Speech Noise Reduction System Based on Adaptive Filter with Variable Step Size Using Cross-Correlation

Koji Shimada*, Naoto Sasaoka*, Takafumi Takemoto*, Yoshio Itoh*, and Kensaku Fujii†

* Graduate school of Engineering, Tottori University
4-101 Koyama-minami, 680-8552, Tottori, Japan
Tel: +81-857-31-5698, Fax: +81-857-31-5698
E-mail: shimada@dacom2.ele.tottori-u.ac.jp
†Graduate school of Engineering, University of Hyogo
2167 Shosha, 671-2201, Himeji, Japan

***Abstract*— In order to reduce background noise in noisy speech, we have investigated a noise reduction method based on a noise reconstruction system (NRS) with an adaptive line enhancer (ALE). The NRS uses a noise reconstruction filter (NRF) estimating the background noise. In case a fixed step size for updating tap coefficients of the NRF is used, it is difficult to estimate the background noise accurately while maintaining the high quality of enhanced speech. In order to improve the estimation accuracy of noise, a variable step size is introduced to the NRF. In a speech section, the variable step size decreases so as to become no sensitivity for speech, on the other hand, increases to track the background noise in a non-speech section. From simulation results, we have verified that the proposed system can reduce the actual noise while maintaining the high quality of enhanced speech.**

I. INTRODUCTION

So far there have been a wide variety of noise reduction methods, such as adaptive microphone array [1], spectral subtraction (SS) [2] and adaptive noise canceling (ANC) [3], which reduce background noise in noisy speech. The adaptive microphone array can be considered as a directional microphone with a blind spot facing to the direction of arrival of noise. The existence of many noise sources causes the number of microphones to increase. On the other hand, SS is known as a method using only one microphone. However, musical noise are caused by residual error. In addition the processing delay occurs due to frame processing. Furthermore, the SS requires a prior estimation of noise spectrum. This implies that the SS requires a speech/non-speech section detector in noisy environments. The ANC should estimate the pitch period of the voice in a low signal to noise ratio (SNR) environment, though it does not need prior estimation of the noise spectrum.

We have proposed a noise reconstruction system (NRS) with an adaptive line enhancer (ALE) [4] as a noise reduction system. The NRS uses a linear prediction error filter (LPEF) and a noise reconstruction filter (NRF). The NRS assumes that the background noise is generated by exciting a linear system with a white signal. First, the white signal is estimated by the ALE and LPEF. Next, the background noise is reconstructed from the white signal by estimating the linear system at the NRF. However, in case a fixed step size for updating tap coefficients of a NRF is used, it is difficult to reduce non-stationary background noise while maintaining the high quality of enhanced speech.

A variable step size is used in order to improve ability of noise reduction more. Many variable step size methods have been proposed [5]-[7]. However, since Shin's method [5] assumes that the disturbance is stationary white Gaussian noise, it cannot be used when disturbance is a non-stationary signal like speech and SNR varies. On the other hand, the variable step size, which is possible to track not only a non-stationary system but also variation of SNR, has been proposed [6],[7]. However, the variable step size can not be introduced to the NRF because the disturbance is assumed to be white Gaussian noise. Therefore, a novel variable step size is required.

In this paper, we introduce a variable step size [8], which can be used in the case that the disturbance is speech, to the NRF. The variable step size makes use of cross-correlations between input signals and an enhanced speech signal. In a speech section, the variable step size decreases so as to become no sensitivity for speech, on the other hand, increases to track the background noise in a non-speech section. Therefore, the proposed system can reduce the background noise effectively.

This paper is organized as follows. In section 2, a conventional noise reconstruction system with ALE is described. A variable step size is introduced to the conventional NRS with ALE in section 3. we then show experimental results of the proposed method and the listening test in section 4 and 5 respectively. In section 6, we conclude our paper.

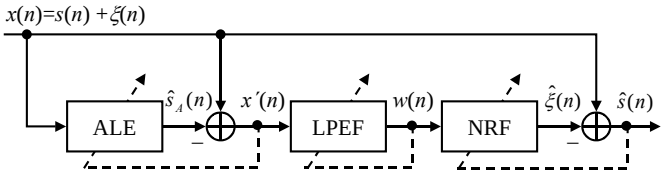

Fig. 1. Structure of noise reconstruction system with ALE.

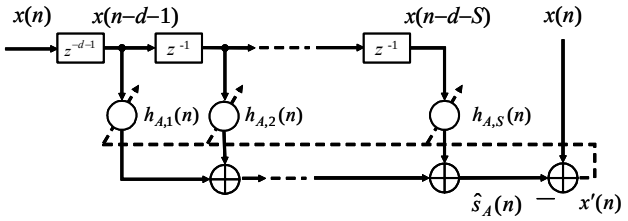

Fig. 2. Adaptive line Enhancer (ALE).

II. Noise Reconstruction System with ALE

A . The principle of noise reconstruction system

A noise reduction method based on the noise reconstruction system with the ALE [4] is shown in Fig. 1. $x(n)$, which is represented as a following equation, is noisy speech.

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

where $s(n)$ and $\xi(n)$ are respectively clean speech and background noise. $\hat{s}_A(n)$ and $x'(n)$ represent an output of the ALE and an input of LPEF respectively. $w(n)$, $\hat{\xi}(n)$ and $\hat{s}(n)$ are output of a LPEF, reconstructed noise and enhanced speech respectively. The transfer function of the ALE, the LPEF and the NRF are respectively represented as $H_{ALE}(z)$, $H_{LPEF}(z)$ and $H_{NRF}(z)$, respectively. The structure of ALE and LPEF are respectively shown in Fig.2 and 3. These transfer functions are given by

$$H_{ALE}(z) = \sum_{k=1}^{S} h_{A,k}(n) z^{-d-k} \tag{2}$$

$$H_{LPEF}(z) = 1 - \sum_{k=1}^{L} h_k(n) z^{-k} \tag{3}$$

$$H_{NRF}(z) = \sum_{k=0}^{M} h'_k(n) z^{-k} \tag{4}$$

where $h_{A,k}(n)$, $h_k(n)$ and $h'_k(n)$ are respectively the k-th tap coefficients of the ALE, the LPEF and the NRF. d is decorrelation parameter of the ALE.

First, a speech signal is estimated by the ALE. Assuming that there is enough decorrelation parameter d to fade the autocorrelation of the background noise and speech

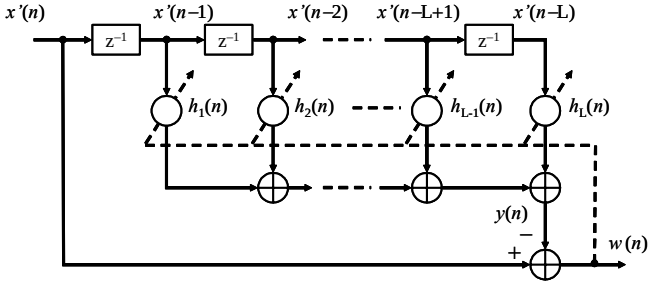

Fig. 3. Linear prediction error filter (LPEF).

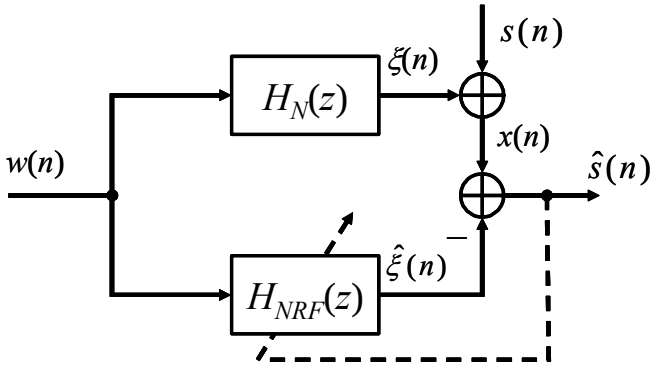

Fig. 4. System identification model.

signals delayed by a pitch period of voiced sound are included in the tap input of the ALE, the ALE can estimate the speech signal. The distribution of pitch frequency is known to be logarithmical Gaussian type. The average and standard deviation of pitch frequency are 125Hz and 20.5Hz for a male respectively. About female pitch frequency, the average and standard deviation are 250Hz and 41.0Hz respectively [9]. Therefore, it is calculated that pitch frequency ranges from 78.2Hz to 399.77Hz for occurring probability 99.7%. Then a pitch period ranges from 20 samples to 102 samples by the signal sampled by 8kHz. Hence, the tap inputs of ALE are set for $x(n-20)$, $x(n-21)$,…, $x(n-102)$ [10]. $x'(n)$ is obtained by subtracting $\hat{s}_A(n)$ from $x(n)$ and occupied by the background noise.

The whitened noise $w(n)$ is obtained by the LPEF. Tap coefficients of a LPEF converge in such a way that a prediction error signal whitens [11]. Since a speech signal can be represented as a stationary and periodic signal in a short time interval, a residual speech signal in $x'(n)$ will be predicted with a linear predictor. On the other hand, assuming that background noise is generated by exciting a linear system $H_N(z)$ by a white signal, the background noise can be made white by a LPEF.

We then consider the background noise is reconstructed from a whitened signal by a NRF. Assuming that a white signal generates background noise by exciting a linear system $H_N(z)$, the background noise can be reconstructed from a whitened signal $w(n)$ by estimating the transfer function of a

noise generating system. This estimation is performed by a system identification model illustrated in Fig.4, where $\xi(n)$, $s(n)$ and $\hat{s}(n)$ are a desired signal, a disturbance and an estimation error signal respectively. Finally, an enhanced speech signal $\hat{s}(n)$ is obtained by subtracting a reconstructed noise $\hat{\xi}(n)$ from $x(n)$.

Normalized Least Mean Square (NLMS) algorithm is used for updating the tap coefficients of the NRF as follows [11]:

$$\mathbf{h}'(n+1) = \mathbf{h}'(n) + \mu' \frac{\hat{s}(n)\mathbf{w}(n)}{\|\mathbf{w}(n)\|^2} \tag{5}$$

$$\mathbf{h}'(n) = [h'_0(n), h'_1(n), ..., h'_M(n)]^T \tag{6}$$

$$\mathbf{w}(n) = [w(n), w(n-1), ..., w(n-M)]^T . \tag{7}$$

where $\mathbf{h}'(n)$, $\mathbf{w}(n)$ and μ' represent a tap coefficient vector, a tap input vector and a step size respectively. T and $\|\cdot\|$ represent a transposition of a vector and a norm respectively. However, residual speech might be included in the whitened noise, which degrades estimation accuracy of a noise generating system in a speech section.

III. Variable Step Size Based on Correlation

In case a fixed step size for updating tap coefficients of a NRF is used, it is difficult to reduce non-stationary background noise while maintaining the high quality of enhanced speech. In this section, we introduce a variable step size to the NRF in order to improve the above-mentioned problem. In a speech section, a variable step size decreases so as not to estimate speech, on the other hand, increase to track background noise in a non-speech section.

From Eq. (5), we get

$$\mathbf{h}'_{\mathrm{o}} - \mathbf{h}'(n+1) = \mathbf{h}'_{\mathrm{o}} - \mathbf{h}'(n) - \mu' \frac{\hat{s}(n)\mathbf{w}(n)}{\|\mathbf{w}(n)\|^2} \tag{8}$$

where $\mathbf{h}'_{\mathrm{o}}$ denotes the ideal tap coefficient vector of NRF. Thus,

$$\begin{aligned} &\{\mathbf{h}'_{\mathrm{o}} - \mathbf{h}'(n+1)\}^T \{\mathbf{h}'_{\mathrm{o}} - \mathbf{h}'(n+1)\} \\ &= \{\mathbf{h}'_{\mathrm{o}} - \mathbf{h}'(n+1)\}^T \{\mathbf{h}'_{\mathrm{o}} - \mathbf{h}'(n) - \mu' \frac{\hat{s}(n)\mathbf{w}(n)}{\|\mathbf{w}(n)\|^2}\} \end{aligned} \tag{9}$$

is obtained. From Eq. (9),

$$\begin{aligned} &\{\mathbf{h}'_{\mathrm{o}} - \mathbf{h}'(n+1)\}^T \{\mathbf{h}'_{\mathrm{o}} - \mathbf{h}'(n+1)\}\|\mathbf{w}(n)\|^2 \\ &= \{\mathbf{h}'_{\mathrm{o}} - \mathbf{h}'(n)\}^T \{\mathbf{h}'_{\mathrm{o}} - \mathbf{h}'(n)\}\|\mathbf{w}(n)\|^2 \\ &\quad - 2\mu'\hat{s}(n)\{\mathbf{h}'_{\mathrm{o}} - \mathbf{h}'(n)\}^T \mathbf{w}(n) + \mu'^2 \hat{s}^2(n) \end{aligned} \tag{10}$$

is given. The following equation is given by traking the statistical expectation of both sides of Eq. (10).

$$\begin{aligned} &E[\|\mathbf{w}(n)\|^2 \{\mathbf{h}'_{\mathrm{o}} - \mathbf{h}'(n+1)\}^T \{\mathbf{h}'_{\mathrm{o}} - \mathbf{h}'(n+1)\}] \\ &= E[\|\mathbf{w}(n)\|^2 \{\mathbf{h}'_{\mathrm{o}} - \mathbf{h}'(n)\}^T \{\mathbf{h}'_{\mathrm{o}} - \mathbf{h}'(n)\}] \\ &\quad + \varepsilon \end{aligned} \tag{11}$$

where $E[\cdot]$ represents an expected value and

$$\begin{aligned} \varepsilon = &-2\mu' E[\hat{s}(n)\{\mathbf{h}'_{\mathrm{o}} - \mathbf{h}'(n)\}^T \mathbf{w}(n)] \\ &+ \mu'^2 E[\hat{s}^2(n)] . \end{aligned} \tag{12}$$

When the tap coefficients are converged adequately, ε is minimized. Therefore, from the following equation:

$$\left.\frac{\partial \varepsilon}{\partial \mu'}\right|_{\mu'=\mu'_0} = 0 , \tag{13}$$

the optimum step size of the NRF is represented as

$$\mu'_0 = \frac{E[\hat{s}(n)\{\mathbf{h}'_{\mathrm{o}} - \mathbf{h}'(n)\}^T \mathbf{w}(n)]}{E[\hat{s}^2(n)]} . \tag{14}$$

Additionally, following equations are given in a noise reconstruction system,

$$\{\mathbf{h}'_{\mathrm{o}} - \mathbf{h}'(n)\}^T \mathbf{w}(n) = \xi(n) - \hat{\xi}(n), \tag{15}$$

and

$$x(n) = s(n) + \xi(n) = \hat{s}(n) + \hat{\xi}(n) . \tag{16}$$

Therefore, from Eqs. (14), (15) and (16), the optimum step size of the NRF is obtained by

$$\mu'_0 = 1 - \frac{E[\hat{s}(n)x(n) - \hat{s}(n)\xi(n)]}{E[\hat{s}^2(n)]} . \tag{17}$$

Since an actual environment is non-stationary, the proposed variable step size at time n is defined by

$$\mu'(n) = 1 - \frac{E[\hat{s}(n)x(n) - \hat{s}(n)\xi(n)]}{E[\hat{s}^2(n)]} . \tag{18}$$

We can see that the variable step size of the NRF becomes small in a speech section and large in a non-speech section respectively. Because there is a possibility that the step size of the NRF becomes too large or too small when the estimation of speech is insufficient, we restrict the range of the variable step size. The range is from minimum step size $\mu_{\min}$ to maximum step size $\mu_{\max}$.

The cross-correlation and power included in Eq. (18) are approximated by

$$E[\hat{s}(n)x(n)] \approx R_{sx}(n)$$
$$= (1-\gamma_{sx})\hat{s}(n)x(n) + \gamma_{sx}R_{sx}(n-1) \quad (19)$$
$$E[\hat{s}(n)\xi(n)] \approx R_{s\xi}(n)$$
$$= (1-\gamma_{s\xi})\hat{s}(n)\xi(n)$$
$$+\gamma_{s\xi}R_{s\xi}(n-1) \quad (20)$$
$$E[\hat{s}^2(n)] \approx P_s(n)$$
$$= (1-\gamma_s)\hat{s}^2(n) + \gamma_s P_s(n-1) \quad (21)$$

where γ represents a forgetting factor. Since the cross-correlation $E[\hat{s}(n)\xi(n)]$ includes input background noise, we cannot obtain $E[\hat{s}(n)\xi(n)]$ in a real environment. Therefore, the estimation of the cross-correlation $E[\hat{s}(n)\xi(n)]$ is needed. In this paper, the input signal of the LPEF $x'(n)$ is used as the background noise $\xi(n)$. Thus, the estimation of the cross-correlation $R_{s\xi}(n)$ is given by

$$R_{s\xi}(n) = (1-\gamma_{s\xi})\hat{s}(n)x'(n) + \gamma_{s\xi}R_{s\xi}(n-1)\,. \quad (22)$$

IV. Simulation Results

A. Experimental conditions

The performance of the proposed noise reduction system was evaluated by computer simulations. All sound data prepared in simulations were sampled by 8kHz in 16bit resolution. The input signals were generated by artificially adding background noise to clean speech. As the speech, the female and male speeches recorded in Acoustic Society of Japan-Japanese Newspaper Article Sentences (ASJ-JNAS) were used. The number of samples of female and male voice were respectively 71,059 and 57,461. On the other hand, the man-made noise and actual noise were used as noise. The man-made noise generated by passing a white noise through a second order infinite impulse response (IIR) filter, which is shown in Fig.5. The transfer function $H_N(z)$ and filter coefficients of the second order IIR filter are given by

$$H_N(z) = \frac{1}{1 + C_1 z^{-1} + C_2 z^{-2}}\,, \quad (23)$$
$$C_1 = -2r\cos\theta \quad (24)$$

and

$$C_2 = r^2\,, \quad (25)$$

where C_1 and C_2 are filter coefficients. As man-made noise, the stationary noise (STN) and non-stationary noise (NSTN) were generated. For generating the stationary noise, r and θ

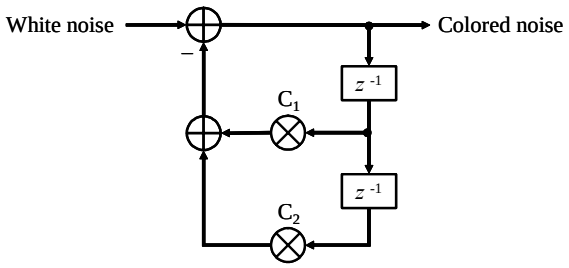

Fig. 5. 2^{nd} order IIR filter

TABLE I. Parameters of Filter

ALE (LMS)	Decorrelation parameter d	19
	Number of tap coefficients S	83
	Step size	0.18
LP (LMS)	Number of tap coefficients L	128
	Step Size	0.2
NRF (NLMS)	Number of tap coefficients M	128
	Forgetting factor γ_{sx}	0.94
	Forgetting factor γ_s	0.96
	Forgetting factor $\gamma_{s\xi}$	0.94
	Minimum step size μ_{min}	0
	Maximum step size μ_{max}	1.0

were set to 0.8 and 0.79, respectively. The non-stationary noise were generated by changing θ continuously as follows:

$$\theta(n) = \frac{1.97 - 0.79}{71059}(j-1) + 0.79, (1 \le j \le N). \quad (26)$$

On the other hand, the tunnel noise (Tunnel) and F16 cockpit noise (F16) were used as the actual noise. The tunnel noise was recorded inside the tunnel of an expressway. F16 cockpit noise is in the Noisex-92 database. Adaptive algorithms for updating tap coefficients are LMS and NLMS algorithm [11]. Table 1 shows each of the parameters that were used in the simulation.

SNR and Quality of Speech (QS) were used to evaluate the noise reduction ability and the quality of enhanced speech respectively. SNR_{in} and SNR_{out} represent the input and the output SNR respectively. These indices are defined as follows:

$$\mathrm{SNR}_{\mathrm{in}} = 10\log_{10}\frac{\sum_{j=1}^{N} s^2(j)}{\sum_{j=1}^{N} \xi^2(j)} \quad (27)$$

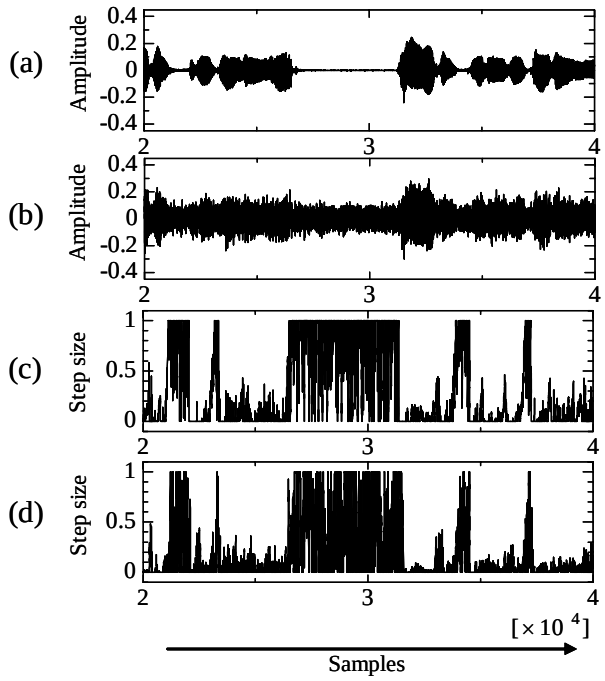

Fig. 6. Variable step size. (a) Clean speech. (b) Noisy speech.(SNR_{in} =0dB) (c) Variable step size with Eq.(20). (d) Variable step size with Eq.(22).

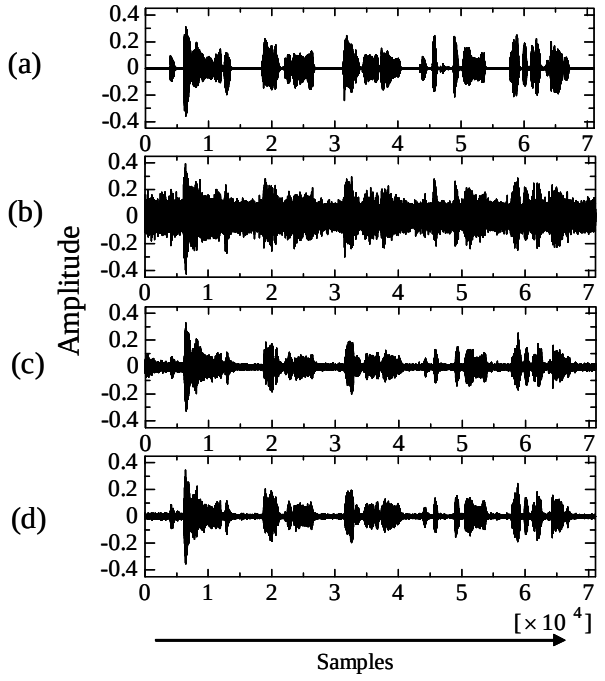

Fig.7. Simulation results of non-stationary noise for female speech. (a)Clean speech. (b)Noisy speech. (SNR_{in} =0dB)(c)Enhanced speech by NRS with ALE [4]. (SNR_{out}=9.2dB, QS=5.9dB) (d)Enhanced speech by proposed system.(SNR_{out} =11.2dB, QS=7.2dB)

$$\mathrm{SNR}_{\mathrm{out}} = 10\log_{10}\frac{\sum_{j=1}^{N}\hat{s}_s^2(j)}{\sum_{j=1}^{N}\hat{s}_\xi^2(j)} \tag{28}$$

and

$$\mathrm{QS} = 10\log_{10}\frac{\sum_{j=1}^{N}s^2(j)}{\sum_{j=1}^{N}\{s(j)-\hat{s}_s(j)\}^2} \tag{29}$$

where $\hat{s}_s(j)$ and $\hat{s}_\xi(j)$ are components of the speech and the noise included in the enhanced speech $\hat{s}(j)$, respectively . In Eq. (29), $\{s(j)-\hat{s}_s(j)\}$ represents the distortion of speech components due to a filter, thus QS increases as the quality of enhanced speech is increased.

B. Waveforms of variable step size

Fig. 6 shows the waveforms of the proposed variable step size when non-stationary noise (NSTN) was used as the background noise. Fig. 6(a) and 6(b) represent clean speech and noisy speech, respectively, in a 0 dB environment. The waveform of a variable step size with Eq. (20) and Eq. (22) are respectively shown in Fig. 6(c) and 6(d). The result of the proposed system is shown in Fig. 6(d). As the estimation of $E[\hat{s}(n)\xi(n)]$, Eq. (20) and Eq. (22) was respectively used in order to indicate the ideal step size and actual step size. Fig. 6(c) and 6(d) show that the step size converges on approximately 1 in a non-speech section and the step size decreases in a speech section.

C. Waveforms of non-stationary noise reduction

Fig. 7 and Fig. 8 show the noise reduction results when non-stationary noise was used in the simulation for female and male speech. The non-stationary noise (NSTN), which has been explained in section ⅣA, was used as the non-stationary noise. (a) and (b) in Fig. 7 and 8 represent clean speech and noisy speech, respectively, in a 0 dB environment. The noise reduction results by the conventional system [4] and a proposed system are illustrated in Fig. (c) and (d), respectively. Comparing Fig. 7(d) with Fig. 7(c), SNR_{out} and the QS improved by about 2.0dB and 1.3 dB, respectively. Comparing Fig. 8(d) with Fig. 8(c), SNR_{out} and the QS improved by about 1.4dB and 1.7dB, respectively. We have therefore verified that the proposed system has the potential for reducing the non- stationary noise while maintaining the high quality of enhanced speech for female and male speech.

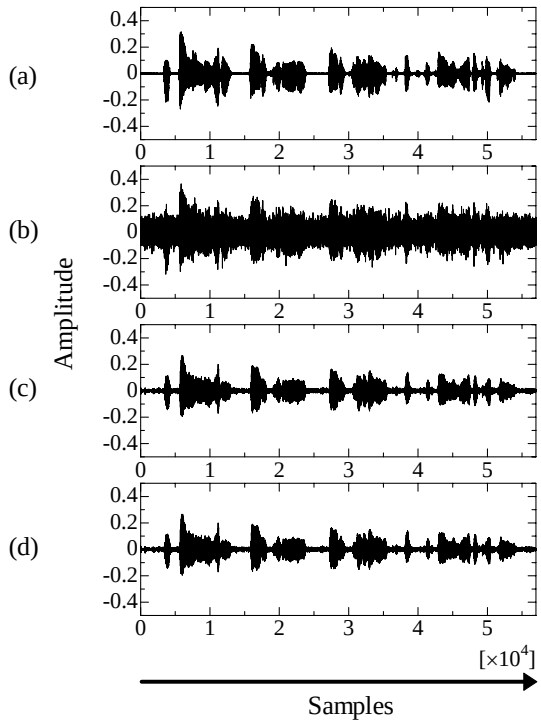

Fig.8. Simulation results of non-stationary noise for male speech. (a)Clean speech. (b)Noisy speech. (SNR_{in} =0dB)(c)Enhanced speech by NRS with ALE [4]. (SNR_{out}=11.5dB, QS=6.0dB) (d)Enhanced speech by proposed system.(SNR_{out} =12.9dB, QS=7.7dB)

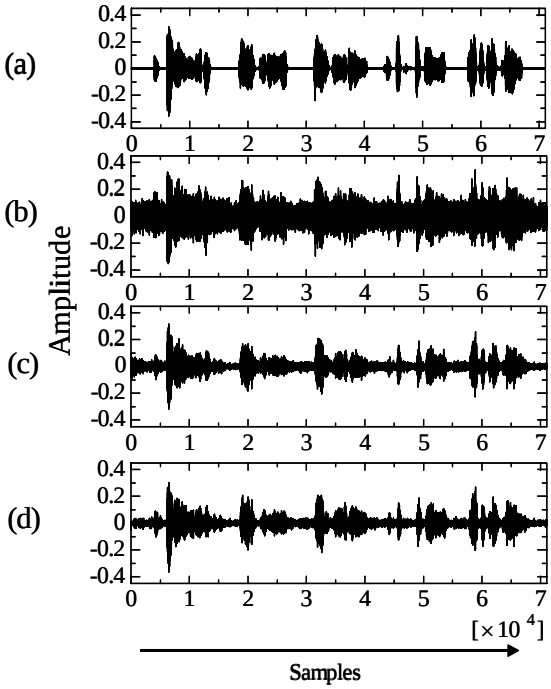

Fig.9. Simulation results of actual noise for female speech. (a)Clean speech. (b)Noisy speech. (c)Enhanced speech by NRS with ALE [4]. (SNR_{out} =6.5dB, QS=5.6dB) (d)Enhanced speech by proposed system. (SNR_{out} =7.3dB, QS=7.0dB

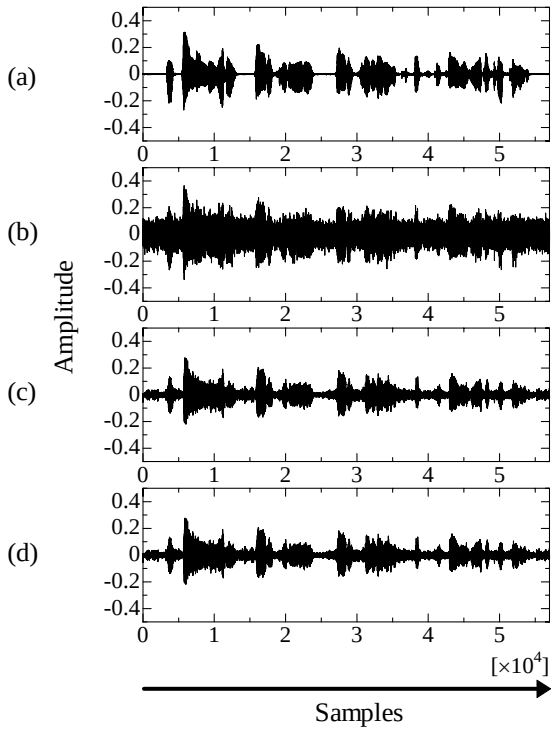

Fig.10. Simulation results of actual noise for male speech. (a)Clean speech. (b)Noisy speech. (c)Enhanced speech by NRS with ALE [4]. (SNR_{out} =6.9dB, QS=5.7dB) (d)Enhanced speech by proposed system. (SNR_{out} =7.5dB, QS=7.3dB

D . Waveforms of actual noise reduction

Fig. 9 and 10 shows the noise reduction results with the actual noise for female and male speech, respectively. As the actual noise, the tunnel noise (Tunnel) which has been explained in section ⅣA, was used. Fig. (a) and (b) in Fig. 9 and 10 represents clean speech and noisy speech , respectively, in a 0 dB environment. The noise reduction results by the conventional system [4] and a proposed system are shown in Fig. (c) with (d), respectively. Comparing Fig. 9(d) with Fig. 9(c), SNR_{out} and the QS are improved by about 0.8dB and 1.4 dB, respectively. Fig. 10(d) with Fig. 10(c), SNR_{out} and the QS are improved by about 0.6dB and 1.6 dB, respectively. We have therefore verified that the proposed system can reduce the actual noise while maintaining the high quality of enhanced speech for female and male voice.

E . Output SNR and quality of speech

Fig. 11, 12, 13 and 14 show noise reduction ability for –5dB to 20dB of input SNR. As the background noise, the stationary noise (STN), non-stationary noise (NSTN), tunnel noise (Tunnel) and F16 cockpit noise (F16) was used, which have been explained in section Ⅳ A, SNR_{out} and QS for –5dB to 20dB of SNR_{in} are respectively shown in Fig.11, 12, 13 and 14 for female and male speech. Comparing the proposed system (PS) with conventional system (CS) [4], SNR_{out} and QS are always higher. In addition, since the NRF in the conventional system

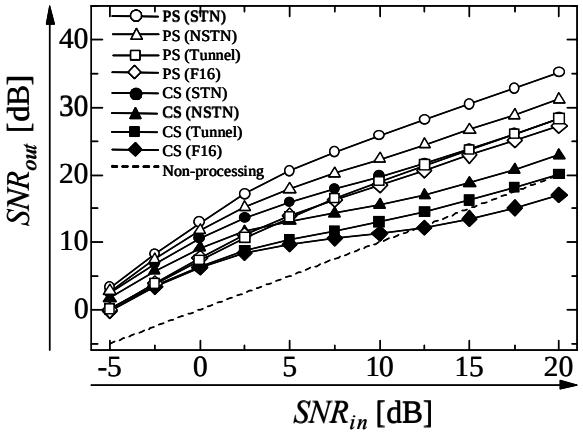

Fig.11. SNR_{out} performances (female).

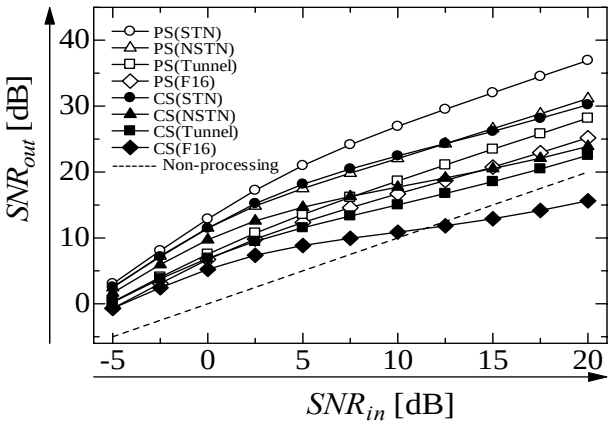

Fig.13. SNR_{out} performances (male).

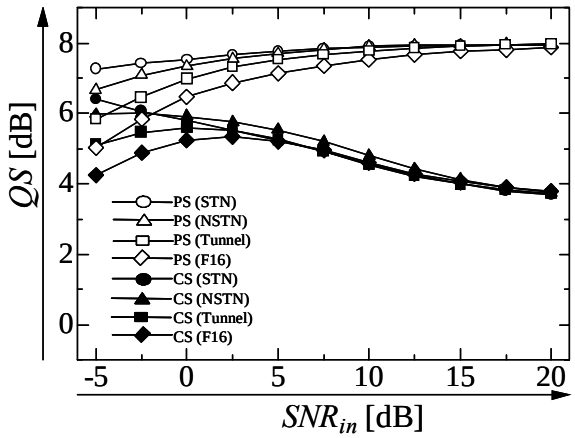

Fig.12. QS performances (female).

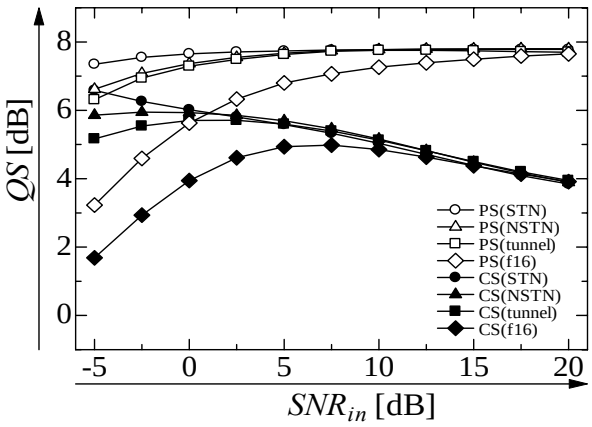

Fig.14. QS performances(male)

tracks a speech component, QS of the conventional system decreases as SNR_{in} increases. On the other hand, QS of the proposed system is improved as SNR_{in} increases. This results represents that the proposed system sets better step size to SNR_{in}. These results show that noise reduction ability of the proposed system is higher than that of conventional system for female and male voice.

V. LISTENING TEST

Subjective measure was also made to evaluate noise reduction ability. A method of pair comparisons between a noisy speech and enhaced speech was carried out as a listening test. The speech signal that was used the female and male speech, which has been explained in section Ⅳ. As the background noise, the stationary noise (STN), non-stationary noise (NSTN), tunnel noise (Tunnel) and F16 cockpit noise (F16) was used, which has also been explained in section Ⅳ.

SNR_{in} was set to 0 dB. In total, 50 subjects, who were students as the Faculty of were used. Each subject was instructed to compare the enhanced speech with the noisy speech, and then evaluate the noise reduction ability and Quality of enhanced speech of the system with the five grades shown in Table Ⅱ. Subjects listened to the noisy speech and the enhanced speech with headphones. The headphones that were used were dynamic closed type headphones (SONY MDR-Z700DJ). The listening test was carried out for NRS with ALE [4] and the proposed method.

Mean opinion score (MOS) is obtained by the method of pair comparisons between the noisy speech and the enhanced speech. The results of the listening test is shown in Table Ⅲ. The noise reduction ability is improved for the stationary noise (STN). About the results of the noise reduction ability, it is confirmed that there is significant difference between the proposed system and the conventional system for the stationary noise (STN) from t-test with a significant level 5%. However, it is not confirmed that there is significant difference between the proposed system and conventional system for other background noise. From the results of quality

TABLE II. SCORING SCALE

Score	Quality
5	Excellent
4	Good
3	Fair
2	Poor
1	Bad

TABLE III. RESULTS OF LISTENING TEST.

NOISE REDUCTION ABILITY

	Conventional system [4]		Proposed system	
	Average	Variance	Average	Variance
STN	3.60	0.52	4.02	0.58
NSTN	3.24	0.66	3.54	0.85
Tunnel	3.48	0.77	3.42	0.92
F16	3.80	0.64	3.96	0.76

QUALITY OF SPEECH

	Conventional system [4]		Proposed system	
	Average	Variance	Average	Variance
STN	2.98	0.99	3.56	1.01
NSTN	2.68	0.66	3.08	0.79
Tunnel	2.76	0.78	3.22	0.69
F16	2.40	0.64	2.96	0.96

of speech, the average scores of proposed system are higher than the conventional system. About the results of the quality of speech, it is confirmed that there is significant difference between the proposed system and conventional system for all the noise from t-test with a significant level 5%. Based on the listening test, we verified that the proposed systems have the potential of reducing the background noise while maintaining the high quality of enhanced speech.

VI. CONCLUSION

We examine the improvement of the noise reduction method using a NRS with an ALE. The variable step size based on cross-correlations between input signals and an enhanced speech signal has been introduced to the NRF. In a speech section, the variable step size decreases so as to become no sensitivity for speech, on the other hand, increases to track the background noise in a non-speech section. Thus, the proposed system can reduce background noise while maintaining the high quality of enhanced speech.

From simulation results, comparing the proposed system with conventional system, SNR_{out} and the QS are improved, for the actual noise. From the listening test, it is confirmed that the proposed system improves the quality of speech. we have therefore verified that the proposed noise reduction method is able to reduce the background noise effectively. In a future work, we will research the improvement of quality of the enhanced speech.

REFERENCES

[1] J. Ohga, Y.Yamazaki and Y. Kaneda, Acoustic System and Digital Processing for Them, IEICE, Tokyo, 1995.

[2] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-27, no. 2, pp.113–120, Apr. 1979.

[3] M.R. Sambur, "Adaptive noise canceling for speech signals," IEEE Trans. Acoust., Speech, Signal Processing, vol.ASSP-26, no.5, pp.419–423, Oct. 1978.

[4] N. Sasaoka, Y. Itoh and K. Fujii, "A noise reduction method for non-stationary noise based on noise reconstruction system with ALE," IEICE Trans. Fundamentals, vol.E88-A, no.2, pp.593–596, Feb. 2005.

[5] H. Shin, A.H. Sayed, W.J. Song, "Varivable step-size NLMS and affine projection algorithms," IEEE Signal Processing letters, vol.11, no.2, pp.132–135, Feb. 2004.

[6] K.Takahashi, I. Sasase and A. Mori, "Estimation of noise level and variable step gain algorithm for improvement of initial convergence characteristics," IEICE Trans. Fundamentals, vol.J76-A, no.12, pp.1704–1713,Dec. 1993.

[7] A. Kawamura, K. Fujii, Y. Itoh and Y.Fukui, "A noise reduction method based on linear prediction with variable step-size," IEICE Trans. Fundamentals, vol.E88-A, no.4, pp. pp.855–861, Apr 2005.

[8] N. Sasaoka, M. Watanabe, Y. Itoh and K. Fujii, "A study on noise reduction method based on LPEF and system identification with step size control," Proc. IEEE 12th Digital Signal Processing Workshop, pp.576–579, Sep. 2006.

[9] S. Furui, Digital Speech Processing, Tokai University Press, Tokyo, 1985.

[10] N. Sasaoka, K. Sumi, Y. Itoh, K. Fujii and A. Kawamura, "A noise reduction system for wideband and sinusoidal noise based on adaptive line enhancer and inverse filter," IEICE Trans. Fundamentals, Vol.E89-A, No.2, pp.503–510, Feb. 2006.

[11] S. Haykin, Adaptive Filter Theory, Prentice-Hall, New Jersey, 1996.