A Noise Spectral Estimation Method Based on 2-Dimension Dynamically VAD Used in Noise Spectral Suppression

Sarayut TUNGPONTAWEE Kenji NAKAYAMA Akihiro HIRANO
Graduate School of Natural Science and Technology, Kanazawa University
Kanazawa, 920-1192, Japan
E-mail: sarayut@leo.ec.t.kanazawa-u.ac.jp, 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. In conventional method, speech, quasi-speech and non-speech frames are detected by using the entropy-based voice activity detector (VAD). A variable two thresholds and a statically variance comparison have been proposed for the conventional VAD to discriminate the noisy speech into three categories. The noise is estimated by using the noisy speech spectrum in the non-speech frames and weighted noisy speech spectrum in the quasi-speech frame and the speech frame.

In proposed method, speech detection using VAD based on two-dimension threshold of entropy and variance spectrum is proposed to improve the performance of speech/quasi-speech/non-speech discrimination. T-shape form discrimination technique and a dynamically calculate instantaneous value of variance spectrum for threshold are newly introduced for the VAD. Furthermore, an adaptive parameter is proposed to stabilize the entropy used in VAD. The weight function used in the quasi-speech frame and the speech frame are 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 proposed method can improve a noise spectral estimation error and a segmental SNR for noise is not changed condition (input SNR 3dB) and noise is changed condition (from babble noise (input SNR 6dB) to six noises (input SNR 2dB))

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. Several noise spectral estimation methods have been proposed for the purpose.

In this paper, a new noise spectral estimation method based on Voice Activity Detection (VAD) is proposed. A two-dimension thresholds of entropy and variance spectrum and a adaptive parameter 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 three kinds of frames. Computer simulations by using speech signal and many kinds of noises will be shown.

2 Spectral Suppression Noise Canceller

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

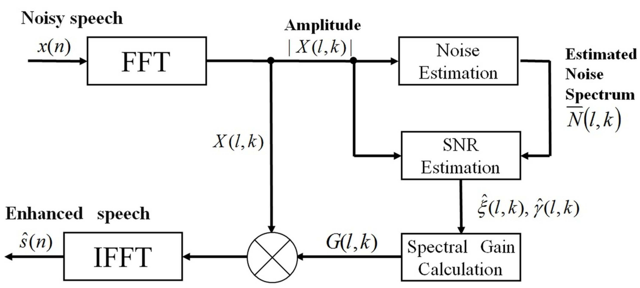

Figure 1: Blockdiagram of spectral suppression noise canceller

and $x(n)$ be clean speech, noise and noisy speech, respectively.

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

The Fourier transforms 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 to this method.

3.1 Voice Activity Detector

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 non-speech frames. Actually, $H_{av}(0)$ is scaled by constant c (<1)

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

$H(l)$ 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:

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

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

4 A New Noise Spectral Estimation Method

4.1 New Adaptive parameter

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 multiply constant α and maximum of $|X(l,k)|$ as follows:

$$P_{new}(l,k) = \frac{X_{energy}(l,k)+C_{new}(l)}{\sum_{k=1}^{2M} X_{energy}(l,k)+C_{new}(l)} \tag{13}$$

$$C_{new}(l) = \alpha \times \max_k\{|X(l,k)|\} \tag{14}$$

$$H_{new}(l) = -\sum_{k=1}^{2M} P_{new}(l,k)\cdot\log(P_{new}(l,k)) \tag{15}$$

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

In this paper, we consider the constant α in 3 different cases for the proposed method, $\alpha = 1, 2$ and 3.5. Simulation results will be shown in Sec.5.3.

4.2 Two-dimension threshold for VAD

The proposed method aims to improve the performance of speech/quasi-speech/non-speech detection. For this purpose, we introduce a new discrimination technique, 2-dimension threshold in T-shape form to classify the noisy speech into three categories. Two kind of thresholds are used for the new VAD, T_h for Entropy $H_{new}(l)$ and T_v for Variance Spectral $V(l)$, as follows:

$$T_h(l) = c_1 E[H_{new}(l)] \tag{16}$$

$$T_v(l) = c_2\frac{\max(V(l)) + median(V(l))}{2} \tag{17}$$

$$V(l) = \log|VAR(X(l,k))| \tag{18}$$

$E[H_{new}(l)]$ means an average over the recent five non-speech frames including the l-th frame. $\max(V(l))$ and $median(V(l))$ mean the maximum and median value of Variance Spectrum over the recent five either quasi-speech frames or non-speech frames including l-th frame. The initial value of $T_h(l)$ and $T_v(l)$ are determined by using an average value in the beginning 5 frames, which are assumed to be the non-speech frame. If $H_{new}(l) > T_h(l-1)$, then $T_h(l)$ is updated by Eqs.(16). If $V(l) < c_3T_v(l-1)$ and if $V(l) > median(V(l))$, then $T_v(l)$ is updated by Eqs.(17). c_1, c_2 and c_3 are 0.98, 1.1 and 1.02 respectively, which are determined by experience.

We consider how to classify the noisy speech into 3 categories in two-dimension of Entropy and Variance Spectral by using T-shape form discrimination technique as

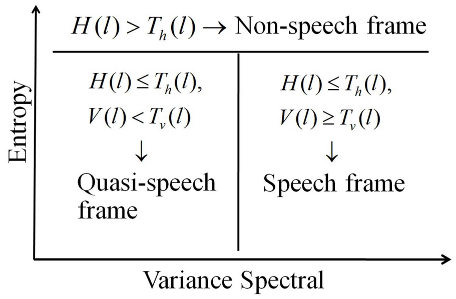

Figure 2: T-shape form discrimination

shown in Fig.2. If $H(l) > T_h$, then current frame is classified as the non speech frame. If $H(l) \leq T_h(l)$ and $V(l) < T_v(l)$, then current frame is classified as the quasi-speech frame. Finally, if $H(l) \leq T_h(l)$ and $V(l) \geq T_v(l)$, then current frame is classified into the speech frame.

4.3 Noise Spectral Estimation Method

4.3.1 *Non-speech frame*

We employ the conventional noise spectral estimation method [12] to estimate the noise spectrum by using the noisy speech spectrum itself in the non-speech frame as follows:

$$\bar{N}(l,k) = |X(l,k)|^2 \tag{19}$$

4.3.2 *Speech Frame*

In conventional method, the weighted noise spectral estimation method [4], [5] is applied for noise spectral estimation in the speech frame. 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.3. The noisy speech spectrum is reduced in the high SNR region in order to avoid over estimation of the noise spectrum. The weighted spectrum is expressed by

$$z(l,k) = W(l,k)|X(l,k)|^2 \tag{20}$$

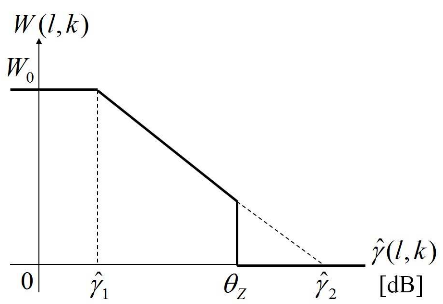

Figure 3: Weight function $W(l,k)$ for noisy speech

In order to improve the noise spectrum estimation in the speech frame from the conventional method [12], the parameters are newly determined by experience as follows: $W_0 = 1, \hat{\gamma}_1 = 0$dB, $\hat{\gamma}_2 = 10$dB and $\theta_z = 9.5$dB.

The beginning several frames are regarded as the non-speech frames, and $|X(l,k)|$ is directly used to estimate the noise spectrum. After that, the noise spectrum is estimated by averaging $z(l,k)$ over several frames. Figure 3 means that if $\hat{\gamma}(l,k) < \hat{\gamma}_1$, then $|X(l,k)|$ is used, and if $\theta_z < \hat{\gamma}(l,k)$, then $|X(l,k)|$ is not used to estimated the noise spectrum. Further more, when $\hat{\gamma}_1 < \hat{\gamma}(l,k) < \theta_z, |X(l,k)|$ is gradually reduced along $\hat{\gamma}(l,k)$ in estimating the noise spectrum [4].

4.3.3 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.3 is modified from the conventional approach in order to suppress over estimation of the noise spectrum more accurately. For this purpose, the parameters are newly determined by experience as follows: $W_0 = 0.8, \hat{\gamma}_1 = 7$dB, $\hat{\gamma}_2 = 10$dB and $\theta_z = 9.5$dB.

5 Simulations and Discussion

5.1 Conventional Methods

The conventional method [12] are employed for comparison. In this method, the non-speech, quasi-speech and the speech frame are discriminated by using Entropy based VAD. The proposed method introduces a new algorithm, two-dimension threshold for VAD, and modifies the weight function in the quasi-speech frame and the speech frame. Furthermore, in both 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*

$$\varepsilon(l) = 10\log_{10}\left(\frac{\sum_{k=0}^{M}\left||N(l,k)|^2 - |\bar{N}(l,k)|^2\right|}{\sum_{k=0}^{M}|N(l,k)|^2}\right) \tag{21}$$

$$\bar{\varepsilon} = \frac{1}{L}\sum_{l=1}^{L}\varepsilon(l) \tag{22}$$

L is the number of all frames. The smaller value of $\bar{\varepsilon}$ 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.

$$SNR_{seg} = \frac{10}{L}\sum_{l=0}^{L-1}\log_{10}\frac{\sum_{n=N_l}^{N_l+N-1}s^2(n)}{\sum_{n=N_l}^{N_l+N-1}(s(n)-\hat{s}(n))^2} \tag{23}$$

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:

$$LSD = \frac{1}{J}\sum_{l=1}^{J}\left(\frac{1}{2M}\sum_{k=1}^{2M}\left(\log\frac{|S(l,k)|+\delta}{|\hat{S}(l,k)|+\delta}\right)^2\right)^{\frac{1}{2}} \quad (24)$$

2M is a frame length, 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 true noise spectrum. The ideal noise suppressed output signal is obtained by

$$\hat{s}(n) = IFFT[G_{tl}(l,k)X(l,k)] \quad (25)$$

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 method and the proposed method. Six kinds of noises are used, that is White, Babble, Car, Pink, Factory and Tank which are provided by [11].

Table 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.

Table 2, 3, 4 and 5 show simulation results for the conventional method [12] and the proposed method when α is set to be 1, 2 and 3.5, respectively. The input SNR_{seg} is set to be 3dB and 9dB. Regarding $\bar{\varepsilon}$ and SNR_{seg}, the proposed method in all cases of α can improve better results when the Input SNR_{seg} is set to be 3dB. Its LSD are almost the same as those of the conventional method. However, when the Input SNR_{seg} is set to be 9dB, the results of the proposed method in all cases of α are almost the same as the results of the conventional method.

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 babble noise (Input SNR_{seg} = 6dB) to six kinds of noises (Input SNR_{seg} = 2dB), are shown in Table 6. 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}	3	9
White	9.42	13.3
Babble	11.9	15.6
Car	15.5	18.4
Pink	12.1	15.6
Factory	12.0	15.4
Tank	15.5	18.5

Table 2: Noise reduction and reduction by conventional method[12]. Noise is not changed.

	$\bar{\varepsilon}$		SNR_{seg}		LSD	
SNR_{seg}	3	9	3	9	3	9
White	-3.66	-2.73	7.06	11.8	0.420	0.345
Babble	-2.05	-1.78	6.10	12.0	0.306	0.242
Car	-2.68	-1.89	9.30	14.1	0.268	0.205
Pink	-3.02	-2.15	7.69	12.8	0.291	0.221
Factory	-2.83	-1.96	7.62	12.8	0.291	0.229
Tank	-3.47	-1.95	9.58	14.0	0.254	0.202

Table 3: Noise reduction and reduction by proposed method. Noise is not changed. $\alpha = 1$

	$\bar{\varepsilon}$		SNR_{seg}		LSD	
SNR_{seg}	3	9	3	9	3	9
White	-3.84	-2.69	7.36	11.9	0.422	0.355
Babble	-2.17	-1.67	6.59	12.2	0.313	0.224
Car	-3.25	-1.84	9.71	14.1	0.271	0.208
Pink	-3.32	-2.02	8.43	13.2	0.281	0.219
Factory	-3.06	-1.98	8.23	13.0	0.286	0.224
Tank	-3.39	-2.19	9.91	14.6	0.260	0.194

Table 4: Noise reduction and reduction by proposed method. Noise is not changed. $\alpha = 2$

	$\bar{\varepsilon}$		SNR_{seg}		LSD	
SNR_{seg}	3	9	3	9	3	9
White	-3.86	-2.63	7.35	11.9	0.422	0.355
Babble	-2.45	-1.79	6.85	12.3	0.314	0.224
Car	-3.30	-1.84	9.61	14.1	0.271	0.208
Pink	-3.41	-2.02	8.46	13.2	0.281	0.219
Factory	-3.11	-1.98	8.24	13.0	0.287	0.224
Tank	-3.43	-2.19	9.88	14.6	0.260	0.194

Table 5: Noise reduction and reduction by proposed method. Noise is not changed. $\alpha = 3.5$

	$\bar{\varepsilon}$		SNR_{seg}		LSD	
SNR_{seg}	3	9	3	9	3	9
White	-3.86	-2.79	7.36	12.0	0.422	0.355
Babble	-2.59	-1.77	6.90	12.3	0.311	0.239
Car	-3.43	-1.81	9.84	14.0	0.268	0.208
Pink	-3.43	-1.99	8.49	13.2	0.281	0.219
Factory	-3.11	-1.91	8.25	13.0	0.286	0.224
Tank	-3.43	-1.94	9.88	14.4	0.261	0.196

Table 6: Noise reduction by ideal method. Noise is changed from Babble(6dB) to six kinds of noises(2dB) in non-speech frame ($\mathrm{SNR}_{seg}(1)$) and in speech frame ($\mathrm{SNR}_{seg}(2)$)

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

The simulation results of the conventional method [12] and the proposed method are shown in Tables 7 and 8. The numbers in the left column are the same numbers shown in the Table 6.

As shown in these tables. $\bar{\varepsilon}$ and SNR_{seg} can be improved by the proposed method in all cases of α. LSD is still the same as those of the conventional method [12].

6 Conclusions

In this paper, a new noise spectral estimation method, 2-dimension dynamically VAD algorithm and new adaptive parameter are proposed, which can correctly discriminate the non-speech, the quasi-speech and the speech frame. Through computer simulations by using six kinds of noises, the proposed method can reduce the normalized noise spectrum estimation error and can increase SNR_{seg} compared with the conventional method.

References

[1] "Minimum performance requirements for noise suppressor application to the AMR speech encode", 3GPPTS 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 er-

Table 7: Performance comparison. Noise is change in non-speech frame

	Conventional Method [12]			Proposed Method $\alpha = 1$		
	$\bar{\varepsilon}$	SNR_{seg}	LSD	$\bar{\varepsilon}$	SNR_{seg}	LSD
1	-2.87	8.76	0.341	-2.97	9.01	0.341
2	-2.07	8.09	0.294	-1.98	8.47	0.297
3	-2.46	9.10	0.280	-2.43	9.56	0.280
4	-2.56	8.83	0.281	-2.58	9.41	0.281
5	-2.20	8.54	0.287	-2.44	9.12	0.281
6	-2.66	9.94	0.268	-2.98	10.6	0.256

	Proposed Method $\alpha = 2$			Proposed Method $\alpha = 3.5$		
	$\bar{\varepsilon}$	SNR_{seg}	LSD	$\bar{\varepsilon}$	SNR_{seg}	LSD
1	-2.99	9.00	0.342	-2.99	9.00	0.342
2	-2.33	8.70	0.292	-2.26	8.70	0.293
3	-2.60	9.63	0.280	-2.62	9.60	0.279
4	-2.68	9.48	0.279	-2.70	9.49	0.279
5	-2.73	9.44	0.281	-2.63	9.24	0.281
6	-2.88	10.4	0.256	-2.86	10.4	0.256

Table 8: Performance comparison. Noise is change in speech frame

	Conventional Method [12]			Proposed Method $\alpha = 1$		
	$\bar{\varepsilon}$	SNR_{seg}	LSD	$\bar{\varepsilon}$	SNR_{seg}	LSD
1	-2.80	8.69	0.336	-2.83	9.01	0.336
2	-1.63	7.63	0.290	-1.84	8.33	0.291
3	-2.33	9.17	0.280	-2.46	9.48	0.281
4	-2.62	8.92	0.279	-2.51	9.26	0.279
5	-2.33	8.62	0.283	-2.17	9.00	0.284
6	-2.47	9.65	0.280	-2.68	10.1	0.274

	Proposed Method $\alpha = 2$			Proposed Method $\alpha = 3.5$		
	$\bar{\varepsilon}$	SNR_{seg}	LSD	$\bar{\varepsilon}$	SNR_{seg}	LSD
1	-2.90	9.05	0.336	-2.92	9.01	0.336
2	-2.16	8.53	0.289	-2.02	8.46	0.289
3	-2.52	9.60	0.281	-2.57	9.69	0.281
4	-2.75	9.38	0.279	-2.81	9.44	0.277
5	-2.56	9.19	0.284	-2.58	9.21	0.281
6	-2.78	10.2	0.274	-2.77	10.2	0.274

ror 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 pharse 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 enviroments", 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.E89A, 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.

[12] K.Nakayama, S.Higashi and A.Hirano, "A noise estimation method based on improved VAD used in noise spectral suppression under highly nonstationary noise environment", Proc. EUSIPCO-09, pp.2494-2498, August 2009.