A NOISE ESTIMATION METHOD BASED ON IMPROVED VAD USED IN NOISE SPECTRAL SUPPRESSION UNDER HIGHLY NON-STATIONARY NOISE ENVIRONMENTS

Kenji Nakayama   *Shoya Higashi*   *Akihiro Hirano*

Graduate School of Natural Science and Technology, Kanazawa University
Kanazawa, 920-1192, Japan
E-mail: nakayama@t.kanazawa-u.ac.jp

ABSTRACT

A noise spectral estimation method, which is used in spectral suppression noise cancellers, is proposed for highly non-stationary noise environments. Speech and non-speech frames are detected by using the entropy-based voice activity detector (VAD). An adaptive parameter and variable thresholds are newly introduced for the VAD. The former can stabilize the entropy used in the VAD. The latter is used to discriminate the noisy speech into three categories, a non-speech frame, a quasi-speech frame and a speech frame. The noise spectrum is estimated by using the noisy speech spectrum in the non-speech frames and the weighted noisy speech spectrum in the quasi-speech frame and the speech frame. The weighting function used in the quasi-speech frame is modified from the conventional approach to suppress over estimation. These proposed techniques are very useful for rapid change in the noise spectrum and power. Simulations are carried out by using many kinds of noises, including white, babble, car, pink, factory and tank. The segmental SNR can be improved by $1.7 \sim 2.8$dB and $0.6 \sim 1.8$dB for the input SNR=3dB and 9dB, respectively. The noise spectral estimation error can be improved by $1.42 \sim 2.4$dB and $0.6 \sim 1.4$dB for the input SNR=3dB and 9dB, respectively.

1. INTRODUCTION

A spectral suppression technique is a hopeful approach to noise cancellers used in a mobile phone [1]. In this approach, it is very important to estimate a spectral gain, used to suppress the noise spectrum. Several methods, including MMSE STSA [2] and Joint MAP [3], have been proposed. Furthermore, performance of the spectral suppression technique is highly dependent on accuracy of the noise spectral estimation [4],[5]. There exist many kinds of noises. In highly non-stationary noise environments, power and spectrum of the noises can be dynamically changed. The noise spectral estimation should adapt this kind of changes quickly. Several noise spectral estimation methods have been proposed for this purpose [6]-[10].

In this paper, a noise spectral estimation method based on Voice Activity Detection (VAD) is proposed. An adaptive parameter and variable thresholds are proposed to improve the VAD performance. The noisy speech is discriminated in three categories, a non-speech frame, a quasi-speech frame and a speech frame. The noise spectrum is optimally estimated in each frames. Computer simulations by using speech signal and many kinds of noises will be shown.

2. SPECTRAL SUPPRESSION NOISE CANCELLER

Figure 1 shows a blockdiagram of the spectral suppression noise canceller. Spectra of speech and noise are assumed to be statistically independent. Let $s(n)$, $n(n)$ and $x(n)$ be clean speech, noise and noisy speech, respectively.

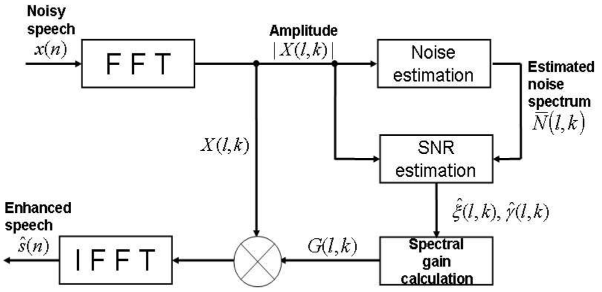

Fig. 1. Blockdiagram of spectral suppression noise canceller.

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

The Fourier transform of $x(n)$, $s(n)$ and $n(n)$ in the lth frame and at the kth frequency bin are expressed by

$$X(l,k) = S(l,k) + N(l,k) \tag{2}$$

The prior SNR $\xi(l,k)$, a ratio of the clean speech power to the noise power, and the posterior SNR $\gamma(l,k)$, a ratio of the noisy speech power to the noise power, are defined by

$$\xi(l,k) = \frac{E[|S(l,k)|^2]}{E[|N(l,k)|^2]} \tag{3}$$

$$\gamma(l,k) = \frac{|X(l,k)|^2}{E[|N(l,k)|^2]} \tag{4}$$

Actually, the noisy speech signal $x(n)$ is only available. The prior SNR $\xi(l,k)$ is estimated as follows [2]:

$$\begin{aligned}\hat{\xi}(l,k) &= \alpha\gamma(l-1,k)G^2(l-1,k) \\ &+ (1-\alpha)P[\gamma(l,k)-1]\end{aligned} \tag{5}$$

where $0 < \alpha < 1$ and $P[x]$ satisfies

$$P[x] = \begin{cases} x, & x > 0 \\ 0, & otherwise \end{cases} \tag{6}$$

The posterior SNR $\gamma(l,k)$ can be estimated by using the noise spectrum estimation $\bar{N}(l,k)$ as follows:

$$\hat{\gamma}(l,k) = \frac{|X(l,k)|^2}{\bar{N}(l-1,k)} \tag{7}$$

How to estimate $N(l,k)$ is a main issue in this paper. A spectral gain $G(l,k)$ is estimated by using the prior SNR $\hat{\xi}(l,k)$ and the posterior SNR $\hat{\gamma}(l,k)$, and is used to suppress the noise spectrum included in the noisy speech. In order to calculate $G(l,k)$, we employ Joint MAP method [3], in which the speech is assumed to follow super Gaussian distribution and is better than MMSE STSA method [2],[5].

3. CONVENTIONAL RAPID ADAPTATION METHOD

In this section, a conventional noise spectral estimation method proposed for non-stationary noise environments is briefly described [7]-[10]. A voice activity detector (VAD) [6] is applied in this method.

3.1. Voice Activity Detector (VAD)

The VAD discriminates the speech frame and the non-speech frame based on the following entropy $H(l)$

$$P_r(l,k) = \frac{X_{energy}(l,k)}{\sum_{k=1}^{2M} X_{energy}(l,k)} \tag{8}$$

$$H(l) = -\sum_{k=1}^{2M} P_r(l,k) \cdot \log(P_r(l,k)) \tag{9}$$

$$X_{energy}(l,k) = |X(l,k)|^2 \tag{10}$$

The entropy $H(l)$ has a large value in the non-speech frame compared to the speech frame. We assume several frames at the beginning to be the non-speech frame. An average of the entropy, estimated in these frames, denoted $H_{av}(0)$ is used as the threshold, with which the following frames are discriminated as the speech or the non-speech frames. Actually, $H_{av}(0)$ is scaled by a constant $c(< 1)$.

$$\begin{array}{lcl} H(l) > cH_{av}(0) & \rightarrow & \text{Non-speech frame} \\ H(l) < cH_{av}(0) & \rightarrow & \text{Speech frame} \end{array}$$

$H(l)$ is not accurate and cannot discriminate the non-speech frame and the speech frame, when the spectra of the speech and the noise are small and large, respectively. In order to improve this problem, a positive constant C has been introduced in $P_r(l,k)$ as follows:

$$P_{rc}(l,k) = \frac{X_{energy}(l,k) + C}{\sum_{k=1}^{2M} X_{energy}(l,k) + C} \tag{11}$$

$$H_c(l) = -\sum_{k=1}^{2M} P_{rc}(l,k) \cdot \log(P_{rc}(l,k)) \tag{12}$$

3.2. Noise Spectral Estimation Method

The conventional method is briefly described here [7],[8]. The noisy speech is discriminated into the non-speech frame or the speech frame by using the VAD. The noise spectrum is estimated in the non-speech frame by

$$\bar{N}(l,k) = \lambda \cdot \bar{N}(l-1,k) + (1-\lambda) \cdot |X(l,k)|^2 \tag{13}$$

On the other hand, in the speech frames, the noise spectrum is estimated by the following recursive equation.

$$\begin{aligned} \bar{N}(l,k) &= \rho(l,k) \cdot \bar{N}(l-1,k) \\ &+ (1-\rho(l,k)) \cdot |X(l,k)|^2 \qquad (14) \\ \rho(l,k) &= a_d + (1-a_d) \cdot P_{sp}(l,k) \qquad (15) \end{aligned}$$

$P_{sp}(l,k)$ is a probability of including the speech in the noisy speech, that is a speech presence probability, given by

$$P_{sp}(l,k) = \frac{|X(l,k)|^2}{P_{min}(l,k)} \tag{16}$$

$P_{min}(l,k)$ is the minimum of the noisy speech spectrum, as shown in the following. First, the averaged spectrum of the noisy speech $P(l,k)$ is obtained by

$$P(l,k) = \eta P(l-1,k) + (1-\eta)|X(l,k)|^2 \tag{17}$$

η is a smoothing factor. Next, $P_{min}(l,k)$ is updated by the following equations.

$$\begin{aligned} &P_{min}(l,k) = \gamma \cdot P_{min}(l-1,k) + \frac{1-\gamma}{1-\beta}(P(l,k) \\ &-\beta \cdot P(l-1,k)), \ \text{if } P_{min}(l-1,k) \le P(l,k) \quad (18) \\ &P_{min}(l,k) = P(l,k), \ \text{if } P_{min}(l-1,k) > P(l,k) (19) \end{aligned}$$

β and γ are determined by experience.

4. A NEW NOISE SPECTRAL ESTIMATION METHOD

4.1. New Adaptive Parameter and Thresholds for VAD

In the conventional method, as shown in Eq.(11), $P_{rc}(l,k)$ is stabilized by using a constant C. The constant C is highly dependent on SNR of the noisy speech, and should be optimized. In this paper, we propose a new adaptive parameter, which is controlled by the maximum of $|X(l,k)|$ as follows:

$$P_{new}(l,k) = \frac{X_{energy}(l,k) + C_{new}(l)}{\sum_{k=1}^{2M} X_{energy}(l,k) + C_{new}(l)} \tag{20}$$

$$C_{new}(l) = \max_k\{|X(l,k)|\} \tag{21}$$

$$H_{new}(l) = -\sum_{k=1}^{2M} P_{new}(l,k) \cdot \log(P_{new}(l,k)) \tag{22}$$

The stabilization parameter $C_{new}(l)$ is adjusted in each frame in order to adapt rapid change in noise spectrum and power.

Furthermore, the proposed method aims to precisely discriminate the non-speech frames in order to estimate the noise spectrum more precisely. For this purpose, we classify the noisy speech into three categories, a non-speech frame, a quasi-speech frame and a speech frame. For this purpose, we introduce two kinds of thresholds, $\delta_1(l)$ and $\delta_2(l)$ for the entropy $H_{new}(l)$, as follows:

$$\delta_1(l) = c_1 E[H_{new}(l)] \tag{23}$$

$$\delta_2(l) = c_2 E[H_{new}(l)], \quad 0 < c_2 < c_1 \tag{24}$$

$E[H_{new}(l)]$ means an average over the recent five non-speech frames including the l-th frame. The initial guess of $\delta_1(l)$ and $\delta_2(l)$ are determined in the begining five frames, which are assumed to be the non-speech frame. If $H_{new}(l) > \delta_1(l-1)$, then $\delta_1(l)$ and $\delta_2(l)$ are updated by Eqs.(23) and (24). After that, if $H_{new}(l) > \delta_1(l)$, then this frame is discriminated as the non-speech frame. If $\delta_2(l) < H_{new}(l) < \delta_1(l)$, then this frame is discriminated as the quasi-speech frame. Finally, if $H_{new}(l) < \delta_2(l)$, then this frame is discriminated as the speech frame. c_1 and c_2 are 0.98 and 0.95, respectively, which are determined by experience.

Furthermore, we introduce another criterion for the condition $\delta_2(l) < H_{new}(l) < \delta_1(l)$. If the covariance of $P_{new}(l,k)$ in the l-th frame is less than the threshold σ_p, then this frame is regarded as the non-speech frame. σ_p is set to be 2 by experience.

The purpose of introducing the quasi-speech frame is to correctly analyze the non-speech frames.

4.2. A New Noise Spectral Estimation Method

4.2.1. Non-Speech Frame

In the proposed VAD, the non-speech frames can be correctly discriminated by using the relatively high threshold. Furthermore, we consider highly non-stationary noise environment and rapid change in noise spectrum and power. Taking these situations into account, the noise spectrum is estimated by using the noisy speech spectrum itself in the non-speech frame as follows:

$$\bar{N}(l,k) = |X(l,k)|^2 \tag{25}$$

In the conventional methods, recursive averaging is employed, which includes the past information, and is difficult to quickly adapt rapid change in noise spectrum and power [7]-[10]. Furthermore, controlling the smoothing parameters highly depends on noise spectrum and power, and is difficult.

4.2.2. Speech Frame

In the conventional method, the recursive averaging estimation was used as shown by Eq.(14). However, this method does not work well for the non-stationary noise environments.

In this paper, we apply the weighted noise spectral estimation method [4],[5]. The noisy speech spectrum is weighted by the weight function $W(l,k)$, which is determined based on the posterior SNR $\hat{\gamma}(l,k)$ as shown in Fig.2. The noisy speech spectrum is reduced in the high SNR region in order to suppress over estimation of the noise spectrum. The weighted spectrum is expressed by

$$z(l,k) = W(l,k)|X(l,k)|^2 \tag{26}$$

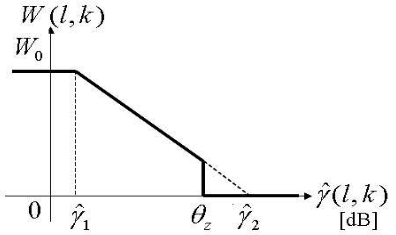

Fig. 2. Weight function $W(l,k)$ for noisy speech.

The parameters are determined by experience as follows: $W_0 = 1$, $\hat{\gamma}_1 = 0$dB, $\hat{\gamma}_2 = 10$dB and $\theta_z = 7$dB. The noise spectrum is estimated by averaging $z(l,k)$ over several frames. Figure 2 means that in the beginning frames and in the low SNR region, that is $\hat{\gamma}(l,k) < \theta_z$, $z(l,k)$ is included as the noise components. On the other hand, after the beginning frames and in the high SNR frames, that is $\hat{\gamma}(l,k) > \theta_z$, $z(l,k)$ is not included as the noise components, rather the previous average of $z(l,k)$ is used [4].

4.2.3. Quasi-Speech Frame

We employ the same approach as in the speech frame to estimate the noise spectrum in the quasi-speech frame. However, since the quasi-speech frames may include a consonant sound, a fricative sound, an explosive sound, the weight function shown in Fig.2 is modified in order to suppress over estimation of the noise spectrum. The parameters are determined by experience as follows: $W_0 = 0.8$, $\hat{\gamma}_1 = 3$dB, $\hat{\gamma}_2 = 10$dB and $\theta_z = 7$dB.

5. SIMULATION AND DISCUSSIONS

5.1. Conventional Methods

The conventional methods [4] and [10] are employed for comparison. In [4], the noise spectrum estimation algorithm described in Sec.4.2.2 is used in all frames. In this method, the non-speech and the speech frames are not discriminated. The proposed method introduces a new algorithm in the non-speech frame, and modifies the weight function in the quasi-speech frame. In [10], VAD is used to discriminate the non-speech and the speech frames. The noise spectrum is estimated by the recursive averaging methods in both the non-speech and the speech frames. Furthermore, in all methods, we apply the lower boundary for the spectral gain $G(l,k)$ and adding the original noisy speech to the output signal $\hat{s}(n)$ in a small rate in order to improve speech quality [5]. $G(l,k)$ is estimated by Joint MAP method [3].

5.2. Evaluation Measures

5.2.1. Normalized Noise Spectrum Estimation Error

$$\varepsilon(l) = 10\log_{10}\left(\frac{\sum_{k=0}^{M}\left||N(l,k)|^2 - |\bar{N}(l,k)|^2\right|}{\sum_{k=0}^{M}|N(l,k)|^2}\right) \tag{27}$$

$$\bar{\varepsilon} = \frac{1}{L}\sum_{l=1}^{L}\varepsilon(l) \tag{28}$$

L is the number of all frames. The smaller value of ε means the higher accurate estimation.

5.2.2. Segmental SNR

A signal to noise ratio at the output is evaluated by the following segmental SNR.

$$SNR_{seg} = \frac{10}{L}\sum_{l=0}^{L-1}\log_{10}\frac{\sum_{n=N_l}^{N_l+N-1}s^2(n)}{\sum_{n=N_l}^{N_l+N-1}(s(n)-\hat{s}(n))^2} \tag{29}$$

N is a length of the interval, where SNR$_{seg}$ is evaluated. The actual length is 12ms. SNR$_{seg}$ at the input is evaluated by using $n^2(n)$ instead of $(s(n)-\hat{s}(n))^2$ in the above equation.

5.2.3. Log-Spectral Distortion (LSD)

The noise suppressed signal $\hat{S}(l,k)$ is compared with the clean speech signal $S(l,k)$ in an amplitude response using a log function as follows:

$$LSD = \frac{1}{J}\sum_{l=1}^{J}\left(\frac{1}{2M}\sum_{k=1}^{2M}\left(\log\frac{|S(l,k)|+\delta}{|\hat{S}(l,k)|+\delta}\right)^2\right)^{1/2} \tag{30}$$

2M is a frame lenght, J is the number of the frames, δ is a positive small value.

5.2.4. Ideal Estimation

In order to evaluate accuracy of the noise spectrum estimation, the true noise spectrum is used. Let $G_{tl}(l,k)$ be the spectral gain obtained by using the true noise spectrum. The ideal noise suppressed output signal is obtained by

$$\hat{s}(n) = IFFT[G_{tl}(l,k)X(l,k)] \tag{31}$$

5.3. Simulation Results

5.3.1. Noise Estimation and Reduction - Noise is not Changed -

The normalized noise spectrum estimation error $\bar{\varepsilon}$, the segmental SNR (SNR$_{seg}$) and the log-spectral distortion (LSD) are evaluated by using the ideal method, the conventional methods [4],[10] and the proposed method. Six kinds of noises are used, that is White, Babble, Car, Pink, Factory, Tank, which are provided by [11].

Tabel 1 shows $\bar{\varepsilon}$, SNR_{seg} and LSD for the ideal method. The input SNR$_{seg}$ is set to be 3dB and 9dB. In this case, $\bar{\varepsilon} = -\infty$ and $LSD = 0$ can be expected.

Tables 2, 3 and 4 show the simulation results for the conventional methods [4], [10] and the proposed method, respectively. The input SNR$_{seg}$ is set to be 3dB and 9dB. Regarding $\bar{\varepsilon}$ and SNR$_{seg}$, the proposed method can provide good results. Its LSD are almost the same as those of the conventional method [4]. Another conventional method [10] is not good in all evaluation measures.

5.3.2. Noise Estimation and Reduction - Noise is Changed -

SNR_{seg} of the ideal method under the dynamical situation, the noise is changed from the babble noise (Input SNR$_{seg}$ = 6dB) to six kinds of noises (Input SNR$_{seg}$ = 2dB), are shown in Table 5. The average input SNR$_{seg}$ is 5dB.

Table 1. Noise reduction by ideal method. Noise is not changed. $\bar{\varepsilon} = -\infty$, LSD=0.

Evaluation Method	SNR$_{seg}$	
Input SNR$_{seg}$[dB]	3	9
White	9.39	13.3
Babble	11.9	15.6
Car	15.5	18.5
Pink	12.1	15.6
Factory	12.0	15.4
Tank	15.5	18.5

Table 2. Noise estimation and reduction by conventional method [4]. Noise is not changed.

	$\bar{\varepsilon}$		SNR$_{seg}$		LSD	
SNR$_{seg}$	3	9	3	9	3	9
White	-1.15	-1.20	5.18	10.7	0.370	0.285
Babble	-0.664	-0.760	4.34	10.3	0.302	0.226
Car	-1.15	-1.54	7.57	13.0	0.250	0.184
Pink	-1.07	-1.12	5.61	11.3	0.288	0.215
Factory	-1.09	-1.14	5.83	11.5	0.284	0.215
Tank	-1.15	-1.13	6.49	12.0	0.278	0.207

Table 3. Noise estimation and reduction by conventional method [10]. Noise is not changed.

	$\bar{\varepsilon}$		SNR$_{seg}$		LSD	
SNR$_{seg}$	3	9	3	9	3	9
White	4.47	9.89	5.09	7.27	0.497	0.468
Babble	8.07	11.8	4.15	6.86	0.425	0.381
Car	7.60	11.6	4.95	7.30	0.393	0.352
Pink	5.96	10.1	4.95	7.45	0.378	0.345
Factory	6.45	10.2	4.59	7.53	0.380	0.350
Tank	6.02	10.3	5.70	7.62	0.373	0.333

Table 4. Noise estimation and reduction by proposed method. Noise is not changed.

	$\bar{\varepsilon}$		SNR$_{seg}$		LSD	
SNR$_{seg}$	3	9	3	9	3	9
White	-3.57	-2.62	7.08	11.8	0.423	0.350
Babble	-2.03	-1.69	6.04	12.0	0.310	0.244
Car	-2.81	-1.75	9.33	13.6	0.273	0.216
Pink	-3.24	-1.92	7.95	12.7	0.289	0.228
Factory	-3.02	-1.72	7.79	12.7	0.297	0.233
Tank	-3.18	-1.73	9.24	13.8	0.271	0.207

Table 5. Noise reduction by ideal method. Noise is changed in non-speech frame (SNR$_{seg}(1)$) and in speech frame (SNR$_{seg}(2)$).

Evaluation Methods	SNR$_{seg}(1)$	SNR$_{seg}(2)$
1. Babble(6dB)→White(2dB)	11.8	11.9
2. Babble(6dB)→Babble(2dB)	13.2	13.1
3. Babble(6dB)→Car(2dB)	14.9	14.8
4. Babble(6dB)→Pink(2dB)	13.2	13.3
5. Babble(6dB)→Factory(2dB)	13.1	13.1
6. Babble(6dB)→Tank(2dB)	15.2	15.2

The simulation results of the conventional method [4] and the proposed method are shown in Tables 6 and 7. The numbers in the left column are the same numbers shown in Table 5. In this case, the simulation results of the conventional method [10] are not shown, because its performance is not good as shown in Table 3 compared with the other methods.

As shown in these tables, $\bar{\varepsilon}$ and SNR$_{seg}$ can be more improved by the proposed method. LSD is still almost the same as those of the conventional method [4]. Since the proposed method can precisely estimate the noise spectrum in the non-speech frames, $\bar{\varepsilon}$ can be well reduced, as a result SNR$_{seg}$ can be improved. Furthermore, in the non-speech frame, the noise spectrum is estimated by using the noisy speech spectrum only in this frame. Therefore, the noise estimation can quickly follow the dynamical change in noise properties and power. The conventional methods do not work well in these situations.

Table 6. Performance comparison. Noise is changed in non-speech frame.

	Conventional Method [4]			Proposed Method		
	$\bar{\varepsilon}$	SNR$_{seg}$	LSD	$\bar{\varepsilon}$	SNR$_{seg}$	LSD
1	-0.428	5.75	0.356	-2.85	8.79	0.343
2	-0.573	6.00	0.283	-2.26	8.19	0.297
3	0.491	5.78	0.286	-2.50	9.12	0.286
4	-0.524	5.94	0.283	-2.65	8.99	0.284
5	-0.546	6.01	0.281	-2.43	8.72	0.289
6	-0.474	5.83	0.285	-2.77	9.91	0.270

Table 7. Performance comparison. Noise is changed in speech frame.

	Conventional Method [4]			Proposed Method		
	$\bar{\varepsilon}$	SNR$_{seg}$	LSD	$\bar{\varepsilon}$	SNR$_{seg}$	LSD
1	-0.435	5.76	0.351	-2.78	8.82	0.339
2	-0.541	5.91	0.283	-2.03	8.06	0.294
3	-0.488	5.77	0.287	-2.39	9.13	0.284
4	-0.517	5.91	0.283	-2.59	8.88	0.288
5	-0.550	5.98	0.282	-2.69	8.95	0.284
6	-0.467	5.80	0.292	-2.59	9.66	0.285

6. CONCLUSIONS

In this paper, a new noise spectral estimation method is proposed. An improved VAD algorithm is proposed, which can correctly distinguish the non-speech frame, where the noise spectrum can be well estimated. Through computer simulations by using six kinds of noises, the proposed method can reduce the normalized noise spectrum estimation error by 0.6∼2.4dB and can increase SNR$_{seg}$ at the output by 0.6∼2.8dB compared with the conventional methods.

7. REFERENCES

[1] ”Minimum performance requirements for noise suppressor application to the AMR speech encode”, 3GPP TS 06.77 V8.1.1, April 2001.

[2] Y.Ephraim and D.Malah, ”Speech enhancement using minimum mean-square error short-time spectral amplitude estimator”, IEEE Trans. vol.ASSP-32, no.6, pp.1109-1121, Dec. 1984. mean-square error log-spectral amplitude estimator”, IEEE Trans. vol.ASSP-33, no.2, pp.443-445, April 1985.

[3] T.Lotter and P.Vary, ”Noise reduction by joint maximum a posteriori spectral amplitude and phase estimation with super-gaussian speech modeling”, Proc. EUSIPCO-04, pp.1447-60, Sep. 2004.

[4] M.Katou, A.Sugiyama and M.Serizawa, ”Noise suppression with high speech quality based on weighted noise estimation and MMSE STSA”, IEICE (Japan) Trans. Fundamental, vol.E85-A, no.7, pp.1710-1718, July 2002.

[5] K.Nakayama, H.Suzuki and A.Hirano, ”Improved methods for noise spectral estimation and adaptive spectral gain control in noise spectral suppressor”, Proc. ISPACS’07, Xiamen, China, pp.97-100, Dec. 2007.

[6] J.Sohn and N.Kim, ”Statistical model-based voice activity detection”, IEEE signal Processing Letter, vol.6, no.1, pp.1-3, 1999.

[7] I.Cohen and B.Berdugo, ”Speech enhancement for non-stationary noise environments”, Signal Processing, vol.81, no.11, pp.2403-2418, Nov. 2001.

[8] I.Cohen and B.Berdugo, ”Noise estimation by minima controlled recursive averaging for robust speech enhancement”, IEEE Signal Processing, Lett. vol.9, no.1, pp.12-15, 2002.

[9] S.Rangachari, P.Loizou and Y.Hu, ”A noise estimation algorithm with rapid adaptation for highly nonstationary environments”, Proc. IEEE ICASSP’04, pp.305-308, 2004.

[10] B.F.Wu and K.C.Wang, ”Noise spectrum estimation with entropy-based VAD in non-stationary environments”, IEICE (Japan) Trans. Fundamentals, vol.E89-A, no2, pp.479-485, Feb. 2006.

[11] A.Varga and H.J.M.Steeneken, ”Assessment for automatic speech recognition:II.NOISEX-92:A database and an experiment to study the effect of additive noise on speech recognition systems”, Speech Commun., vol.12, no.3, pp.247-251, July 1993.