THE JOURNAL OF THE ACOUSTICAL SOCIETY OF KOREA VOL.26, NO.3E 2007, 12 pp. 87-96

Noise Reduction Using the Standard Deviation of the Time-Frequency Bin and Modified Gain Function for Speech Enhancement in Stationary and Nonstationary Noisy Environments

Soojeong Lee*, Soonhyob Kim*
*Department of Computer Engineering, Kwangwoon University
(Received October 29 2007; Revised November 28 2007; Accepted December 14 2007)

Abstract

In this paper we propose a new noise reduction algorithm for stationary and nonstationary noisy environments. Our algorithm classifies the speech and noise signal contributions in time-frequency bins, and is not based on a spectral algorithm or a minimum statistics approach. It relies on calculating the ratio of the standard deviation of the noisy power spectrum in time-frequency bins to its normalized time-frequency average. We show that good quality can be achieved for enhancement speech signal by choosing appropriate values for δ_t and δ_f. The proposed method greatly reduces the noise while providing enhanced speech with lower residual noise and somewhat higher mean opinion score (MOS), background intrusiveness (BAK) and signal distortion (SIG) scores than conventional methods.
Keywords: *Speech enhancement, Noise reduction, Noise estimator*

I. Introduction

The noise estimation algorithm is an essential component of many modern communications systems. Generally included as part of the noise reduction component, it improves the performance of the system by improving the speech quality or intelligibility for signals corrupted by noise. Since it is difficult to reduce noise without distorting the speech, the performance of any noise estimation algorithm is usually a trade-off between speech distortion and noise reduction [1].

The spectral subtraction (SS) method is one of the best-known techniques for noise reduction [2]. It is computationally efficient and has a simple mechanism to control the trade-off between speech distortion and residual noise, although it does suffer from a notorious artifact known as "musical noise" [3, 4]. The minimum mean square error (MMSE) [5] class of estimators and the Wiener estimator present a moderate computational load, but have no mechanism to control the balance between speech distortion and residual noise [3, 4]. The common feature of all these methods is that they first estimate the spectrum of the noise during nonspeech periods. This is valid for the case of stationary noise in which the noise spectrum does not vary much over time. However, it is much less effective for nonstationary noise in which the noisy power spectrum varies during speech. In addition, voice activity detectors are generally difficult to tune and very unreliable for low signal-to-noise ratios (SNRs) [6, 7].

Several recent studies have proposed noise estimation techniques [6-9] designed for unknown nonstationary noise signals using minimum statistics (MS). The ability

Corresponding author: Soojeong Lee (leesoo86@kw.ac.kr)
Department of Computer Engineering, Kwangwoon University.

to track varying noise levels is a prominent feature of such methods. Martin [6] proposed an algorithm in which the noise estimate is obtained as the minimum value of a smoothed power estimate of the noisy signal, multiplied by a factor that compensates the bias. The main drawback of this method is that it takes somewhat more than the duration of the minimum-search windows to update the noise spectrum when the noise level increases suddenly [7]. Cohen [9] proposed a minima controlled recursive algorithm (MCRA), which updates the noise estimate by tracking the noise-only regions of the noisy speech spectrum. These regions are found by comparing the ratio of the noisy speech to the local minimum against a threshold. However, the noise estimate introduces a delays of at most twice that window length when the noise spectrum increases suddenly [7]. A disadvantage to most of the noise-estimation schemes mentioned is that residual noise is still present in frames in which speech is absent. In addition, the conventional noise estimation algorithms are combined with a noise reduction algorithm such as the SS and MMSE [2, 5].

In this paper, we describe a method to enhance speech by improving its overall quality while minimizing residual noise. The proposed algorithm is based on calculating the ratio of the standard deviation (STD) of the noisy power spectrum in the time-frequency bin to its normalized time-frequency average and a sigmoid function (NTFAS). This technique, which we call the "NTFAS noise reduction algorithm", determines that speech is present only if the ratio is greater than the adaptive threshold using the sigmoid function. In the case of a region where a strong speech signal is present, the ratio of STD will be high. This is not true for a region without a speech signal. Specifically, our method uses an adaptive scheme for tracking the threshold in a nonstationary noisy environment to control the trade-off between speech distortion and residual noise.

The estimated clean speech power spectrum is obtained by the modified gain function and the updated noisy power spectrum of the time-frequency bin. We tested the algorithm's performance with the

[10] database, using the segment signal-to-noise ratio (SNR) and ITU-T P.835 [11] as evaluation criteria. We also examined its adaptive tracking capability in nonstationary environments. We show that the performance

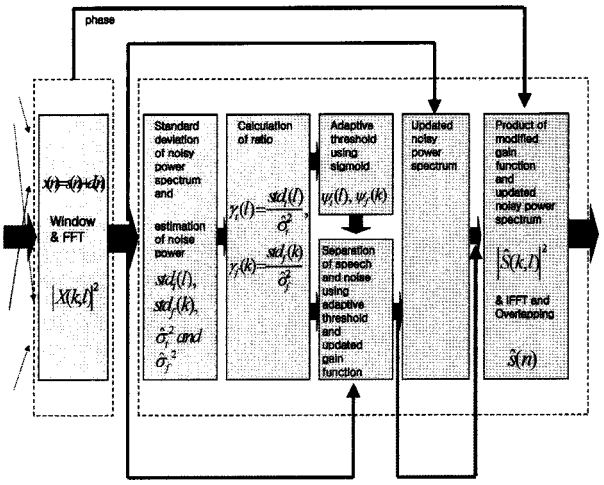

Fig. 1. Flow diagram of proposed speech enhancement algorithm.

of the proposed algorithm is superior to that of the conventional methods. Moreover, this algorithm produces a significant reduction in residual noise ("musical") noise.

The structure of the paper is as follows. Section 2 introduces the overall signal model. Section 3 describes the proposed noise reduction algorithm, while Section 4 contains the experimental results and discussion. The conclusion in Section 5 looks at future research directions for the algorithm.

II. System model

Assuming that speech and noise are uncorrelated, the noisy speech signal $x(n)$ can be represented as

$$x(n) = s(n) + d(n) \tag{1}$$

where $s(n)$ 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) to each frame gives the time-frequency representation $X(k,l) = S(k,l) + D(k,l)$, where k is the frequency bin index and l is the frame index [12]. The power spectrum of the noisy speech $|X(k,l)|^2$ can then be represented as

$$|X(k,l)|^2 = |S(k,l)|^2 + |D(k,l)|^2 \tag{2}$$

where $|S(k,l)|^2$ is the power spectrum of the clean speech signal and $|D(k,l)|^2$ is the power spectrum of

the noise signal. The proposed algorithm is summarized in the block diagram shown in Fig. 1. It consists of seven main components: window and fast Fourier transform (FFT), standard deviation (STD) of the noisy power spectrum and estimation of noise power, calculation of the ratio, adaptive threshold using a sigmoid function, separation of speech presence and absence in time-frequency bins and updated gain function, updated noisy power spectrum, and product of the modified gain function and updated noisy power spectrum.

III. Proposed noise reduction algorithm

The noise reduction algorithm is based on the STD of the noisy power spectrum in a time and frequency-dependent manner as follows:

$$\overline{x}_t(l) = \frac{1}{L}\sum_{k=1}^{L}|X(k,l)|^2 \tag{3}$$

$$\overline{x}_f(k) = \frac{1}{M}\sum_{l=1}^{M}|X(k,l)|^2 \tag{4}$$

$$std_t(l) = \sqrt{\frac{1}{L}\sum_{k=1}^{L}(x_k - \overline{x}_t(l))^2} \tag{5}$$

$$std_f(k) = \sqrt{\frac{1}{M}\sum_{l=1}^{M}(x_l - \overline{x}_f(k))^2} \tag{6}$$

$$\sigma_t = \frac{1}{M}\sum_{l=1}^{M} std_t(l) \tag{7}$$

$$\sigma_f = \frac{1}{L}\sum_{k=1}^{L} std_f(k) \tag{8}$$

$$\gamma_t(l) = \frac{std_t(l)}{\sigma_t} \tag{9}$$

$$\gamma_f(k) = \frac{std_f(k)}{\sigma_f} \tag{10}$$

where $\overline{x}_t(l)$ is the average noisy power spectrum in the frequency bin, $\overline{x}_f(k)$ is the average noisy power spectrum for the frame index, and σ_t and σ_f are the assumed of noise power estimates. Fig. 2 shows the procedure for estimating the noise power using the noisy power spectrum. Eq. 9 and 10 give the ratio of the (STD) for the noisy power spectrum in the time-frequency bin to its normalized time-frequency average. In the case of a region in which a strong speech signal is present, the STD ratio calculated by Eq. 9 and 10 will be high. This is generally not true for a region without a speech signal. Therefore, we can use the ratio in Eq. 9 and 10 to determine speech-presence or speech-absence in the time-frequency bins.

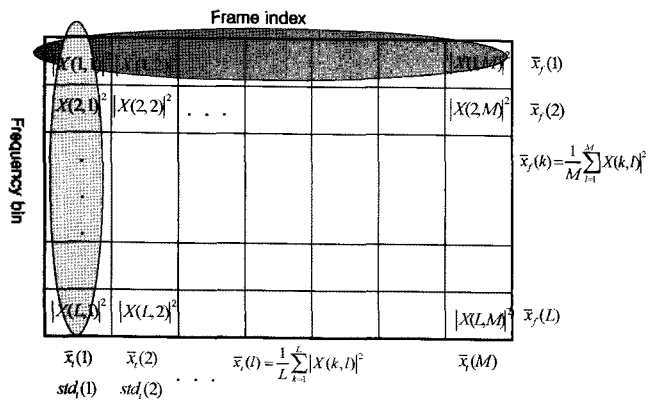

Fig. 2. Procedure for estimating noise power using the noisy power spectrum.

3.1. Separation of speech and noise in frames using an adaptive sigmoid function

Our method uses an adaptive algorithm with a sigmoid function to track the threshold and control the trade-off between speech distortion and residual noise:

$$\Psi_t(l) = \left[\frac{1}{1+\exp(10*(\gamma_t(l)-\delta_t))}\right] \tag{11}$$

where $\psi_t(l)$ is the adaptive threshold using the sigmoid function and δ_t is a defined control parameter. This threshold $\psi_t(l)$ is adaptive in the sense that it changes depending on the control parameter δ_t.

Figure 3 shows the effect of δ_t on SNR gains. The output SNR is calculated in a manner similar to the input SNR. The noise power is calculated as the power of the speech signal obtained by subtracting the filtered speech signal from the clean speech signal. Simulation results

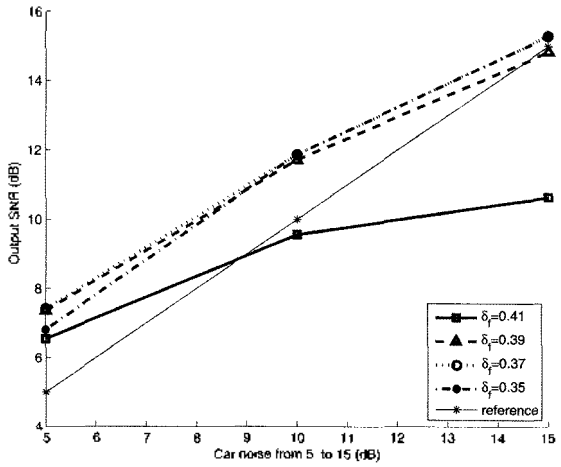

Fig. 3. Effect of various δ_t values on SNR gains.

show that an increase in the δ_t parameter is good for noisy signals with a low SNR of less than 5 dB, and that a decrease in δ_t is good for noisy signals with a relatively high SNR of greater than 15 dB. The δ_t parameter is set to a constant of 0.5 based on initial experiments, but a fixed δ_t will clearly not be optimal over a wide range of SNRs. For example, setting δ_t to 0.475 yields high SNR gain at a low input SNR of 5 dB; however, it also degrades the input speech signal at a high SNR of 15 dB. Distortion of the original speech signal is extremely undesirable in real practical environments. Second, Fig. 4 shows the effect of δ_t on signal distortion (SIG) scores. Simulation results show that the increase in δ_t is somewhat beneficial for noisy signals with low SNRs about 5 dB and high SNRs of about 15 dB. We can thus control the trade-off between speech distortion and residual noise in the frame index using δ_t. Fig. 5 shows that the adaptive threshold using the sigmoid function allows for a trade-off between speech distortion and residual noise by controlling δ_t. If a speech signal is present, the $\psi_t(l)$ calculated by Eq. 11

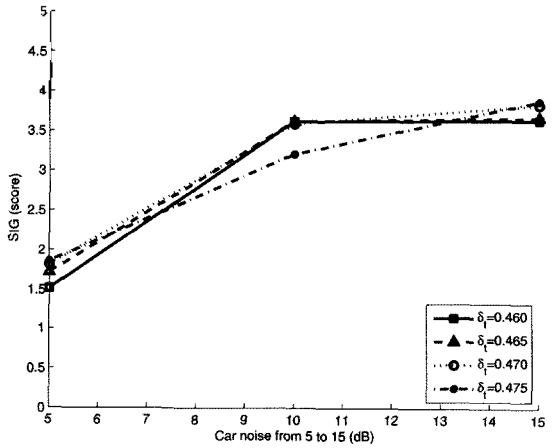

Fig. 4. Effect of various δ_t values on SIG.

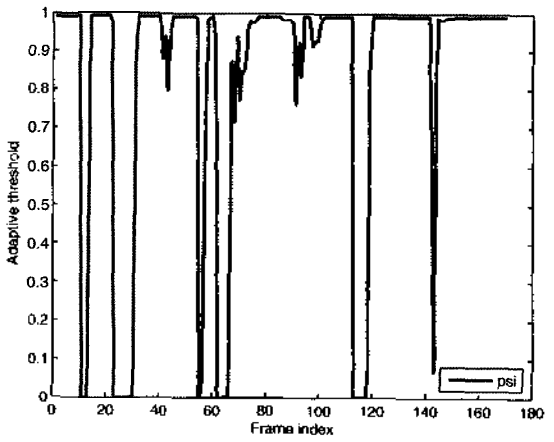

Fig. 5. Adaptive thresholds using a sigmoid function on the time index for car noise 5 dB.

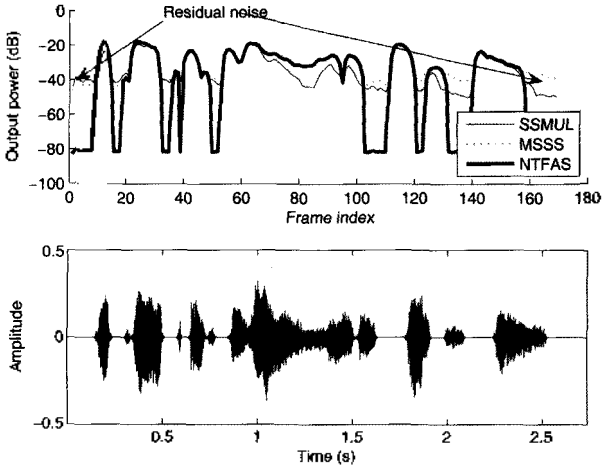

Fig. 6. Example of noise reduction by three enhancement algorithms with 5dB car noise for the sp12.wav female speech sample of "The drip of the rain made a pleasant sound" from the NOIZEUS database. Top panel: output power for car noise 5dB using the SSMUL method (solid line), the MSSS method (dotted line), and NTFAS method (heavy line). Bottom panel : enhanced speech signal using NTFAS.

will be extremely small (i.e., very close to 0). Otherwise if speech is absent, the value of $\psi_t(l)$ calculated by Eq. 11 will be approximately 1. Fig. 6 is a good illustration of Fig. 5.

3.2. Updated noisy power spectrum using separation of speech-presence and absence in frames

The separation rule for determining whether speech is present or absent in a frame is based on the following algorithm:

$$If\ \psi_t(l) > \phi_t \tag{12}$$

$$\widehat{D^2_{level}}(k,l) = |X(k,l)|^2 \tag{13}$$

$$\widehat{D^2_{mean}}(k,1) = \frac{1}{M}\sum_{l=1}^{M}\widehat{D^2_{level}}(k,l) \tag{14}$$

$$G_{update}(k,l) = G(k,l) * \alpha \tag{15}$$

else

$$\widehat{D^2_{level}}(k,l) = \widehat{D^2_{mean}}(k,1) \tag{16}$$

$$G_{update}(k,l) = G(k,l) * (1-\alpha) \tag{17}$$

where decision parameter ϕ_t and constant α are initially 0.99 and the gain function $G(k,l)$ is 1.0. The threshold $\psi_t(l)$ is compared to the decision parameter ϕ_t. If it is greater than ϕ_t, then speech is declared to be absent in the l frames; otherwise speech is present. Then, the l

frames of the noisy spectrum $|X(k,l)|^2$ are set to $\widehat{D^2_{level}}(k,l)$. We estimate $\widehat{D^2_{level}}(k,l)$ frames of the noise power spectrum, and $\widehat{D^2_{mean}}(k,1)$ is calculated by averaging over the frames without speech. The $\widehat{D^2_{mean}}(k,1)$ is the assumed estimate of the residual noise of the frames in the presence of speech. We refer to this value as the "sticky noise" of the speech-presence index. Then we represent $G_{update}(k,l)$, the updated gain function in a frame index using the gain function $G(k,l)$ and the constant α for the frames in which speech is absent. If the l frames are considered to be frames in which speech is present, then $\widehat{D^2_{mean}}(k,1)$ is set to $\widehat{D^2_{level}}(k,l)$, and $\widehat{D^2_{mean}}(k,1)$ is used to reduce the sticky noise of the frames of in the presence of speech. We can see the sticky noise in the the square region and residual noise in the random peak region in Fig. 7.

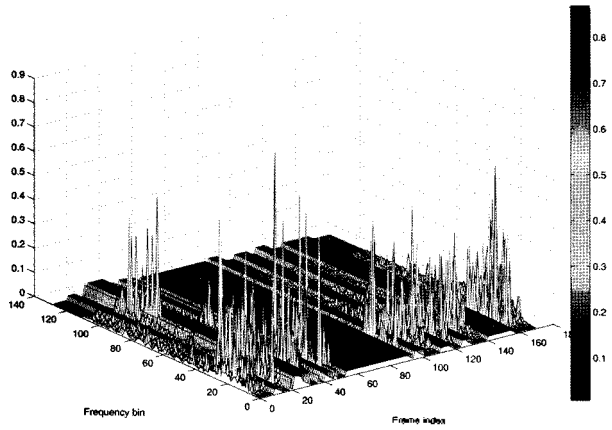

Fig. 7. Estimated noise power spectrum at car noise 10 dB sp12.wav of female "The drip of the rain made a pleasant sound" from the NOIZEUS database.

As a noted above, $G_{update}(k,l)$ is the updated gain function in a frame index using the gain function $G(k,l)$ and the

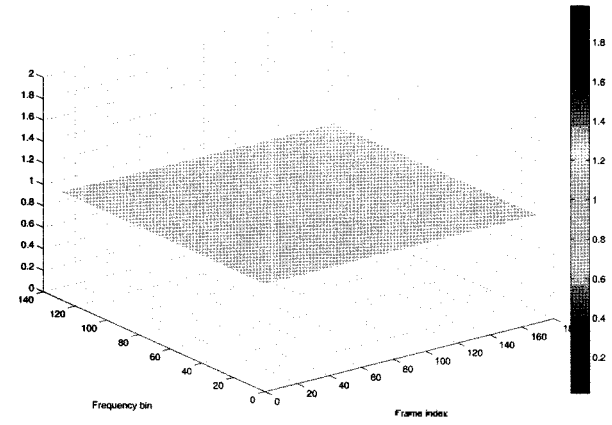

Fig. 8. Gain function.

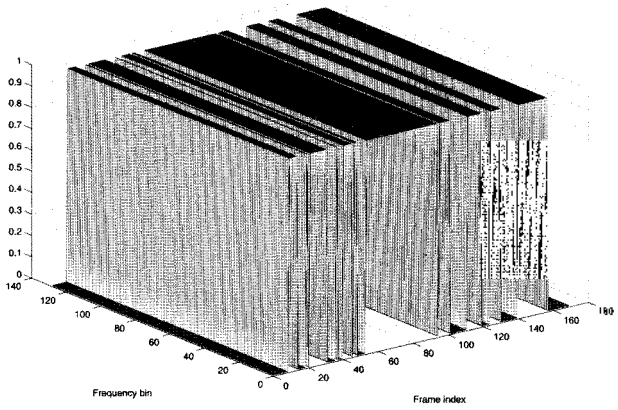

Fig. 9. Updated gain function.

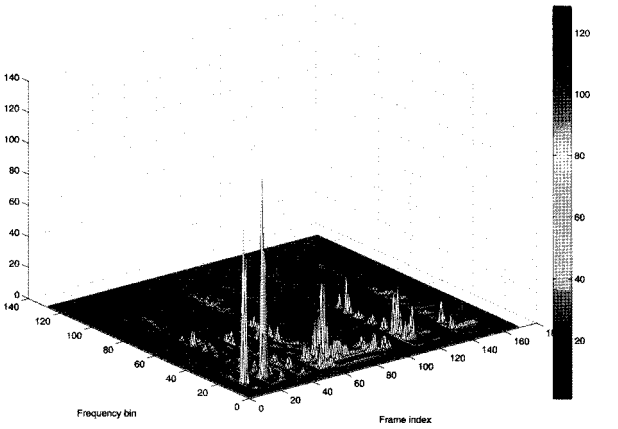

Fig. 10. Updated noisy power spectrum with 10dB car noise for the female sp12.wav speech sample "The drip of the rain made a pleasant.

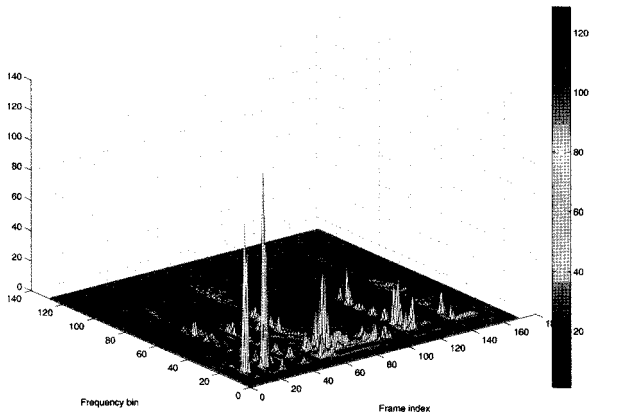

Fig. 11. Noisy power spectrum with 10dB car noise for the female sp12.wav speech sample "The drip of the rainmade a pleasant.

constant $1-\alpha$ for the frames in which speech is present. Figures 8 and 9 show the gain function $G(k,l)$ and the updated gain function $G_{update}(k,l)$, respectively.

The updated noisy power spectrum of the frame index $|X_{update}(k,l)|^2$ is the difference between the noisy power spectrum $|X(k,l)|^2$ and the frames in which speech is absent. $\widehat{D^2_{level}}(k,l)$, as shown in Fig. 10, 11 and 7,

respectively:

$$|X_{update}(k,l)|^2 = |X(k,l)|^2 - \widehat{D^2_{level}}(k,l) \quad (18)$$

$$|X_{update}(k,l)|^2 = MAX(|X_{update}(k,l)|^2,\ \alpha) \quad (19)$$

Eq. 18 reduces the noise of the frames in which speech is absent, and Eq. 19 is used to avoid negative values [13].

3.3. Separation of speech and noise in frequency bins using adaptive thresholds

In a manner parallel to that described bins in the previous subsection, our method uses an adaptive algorithm with a sigmoid function to track the threshold in a frequency bins:

$$\Psi_f(k) = \left[\frac{1}{1 + \exp(10 * (\gamma_f(k) - \delta_f))} \right] \quad (20)$$

where $\psi_f(k)$ is the adaptive threshold using the sigmoid function in the frequency bins and δ_f is a control parameter. The threshold $\psi_f(k)$ is adaptive in the sense that it changes depending on the control parameter δ_f. Figs. 12 and 13 show the effect of δ_f on SNR gains and scale of the SIG. Simulation results indicate that the optimal value of δ_f is 0.39 for noisy signals with SNR 5 dB and is 0.35 for noisy signal with SNR 15 dB. A fixed value of δ_f will not be optimal over a wide range of SNRs. Fig. 14 shows that the adaptive threshold accounts for the frequency bin index by controlling δ_f.

Consequently, we can control the trade-off between speech distortion and residual noise in the frequency bins using δ_f.

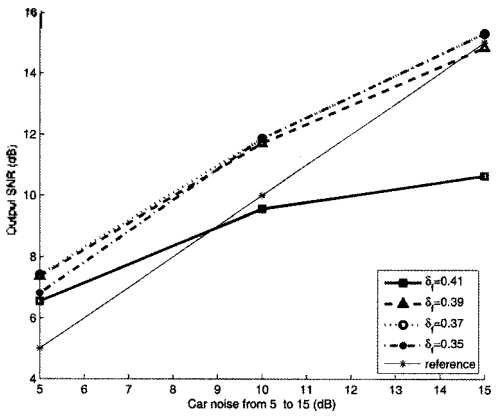

Fig. 12. Effect of various δ_f values on SNR gains.

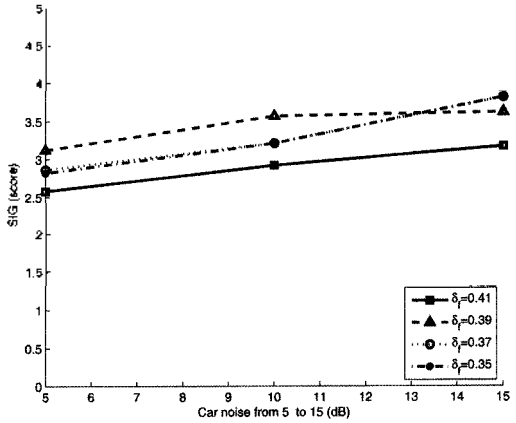

Fig. 13. Effect of various δ_f values on SIG.

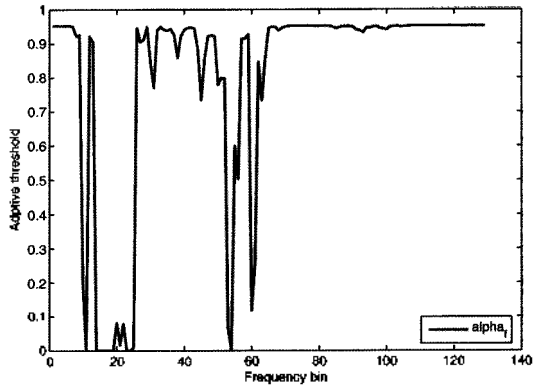

Fig. 14. Adaptive thresholds using a sigmoid function on the frequency bin index for 5 dB car noise.

3.4. Noise reduction using a modified gain function and updated noisy power

The separation algorithm for determining whether speech is present or absent in a frequency bin is

$$If\ \ \psi_f(k) > \phi_f \quad (21)$$

$$G_{\mathrm{modi}}(k,l) = G_{update}(k,l) * \alpha \quad (22)$$

else

$$G_{\mathrm{modi}}(k,l) = G_{update}(k,l) * (1-\alpha) \quad (23)$$

In the same manner as for the time index, where decision parameter ϕ_f is initially 0.95, this threshold $\psi_f(k)$ is compared to the decision parameter ϕ_f. If it is greater than ϕ_f, then speech is declared to be absent in the frequency bin k; otherwise speech is present. The $G_{\mathrm{modi}}(k,l)$ represents the modified gain function for the time and frequency bins using the gain function $G_{update}(k,l)$, the constant α, and $1-\alpha$.

$$|\hat{S}(k,l)|^2 = G_{\mathrm{modi}}(k,l) * |X_{update}(k,l)|^2 \quad (24)$$

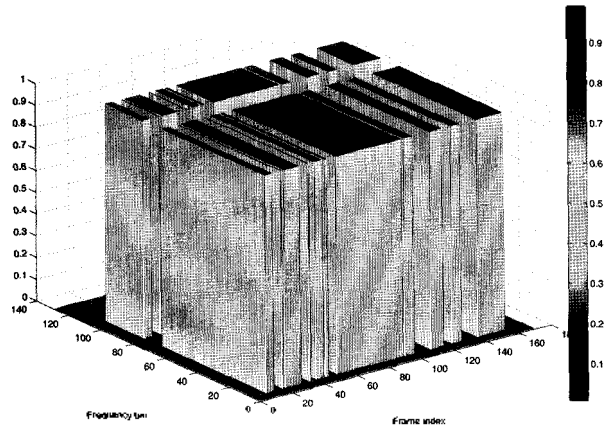

Fig. 15. Modified gain function.

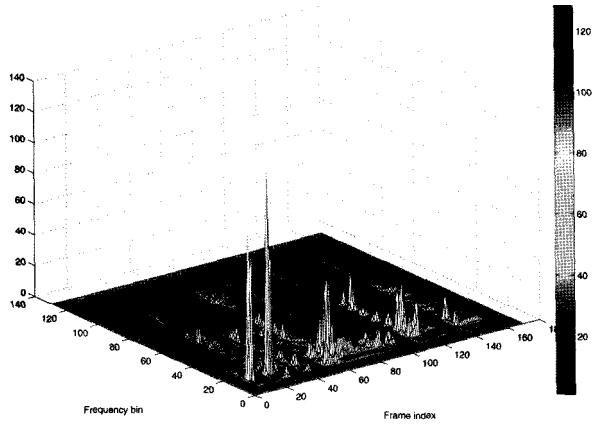

Fig. 16. Estimated clean speech power spectrum with 10 dB car noise for the female sp12.wav speech sample "The drip of the rain made a pleasant sound" from the NOIZEUS database.

Finally, the estimated clean speech power spectrum $|\hat{S}(k,l)|^2$ can be represented as a product of the modified gain function for the time-frequency bins and the updated noisy power spectrum of the time-frequency bins. The estimated clean speech signal can then be transformed back to the time domain using the inverse short-time Fourier transform (STFT) and synthesis with the overlap-add method. We can see the modified gain function $G_{mods}(k,l)$ and the updated noisy power spectrum $|X_{update}(k,l)|^2$ in Figs. 15 and 16, respectively.

IV. Experimental results and discussion

For our evaluation, we selected three male and three female noisy speech samples from the NOIZEUS database [10]. The signal was sampled at 8 kHz and transformed by the STFT using 50% overlapping Hamming windows of

Table 1. Segmental SNR at white, babble and car noise.

	Noise (dB)	white	babble	car
SSMUL	5	4.96	5.89	7.08
	10	8.13	9.28	8.05
	15	10.05	9.89	10.35
MSSS	5	6.83	5.41	6.71
	10	11.20	9.65	10.96
	15	15.23	14.11	14.91
NTFAS	5	9.98	6.44	7.58
	10	11.93	10.68	11.87
	15	16.53	14.49	15.70

256 samples. Evaluating of the new algorithm and a comparing it to the multi band spectral subtraction (SSMUL) and MS with spectral subtraction (MSSS) methods [6, 14] consisted of two parts. First, we tested the segment SNR. This provides a much better quality measure than the classical SNR since it indicates an average error over time and frequency for the enhanced speech signal. Thus, a higher segment SNR value indicates better intelligibility. Second, we used ITU-T P.835 as a subjective measure of quality [11]. This standard is designed to include the effects of both the signal and background distortion in ratings of overall quality [10].

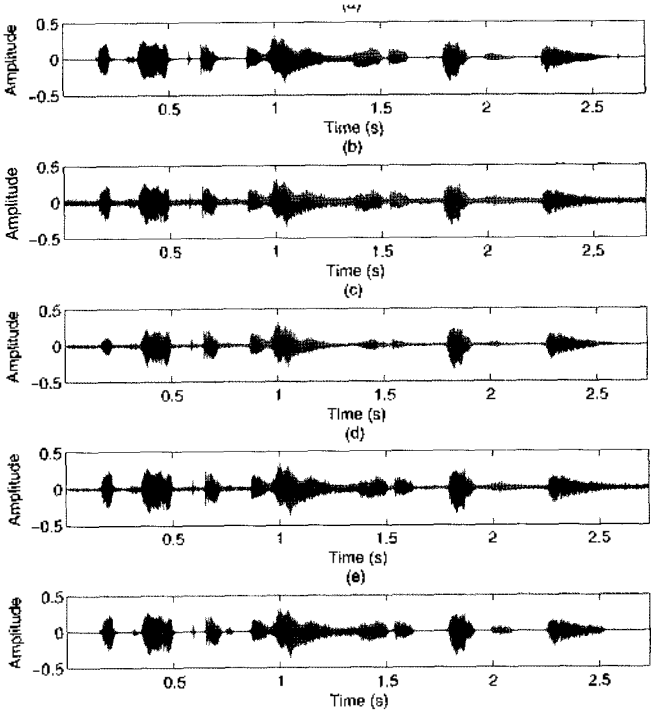

Fig. 17. Example of noise reduction with 10 dB car noise with female sp12.wav speech sample "The drip of the rain made a pleasant sound" from the NOIZEUS database for the three enhancement algorithms. (a) original signal, (b) noisy signal, (c) signal enhanced using the SSMUL method, (d) signal enhanced using the MSSS method, and (e) signal enhanced using the NTFAS method.

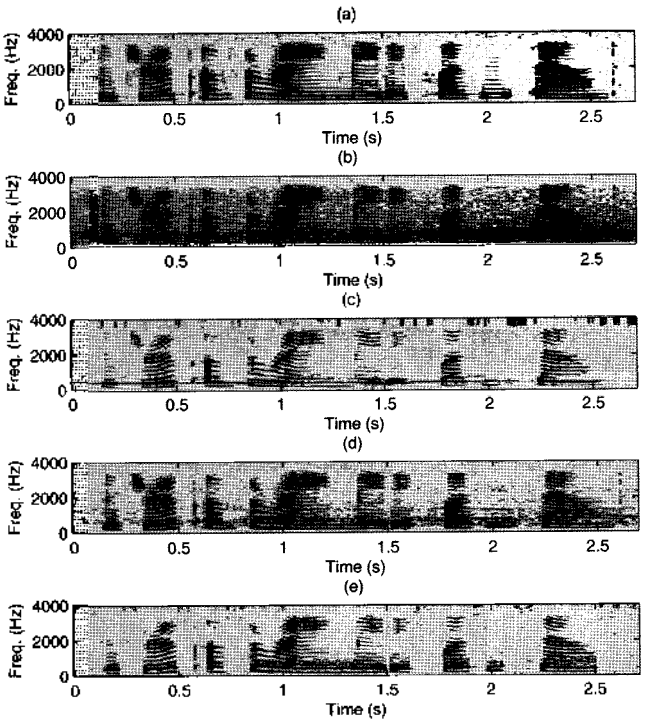

Fig. 18. Example of noise reduction with 10 dB car noise with female sp12.wav speech sample "The drip of the rain made a pleasant sound" from the NOIZEUS database for the three enhancement algorithms. (a) original spectrogram, (b) noisy spectrogram, (c) spectrogram using the SSMUL method, (d) spectrogram using the MSSS method, and (e) spectrogram using the NTFAS method.

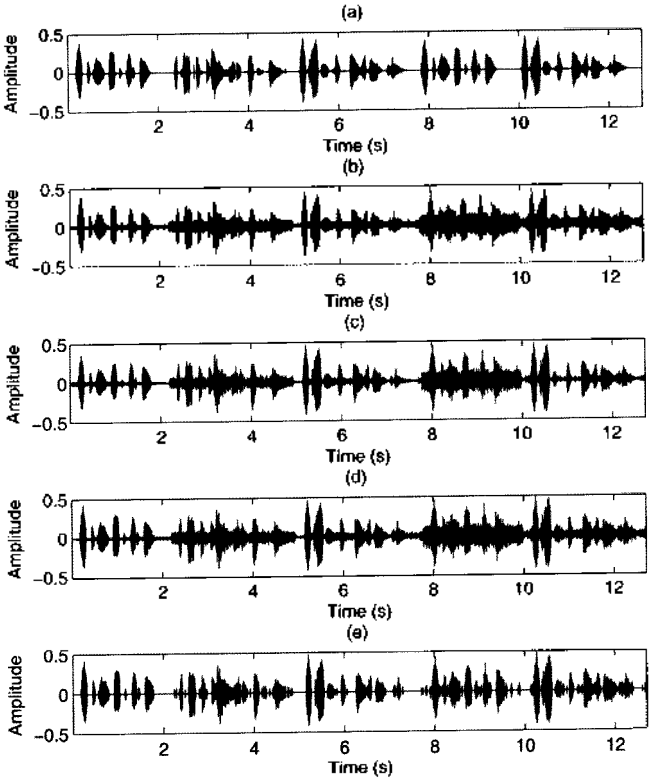

Fig. 19. Time domain results of speech enhancement for 15 dB car noise, 5 dB car noise, 10 dB babble noise, 0 dB white noise, and 5 dB SNR babble noise in a nonstationary environment. The noisy signal comprises five concatenated sentences from the NOIZEUS database. The speech signal were two male and one female sentences from the AURORA 2 corpus. (a) original speech, (b) noisy speech, (c) speech enhanced using SSMUL method, (d) speech enhanced using the MSSS method, (e) speech enhanced using the NTFAS method.

4.1. Segment SNR and speech signal

We measured the segment SNR over short frames and obtained the final result by averaging the value of each frame over all the segments.

Table 1 shows the segment SNR improvement for each speech enhancement algorithm. For the input SNR in the range 5−15 dB for white Gaussian noise, car noise, and babble noise, we noted that the segment SNR after processing was clearly better for the proposed algorithm than for the SSMUL and the MMSE methods [6,14]. The proposed algorithm yields a bigger improvement in the segment SNR with lower residual noise than the conventional methods. The NTFAS algorithm in particular produces good results for white Gaussian noise in the range 5 to 15 dB. Figs. 17 and 18 show the NTFAS algorithm's clear superiority in the 10 dB car noise environment.

For nonstationary noisy environments, the conventional methods worked well for high input SNR values of 10 and 15 dB; however, the output they produced could not be easily understood for low SNR values of car noise (5 dB) and white noise (0 dB), and they produced residual noise and distortion as shown in Fig. 19. This outcome is also confirmed by time domain results of speech enhancement methods illustrated in Figs. 19 and 20. A different result is clear in Fig. 19 (a) and (b) for the waveforms of the clean and noisy speech signals, respectively, (c) the waveforms of speech enhancement using the SSMUL method, (d) the MSSS method, and (e) the proposed NTFAS method. Fig. 19 (c) and (d) show that the presence of residual noise at $t > 7.8$ s is due partly to the inability of the speech enhancement algorithm to track the sudden appearance of a low SNR. In contrast, panel (e) shows that the residual noise is clearly reduced with the proposed NTFAS algorithm.

4.2. The ITU-T P.835 Standard

Noise reduction algorithms typically degrade the speech component in the signal while suppressing the background

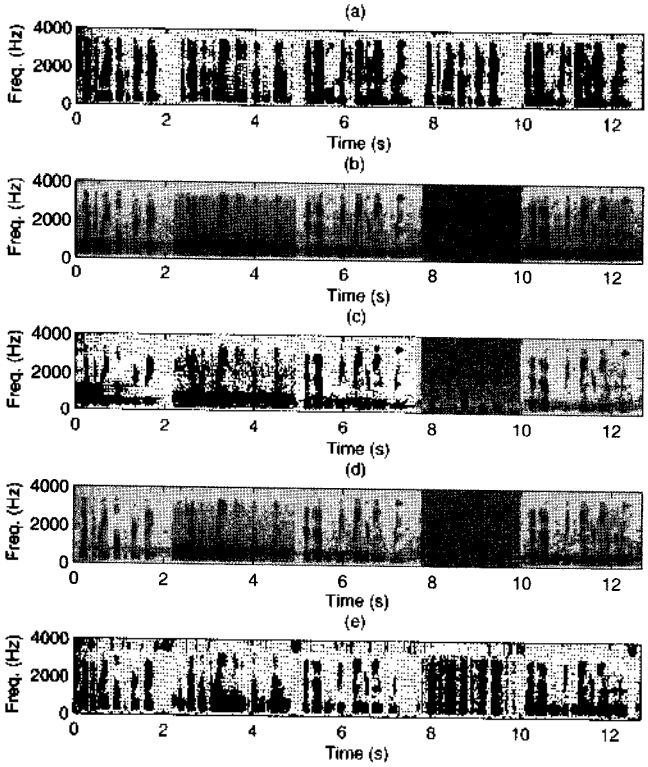

Fig. 20. Frequency domain results of speech enhancement for 15 dB car noise, 5 dB car noise, 10 dB babble noise, 0 dB white noise, and 5dB SNR babble noise in a nonstationary environment. The noisy signal comprises five concatenated sentences from the NOIZEUS database. The speech signal were two male and one female sentences from the AURORA 2 corpus. (a) original spectrogram, (b) noisy spectrogram, (c) spectrogram using the SSMUL method, (d) spectrogram using the MSSS method, (e) spectrogram using the NTFAS method.

noise, particularly under low-SNR conditions. This situation complicates the subjective evaluation of algorithms as it is not clear whether listeners base their overall quality judgments on the distortion of the speech or the presence of noise. The overall effect of speech and noise together was rated using the scale of the Mean Opinion Score (MOS), scale of background intrusiveness (BAK), and the SIG [10]. The proposed method resulted in a great reduction in noise, while providing enhanced

Table 2. The overall effect (OVL) using the Mean Opinion Score (MOS), 5= excellent, 4= good, 3= fair, 2= poor, 1= bad.

	Noise (dB)	white	babble	car
SSMUL	5	1.84	2.47	2.79
	10	3.14	2.96	3.04
	15	3.57	3.49	2.91
MSSS	5	2.98	2.68	2.74
	10	4.41	3.16	3.04
	15	4.43	5.00	3.30
NTFAS	5	4.54	2.55	2.31
	10	5.00	2.67	2.87
	15	5.00	4.56	5.00

Table 3. Scale of Background Intrusiveness (BAK), 5= not noticeable, 4= somewhat noticeable, 3= noticeable but not intrusive, 2= fairly conspicuous, somewhat intrusive, 1= very intrusive.

	Noise (dB)	white	babble	car
SSMUL	5	3.59	2.21	2.82
	10	3.31	2.37	3.01
	15	5.00	1.01	1.79
MSSS	5	3.38	1.63	2.18
	10	4.12	2.46	2.69
	15	3.54	1.00	2.60
NTFAS	5	3.25	2.52	2.17
	10	3.63	2.82	3.07
	15	4.58	5.00	5.00

Table 4. Scale of Signal Distortion (SIG), 5=no degradation, 4=little degradation,3=somewhat degraded, 2=fairly degraded, 1= very degraded.

	Noise (dB)	white	babble	car
SSMUL	5	1.79	2.81	3.74
	10	2.69	3.26	3.75
	15	3.15	3.37	2.87
MSSS	5	1.93	3.25	3.92
	10	2.96	3.63	3.92
	15	4.53	3.87	4.01
NTFAS	5	2.68	3.27	3.62
	10	4.08	3.29	3.62
	15	4.74	3.74	3.83

speech with lower residual noise and somewhat higher MOS, BAK, and SIG scores than the conventional methods. It also degraded the input speech signal in highly nonstationary noisy environments. Degradation of the speech signal is extremely undesirable in real speech recognition systems. Consequently, an automatic noise estimation and separation algorithm is required.

This is confirmed by an enhancement signal and ITU-T P.835 test [11]. The results of the evaluation are shown in Table 2, 3 and 4. The best result for each speech enhancement algorithms is shown in bolds.

V. Conclusions

In this paper, we proposed a new approach to the enhancement of speech signals that have been corrupted by stationary and nonstationary noise. This approach is not a conventional spectral algorithm, but uses a method that separates the speech-presence and

absence contributions in time-frequency bins. We call this technique the NTFAS speech enhancement algorithm. We showed that appropriate choices of δ_t and δ_f produced

good-quality enhanced speech signal. The proposed method resulted in a great reduction in noise while providing enhanced speech with lower residual noise and somewhat MOS, BAK and SIG scores than the conventional methods. It also degraded the input speech signal in highly nonstationary noisy environments. Degradation of the speech signal is very undesirable in real speech recognition systems, and thus automatic noise estimation and separation algorithms are required.

Acknowledgment

The present research has been conducted by the Research Grant of Kwangwoon University in 2006.

References

1. M. Bhatnagar, *A modified spectral subtraction method combined with perceptual weighting for speech enhancement*, (Master's thesis, University of Texas at Dallas, 2003)
2. S. F. Boll,, "Suppression of acoustic noise in speech using spectral subtraction", IEEE Trans. Acoust. Speech Signal Processing, 27 (2), 113-120, 1979.
3. O. Cappe, "Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor", IEEE Trans. Speech Audio Processing, 2 (2), 346-349, 1994.
4. Y. Hu, *Subspace and multitaper methods for speech enhancement*, (Ph.D. dissertation, University of Texas at Dallas, 2003)
5. Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square errror short-time spectral amplitude estimator", IEEE Trans. Acoust. Speech Signal Processing, 32 (6), 1109-1121, 1984.
6. R.Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics", IEEE Trans. Speech Audio Processing, 9 (5), 504-512, 2001.
7. R.Sundarrajan and C.L.Philipos, "A noise-estimation algorithm for highly non-stationary environments",Speech Communication, 48, 220-231,2006.
8. I. Cohen, "Noise spectrum in adverse environments: improved minima controlled recursive averaging", IEEE Trans. SpeechAudio Processing, 11 (5), 466-475, 2003.
9. I. Cohen, "Speech enhancement using a noncausal a priori SNR estimator", IEEE Signal Processing Letters, 11 (9), 725-728, 2004.
10. C. L. *Philipos, Speech Enhancement* (Theory and Practice, 1st edition, CRC Press, Boca Raton, FL,2007)
11. ITU-T, *Subjective test methodology for evaluating speech communication systems that include noise suppression algorithm*, (ITU-T Recommendation 835)
12. A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, (2nd edition, Prentice Hall, Upper Saddle River, NJ, 1999)
13. L. Lin, W. H. Holmes and E. Ambikairajah, "Adaptive noise estimation algorithm for speech enhancement", Electronic Letters, 39 (9), 754-755, 2003.
14. D. K. Sunil and and C. L. Philipos C. L. "A multi-band spectral subtraction method for enhancing speech corrupted by colored noise", In Proceedings International Conference on Acoustics, Speech and Signal Processing, 2002.

[Profile]

• **Soojeong Lee**

1992.3~1997.2: B.S. degree in Computer science, Korea National Open University
1997.3~2000.2: M.S. degree in Computer Engineering, Kwangwoon University
2004.3~2008.2: Ph.D. degree in Computer Engineering, Kwangwoon University

• **Soonhyob KIM**

B.S. degree in electronics engineering from the Ulsan University, Korea, 1970.3~1974.2
M.S. degree in electronics engineering from the Yonsei University, Korea, 1974.3~1976.2
Ph.D. degree in electronics engineering from the Yonsei University, Korea, 1976.3~1983.2
1979. 3~present Professor, Dept. of Computer Engineering, Kwangwoon University
1986. 8~1987. 8 Visiting professor, Dept. of Electrical & Computer Eng. Univ. of Texas at Austin
1998. 1~2000. 12 President, the Acoustical Society of Korea
2000. 10~2004.10 Chairman of the invitation committee ICSLP 2004
2001. 1~present President Emeritus, the Acoustical Society of Korea