International Symposium on Communications
and Information Technologies 2004 (ISCIT 2004)
Sapporo, Japan, October 26- 29, 2004

A Noise Reduction Method Using Linear Predictor with Variable Step-Size

Arata Kawamura*, Youji Iiguni*, and Yoshio Itoh†
*Graduate School of Engineering Science
Osaka University, Toyonaka, Osaka, 560-8531, Japan
Tel: +81-6-6850-6377, Fax: +81-6-6850-6377
E-mail: kawamura@sys.es.osaka-u.ac.jp
†Faculty of Engineering
Tottori University, Tottori, 680-8552, Japan
Tel: +81-857-31-5698, E-mail: itoh@ele.tottori-u.ac.jp

***Abstract*—A white noise reduction method based on linear prediction has been proposed previously. The linear predictor can reduce a white noise from a noisy speech, because the linear predictor converges such that the prediction error signal becomes white. In the conventional method, the linear predictor is updated by using an algorithm with a . xed step-size for estimating the speech signal. However the optimal step-size, which can reduce the estimation error of the linear prediction, varies with the coef. cients of the linear predictor. In this paper, we derive the optimal step-size from the convergence condition of the linear predictor, and propose a variable step-size so that the estimation error of the linear prediction is reduced. Experimental results show that the proposed noise reduction system can improve the noise reduction property in comparison to the conventional one.**

I. Introduction

In recent years, research on methods of noise reduction from a speech degraded by an additive background noise is actively being done by the use of microphone array [1], spectrum subtraction (SS)[2], etc. Imperfection can be seen in the method of the noise reduction using two microphones which can be considered as a directional microphone with a blind spot in the arrival bearing of noise. When many noise sources exist, an increase in number of microphones cannot be avoided. It is therefore important to develop a noise reduction method which uses a single microphone, and which can cancel multiple noise sources. In the systems with only one microphone, extracting a speech from a speech degraded by an additive background noise requires the use of SS method. One of the SS methods [2] improves the signal to noise ratio (SNR) at the expense of processing delay, signal distortion and musical noises that arise due to the residual noise.

Another noise reduction method using one microphone has been proposed as the adaptive noise canceling (ANC) [3], [4], [5]. ANC methods have the common feature that they are based only on the periodicity of voiced speech and require accurate pitch information. Techniques for extracting the pitch information have been proposed [6], [7], [8]. Unfortunately, the accurate pitch extraction method from a speech degraded by a background noise has been not obtained as yet.

In order to solve these problems, we have investigated a noise reduction method based on linear prediction [9]. The linear predictor can reduce a white noise from a noisy speech, because the linear predictor converges such that the prediction error signal becomes white [10]. The linear predictor used in the noise reduction system is updated by using a . xed step-size for estimating the speech signal. However, the optimal step-size which can reduce the estimation error of the linear prediction varies with the coef. cients of the linear predictor. In this paper, we derive the optimal step-size from the convergence condition of the linear predictor, and propose a variable step-size so that the estimation error is reduced. Experimental results show that the proposed noise reduction system can improve the noise reduction capability in comparison to the conventional one.

II. Noise Reduction Based on Linear Prediction

The noise reduction system based on the linear prediction [9] is shown in Fig.1. An observed signal $x(n)$ is a speech signal $s(n)$ mixed with a white noise $\xi(n)$. The transfer functions $H(z)$ and $H'(z)$ are of the linear predictors in the . rst and second stages, respectively. The respective predicted signals $y(n)$ and $y'(n)$ are given by

$$y(n) = \sum_{m=1}^{M} h_m(n)x(n-m) \tag{1}$$

$$y'(n) = \sum_{m=1}^{M} h'_m(n)x(n-m), \tag{2}$$

where $h_m(n)$ and $h'_m(n)$ $(m = 1, \cdots, M)$ show the coef. - cients of the $H(z)$ and the $H'(z)$, respectively. The signals $\hat{s}(n)$ and $\hat{s}'(n)$ are the extracted speech signals of the . rst and second stages, respectively. The prediction error signal $e(n)$ is the difference between $x(n)$ and $y(n)$, and $e'(n)$ is the difference between $\hat{s}(n)$ and $y'(n)$. The multiplier K can improve the quality of the extracted speech, and the cascade system can reduce the noise more than the single system [9].

Since the linear predictor converges such that the prediction error signal becomes white, the linear predictor only passes the speech spectra. In general, the linear predictor is updated by using an algorithm with a step size. Although the step-size is

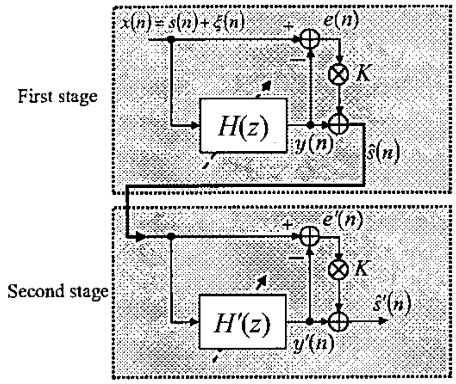

Fig. 1. Noise Reduction System based on Linear Predictor.

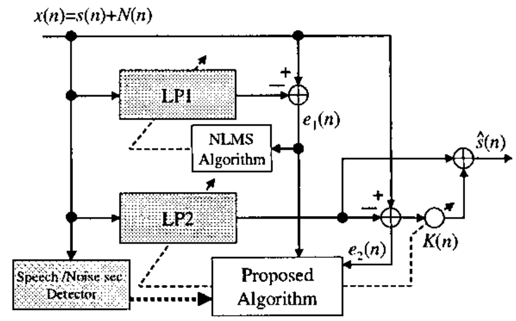

Fig. 2. Noise Reduction System Using Variable Step-Size.

the parameter to adjust a trade-off between the convergence speed and the estimation accuracy, there is no discussion about the method to optimize the step-size in [9]. In the next section, we discuss the optimal value of the step-size and propose a variable step-size.

III. Noise Reduction Using Variable Step-Size

A. Optimal Step-Size

In this section, we discuss the optimal step-size of the NLMS (Normalized Least Mean Square) algorithm [10] for updating the linear predictor in the noise reduction system.

NLMS algorithm is given by

$$\mathbf{H}(n+1) = \mathbf{H}(n) + \mu \frac{\mathbf{X}(n)}{\| \mathbf{X}(n) \|^2} e(n), \quad (3)$$

and

$$\begin{aligned} \mathbf{X}(n) &= [x(n-1), x(n-2), \cdots, x(n-M)]^{\mathrm{T}} \quad (4) \\ \mathbf{H}(n) &= [h_1(n), h_2(n), \cdots, h_M(n)]^{\mathrm{T}}, \quad (5) \end{aligned}$$

where T represents the transpose of a matrix, and μ is the step-size for adaptation. Consider the . rst stage of the noise reduction system in Fig.1. Under the assumption that the linear predictor with the optimal coef. cients can completely predict the speech $s(n)$, we can write

$$\begin{aligned} s(n) &= \mathbf{H}_{\mathrm{opt}}^{\mathrm{T}} \mathbf{X}(n) \quad (6) \\ \mathbf{H}_{\mathrm{opt}} &= [h_{\mathrm{opt},1}, h_{\mathrm{opt},2}, \cdots, h_{\mathrm{opt},M}]^{\mathrm{T}}, \quad (7) \end{aligned}$$

where $\mathbf{H}_{\mathrm{opt}}$ shows the optimal coef. cients vector. Since $s(n)$ is uncorrelated with $\xi(n)$, then we obtain

$$\begin{aligned} & E[\mathbf{H}_e^{\mathrm{T}}(n+1)\mathbf{H}_e(n+1)] \\ &= E[\mathbf{H}_e^{\mathrm{T}}(n)\mathbf{H}_e(n)] \\ &- \frac{(2\mu - \mu^2) E\left[\{s(n) - y(n)\}^2\right] - \mu^2 \sigma_\xi^2(n)}{M \sigma_x^2(n)} \end{aligned} \quad (8)$$

$$\mathbf{H}_e(n) = \mathbf{H}(n) - \mathbf{H}_{\mathrm{opt}}, \quad (9)$$

where $\sigma_x^2(n)$ and $\sigma_\xi^2(n)$ are the variances of $x(n)$ and $\xi(n)$, respectively. The second term in the right hand side in Eq. (8) must be negative to decrease the mean square error of the coef. cients, and thereby we obtain the following condition

$$0 < \mu < \frac{2E[\{s(n) - y(n)\}^2]}{\sigma_e^2(n)}, \quad (10)$$

where $\sigma_e^2(n)$ is the variance of $e(n)$. From Eq. (8), we . nd the optimal step-size which can most reduce the estimation error. The optimal step-size is given by

$$\mu_{\mathrm{opt}}(n) = \frac{E\left[\{s(n) - y(n)\}^2\right]}{\sigma_e^2(n)} = 1 - \frac{\sigma_\xi^2(n)}{\sigma_e^2(n)}. \quad (11)$$

This result corresponds to the result derived from a system identi. cation model [11].

Next, we consider the multiplier K in Fig.1. The extracted speech is given by

$$\hat{s}(n) = y(n) + Ke(n), \quad (12)$$

In conventional method [9], K is . xed to a positive value $(0 < K < 1)$. Since a role of K is to decide a rate of adding the unpredicted speech to the predicted speech, K should be optimized at each time. Consider $y(n) + K(n)e(n) = s(n)$ as the cost function, then we can write

$$K_{\mathrm{opt}}^2(n) = \frac{E\left[\{s(n) - y(n)\}^2\right]}{\sigma_e^2(n)}, \quad (13)$$

where $K_{\mathrm{opt}}(n)$ denotes the optimized K at time n. The right hand side of Eq. (13) corresponds to $\mu_{\mathrm{opt}}(n)$ in Eq. (11), and thus the optimized K is written by

$$K_{\mathrm{opt}}(n) = \sqrt{\mu_{\mathrm{opt}}(n)}. \quad (14)$$

B. Estimation of Noise Variance

In section 3.1, we have obtained the optimized step-size and the optimized multiplier K. When the results are actually used, $\sigma_\xi^2(n)$ needs to be required. In this section, we discuss about a method to estimate $\sigma_\xi^2(n)$.

Since the co-variance matrix of the coef. cients error in the linear predictor becomes a diagnostic matrix after convergence, we obtain

$$\begin{aligned} & E\left[\{s(n)-y(n)\}^2\right] \\ & \quad = E\left[\mathbf{H}_e(n)^{\mathrm{T}}\mathbf{X}(n)\mathbf{X}(n)^{\mathrm{T}}\mathbf{H}_e(n)\right] \\ & \quad = \sigma_x^2 E\left[\mathbf{H}_e(n)^{\mathrm{T}}\mathbf{H}_e(n)\right], \end{aligned} \tag{15}$$

and

$$\begin{aligned} E\left[\mathbf{H}_e^{\mathrm{T}}(n+1)\mathbf{H}_e(n+1)\right] &= \frac{\mu^2\sigma_\xi^2(n)}{M\sigma_x^2(n)} \\ &+ \left(1+\frac{\mu^2-2\mu}{M}\right)E\left[\mathbf{H}_e^{\mathrm{T}}(n)\mathbf{H}_e(n)\right], \end{aligned} \tag{16}$$

where Eq. (16) is given by substituting Eq. (15) into Eq. (8). Assuming that $x(n)$ and $\xi(n)$ are stationary from $(n-L)$ to n, Eq. (16) is rewritten as

$$\begin{aligned} & E\left[\mathbf{H}_e^{\mathrm{T}}(n)\mathbf{H}_e(n)\right] \\ & = \frac{\mu}{(2-\mu)}\frac{\sigma_\xi^2(n)}{\sigma_x^2(n)}\left\{1-\left(1+\frac{\mu^2-2\mu}{M}\right)^{L+1}\right\} \\ & \quad + \left(1+\frac{\mu^2-2\mu}{M}\right)^{L+1} E\left[\mathbf{H}_e^{\mathrm{T}}(n-L)\mathbf{H}_e(n-L)\right]. \end{aligned} \tag{17}$$

If $\{1+(\mu^2-2\mu)/M\}^{L+1}\approx 0$, then

$$E\left[\mathbf{H}_e^{\mathrm{T}}(n)\mathbf{H}_e(n)\right] \approx \frac{\mu}{(2-\mu)}\frac{\sigma_\xi^2(n)}{\sigma_x^2(n)}. \tag{18}$$

From Eq. (15) and Eq. (18), the following equation is obtained.

$$E\left[\{s(n)-y(n)\}^2\right] \approx \frac{\mu}{2-\mu}\sigma_\xi^2(n) \tag{19}$$

Substituting the relation $E\left[\{s(n)-y(n)\}^2\right] = \sigma_e^2(n)-\sigma_\xi^2(n)$ into Eq. (19), we can write

$$\sigma_\xi^2(n) \approx \left(1-\frac{\mu}{2}\right)\sigma_e^2(n). \tag{20}$$

Eq. (20) with $\mu=1$ also corresponds to the relation which has been derived from system identi. cation theory in the paper [11].

We therefore adopts the algorithm described in [11] to the noise reduction system. The algorithm is given by

$$\mu(n) = 1-\frac{\hat{\sigma}_\xi^2(n)}{\tilde{\sigma}_{e2}^2(n)} \tag{21}$$

$$\hat{\sigma}_\xi^2(n) = \alpha\tilde{\sigma}_\xi^2(n)+(1-\alpha)\hat{\sigma}_\xi^2(n-1) \tag{22}$$

$$\tilde{\sigma}_\xi^2(n) = \tilde{\sigma}_{e1}^2(n)-0.5\tilde{\sigma}_{e1}^2(n-M) \tag{23}$$

if $\tilde{\sigma}_{e1}^2(n)-0.5\tilde{\sigma}_{e1}^2(n-M)<0$ then $\tilde{\sigma}_\xi^2(n)=0$.

$$\tilde{\sigma}_{e1}^2(n) = \frac{1}{L}\sum_{j=0}^{L-1}e_1^2(n-j) \tag{24}$$

$$\tilde{\sigma}_{e2}^2(n) = \frac{1}{L}\sum_{j=0}^{L-1}e_2^2(n-j), \tag{25}$$

where $e_1(n)$ is the prediction error signal of an additive linear predictor LP1 for estimating the variance of the white noise, and $e_2(n)$ is the prediction error signal of the linear predictor LP2 for the noise reduction. L is the number of summation of the signals used for estimating the time average, M is the number of tap coef. cients of the LP1, α is the forgetting factor, and $\mu(n)$ is restricted in $0<\mu(n)<1$. The proposed noise reduction system is shown in Fig.2. The speech/non-speech detector in Fig.2 is explained in the next section.

C. Speech/Non-Speech Detector

For estimating the optimal step-size, the white noise variance is necessary, and it is calculated on the non-speech section where $x(n)=\xi(n)$. We therefore propose a speech/non-speech detector using an one order linear predictor $\mathrm{LP_J}$. The predicted signal $y_J(n)$ and the prediction error signal $e_J(n)$ of the $\mathrm{LP_J}$ are given by

$$y_J(n) = h_J(n)x(n-1) \tag{26}$$

$$e_J(n) = x(n)-y_J(n), \tag{27}$$

respectively. Here $h_J(n)$ is the coef. cient of the $\mathrm{LP_J}$. The speech section or the non-speech section is destinguished by using the sign of $D(n)$ de. ned by

$$\begin{aligned} D(n) &= 2E\left[y_J(n)\{y_J(n)-x(n)\}\right] \\ &= E[e_J^2(n)]+E[y_J^2(n)]-E[x^2(n)]. \end{aligned} \tag{28}$$

When $x(n)=\xi(n)$, we obtain $D(n)=2\sigma_y^2(n)>0$. On the other hand, when $x(n)=s(n)+\xi(n)$, we obtain $D(n)=\sigma_y^2(n)-E[y(n)s(n)]\le 0$, under the assumption of that the signs of $y(n)$ and $s(n)$ are the same, and $|s(n)|>|y(n)|$. Therefore, we can distinguish the speech or non-speech section by using the sign of $D(n)$.

In simulations in the next section, we use

$$\begin{aligned} & \hat{D}(n) \\ & \quad = \lambda\hat{D}(n-1)+(1-\lambda)\{e_J^2(n)+y_J^2(n)-x_J^2(n)\}, \end{aligned} \tag{29}$$

where λ is the forgetting factor.

IV. Simulation Results

Some simulations had been carried out for con. rming the effectiveness of the proposed method. The signals used in the simulations were sampled by 8kHz. In the . rst simulation, we used a pseudo speech signal given by

$$s(n)=\frac{1}{I}\sum_{i=1}^{I}\cos(i\omega_1 n+\phi_i), \tag{30}$$

where I is the number of sinusoidal signals, ω_1 and ϕ_i are its frequency and phase, respectively In the simulation, we set $I=20$, $\omega_1=120$, and SNR= 0dB. Speech section or non-speech section was changed to each other at every 8000 samples. The step-size of the LP1 was set to 1. The order of the LP1 and LP2 were set to 200. Other parameters were set to $\alpha=0.001$, $L=300$, and $\lambda=0.99$. The LMS algorithm [10]

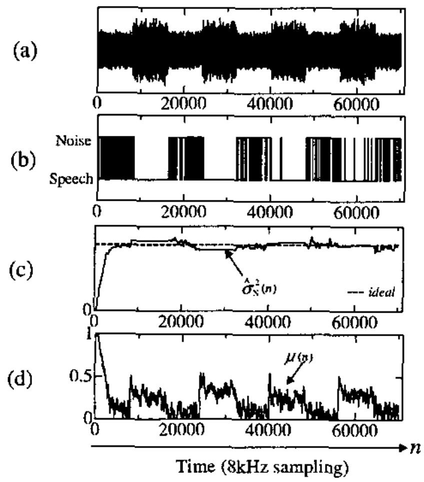

Fig. 3. Simulation Results.(a)input signal $x(n)$, (b)results of speech/non-speech detection, (c)estimation results of white noise variance, (d)variable step-size$\mu(n)$

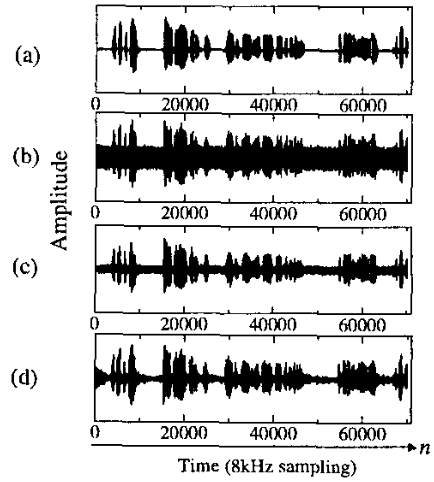

Fig. 4. Noise Reduction Results for human speech signal mixed with white noise.(a)speech signal $s(n)$, (b)input signal $x(n)$, (c)results of conventional method, (d)results of proposed method

was used for updating the speech/non-speech detector LP_J, and its step-size was set to 0.001.

The results are summarized in Fig.3(a)–(d). Fig.3(a) shows the input signal with SNR of 0dB. In Fig.3(b), the upper and lower sides show the section detected as the non-speech and the speech section, respectively. From the results, we see that the non-speech section was accurately detected. In Fig.3 (c), the solid line shows the estimated white noise variance, and the dotted line shows the desired value of the white noise variance. We see that the proposed method can estimate the white noise variance with accuracy. From the results of Fig.3 (d), we see that the variable step-size increases on the change of the signal state. The proposed noise reduction system therefore achieves both the convergence speed and the estimation accuracy.

Next, we performed computer simulation using a real speech. The proposed method is compared with the conventional one [9]. In the conventional method, the order of the linear predictor was set to 200, the step-size was set to 0.2, and $K = \sqrt{0.2}$. The results are shown in Fig.4(a)–(d). Fig.4(a) shows the speech signal $s(n)$, (b) shows the obserbed signal $x(n)$, (c) shows the signals extracted by the conventional method, and (d) shows the signals extracted by the proposed method. From the results, we see that the proposed method reduces the white noise more than the conventional method in the non-speech section, and extracts the desired signal more than the conventional method in the speech section.

V. Conclusion

A white noise reduction method using the linear predictor with the variable step-size has been proposed. The variable step-size can reduce the estimation error of the linear prediction. To calculate the variable step-size, the white noise variance needs to be required. Since the white noise variance is obtained on the non-speech section, a speech/non-speech detector has also been proposed. In particular, the speech/non-speech detector estimates the speech section with accuracy. From simulation results of the proposed system with the variable step-size, we saw that the proposed system can improve the noise reduction property in comparison to the conventional method.

References

[1] Y. Kaneda, and J. Ohga, "Adaptive microphone-array system for noise reduction," *IEEE Trans. on Acoustics, speech, and signal processing*, vol. ASSP-34, Dec. 1986.

[2] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-27, pp.113-120, April 1979.

[3] M. R. Sambur,"Adaptive noise canceling for speech signals," *IEEE Trans. Acoust., Speech, Signal Proc.*, vol. ASSP-26, pp. 419-423, Oct. 1978.

[4] J. R. Deller Jr., J. G. Proakis, and J. H. L. Hansen, *Discreate-Time Processing of Speech Signals*, Chap.8, NY:Macmillan Publishing co., 1993.

[5] J. S. Lim and A.V. Oppenhaim,"Enhancement and Bandwidth Compression of Noisy Speech", *Proc. IEEE*, vol.67, Dec. 1979

[6] L. R. Rabiner, M. J. Cheng, A. E. Rosenberg, and C. A. McGonegal,"A comparative performance study of several pitch detection algorithms," *IEEE Trance. Acoust., Speech, Signal Processing*, vol. ASSP-24, pp. 399-418, Oct. 1976.

[7] A. M. Noll,"Cepstrum pitch determination," *J. Acoust. Soc. Am.*, vol. 41, pp. 293-309, Feb. 1967.

[8] B. G. Secrest, and G. R. Doddington,"An integrated pitch tracking algorithm for speech systems," *Proc. ICASSP 83 (Boston)*, pp. 1352-1355, May 1983.

[9] A. Kawamura, K. Fujii, Y. Itoh, and Y. Fukui,"A noise reduction method based on linear prediction analysis," *IEICE Trans. A*, vol. J85-A, pp. 415-423, April 2002.

[10] S. Haykin, *Introduction to Adaptive Filters*, New York: Macmillan publishing co., 1984.

[11] K. Takahashi, I. Sasase, and S. Mori,"Estimation of Noise Level and Variable Step Gain Algorithm for Improvement of Initial Convergence Characteristics," *IEICE Trans. A*, vol. J76-A, pp. 1704-1713, Dec. 1993.