Hindawi Publishing Corporation
Journal of Sensors
Volume 2016, Article ID 5352437, 7 pages
http://dx.doi.org/10.1155/2016/5352437

Research Article

Noise Estimation and Suppression Using Nonlinear Function with *A Priori* Speech Absence Probability in Speech Enhancement

Soojeong Lee[1] and Gangseong Lee[2]

[1]*School of Electronic Engineering, Hanyang University, 222 Wangsimni-ro, Seongdong, Seoul 133-791, Republic of Korea*
[2]*Kwangwoon University, 20 Kwangwoon-ro, Nowon-gu, Seoul, Republic of Korea*

Correspondence should be addressed to Soojeong Lee; leesoo86@hanyang.ac.kr

Received 7 December 2015; Accepted 6 April 2016

Academic Editor: Marco Anisetti

This paper proposes a noise-biased compensation of minimum statistics (MS) method using a nonlinear function and *a priori* speech absence probability (SAP) for speech enhancement in highly nonstationary noisy environments. The MS method is a well-known technique for noise power estimation in nonstationary noisy environments; however, it tends to bias noise estimation below that of the true noise level. The proposed method is combined with an adaptive parameter based on a sigmoid function and *a priori* SAP for residual noise reduction. Additionally, our method uses an autoparameter to control the trade-off between speech distortion and residual noise. We evaluate the estimation of noise power in highly nonstationary and varying noise environments. The improvement can be confirmed in terms of signal-to-noise ratio (SNR) and the Itakura-Saito Distortion Measure (ISDM).

1. Introduction

Noise estimation algorithms are essential components of many modern mobile communication, speech recognition, and human computer interaction systems for speech enhancement [1, 2]. It is generally included as a part of the speech enhancement to improve the speech intelligibility or quality of a signal corrupted by noise. However, it is difficult to reduce noise without distorting speech because the performance of any noise estimation algorithm usually depends on a trade-off between speech distortion and noise reduction.

Current single microphone speech enhancement methods belong to two groups, namely, time domain methods such as the subspace method and frequency domain methods such as the spectral subtraction (SS) [3] and minimum mean square error (MMSE) estimator [4]. Both methods have their own advantages and drawbacks. Subspace methods provide a mechanism to control the trade-off between speech distortion and residual noise, but with the cost of a heavy computational load [5]. Frequency domain methods, on the other hand, usually consume less computational resources but do not have a theoretically established mechanism to control trade-off between speech distortion and residual noise. Among them, spectral subtraction (SS) is computationally efficient and has a simple mechanism to control trade-off between speech distortion and residual noise but suffers from a notorious artifact known as musical noise [6]. These spectral noise reduction algorithms require an estimate of the noise spectrum, which can be obtained from speech absence frames indicated by a voice activity detector (VAD) or, alternatively, with the minimum statistic (MS) methods [7], that is, by tracking spectral minima in each frequency band.

Several recent studies have proposed noise estimation schemes for unknown noise signals [1–14]. The minimum statistics (MS) noise estimation scheme [7] is one that works well in nonstationary noisy environments. Martin proposed an algorithm for noise estimation based on minimum statistics [7]. The ability to track varying noise levels is a prominent feature of the minimum statistics (MS) algorithm [7]. The noise estimate is obtained as the minima values of a smoothed power estimate of the noisy signal, multiplied by a factor that compensates the bias. However, the MS algorithm still has

a tendency to bias the noise estimate below that of the true noise level, regardless of the number of frames [8]. Therefore, it leaves residual noise in the frames of speech absence and in the frames of variation of noise characteristic in highly nonstationary noisy environments.

To solve this problem, we propose a combined adaptive factor based on a sigmoid function and *a priori* speech absence probability (SAP) estimation [9] for biased compensation. Specifically, we apply the adaptive factor δ as *a posteriori* SNR. When the *a posteriori* SNR decreases, δ increases but is constrained to take a value between $\delta_{\min}$ and $\delta_{\max}$. Thus, the proposed adaptive biased compensation factor δ approaches $\delta_{\max}$ at times when the SNR is low. In addition, when the *a priori* SAP equals unity, the adaptive biased compensation factor δ also approaches $\delta_{\max}$ in each frequency bin and vice versa. Furthermore, our method uses another adaptive parameter to control the trade-off between speech distortion and residual noise for suppressing the estimated noise in highly nonstationary and various noisy environments. The autocontrol parameter is controlled by *a posteriori* signal-to-noise ratio (SNR) as the variation of the noise level.

We evaluate the performance of the proposed algorithm for nonstationary noise and various noise environments. The improvement can be confirmed in the segmental SNR and the Itakura-Saito Distortion Measure (ISDM) [15]. The results show that our proposed method is superior to the conventional MS approach. The structure of the paper is as follows. Section 2 reviews the minimum statistics and the *a priori* SAP estimation algorithms. Section 3 addresses noise estimation and suppression using a linear and a nonlinear function. In Section 4, we express the combined sigmoid function using the *a posteriori* SNR and *a priori* SAP estimation for robust biased compensation. In Section 5, we discuss the experimental results.

2. Minimum Statistics (MS) and Speech Absence Probability (SAP)

2.1. Review of MS. The noisy speech signal $y(n)$ can be represented as $y(n) = x(n) + d(n)$, where $x(n)$ is the clean speech signal and $d(n)$ is the noise signal. Dividing the signal into overlapping frames using a window function and applying the short-time Fourier transform (STFT) [16] to each frame yield the time-frequency representation $Y(k,l) = X(k,l) + D(k,l)$, where $k = 1, 2, \ldots, K$ is the frequency bin index and $l = 1, 2, \ldots, L$ is the time frame index. It can be shown that

$$\left|Y_k(l)\right|^2 \approx \left|X_k(l)\right|^2 + \left|D_k(l)\right|^2, \tag{1}$$

where $|Y_k(l)|^2$, $|X_k(l)|^2$, and $|D_k(l)|^2$ are the power spectrum of the noisy speech signal, clean speech, and noise, respectively.

The MS algorithm relies on the fact that the noisy power spectrum often becomes equal to the noise power spectrum during periods of speech pauses [7, 13, 17]. Therefore, an estimate of the noise power spectrum is obtained by separately tracking the minimum of the noisy speech in each frequency bin. In addition, because the minimum is biased towards lower values, an unbiased estimate may be obtained through multiplication by a bias factor, which is derived from the statistics of the local minimum. To search for the minimum, we take the first-order recursive of the noisy power spectrum:

$$S_k(l) = \alpha S_k(l-1) + (1-\alpha)\left|Y_k(l)\right|^2, \tag{2}$$

where $S_k(l)$ is the smoothed periodogram and α is the smoothing factor. The smoothing factor used in (2) must be close to 1 to keep the variance of the minimum tracking as small as possible. Hence, time and frequency dependence are required to determine if speech is present or absent. The smoothing factor is therefore derived by minimizing the mean square error between $S_k(l)$ and $\sigma_{d,k}^2(l)$:

$$E\left\{\left(S_k(l) - \sigma_{d,k}^2(l)\right)^2 \mid S_k(l-1)\right\}, \tag{3}$$

where $\sigma_{d,k}^2(l)$ is the noise variance:

$$S_k(l) = \alpha_k(l) S_k(l-1) + \left(1 - \alpha_k(l)\right)\left|Y_k(l)\right|^2. \tag{4}$$

In (4), the time-frequency dependent smoothing factor $\alpha_k(l)$ is used instead of the fixed α defined in (2). Substituting (4) into (3) and setting the first derivative to 0, we find the optimum value for $\alpha_k(l)$

$$\alpha_{\text{opt},k}(l) = \frac{1}{1 + \left(S_k(l-1)/\sigma_{d,k}^2(l) - 1\right)}. \tag{5}$$

According to (5), the smoothing factor can vary between 0 and 1, but such a smoothing factor is not practical [15]. The value of α_{opt} becomes progressively smaller for a large *a posteriori* SNR $\overline{\gamma} \approx (S_k, (l-1)/\sigma_{d,k}^2(l))$ (speech present). However, smoothing is required even during periods of speech because the speech power spectrum also contains a percentage of noise. Hence, the smoothing factor has a floor of (0.3), which results in a maximum of only (70%) of the original spectrum remaining within any one frame. Conversely, when the *a posteriori* SNR $\overline{\gamma}$ is low (speech is absent) α tends towards 1, which causes the smoothed output to lock onto the previous value. To eliminate this, (5) is multiplied by $\alpha_{\max} = 0.96$. From (5), we note that $\alpha_{\text{opt},k}(l)$ depends on the true noise variance $\sigma_{d,k}^2(l)$, which is unknown. In practice, we can replace $\sigma_{d,k}^2(l)$ with the latest estimated value $\widehat{\sigma}_{d,k}^2(l-1)$. In general, however, this lags the true noise variance, and hence the estimated smoothing factor may be too small or large. Problems may arise when $\alpha_{\text{opt},k}(l)$ is close to 1 because $S_k(l)$ will not respond fast enough to changes in the noise. Thus, tracking errors were monitored in [7] by comparing the average short-term smoothed periodogram to the estimated noise variance. After including the correction factor [7]

$$\alpha_c(l) = \frac{1}{1 + \left(\sum_{k=1}^{K} S_k(l-1) \,/\, \sum_{k=1}^{K}\left|Y_k(l)\right|^2 - 1\right)^2}, \tag{6}$$

the final factor

$$\alpha_{\text{opt},k}(l) = \frac{\alpha_{\max} \cdot \alpha_c(l)}{1 + \left(S_k(l-1)/\widehat{\sigma}^2_{d,k}(l-1) - 1\right)^2} \quad (7)$$

is also smoothed over time [7].

The estimated noise power based the MS algorithm [7] is obtained by searching for a minimum within a finite window length C of the smoothed power estimates $P(k,l)$:

$$S_{\min,k}(l) = \min\left\{S_k(l), S_k(l-1), \ldots, S_k(l-C)\right\}. \quad (8)$$

Because the minimum power estimate obtained through the time-varying smoothing factor is smaller than the mean value, the MS algorithm requires a bias compensation for the unbiased noise power estimate as detailed in the following [7]:

$$\widehat{\sigma}^2_{d,k}(l) = \beta_{\min,k}(l) \cdot S_{\min,k}(l), \quad (9)$$

where $\widehat{\sigma}^2_{d,k}(l)$ is the unbiased noise power estimate. The quantity $\beta_{\min,k}(l)$ is the bias compensation factor.

2.2. Review of Speech Absence Probability. The two-state model of speech events can be represented as a binary hypothesis model [9, 15, 17]:

$$\begin{aligned} H_0(k,l) &: Y(k,l) = D(k,l), \\ H_1(k,l) &: Y(k,l) = X(k,l) + D(k,l), \end{aligned} \quad (10)$$

where $H_0(k,l)$ and $H_1(k,l)$ represent the absence and presence of speech, in the kth frequency bin of the lth frame, respectively, and where

$$P\left(H_0(k,l) \equiv q(k,l)\right) \quad (11)$$

is the *a priori* probability that speech will be absent. An efficient estimator is derived for the *a priori* SAP using a soft-decision approach based on the estimated *a priori* SNR [9]. A recursive average of this can be defined as

$$\zeta(k,l) = \beta\zeta(k,l-1) + (1-\beta)\widehat{\xi}(k,l-1), \quad (12)$$

where β is a time constant. The decision-directed method proposed by Ephraim and Malah [4] provides a useful estimation scheme for the *a priori* SNR:

$$\widehat{\xi}(k,l) = a\frac{\widehat{X}^2(k,l-1)}{\widehat{\sigma}^2_d(k,l)} + (1-a)\max\left[\gamma(k,l) - 1, 0\right], \quad (13)$$

where a $(0 < a < 1)$ is a smoothing factor, max is a function that prevents negative values, and $\gamma(k,l) \approx |Y(k,l)|^2/\widehat{\sigma}^2_d(k,l)$ represents the *a posteriori* SNR [9]. The local and global averaging window are then applied to (13) [9], resulting in

$$\zeta_\lambda(k,l) = \sum_{i=-w_\lambda}^{w_\lambda} h_\lambda(i)\zeta(k-i,l), \quad (14)$$

where the subscript λ may denote either "local" or "global" window and h_λ is a normalized window of size $2w_\lambda + 1$. We define two parameters P_{local} and P_{global}, which represent the relationship between the above averages and the likelihood of speech in the kth frequency bin of the lth frame. These parameters are given as [9]

$$P_\lambda(k,l) = \begin{cases} 0, & \text{if } \zeta_\lambda(k,l) \le \zeta_{\min}, \\ 1, & \text{if } \zeta_\lambda(k,l) \ge \zeta_{\max}, \\ \dfrac{\log\left(\zeta_\lambda(k,l)/\zeta_{\min}\right)}{\log\left(\zeta_{\max}/\zeta_{\min}\right)}, & \text{otherwise}, \end{cases} \quad (15)$$

where $\zeta_{\min}$ and $\zeta_{\max}$ are empirical constants, maximized to attenuate noise while leaving weak speech components unaffected. The third parameter $P_{\text{frame}}(l)$, which is required to attenuate more noise in speech-absent frames, is based on the speech energy in neighboring frames [9]:

If $\zeta_{\text{frame}}(l) > \zeta_{\min}$ then

if $\zeta_{\text{frame}}(l) > \zeta_{\text{frame}}(l-1)$ then

$P_{\text{frame}}(l) = 1$

$\zeta_{\text{peak}}(l) = \min\{\max[\zeta_{\text{frame}}, \zeta_{p\min}], \zeta_{p\max}\}$

else

$P_{\text{frame}}(l) = \mu(l)$

Else

$P_{\text{frame}}(l) = 0,$

where $\zeta_{\text{frame}}(l) = (1/(K/2+1))\sum_{k=1}^{K/2+1}\zeta(k,l)$ is an average in the frequency domain, $\mu(l)$ represents a soft transition from speech to noise, ζ_{peak} is a confined peak value of ζ_{frame}, and $\zeta_{p\min}$ and $\zeta_{p\max}$ are empirical constants that determine the delay of the transition, as defined in [9]. Finally, the *a priori* SAP can be defined as [9]

$$\widehat{q}(k,l) = 1 - P_{\text{local}}(k,l) \cdot P_{\text{global}}(k,l) \cdot P_{\text{frame}}(l). \quad (16)$$

Accordingly, $\widehat{q}(k,l)$ is larger if either previous frames or recent neighboring frequency bins do not contain speech. Therefore, when SAP goes to 1, the speech presence probability goes to 0.

3. Noise Estimation and Suppression Using Linear and Nonlinear Function

3.1. Combining Adaptive Factor Based on Sigmoid Function and A Priori SAP. In this section, we propose a method that combines the adaptive factor based on the sigmoid function and the *a priori* SAP estimation [9] to achieve biased compensation.

First, we can detect the adaptive factor by requiring the smoothed power spectrum $P(k,l)$ be equal to the updated noise power estimator $\overline{\sigma}^2_d$ during speech absence region. In particular, we can determine the adaptive factor by minimizing the mean squared error (MSE) between $P(k,l)$ and $\overline{\sigma}^2_d$ as follows:

$$E\left\{\left(P(k,l) - \overline{\sigma}^2_d(k,l)\right)^2 \mid P(k,l-1)\right\}, \quad (17)$$

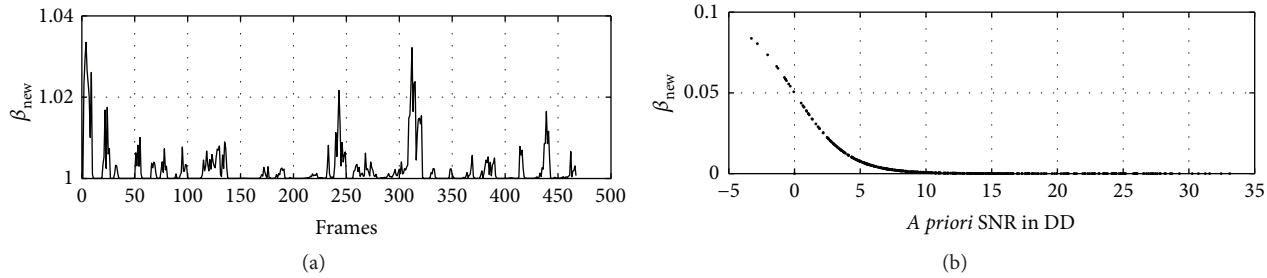

FIGURE 1: (a) Plot of the adaptive factor δ in the frame index. (b) Adaptive factor δ using a sigmoid function based on the *a posteriori* SNR.

where we assume that the updated noise power estimator $\overline{\sigma}_d^2$ during the speech absence region is

$$\overline{\sigma}_d^2(k,l)^2 \approx \widehat{\sigma}_d^2(k,l)^2 + \delta(k,l). \tag{18}$$

Substituting (18) into (17) then after taking the first derivative of the MSE with respect to $\delta(l)$ and setting it equal to zero, we get the adaptive factor for $\delta(l)$:

$$\delta(l) = P(k,l) - \widehat{\sigma}_d^2(k,l), \tag{19}$$

where $\widehat{\sigma}_d^2$ is the unbiased noise power estimate in (9). We apply the adaptive factor based on the sigmoid function to the biased compensation factor of the MS algorithm according to the *a posteriori* SNR:

$$\delta(l) \approx \eta \cdot \frac{1}{\left(1 + \exp\left(-\left(-\rho \cdot \text{SNR}(l)\right)\right)\right)}, \tag{20}$$

where $\delta(l)$ is derived from the slope factor $\rho = 0.5$ and the empirical constant $\eta = 0.1$ for $\delta_{\max}$. The *a posteriori* SNR is

$$\text{SNR}(l) = 10 \cdot \log\left(\frac{\left\|\left|Y(k,l)\right|^2\right\|}{\left\|\widehat{\sigma}_d^2(k,l)\right\|}\right), \tag{21}$$

where $\|\cdot\|$ is the Euclidean length of a vector. The adaptive factor $\delta(l)$ is controlled by the *a posteriori* SNR. When the *a posteriori* SNR decreases, $\delta(l)$ increases but is constrained to take a value between $\delta_{\min}$ and $\delta_{\max}$. Thus, the proposed adaptive biased compensation factor $\delta(l)$ approaches $\delta_{\max}$ at times when the SNR is low. In addition, when the *a priori* SAP equals unity, the adaptive biased compensation factor $\delta(l)$ is also equal to $\delta_{\max}$ in each frequency bin and vice versa. The adaptive factor is shown to be a biased compensation in Figure 1. It shows, as suggested by (20) and (21), that as the *a posteriori* SNR increases, $\delta(l)$ decreases but $\delta(l)$ maintains a value between $\delta_{\max}$ ($\delta_{\max} \ll 0.1$) and $\delta_{\min}$. Thus, the adaptive factor $\delta(l)$ approaches $\delta_{\min}$ when the SNR is close to 20 dB. Simulation results show that an increase in the $\delta(l)$ is good for noisy signals with a low SNR of less than 5 dB and that a decrease in $\delta(l)$ is good for noisy signals with a relatively high SNR greater than 10 dB. We can thus control the trade-off between speech distortion and residual noise in the frame index using $\delta(l)$. In (22), let $\overline{\sigma}_d^2(k,l)$ be the updated noise power estimate according to the combined *a priori* SAP and the adaptive factor:

$$\overline{\sigma}_d^2(k,l) = \widehat{\sigma}_d^2(k,l) + \delta(l) \cdot \widehat{q}(k,l). \tag{22}$$

The term $\widehat{q}(k,l)$ is the *a priori* SAP in (16). When $\widehat{q}(k,l)$ becomes 1, the adaptive biased compensation factor $\delta(l)$ is equal to $\delta_{\max}$. Therefore, the speech absence region is efficiently compensated by combining the *a priori* SAP and the adaptive factor in the kth frequency bin of the lth frame. As a result, the updated noise power estimator for the optimal smoothing factor $\widehat{\alpha}_{\text{opt}}(k,l)$ of $\alpha_{\text{opt}}(k,l)$ is deduced from (7) as

$$\widehat{\alpha}_{\text{opt}}(k,l) = \frac{\alpha_{\max} \cdot \alpha_c(l)}{1 + \left(P(k,l-1)/\overline{\sigma}_d^2(k,l-1) - 1\right)^2}. \tag{23}$$

3.2. Estimated Noise Suppression Using Linear Function. In this subsection, our method uses another adaptive parameter to control the trade-off between speech distortion and residual noise for suppressing the estimated noise in a highly nonstationary and varying noisy environment. The autocontrol parameter is controlled by *a posteriori* signal-to-noise ratio (SNR) as the variation of the noise level.

The estimated clean speech power spectrum can be represented as shown in (28). One has

$$\text{SNR}(k) = \log\left(\frac{P(k,l)}{\overline{\sigma}_d^2(k,l)}\right), \tag{24}$$

$$\zeta_s = \frac{\zeta_{\min} - \zeta_{\max}}{\text{SNR}_{\max} - \text{SNR}_{\min}}, \tag{25}$$

$$\zeta_o = \zeta_{\max} - \zeta_s \cdot \text{SNR}_{\max}, \tag{26}$$

$$\zeta(k) = \zeta_s \cdot \text{SNR}(k) + \zeta_o, \tag{27}$$

$$\left|\widehat{X}(k,l)\right|^2 \approx |Y(k,l)|^2 - \zeta(k) \cdot \overline{\sigma}_d^2(k,l), \tag{28}$$

where $\zeta(k)$ is the oversubtraction factor, ζ_s is the slope, and ζ_o is the offset. The constants $\zeta_{\min} = 1$, $\zeta_{\max} = 3$, $\text{SNR}_{\max} = 20$ dB, and $\text{SNR}_{\min} = -5$ dB, respectively [3]. The adaptive linear factor $\zeta(k)$ affects the amount of speech distortion caused by the spectral subtraction in (28). The factor $\zeta(k)$ offers a large amount of flexibility to the modified spectral subtraction (MSS) scheme. The SNR(k) in (24) is the *a posteriori* SNR in frequency bin. The estimated clean speech signal can then be transformed back to the time domain by taking the inverse STFT and synthesizing using the overlap-add method.

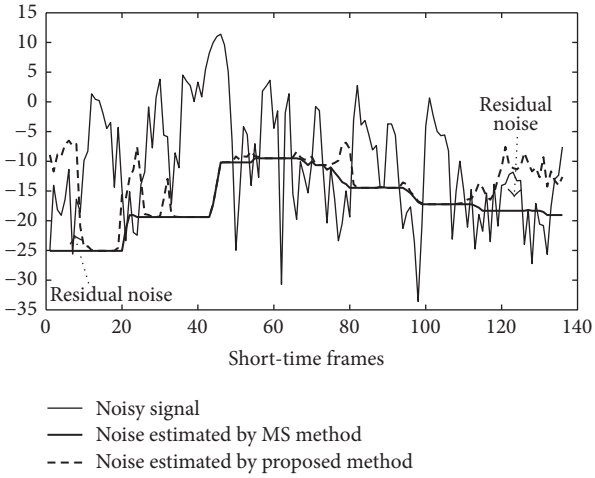

Figure 2: Comparison between the noisy signal, noise estimated by MS, and noise estimated by the proposed method in restaurant 5 dB noisy environment.

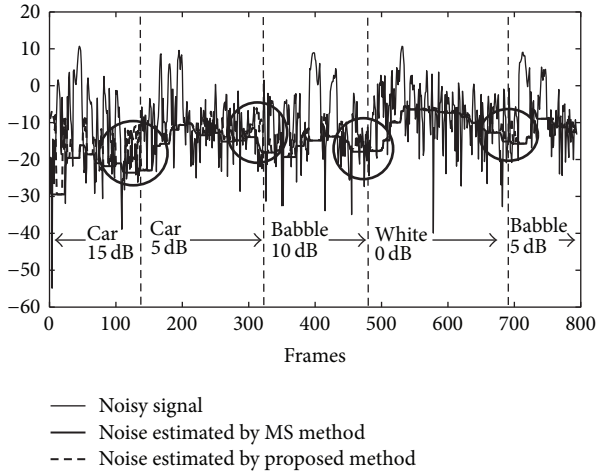

Figure 3: Comparison between the noisy signal, noise estimated by minimum statistics (MS), and noise estimated by the proposed method in highly nonstationary noisy environments.

4. Experimental Results and Discussion

The noisy signals used in our evaluation were taken from the NOIZEUS database [15]. We used 30 test utterances, of which three each were from male and female speech signals. The analyzed signal was sampled at 8 kHz and short-time Fourier-transformed using 50% overlapping Hamming windows of 256 samples. Both the MS [7] and proposed methods track the minimum of the noisy speech to update the noise estimate in Figure 2. The MS method is obtained by tracking the minimum of the noisy power spectrum over a specified number of frames. Thus, the MS algorithm noise estimate tends to be biased below the true noise level, regardless of the number of frames. Our proposed method efficiently compensates the speech absence region by combining the adaptive bias compensation factor and *a priori* SAP. This implies that the proposed method is more accurate than the conventional one and could improve residual noise reduction.

Figure 3 shows the clear superiority of the proposed method in highly nonstationary noisy environments. The conventional method [7] does not work well from initial frame to 20 frames of car noise (15 dB) and from 110 frames to 130 frames of car (15 dB) and also suffered from residual noise. A different outcome is observed in the red circle of Figure 3. Particularly, the robust characteristics of the proposed method in spite of the variation of the noisy environments are well demonstrated. Thus, we can estimate more exactly the noise level to reduce a residual noise when compared with conventional method in highly nonstationary noisy environments.

The spectrum of the clean signal is given in Figure 4(a), and the spectrum of the noisy speech signal for speech enhancement using the MS plus spectral subtraction (SS) (MS + SS) [3, 7] method is given in Figure 4(b). We can also observe the minimum controlled recursive averaging (MCRA) with SS in Figure 4(c). There is residual noise in Figure 4(c) from 0 s $< t <$ 0.15 s and at $t >$ 1.8 s, partly because of the inability of the noise estimation algorithm to bias below the true noise level. The spectrogram of the proposed methods for noise reduction is shown in Figure 4(d). In contrast, panel Figure 4(d) shows that the residual noise is more clearly reduced than the conventional methods.

Table 1: Objective evaluation and comparison of the proposed method segmental SNR values.

Noise	Method	SNR			
		0 (dB)	5 (dB)	10 (dB)	15 (dB)
White	MS	4.27	8.77	12.83	16.57
	MCRA	5.08	9.99	13.56	17.15
	Proposed	**5.69**	**10.64**	**14.09**	**17.40**
Car	MS	3.44	7.48	12.01	16.10
	MCRA	4.92	7.93	11.85	16.42
	Proposed	**5.39**	**8.84**	**12.45**	**17.06**
Babble	MS	1.83	6.00	11.17	15.03
	MCRA	3.73	6.79	10.52	16.22
	Proposed	**3.83**	**7.05**	**11.60**	**16.23**
Airport	MS	1.75	7.04	9.85	14.73
	MCRA	1.64	7.54	9.66	15.15
	Proposed	**2.17**	**8.16**	**10.10**	**15.73**
Street	MS	2.77	6.75	**10.88**	**15.31**
	MCRA	2.34	7.40	9.88	14.36
	Proposed	**3.18**	**8.24**	10.62	15.03
Restaurant	MS	0.31	4.48	9.40	14.74
	MCRA	0.27	5.47	9.20	15.24
	Proposed	**0.28**	**5.57**	**9.43**	**15.58**

Tables 1 and 2 summarize the averaged results of the segmental SNR and the Itakura-Saito Distortion Measure (ISDM) [15]. The segmental SNR can be evaluated in either the time or frequency domain. The time domain measure is perhaps one of the simplest objective measures used to evaluate speech enhancement method. For this measure to

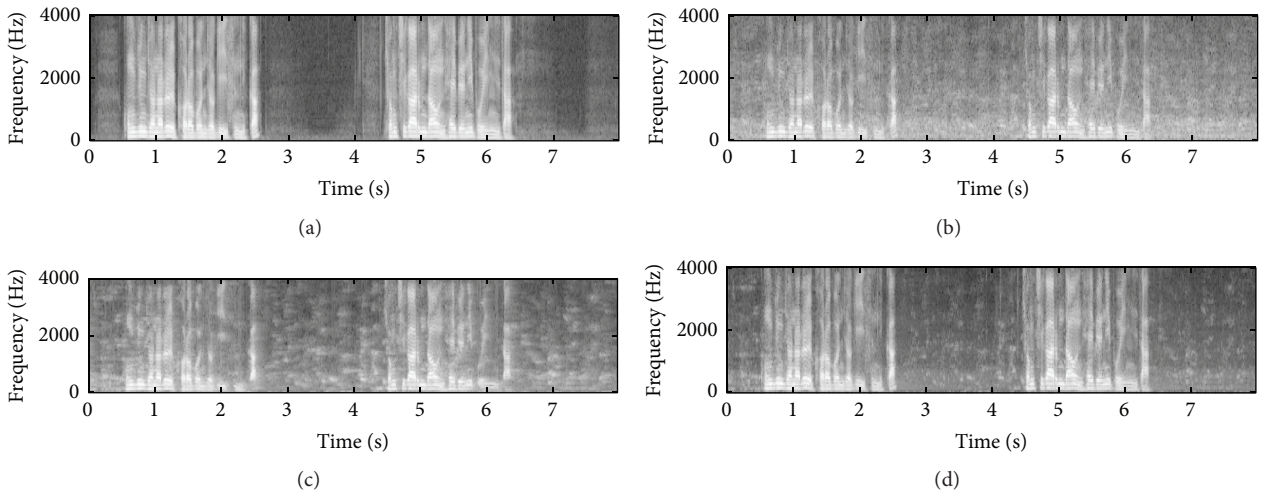

Figure 4: Frequency domain results of speech enhancement for exhibition noise 5 dB SNRs in noisy environments. (a) Original spectrogram, (b) spectrogram using MS with SS method, (c) spectrogram using the MCRA with SS method, and (d) spectrogram using the proposed method.

Table 2: Objective evaluation and comparison of the Itakura-Saito Distortion Measure (ISDM).

Noise	Method	ISDM			
		0 (dB)	5 (dB)	10 (dB)	15 (dB)
White	MS	1.20	0.87	0.60	0.42
	MCRA	0.92	0.44	0.55	0.37
	Proposed	**0.84**	**0.43**	**0.38**	**0.34**
Car	MS	0.15	0.16	0.02	0.01
	MCRA	0.21	0.05	0.02	0.01
	Proposed	**0.09**	**0.02**	0.02	0.02
Babble	MS	0.15	0.06	0.02	0.01
	MCRA	0.12	0.02	0.01	0.01
	Proposed	**0.08**	0.02	0.01	0.01
Airport	MS	0.19	0.06	0.02	0.02
	MCRA	0.14	0.04	0.02	0.01
	Proposed	**0.10**	0.04	**0.01**	0.01
Street	MS	0.16	0.18	**0.06**	**0.03**
	MCRA	0.55	0.13	0.09	0.05
	Proposed	0.17	**0.09**	0.08	0.04
Restaurant	MS	0.10	0.05	0.01	0.01
	MCRA	0.10	0.03	0.01	0.01
	Proposed	0.10	**0.01**	0.01	0.01

be meaningful it is important that the original and processed signals be aligned in time and that any phase error present be corrected [15]. For various noise types with an input SNR ranging from 0 to 15 dB, the segmental SNR after processing was clearly better for the proposed method compared to conventional ones [7], except for the case of (highlighted in bold). We can also confirm that our methods work well to control the trade-off between speech distortion and residual noise for suppressing the estimated noise in highly nonstationary and various noisy environments.

The ISDM was shown to give a good correlation with subjective intelligibility measures specifically the diagnostic acceptability measure (DAM). This results in an objective test that can be used to produce a good meaningful result. This also results in a test that shows the distortion and noise reduction [15]. Here, we can confirm that the results of the ISDM with the proposed method produce good results of ISDM when compared with the conventional methods except for the case of the the MS method with SS in street 10 dB noisy signal.

5. Conclusion

We presented a modified noise estimation and suppression algorithm that combined the nonlinear function and *a priori* SAP estimation for biased compensation. Moreover, our method uses another adaptive parameter to control the trade-off between speech distortion and residual noise for suppressing the estimated noise in highly nonstationary and various noisy environments. The performance of the new algorithm was evaluated by measuring the segment SNR and the ISDM. We showed that the proposed algorithm was generally superior to conventional methods, reducing both residual noise and speech distortion in nonstationary and noisy environments. In the future, we plan to evaluate its possible application in preprocessing for signal processing area.

Competing Interests

The authors declare no competing interests.

Acknowledgments

This research was supported by NRF (2013R1A1A2012536).

References

[1] S. Lee, C. Lim, and J.-H. Chang, "A new a priori SNR estimator based on multiple linear regression technique for speech enhancement," *Digital Signal Processing*, vol. 30, no. 7, pp. 154–164, 2014.

[2] S. Lee and J. Chang, "On using multivariate polynomial regression model with spectral difference for statistical model-based speech enhancement," *Journal of Systems Architecture*, In press.

[3] M. Berouti, M. Schwartz, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 208–211, Washington, DC, USA, April 1979.

[4] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 32, no. 6, pp. 1109–1121, 1984.

[5] Y. Hu, *Subspace and multitaper methods for speech enhancement [Ph.D. dissertation]*, University of Texas at Dallas, Richardson, Tex, USA, 2003.

[6] O. Cappe, "Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor," *IEEE Transactions on Speech and Audio Processing*, vol. 2, no. 2, pp. 345–349, 1994.

[7] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 5, pp. 504–512, 2001.

[8] S. Rangachari and P. C. Loizou, "A noise-estimation algorithm for highly non-stationary environments," *Speech Communication*, vol. 48, no. 2, pp. 220–231, 2006.

[9] I. Cohen, "Optimal speech enhancement under signal presence uncertainty using log-spectral amplitude estimator," *IEEE Signal Processing Letters*, vol. 9, no. 4, pp. 113–116, 2002.

[10] I. Cohen, "Noise spectrum estimation in adverse environments: improved minima controlled recursive averaging," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 5, pp. 466–475, 2003.

[11] L. Lin, W. H. Holmes, and E. Ambikairajah, "Adaptive noise estimation algorithm for speech enhancement," *Electronics Letters*, vol. 39, no. 9, pp. 754–755, 2003.

[12] S. J. Lee and S. H. Kim, "Noise estimation based on standard deviation and sigmoid function using a posteriori signal to noise ratio in nonstantionary noisy environments," *International Journal of Control, Automation, and Systems*, vol. 6, no. 6, pp. 818–827, 2008.

[13] R. Martin, "Bias compensation methods for minimum statistics noise power spectral density estimation," *Signal Processing*, vol. 86, no. 6, pp. 1215–1229, 2006.

[14] D. Mauler and R. Martin, "Improved reproduction of stops in noise reduction systems with adaptive windows and nonstationarity detection," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, Article ID 469480, 17 pages, 2009.

[15] P. C. Loizou, *Speech Enhancement: Theory and Practice*, CRC Press, Boca Raton, Fla, USA, 2007.

[16] J. R. Deller, J. H. L. Hansen, and J. G. Proakis, *Discrete Time Processing of Speech Signals*, IEEE Press, New York, NY, USA, 1999.

[17] Y.-S. Park and J.-H. Chang, "A probabilistic combination method of minimum statistics and soft decision for robust noise power estimation in speech enhancement," *IEEE Signal Processing Letters*, vol. 15, pp. 95–98, 2008.