automatic speech recognition (ASR) systems [4]. The algorithms can be summarized into four classes: spectral subtractive [5-8], sub-space [9,10], statistical model-based [11-13], and Wiener-type [3,14-16] algorithms.

Much progress has been made in the development of SE algorithms capable of improving speech quality [17,18] which was evaluated mainly by the objective performance criteria such as signal-to-noise ratio (SNR) [19]. However, SE algorithm that improves speech quality may not perform well in real-world listening situations where background noise level and characteristics are constantly changing [20]. The first intelligibility study done by Lim [21] in the late 1970s found no intelligibility improvement with the spectral subtraction algorithm for speech corrupted in white noise at −5 to 5 dB SNR. Thirty years later, a study conducted by Hu and Loizou [1] found that none of the examined eight different algorithms improved speech intelligibility relative to unprocessed (corrupted) speech. Moreover, according to [1], the algorithms with the highest overall speech quality may not perform the best in terms of speech intelligibility (e.g., logMMSE [12]). And the algorithm which performs the worst in terms of overall quality may perform well in terms of preserving speech intelligibility (e.g., KLT [9]). To our knowledge, very few speech enhancement algorithms [22-25] claimed to improve speech intelligibility by subjective tests for either normal-hearing listeners

*Correspondence: mayanna@nh.cradle.titech.ac.jp
Department of Communications and Integrated Systems, Tokyo Institute of Technology, Meguro-ku, Tokyo 152-8552, Japan

or hearing-impaired listeners. Hence, we focused in this paper on improving performance on speech intelligibility of the SE algorithm.

From [19], we know that the perceptual effects of attenuation and amplification distortion on speech intelligibility are not equal. Amplification distortion in excess of 6.02 dB (region III) bears the most detrimental effect on speech intelligibility, while the attenuation distortion (region I) was found to yield the least effect on intelligibility. Region I+II constraints are the most robust in terms of yielding consistently large benefits in intelligibility independent of the SE algorithm used. However, in order to divide those three regions [19], the estimated magnitude spectrum needs to be compared with the clean spectrum which we usually do not have in real circumstances.

In this paper, we explored the multitaper spectrum which was shown in [26] to have good bias and variance properties. The spectral estimate was further refined by wavelet thresholding the log multitaper spectrum in [16]. The refined spectrum was proposed in this paper to be used as an alternative of the clean spectrum. Then, the region I+II constraints were imposed and incorporated in the derivation of the gain function of the Wiener algorithm based on *a priori* SNR [3]. We have experimentally evaluated its performance under a variety of noise types and SNR conditions.

The structure of the rest of this paper is organized as follows. Section 2 provides the background information on wavelet thresholding the multitaper spectrum, and Section 3 presents the proposed approach which imposes constraints on the Wiener filtering gain function. Section 4 contains the speech and noise database and metrics used in the evaluation. The simulation results are given in Section 5. Finally, a conclusion of this work and the discussion are given in Section 6.

2 Wavelet thresholding the multitaper spectrum

In real-world scenarios, the background noise level and characteristics are constantly changing [20]. Better estimation of the spectrum is required to alleviate the distortion caused by SE algorithms. For speech enhancement, the most frequently used power spectrum estimator is direct spectrum estimation based on Hann windowing. However, windowing reduces only the bias not the variance of the spectral estimate [27]. The multitaper spectrum estimator [26], on the other hand, can reduce this variance by computing a small number (L) of direct spectrum estimators (eigenspectra) each with a different taper (window) and then averaging the L spectral estimates. The underlying philosophy is similar to Welch's method of modified periodogram [27].

The multitaper spectrum estimator is given by

$$\widehat{S}^{mt}(\omega) = \frac{1}{L}\sum_{k=0}^{L-1}\widehat{S}_k^{mt}(\omega) \tag{1}$$

with

$$\widehat{S}_k^{mt}(\omega) = \left|\sum_{m=0}^{N-1} a_k(m)x(m)e^{-j\omega m}\right|^2, \tag{2}$$

where N is the data length, and a_k is the kth sine taper used for the spectral estimate $\widehat{S}_k^{mt}(\cdot)$, which is proposed by Riedel and Sidorenko [28] and defined by

$$a_k(m) = \sqrt{\frac{2}{N+1}}sin\frac{\pi k(m+1)}{N+1}, m = 0,\cdots,N-1. \tag{3}$$

The sine tapers were proved in [28] to produce smaller local bias than the Slepian tapers, with roughly the same spectral concentration.

The multitaper estimated spectrum can be further refined by wavelet thresholding techniques [29-31]. Improved periodogram estimates were proposed in [29], and improved multitaper spectrum estimates were proposed in [30,31]. The underlying idea behind those techniques is to represent the log periodogram as 'signal' plus the 'noise', where the signal is the true spectrum and the noise is the estimation error [32]. It was shown in [33] that if the eigenspectra defined in Equation 2 are assumed to be uncorrelated, the ratio of the estimated multitaper spectrum $\widehat{S}^{mt}(\omega)$ and the true power spectrum $S(\omega)$ conforms to a chi-square distribution with $2L$ degrees of freedom, i.e.,

$$\upsilon(\omega) = \frac{\hat{S}^{mt}(\omega)}{S(\omega)} \sim \frac{\chi^2_{2L}}{2L}, \quad 0 < \omega < \pi. \tag{4}$$

Taking the log of both sides, we get

$$\log\hat{S}^{mt}(\omega) = \log S(\omega) + \log\upsilon(\omega). \tag{5}$$

From Equation 5, we know that the log of the multitaper spectrum can be represented as the sum of the true log spectrum plus a $\log\chi^2$ distributed noise term. It follows from Bartlett and Kendall [34] that the distribution of $\log\upsilon(\omega)$ is with mean $\phi(L) - \log(L)$ and variance $\phi'(L)$, where $\phi(\cdot)$ and $\phi'(\cdot)$ denote, respectively, the digamma and trigamma functions. For $L \geq 5$, the distribution of $\log\upsilon(\omega)$ will be close to a normal distribution [35]. Hence, provided L is at least 5, the random variable $\eta(\omega)$

$$\eta(\omega) = \log\upsilon(\omega) - \phi(L) + \log(L) \tag{6}$$

will be approximately Gaussian with zero mean and variance $\sigma_\eta^2 = \phi'(L)$. If $Z(\omega)$ is defined as

$$Z(\omega) = \log\hat{S}^{mt}(\omega) - \phi(L) + \log(L), \tag{7}$$

then we have

$$Z(\omega) = \log S(\omega) + \eta(\omega), \tag{8}$$

i.e., the log multitaper power spectrum plus a known constant ($\log(L) - \phi(L)$) can be written as the true log power spectrum plus approximately Gaussian noise $\eta(\omega)$ with zero mean and known variance σ_η^2 [30].

The model in Equation 8 is well suited for wavelet denoising techniques [36-39] for eliminating the noise $\eta(\omega)$ and obtaining a better estimate of the log spectrum. The idea behind refining the multitaper spectrum by wavelet thresholding can be summarized into four steps [16].

- Obtain the multitaper spectrum using Equations 1 to 3 and calculate $Z(w)$ using Equation 7.
- Apply a standard periodic discrete wavelet transform (DWT) out to level q_0 to $Z(w)$ to get the empirical DWT coefficients $z_{j,k}$ at each level j, where q_0 is specified in advance [40].
- Apply a thresholding procedure to $z_{j,k}$.
- The inverse DWT is applied to the thresholded wavelet coefficients to obtain the refined log spectrum.

3 Speech enhancement based on constrained Wiener filtering algorithm

Among the numerous techniques that were developed, the Wiener filter can be considered as one of the most fundamental SE approaches, which has been delineated in different forms and adopted in various applications [4]. The Wiener gain function is the least aggressive, in terms of suppression, providing small attenuation even at extremely low SNR levels.

A block diagram of the proposed SE algorithm is shown in Figure 1. The initial four frames are assumed to be noise only. The algorithm can be described as follows. The input noisy speech signal is decomposed into frames of 20-ms length with an overlap of 10 ms by the Hann window. Each segment was transformed using a 160-point discrete Fourier transform (DFT). The spectrum of the segmented noisy and noise signal are estimated by the multitaper method and then further refined by wavelet thresholding technique. The estimated 'clean' spectrum was gotten from the refined multitaper estimated noisy and noise spectrum. On the other hand, the noise-corrupted sentences were enhanced by the Wiener algorithm based on *a priori* SNR estimation [3]. The region I+II constraints were then imposed on the enhanced spectrum. Finally, the inverse fast Fourier transform (FFT) was applied to obtain the enhanced speech signal.

The implementation details of the proposed method can be described in the following four steps. For each speech frame,

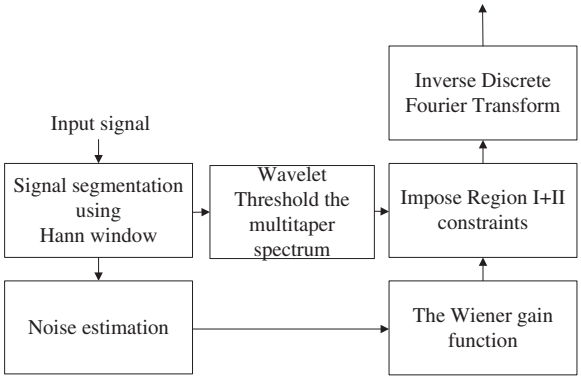

Figure 1 Block diagram of the proposed speech enhancement algorithm.

- compute the multitaper power spectrum $\widehat{S}_{\mathbf{y}}^{mt}$ of the noisy speech $\mathbf{y}$ using Equation 1 and estimate the multitaper power spectrum $\widehat{S}_{\mathbf{x}}^{mt}$ of the clean speech signal by $\widehat{S}_{\mathbf{x}}^{mt} = \widehat{S}_{\mathbf{y}}^{mt} - \widehat{S}_{\mathbf{n}}^{mt}$, where $\widehat{S}_{\mathbf{n}}^{mt}$ is the multitaper power spectrum of the noise. $\widehat{S}_{\mathbf{n}}^{mt}$ can be obtained using noise samples collected during speech absent frames. Here, L is set to 16. Any negative elements of $\widehat{S}_{\mathbf{x}}^{mt}$ are floored as follows:

$$\widehat{S}_{\mathbf{x}}^{mt} = \begin{cases} \widehat{S}_{\mathbf{y}}^{mt} - \widehat{S}_{\mathbf{n}}^{mt}, & \text{if } \widehat{S}_{\mathbf{y}}^{mt} > \widehat{S}_{\mathbf{n}}^{mt} \\ \beta \widehat{S}_{\mathbf{n}}^{mt}, & \text{if } \widehat{S}_{\mathbf{y}}^{mt} \leq \widehat{S}_{\mathbf{n}}^{mt}, \end{cases} \tag{9}$$

  where β is the spectral floor set to $\beta = 0.002$.
- compute $Z(\omega) = log\hat{S}_{\mathbf{y}}^{mt}(\omega) - \phi(L) + log(L)$ and then apply the DWT of $Z(\omega)$ out to level q_0 to obtain the empirical DWT coefficients $z_{j,k}$ for each level j, where q_0 is specified to be 5 [40]. Threshold the wavelet coefficients $z_{j,k}$ and apply the inverse DWT to the thresholded wavelet coefficients to obtain the refined log spectrum, $log\widehat{S}_{\mathbf{y}}^{\omega mt}(\omega)$, of the noisy singal. Repeat the above procedure to obtain the refined log spectrum, $log\widehat{S}_{\mathbf{n}}^{\omega mt}(\omega)$, of the noise signal. The estimated power spectrum $\widehat{S}_{\mathbf{x}}^{\omega mt}(\omega)$ of the clean speech signal can be estimated using

$$\widehat{S}_{\mathbf{x}}^{\omega mt}(\omega) = \widehat{S}_{\mathbf{y}}^{\omega mt}(\omega) - \widehat{S}_{\mathbf{n}}^{\omega mt}(\omega) \tag{10}$$

- let $Y(\omega, t)$ denote the magnitude of the noisy spectrum at time frame t and frequency bin ω estimated by the method in [41]. Then, the estimate of the signal spectrum magnitude is obtained by multiplying $Y(\omega, t)$ with a gain function $G(\omega, t)$ as $\hat{X}(\omega, t) = G(\omega, t) \cdot Y(\omega, t)$. The Wiener gain function is based on the *a priori* SNR and is given by

$$G(\omega, t) = \sqrt{\frac{\mathrm{SNR}_{\mathrm{prio}}(\omega, t)}{1 + \mathrm{SNR}_{\mathrm{prio}}(\omega, t)}}, \tag{11}$$

where $\mathrm{SNR}_{\mathrm{prio}}$ is the *a priori* SNR estimated using the decision-directed approach [3,19] as follows:

$$\mathrm{SNR}_{\mathrm{prio}} = \alpha \cdot \frac{X_M^2(\omega, t-1)}{\hat{P}_D^2(\omega, t-1)} + (1-\alpha) \cdot \max\left[\frac{Y^2(\omega, t)}{\hat{P}_D^2(\omega, t)} - 1,\ 0\right], \tag{12}$$

where $\hat{P}_D^2(\omega, t)$ is the estimate of the power spectral density of background noise, and α is a smoothing constant (typically set to $\alpha = 0.98$).

- to maximize speech intelligibility, the final enhanced spectrum, $X_M(\omega, t)$, can be obtained by utilizing the region I+II constraints to the enhanced spectrum $\hat{X}(\omega, t)$ as follows:

$$X_M(\omega, t) = \begin{cases} \hat{X}(\omega, t), & \text{if } \hat{X}(\omega, t) < 2\widehat{S}_{\mathbf{x}}^{\omega mt}(\omega) \\ 0 & \text{else} \end{cases} \tag{13}$$

Finally, the enhanced speech signal can be obtained by apply the inverse FFT of $X_M(\omega, t)$.

The above estimator was applied to 20-ms duration frames of the noisy signal with 50% overlap between frames. The enhanced speech signal was combined using the overlap and add method.

4 Evaluation setup

The proposed SE algorithm was tested using a speech database that was corrupted by eight different real-world noises at different SNRs. The system was evaluated using both the composite evaluation measures proposed in [42] and the $\mathrm{SNR}_{\mathrm{LOSS}}$ measure proposed in [43].

4.1 Database description

For the evaluation of SE algorithms, NOIZEUS [44] is preferred since it is a noisy speech corpus recorded by [18] to facilitate comparison of SE algorithms among different research groups [20]. The noisy database contains thirty IEEE sentences [45] which were recorded in a sound-proof booth using Tucker Davis Technologies (TDT; Alachua, FL, USA) recording equipment. The sentences were produced by three male and three female speakers (five sentences/speaker). The IEEE database was used as it contains phonetically balanced sentences with relatively low word-context predictability. The 30 sentences were selected from the IEEE database so as to include all phonemes in the American English language. The sentences were originally sampled at 25 kHz and downsampled to 8 kHz.

To simulate the receiving frequency characteristics of the telephone handsets, the intermediate reference system (IRS) filter used in ITU-T P.862 [46] for evaluation of the perceptual evaluation of speech quality (PESQ) measures was independently applied to the clean and noise signal [17]. Then, noise segment of the same length as the speech signal was randomly cut out of the noise recordings, appropriately scaled to reach the desired SNR levels ($-8, -5, -2, 0, 5, 10$, and 15 dB) and finally added to the filtered clean speech signal. Noise signals were taken from the AURORA database [47] and included the following recordings from different places: train, babble (crowd of people), car, exhibition hall, restaurant, street, airport, and train station. Therefore, in total, there are 1,680 (30 sentences $\times$ 8 noises $\times$ 7 SNRs) noisy speech segments in the test set.

4.2 Performance evaluation

The performance of an SE algorithm can be evaluated both subjectively and objectively. In general, subjective listening test is the most accurate and preferable method for evaluating speech quality and intelligibility. However, it is time consuming and cost expensive. Recently, many researchers have placed much effort on developing objective measures that would predict subjective quality and intelligibility with high correlation [42,43,48,49] with subjective listening test. Among them, the composite objective measures [42] were proved to have high correlation with subjective ratings and, at the same time, capture different characteristics of the distortions present in the enhanced signals [35], while the $\mathrm{SNR}_{\mathrm{LOSS}}$ measure [43] was found appropriate in predicting speech intelligibility in fluctuating noisy conditions by yielding a high correlation for predicting sentence recognition. Therefore, the composite objective measures and the $\mathrm{SNR}_{\mathrm{LOSS}}$ measure were adopted to predict the performance of the proposed SE algorithm on subjective quality and speech intelligibility, respectively.

4.2.1 *The composite measures to predict subjective speech quality*

The composite objective measures are obtained by linearly combining existing objective measures that highly correlate with subjective ratings. The objective measures include segmental SNR (*segSNR*) [18], weighted-slope spectral (*WSS*) [50], *PESQ* [51], and log likelihood ration (*LLR*) [18].

The three new composite measures obtained from multiple linear regression analysis are given below:

- C_{sig}: A five-point scale of signal distortion (SIG) formed by linearly combining the *LLR*, *PESQ*, and *WSS* measures (Table 1).
- C_{bak}: A five-point scale of noise intrusiveness (BAK) formed by linearly combining the *segSNR*, *PESQ*, and *WSS* measures (Table 1).
- C_{ovl}: The mean opinion score of overall quality (OVRL) formed by linearly combining the *PESQ*, *LLR*, and *WSS* measures.

Table 1 Scale of signal distortion, background intrusiveness and overall quality

Scale of signal distortion	Scale of background intrusiveness	Scale of overall quality
5- very natural, no degradation	5- not noticeable	5- excellent
4- fairly natural, little degradation	4- somewhat noticeable	4- good
3- somewhat natural, somewhat degraded	3- noticeable but not intrusive	3- fair
2- fairly unnatural, fairly degraded	2- fairly conspicuous, somewhat intrusive	2- poor
1- very unnatural, very degraded	1- very conspicuous, very intrusive	1- bad

The three new composite measures obtained from multiple linear regression analysis are given below:

$$C_{\text{sig}} = 3.093 - 1.029 \cdot \text{LLR} + 0.603 \cdot \text{PESQ} - 0.009 \cdot \text{WSS} \tag{14}$$

$$C_{\text{bak}} = 1.634 + 0.478 \cdot \text{PESQ} - 0.007 \cdot \text{WSS} + 0.063 \cdot \text{segSNR} \tag{15}$$

$$C_{\text{ovl}} = 1.594 + 0.805 \cdot \text{PESQ} - 0.512 \cdot \text{LLR} - 0.007 \cdot \text{WSS} \tag{16}$$

The correlation coefficients between the three composite measures and real subjective measures are given in Table 2 [42]. All three parameters should be maximized in order to get the best performance.

4.2.2 The SNR_{LOSS} measure to predict speech intelligibility

The SNR loss in band j and frame m is defined as follows [43]:

$$SL(j,m) = \text{SNR}_X(j,m) - \text{SNR}_{\hat{X}}(j,m), \tag{17}$$

where $\text{SNR}_X(j,m)$ is the input SNR in band j, $\text{SNR}_{\hat{X}}(j,m)$ is the SNR of the enhanced signal in the jth frequency band at the mth frame.

Assuming the SNR range is restricted to $[-\text{SNR}_{\text{Lim}}, \text{SNR}_{\text{Lim}}]$ dB ($\text{SNR}_{\text{Lim}} = 3$ in this paper), the $SL(j,m)$ term is then limited as follows:

$$\hat{SL}(j,m) = \min(\max(SL(j,m), -\text{SNR}_{\text{Lim}}), \text{SNR}_{\text{Lim}}) \tag{18}$$

and subsequently mapped to the range of [0, 1] using the following equation:

Table 2 Correlation coefficients between the composite measures and subjective measure

	C_{sig}	C_{bak}	C_{ovl}
SIG	0.7		
BAK		0.58	
OVRL			0.73

$$\text{SNR}_{\text{LOSS}}(j,m) = \begin{cases} -\dfrac{C_-}{\text{SNR}_{\text{Lim}}}\hat{SL}(j,m), & \text{if } \hat{SL}(j,m) < 0 \\ \dfrac{C_+}{\text{SNR}_{\text{Lim}}}\hat{SL}(j,m), & \text{if } \hat{SL}(j,m) \geq 0 \end{cases} \tag{19}$$

where C_- and C_+ are parameters (fixed to be 1 in this paper) controlling the slopes of the mapping function which was defined in the range of [0, 1]; therefore, the frame SNR_{LOSS} is normalized to the range of $0 \leq \text{SNR}_{\text{LOSS}}(j,m) \leq 1$. The average SNR_{LOSS} is finally computed by averaging $\text{SNR}_{\text{LOSS}}(j,m)$ over all frames in the signal as follows:

$$\overline{\text{SNR}}_{\text{LOSS}} = \frac{1}{M}\sum_{m=0}^{M-1} f\text{SNR}_{\text{LOSS}}(m), \tag{20}$$

where M is the total number of data segments in the signal, and $f\,\text{SNR}_{\text{LOSS}}(m)$ is the average (across bands) SNR loss computed as follows:

$$f\text{SNR}_{\text{LOSS}}(m) = \frac{\sum_{j=1}^{K} W(j) \cdot \text{SNR}_{\text{LOSS}}(j,m)}{\sum_{j=1}^{K} W(j)}, \tag{21}$$

where $W(j)$ is the weight (i.e., band importance function [52]) placed on the jth frequency band and was taken from Table B.1 in the ANSI standard [52].

The implementation of the SNR_{LOSS} measure was supplied in the website of the authors in [43]. The smaller the value of the SNR_{LOSS} measure is, the better performance of the SE algorithm is achieved.

5 Simulation results

The evaluation of the subjective quality and intelligibility of the speech enhanced by our proposed SE algorithm are reported in this section. Three other SE schemes, namely, wavelet thresholding (WT) [16], KLT [9], and Wiener algorithm with clean signal present (Wiener_Clean) [19], were also evaluated in order to gain a comparative analysis of the proposed SE algorithm. The KLT algorithm was proved in [1] and [22] by subjective tests to perform well in terms of preserving speech intelligibility for normal hearing listeners and improving speech intelligibility significantly for cochlear implant users in regard to recognition of sentences corrupted by stationary noises,

respectively. The Wiener_Clean algorithm was taken as the ground truth in this paper because there is clean signal used in the algorithm. The unprocessed noisy signal (UP) was also evaluated by the SNR_{LOSS} measure for comparison purposes. The implementations of these three schemes were taken from the implementations in [18].

5.1 Performance of predicting subjective quality

5.1.1 Performance average over all eight kinds of noise

In Figure 2, the proposed algorithm is compared with WT and KLT algorithms in terms of the composite measures averaging over all eight noises for seven SNRs. The four objective measures (*LLR*, *segSNR*, *WSS*, and *PESQ*) that composed the composite measures were also given in the first row for reference. The Wiener_Clean algorithm, as the ground truth, performed the best for all four objective evaluation measures. According to [42], the *LLR* measure performed the best in terms of predicting signal distortion, while the *PESQ* measure gave the best prediction for both noise intrusiveness and overall speech quality. From the first row of Figure 2, we can notice that our proposed algorithm gives better performance than both WT and KLT in terms of the *LLR* measure for all seven SNRs tested. Moreover, when SNR is smaller than 5 dB, our proposed algorithm also performed better than both WT and KLT for the *PESQ* measure.

The second row of Figure 2 shows the composite measures, which include C_{sig}, C_{bak}, and C_{ovl}, estimated by the combination of all those four objective measures expressed in the first row. In terms of both signal distortion C_{sig} and overall quality C_{ovl}, our proposed method performs the best when SNR is less than 10 dB. Specifically speaking, for overall quality measure C_{ovl}, the proposed algorithm improved 10.94%, 18.94%, 21.63%, 23.66%, and 6.67% for $-8, -5, -2, 0,$ and 5 dB, respectively, when compared with the KLT method. In general, the proposed algorithm achieved 13.88% and 6.40% improvement for C_{sig} and C_{ovl}, respectively, when average over all seven tested SNR levels. However, for C_{bak}, the WT and KLT algorithms give similar and better results than our proposed one when SNR is no smaller than 0 dB. The improvement was 0.98%, 6.98%, 11.11%, and 16.55% for 0, 5, 10, and 15 dB, respectively. In average, the WT and KLT methods were 5.14% better than our proposed algorithm in terms of background intrusiveness C_{bak}.

5.1.2 Performance average over seven SNRs

Figure 3 shows the three different composite measures averaged over seven SNRs for eight kinds of noise computed for WT, KLT, Wiener_Clean, and proposed SE algorithms. The Wiener_Clean algorithm still works as the ground truth here. From Figure 3, it is clear that in terms of C_{sig}, the KLT works much better than WT. Hence, the proposed algorithm is compared with only the KLT method here. We observe that on average, the proposed algorithm is better than the KLT method in terms of C_{sig} for train (9.19%), babble (15.93%), car (14.51%), exhibition hall (7.74%), restaurant (16.74%), street (13.42%), airport (16.64%), and train station (16.23%) noises. The number in the bracket indicates the C_{sig} by which our proposed

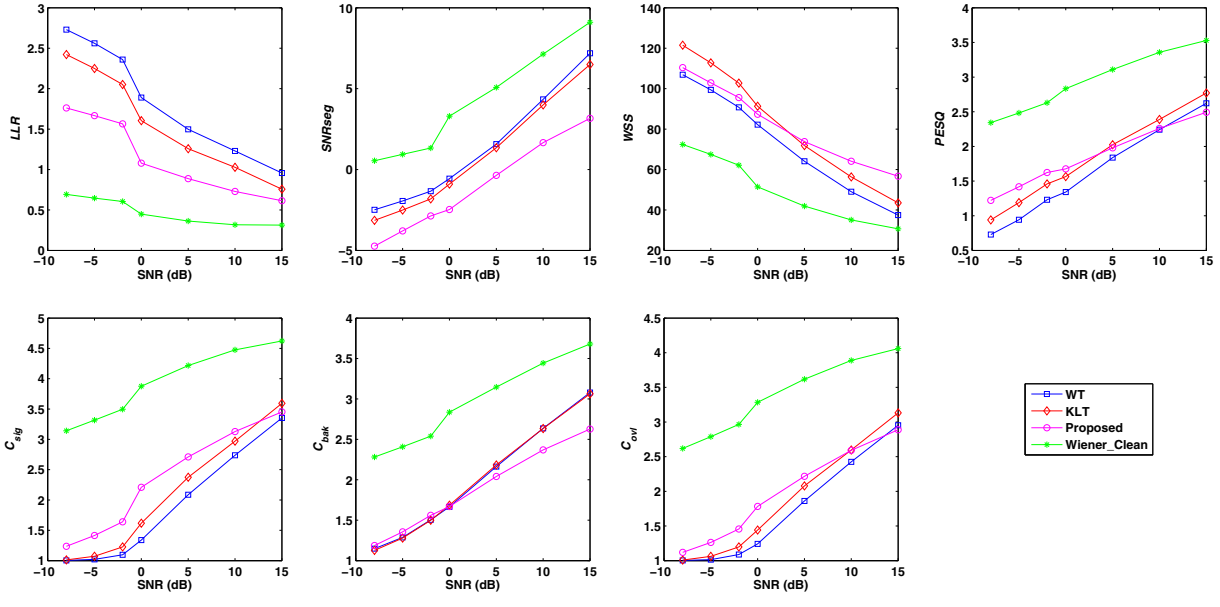

Figure 2 Averaged over seven SNRs. The composite measure comparisons for four SE schemes (WT, KLT, proposed, and Wiener_Clean) averaged over seven SNRs ($-8, -5, -2, 0, 5, 10,$ and 15 db) for eight kinds of noise.

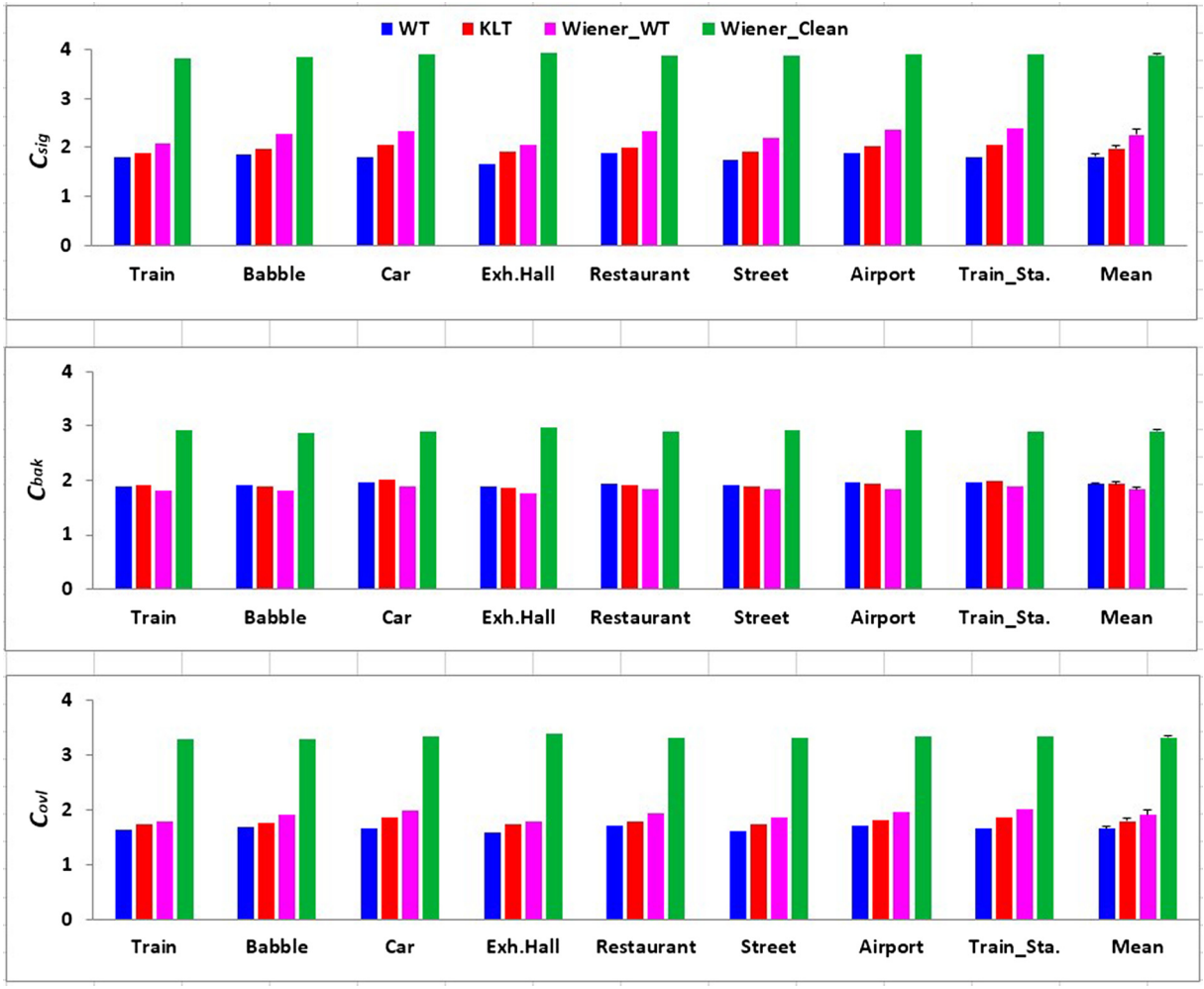

Figure 3 Averaged over eight kinds of noise. The composite measure comparisons for four SE schemes (WT, KLT, proposed, and Wiener_Clean) averaged over eight kinds of noise for seven SNRs (−8, −5, −2, 0, 5, 10, and 15 dB).

algorithm is better than the KLT method. The mean C_{sig} over all eight noise types of our proposed SE algorithm is 13.88% better than that of the KLT method. Furthermore, the proposed SE algorithms outperforms the KLT in terms of C_{ovl} by an average of 6.40% over all eight kinds of noise that were considered. However, in terms of background intrusiveness C_{bak}, the KLT algorithm gives an average of 5.14% better results than our proposed algorithm.

Thus, in conclusion, the proposed SE algorithm was predicted to be able to achieve the best overall subjective quality for most SNRs and all noise types considered when comparing with WT and KLT algorithms.

5.2 Performance of predicting speech intelligibility

The SNR_{LOSS} measure values obtained from each algorithm (include UP) were subjected to statistical analysis in order to assess their significant differences. A highly significant effect ($p < 0.005$) was found in all SNR levels and all types of noise by analysis of variance (ANOVA). Following the ANOVA, multiple comparison statistical tests according to Tukey's HSD test were done to assess the significance between algorithms. The difference was deemed significant if the p value was smaller than 0.05.

Table 3 gives the statistical comparisons of the SNR_{LOSS} measure between unprocessed noisy sentences (UP) and enhanced sentences by four SE algorithms (WT, KLT, Wiener_Clean, and proposed). At the same time, the comparisons between our proposed SE algorithm and the other three algorithms were also given. From Table 3, we know that when compared with the UP, our proposed algorithm was predicted by the SNR_{LOSS} measure to be able to improve the intelligibility in low SNRs for most noises tested (italicized). The R in the table gives the

Table 3 Statistical comparisons of the SNR_{LOSS} measure between unprocessed sentences and enhanced sentences by SE algorithms

		WT-UP		KLT-UP		Wiener_Clean-UP		Proposed-UP		Proposed-WT		Proposed-KLT		Proposed-Wiener_Clean	
Noise	**SNR (dB)**	***R* (%)**	***p* value**	***R* (%)**	***p* value**	***R* (%)**	***p* value**	***R* (%)**	***p* value**	***R* (%)**	***p* value**	***R* (%)**	***p* value**	***R* (%)**	***p* value**
Train	−8	1.30	0.000	0.98	0.011	−7.05	0.000	*−0.87*	*0.031*	*−2.15*	*0.000*	*-1.83*	*0.000*	6.64	0.000
	−5	1.61	0.000	1.26	0.004	−6.87	0.000	−0.91	0.076	*−2.47*	*0.000*	*−2.13*	*0.000*	6.41	0.000
	−2	1.93	0.000	1.43	0.005	−6.74	0.000	−0.96	0.127	*−2.83*	*0.000*	*−2.35*	*0.000*	6.20	0.000
	0	3.41	0.000	2.33	0.002	−9.83	0.000	−1.67	0.054	*−4.92*	*0.000*	*−3.92*	*0.000*	9.04	0.000
	5	3.44	0.001	2.12	0.084	−8.95	0.000	0.45	0.983	*−2.89*	*0.004*	−1.63	0.264	10.32	0.000
	10	2.96	0.024	2.47	0.091	−7.06	0.000	3.74	0.002	0.76	0.930	1.25	0.687	11.63	0.000
	15	3.54	0.034	4.96	0.001	−1.85	0.553	10.76	0.000	6.97	0.000	5.52	0.000	12.85	0.000
Babble	−8	1.61	0.000	0.73	0.128	−7.69	0.000	*−1.16*	*0.002*	*−2.73*	*0.000*	*−1.88*	*0.000*	7.07	0.000
	−5	2.25	0.000	1.21	0.010	−7.23	0.000	*−1.11*	*0.022*	*−3.28*	*0.000*	*−2.30*	*0.000*	6.60	0.000
	−2	2.42	0.000	1.22	0.047	−6.90	0.000	−1.06	0.118	*−3.39*	*0.000*	*−2.25*	*0.000*	6.28	0.000
	0	3.25	0.000	1.55	0.046	−9.57	0.000	*−1.85*	*0.009*	*−4.93*	*0.000*	*−3.34*	*0.000*	8.54	0.000
	5	4.88	0.000	2.76	0.002	−8.06	0.000	0.10	1.000	*−4.56*	*0.000*	*−2.59*	*0.004*	8.87	0.000
	10	6.11	0.000	4.74	0.000	−4.76	0.000	4.72	0.000	−1.31	0.684	−0.02	1.000	9.95	0.000
	15	6.84	0.000	7.72	0.000	1.91	0.687	12.76	0.000	5.54	0.001	4.67	0.006	10.65	0.000
Car	−8	1.18	0.000	0.12	0.984	−8.38	0.000	*−1.89*	*0.000*	*−3.04*	*0.000*	*−2.01*	*0.000*	7.09	0.000
	−5	1.90	0.000	0.46	0.424	−8.20	0.000	*−2.11*	*0.000*	*−3.94*	*0.000*	*−2.56*	*0.000*	6.64	0.000
	−2	2.25	0.000	0.56	0.321	−8.20	0.000	*−2.48*	*0.000*	*−4.62*	*0.000*	*−3.03*	*0.000*	6.23	0.000
	0	3.38	0.000	0.78	0.373	−11.23	0.000	*−3.45*	*0.000*	*−6.61*	*0.000*	*−4.19*	*0.000*	8.77	0.000
	5	3.46	0.000	−0.21	0.997	−10.70	0.000	*−2.33*	*0.001*	*−5.59*	*0.000*	*−2.12*	*0.005*	9.38	0.000
	10	5.32	0.000	1.90	0.345	−7.52	0.000	1.55	0.554	*−3.58*	*0.003*	−0.34	0.997	9.81	0.000
	15	5.10	0.000	2.78	0.075	−2.68	0.093	8.25	0.000	2.99	0.030	5.32	0.000	11.22	0.000
Exhibition Hall	−8	0.58	0.026	−0.42	0.201	−7.79	0.000	*−1.44*	*0.000*	*−2.01*	*0.000*	*−1.03*	*0.000*	6.88	0.000
	−5	1.01	0.000	−0.17	0.947	−7.70	0.000	*−1.61*	*0.000*	*−2.60*	*0.000*	*−1.45*	*0.000*	6.60	0.000
	−2	1.41	0.000	0.03	1.000	−7.48	0.000	*−1.83*	*0.000*	*−3.20*	*0.000*	*−1.86*	*0.000*	6.10	0.000
	0	4.00	0.000	0.77	0.738	−10.52	0.000	*−1.95*	*0.020*	*−5.72*	*0.000*	*−2.70*	*0.000*	9.58	0.000
	5	3.68	0.000	−0.62	0.889	−9.69	0.000	0.15	0.999	*−3.40*	*0.000*	0.77	0.779	10.90	0.000
	10	4.86	0.001	2.14	0.393	−6.57	0.000	4.34	0.004	−0.49	0.993	2.15	0.365	11.67	0.000
	15	6.17	0.000	4.88	0.003	−1.24	0.880	11.13	0.000	4.67	0.002	5.96	0.000	12.53	0.000

Table 3 Statistical comparisons of the SNR_{LOSS} measure between unprocessed sentences and enhanced sentences by SE algorithms *(Continued)*

Restaurant	−8	1.81	0.000	0.95	0.033	−7.42	0.000	*−1.14*	*0.005*	*−2.90*	*0.000*	*−2.07*	*0.000*	6.78	0.000
	−5	2.15	0.000	1.10	0.044	−7.09	0.000	−1.07	0.053	*−3.16*	*0.000*	*−2.15*	*0.000*	6.47	0.000
	−2	2.18	0.000	1.12	0.113	−6.67	0.000	−1.05	0.159	*−3.16*	*0.000*	*−2.14*	*0.000*	6.03	0.000
	0	4.76	0.000	3.21	0.000	−8.82	0.000	−0.88	0.746	*−5.38*	*0.000*	*−3.96*	*0.000*	8.70	0.000
	5	4.06	0.002	2.53	0.142	−7.90	0.000	0.83	0.939	*−3.10*	*0.028*	−1.65	0.525	9.49	0.000
	10	7.40	0.000	6.27	0.000	−3.21	0.053	6.53	0.000	−0.81	0.945	0.24	1.000	10.06	0.000
	15	8.27	0.000	9.01	0.000	−3.77	0.101	15.18	0.000	6.38	0.000	5.67	0.001	11.00	0.000
Street	−8	1.11	0.045	0.72	0.363	−7.26	0.000	*−1.27*	*0.015*	*−2.35*	*0.000*	*−1.98*	*0.000*	6.46	0.000
	−5	1.48	0.017	0.89	0.329	−7.24	0.000	*−1.32*	*0.044*	*−2.76*	*0.000*	*−2.19*	*0.000*	6.39	0.000
	−2	1.58	0.030	0.86	0.502	−7.12	0.000	−1.36	0.088	*−2.90*	*0.000*	*−2.20*	*0.001*	6.21	0.000
	0	4.76	0.000	2.90	0.004	−9.14	0.000	−1.13	0.622	*−5.62*	*0.000*	*−3.91*	*0.000*	8.82	0.000
	5	4.64	0.000	2.85	0.036	−8.59	0.000	0.74	0.944	*−3.73*	*0.001*	−2.05	0.210	10.20	0.000
	10	6.28	0.001	5.29	0.007	−5.61	0.003	5.36	0.006	−0.86	0.976	0.07	1.000	11.62	0.000
	15	7.55	0.000	7.49	0.000	0.47	0.998	12.22	0.000	4.34	0.014	4.41	0.012	11.70	0.000
Airport	−8	1.80	0.000	0.96	0.147	−7.89	0.000	*−1.51*	*0.004*	*−3.25*	*0.000*	*−2.44*	*0.000*	6.93	0.000
	−5	2.31	0.000	1.37	0.048	−7.42	0.000	*−1.37*	*0.048*	*−3.60*	*0.000*	*−2.70*	*0.000*	6.53	0.000
	−2	2.65	0.000	1.60	0.035	−6.78	0.000	−1.13	0.249	*−3.68*	*0.000*	*−2.68*	*0.000*	6.06	0.000
	0	4.53	0.000	1.72	0.225	−9.43	0.000	−1.45	0.392	*−5.72*	*0.000*	*−3.12*	*0.001*	8.81	0.000
	5	5.26	0.000	3.36	0.007	−7.69	0.000	0.60	0.972	*−4.42*	*0.000*	*−2.67*	*0.042*	8.99	0.000
	10	7.72	0.000	5.88	0.000	−2.90	0.113	6.58	0.000	−1.05	0.876	0.66	0.977	9.76	0.000
	15	7.76	0.000	7.43	0.000	2.61	0.270	13.81	0.000	5.62	0.000	5.95	0.000	10.92	0.000
Train station	−8	1.67	0.000	0.89	0.034	−8.37	0.000	*−1.91*	*0.000*	*−3.52*	*0.000*	*−2.78*	*0.000*	7.04	0.000
	−5	2.30	0.000	1.22	0.010	−7.89	0.000	*−1.90*	*0.000*	*−4.11*	*0.000*	*−3.09*	*0.000*	6.50	0.000
	−2	2.64	0.000	1.47	0.006	−7.53	0.000	*−1.88*	*0.000*	*−4.41*	*0.000*	*−3.30*	*0.000*	6.11	0.000
	0	3.19	0.000	0.58	0.857	−10.72	0.000	*−2.97*	*0.000*	*−5.98*	*0.000*	*−3.54*	*0.000*	8.67	0.000
	5	4.25	0.003	1.24	0.812	−9.47	0.000	−1.63	0.608	*−5.64*	*0.000*	−2.84	0.092	8.65	0.000
	10	5.43	0.000	2.38	0.192	−5.94	0.000	2.79	0.082	−2.50	0.115	0.41	0.995	9.29	0.082
	15	7.60	0.000	6.30	0.000	0.38	0.999	11.19	0.000	3.33	0.065	4.60	0.004	10.76	0.000

Four SE algorithms: WT, KLT, Wiener_Clean, and proposed. The comparisons between our proposed SE algorithm and the other three were also given. The results in italics show the SNR_{LOSS} measure by which our proposed SE algorithm is better than others, and the difference was significant ($p<0.05$).

percentage by which our algorithm is better than others; the value is negative because better performance gave smaller SNR_{LOSS} measure. Furthermore, our proposed SE algorithm was also compared with the WT and KLT algorithms and was proved to supply better performance for most conditions tested.

6 Conclusions

The main contribution of this paper was the introduction of a new SE algorithm based on imposing constraint on Wiener gain function. Experiments were done on NOIZEUS database for eight kinds of noise (AURORA database) across seven different SNRs ranging from −8 to 15 dB. The Wiener_Clean algorithm was taken as the ground truth. The performance of our proposed algorithm was compared with WT and KLT methods. The results were analyzed mainly by three composite measures and the SNR_{LOSS} measure to predict the performance on subjective quality and speech intelligibility, respectively. Through extensive experiments, we showed that when averaged over all eight kinds of noises, our proposed SE algorithm achieved the best results in terms of predicting signal distortion C_{sig} and overall quality C_{ovl} when SNR is no more than 10 dB. Furthermore, we investigated the individual performance on each noise type. Our proposed SE algorithm outperformed the KLT algorithm for all noise types tested in terms of both C_{sig} and C_{ovl}. On the other hand, the SNR_{LOSS} measure comparisons with both the UP and other SE algorithms predicted that our proposed algorithm was able to improve speech intelligibility for low SNR levels and outperform WT and KLT algorithms for most conditions examined.

It is important to point out that the three composite measures and the SNR_{LOSS} measure used in this paper are adopted for predicting the subjective quality and intelligibility of noisy speech enhanced by noise suppression algorithms because of their high correlation with real subjective tests [42,43]. Further subjective tests on both normal-hearing listeners and hearing-impaired listeners are needed to verify the effectiveness of the proposed algorithm on improving both subjective quality and speech intelligibility. It is also worth mentioning that depending on the nature of the application, some practical SE systems may require very high quality speech but can tolerate a certain amount of noise, while other systems may want speech as clean as possible even with some degree of speech distortion. Therefore, it should be noted that according to different applications, different SE algorithms should be chosen to meet the variant requirement.

Competing interests
The authors declare that they have no competing interests.

Received: 27 December 2013 Accepted: 18 July 2014
Published: 27 August 2014

References

1. Y Hu, PC Loizou, A comparative intelligibility study of single-microphone noise reduction algorithms. J. Acous. Soc. Am. **122**(3), 1777–1786 (2007)
2. MR Schroeder, Apparatus for suppressing noise and distortion in communication signals. U.S Patent 3180936 (April27, 1965)
3. P Scalart, J Vieira-Filho, Speech enhancement based on a priori signal to noise estimation, in *Proceedings of the 21st IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP-96)* (Atlanta Georgia, 1996), pp. 629–632
4. J Chen, J Benesty, Y Huang, New insights into the noise reduction wiener filter. IEEE Trans. Audio Speech Lang. Process. **14**(4), 1218–1234 (2006)
5. S Boll, Suppression of acoustic noise in speech using spectral subtraction. IEEE Trans. Acoust. Speech Signal Proces. **27**(2), 113–120 (1979)
6. M Berouti, R Schwartz, J Makhoul, Enhancement of speech corrupted by acoustic noise, in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 4, (1979), pp. 208–211
7. H Gustafsson, SE Nordholm, I Claesson, Spectral subtraction using reduced delay convolution and adaptive averaging. IEEE Trans. Speech Audio Proces. **9**(8), 799–807 (2001)
8. S Kamath, PC Loizou, A multi-band spectral subtraction method for enhancing speech corrupted by colored noise, in *Student Research Abstracts of Proc. IEEE International Conference On Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 4 (Orlando, FL, USA, 2002), pp. IV-4164
9. Y Hu, PC Loizou, A generalized subspace approach for enhancing speech corrupted by colored noise. IEEE Trans. Speech Audio Proces. **11**(4), 334–341 (2003)
10. F Jabloun, B Champagne, Incorporating the human hearing properties in the signal subspace approach for speech enhancement. IEEE Trans. Speech Audio Proces. **11**(6), 700–708 (2003)
11. Y Ephraim, D Malah, Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator. IEEE Trans. Acoust. Speech Signal Proces. **32**(6), 1109–1121 (1984)
12. Y Ephraim, D Malah, Speech enhancement using a minimum mean-square error log-spectral amplitude estimator. IEEE Trans. Acoust. Speech Signal Proces. **33**(2), 443–445 (1985)
13. PC Loizou, Speech enhancement based on perceptually motivated Bayesian estimators of the magnitude spectrum. IEEE Trans. Speech Audio Proces. **13**(5), 857–869 (2005)
14. JS Lim, AV Oppenheim, Enhancement and bandwidth compression of noisy speech. Proc. IEEE. **67**(12), 1586–1604 (1979)
15. JS Lim, AV Oppenheim, All-pole modeling of degraded speech. IEEE Trans. Acoust. Speech Signal Proces. **26**(3), 197–210 (1978)
16. Y Hu, PC Loizou, Speech enhancement based on wavelet thresholding the multitaper spectrum. IEEE Trans. Speech Audio Proces. **12**(1), 59–67 (2004)
17. Y Hu, PC Loizou, Subjective comparison and evaluation of speech enhancement algorithms. Speech Commun. **49**(7), 588–601 (2007)
18. PC Loizou, *Speech Enhancement: Theory and Practice*. (CRC, Boca Raton, 2007)
19. PC Loizou, G Kim, Reasons why current speech-enhancement algorithms do not improve speech intelligibility and suggested solutions. IEEE Trans. Audio Speech Lang. Process. **19**(1), 47–56 (2011)
20. Y Hu, PC Loizou, Subjective comparison of speech enhancement algorithms, in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 1, (2006), pp. 153–156
21. JS Lim, Evaluation of a correlation subtraction method for enhancing speech degraded by additive white noise. IEEE Trans. Acoust. Speech Signal Proces. **26**(5), 471–472 (1978)
22. PC Loizou, A Lobo, Y Hu, Subspace algorithms for noise reduction in cochlear implants. J. Acoust. Soc. Am. **118**(5), 2791–2793 (2005)
23. G Kim, Y Lu, Y Hu, PC Loizou, An algorithm that improves speech intelligibility in noise for normal-hearing listeners. J. Acoust. Soc. Am. **126**(3), 1486–1494 (2009)
24. EW Healy, SE Yoho, Y Wang, DL Wang, An algorithm to improve speech recognition in noise for hearing-impaired listeners. J. Acoust. Soc. Am. **134**(4), 3029–3038 (2013)
25. K Hu, P Divenyi, D Ellis, Z Jin, BG Shinn-Cunningham, DL Wang, Preliminary intelligibility tests of a monaural speech segregation system, in *ISCA Tutorial and Research Workshop on Statistical and Perceptual Audition (SAPA)*, (2008), pp. 11–16

26. DJ Thomson, Spectrum estimation and harmonic analysis. Proc. IEEE. **70**(9), 1055–1096 (1982)
27. SM Kay, *Modern Spectral Estimation*. (Prentice-Hall, Englewood Cliffs, 1988)
28. KS Riedel, A Sidorenko, Minimum bias multiple taper spectral estimation. IEEE Trans. Signal Process. **43**(1), 188–195 (1995)
29. P Moulin, Wavelet thresholding techniques for power spectrum estimation. IEEE Trans. Signal Process. **42**(11), 3126–3136 (1994)
30. AT Walden, DB Percival, EJ McCoy, Spectrum estimation by wavelet thresholding of multitaper estimators. IEEE Trans. Signal Process. **46**(12), 3153–3165 (1998)
31. AC Cristan, AT Walden, Multitaper power spectrum estimation and thresholding: wavelet packets versus wavelets. IEEE Trans. Signal Process. **50**(12), 2976–2986 (2002)
32. G Wahba, Automatic smoothing of the log periodogram. J. Am. Stat. Assoc. **75**(369), 122–132 (1980)
33. DB Percival, *Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques*. (Cambridge University Press, Cambridge, 1993)
34. MS Bartlett, DG Kendall, The statistical analysis of variance-heterogeneity and the logarithmic transformation. Suppl. J. R. Stat. Soc. **8**(1), 128–138 (1946)
35. SR Quackenbush, TP Barnwell, MA Clements, *Objective Measures of Speech Quality*. (Prentice-Hall, Englewood Cliffs, 1988)
36. DL Donoho, IM Johnstone, Ideal spatial adaptation by wavelet shrinkage. Biometrika. **81**(3), 425–455 (1994)
37. DL Donoho, De-noising by soft-thresholding. IEEE Trans. Inform. Theory. **41**(3), 613–627 (1995)
38. DL Donoho, IM Johnstone, Adapting to unknown smoothness via wavelet shrinkage. J. Am. Stat. Assoc. **90**(432), 1200–1224 (1995)
39. IM Johnstone, BW Silverman, Wavelet threshold estimators for data with correlated noise. J. Roy. Stat. Soc. B. **59**(2), 319–351 (1997)
40. SG Mallat, A theory for multiresolution signal decomposition: the wavelet representation. IEEE Trans. Pattern Anal. Mach. Intell. **11**(7), 674–693 (1989)
41. S Rangachari, PC Loizou, A noise-estimation algorithm for highly non-stationary environments. Speech Commun. **48**(2), 220–231 (2006)
42. Y Hu, PC Loizou, Evaluation of objective measures for speech enhancement, in *Proceedings of INTERSPEECH*, (2006), pp. 1447–1450
43. J Ma, PC Loizou, SNR loss: a new objective measure for predicting the intelligibility of noise-suppressed speech. Speech Commun. **53**(3), 340–354 (2011)
44. Y Hu, PC Loizou, Subjective comparison and evaluation of speech enhancement algorithms. Speech Commun. **49**(7), 588–601 (2007)
45. EH Rothauser, WD Chapman, N Guttman, MHL Hecker, KS Nordby, HR Silbiger, GE Urbanek, M Weinstock, IEEE recommended practice for speech quality measurements. IEEE Trans. Audio Electroacoust. **17**(3), 225–246 (1969)
46. International Telecommunication Union, P.862: Perceptual Evaluation of Speech Quality (PESQ): An Objective Method for End-to-End Speech Quality Assessment of Narrow-Band Telephone Networks and Speech Codecs. (ITU-T Recommendation P. 862, 2000)
47. H Hirsch, D Pearce, The AURORA experimental framework for the performance evaluation of speech recognition systems under noisy conditions, in *ASR-2000* (Paris, France, 2000), pp. 181–188
48. Y Hu, PC Loizou, Evaluation of objective quality measures for speech enhancement. IEEE Trans. Audio Speech Lang. Process. **16**(1), 229–238 (2008)
49. Y Hu, J Ma, PC Loizou, Objective measures for predicting speech intelligibility in noisy conditions based on new band-importance functions. J. Acoust. Soc. Am. **125**(5), 3387–3405 (2009)
50. D Klatt, Prediction of perceived phonetic distance from critical-band spectra: a first step, in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 7, (1982), pp. 1278–1281
51. International Telecommunication Union, ITU-T Recommendation P.835: Subjective Test Methodology for Evaluating Speech Communication Systems that Include Noise Suppression Algorithms. (ITU-T Recommendation P.835, 2003)
52. ANSI, Methods for Calculation of the Speech Intelligibility Index. Technical Report S3.5-1997, American National Standards Institute (1997)

doi:10.1186/s13636-014-0032-7
Cite this article as: Ma and Nishihara: **A modified Wiener filtering method combined with wavelet thresholding multitaper spectrum for speech enhancement.** *EURASIP Journal on Audio, Speech, and Music Processing* 2014 **2014**:32.