A NEW NOISE REDUCTION METHOD USING LINEAR PREDICTION AND SYSTEM IDENTIFICATION

Arata Kawamura† *Kensaku Fujii*†† *Yoshio Itoh*† *Yutaka Fukui*†

†Faculty of Engineering, Tottori University, Tottori, Japan
4-101, Koyama-Minami, Tottori, Tottori, 680-8552 Japan
TEL: +81-857-31-5690, FAX: +81-857-31-5690
E-mail: m99t3012@maxwell.ele.tottori-u.ac.jp

††Faculty of Engineering, Himeji Institute of Technology,
2167 Shosya, Himeji, 671-2201, Japan

ABSTRACT

A technique that uses linear prediction and system identification to achieve noise reduction in a voice signal mixed with a background noise is proposed. In this method, the linear prediction error filter (LPEF) whose coefficients will converge such that the prediction error signal becomes white, removes the spectrum of the voice from a voice mixed with noise. Thereby, the prediction error signal has not cross-correlation to the voice. The background noise can be reconstructed from the prediction error signal by using the adaptive digital filter (ADF) which is used as one for the system identification. The coefficients of the ADF will converge such that the error signal which is obtained by subtracting the output of the ADF from a voice mixed with noise, is minimized. Then, the ADF estimates only the background noise since that the input of the ADF (the prediction error signal) has not cross-correlation to the voice. Noise reduction is achieved by subtracting the output of the ADF from the voice mixed with noise.

1. INTRODUCTION

In recent years, research on methods of noise reduction from a voice mixed with background noise is actively being done by the use of microphone array [1], spectrum subtraction [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 voice mixed with background noise requires the use of spectrum subtraction method for noise reduction. One of the spectrum subtraction method [2] improves the signal to noise ratio (SNR) at the expense of processing delay, signal distortion and musical tones that arise due to the residual noise.

In order to improve on these negative effects, we had investigated a noise reduction method based on linear prediction [3]. In this method, when a background noise is white, the noise reduction is efficiently performed because the linear predictor estimates only a voice spectrum. However, when a background noise is colored, effectiveness of the noise reduction decreases since the linear predictor estimates both a voice and a colored noise spectrum.

In this paper, we propose a new noise reduction method using linear prediction and system identification to achieve noise reduction in a voice mixed with white noise or colored noise. The linear prediction error filter (LPEF) whose coefficients will converge such that the prediction error signal becomes white, removes the spectrum of the voice from a voice mixed with noise. Thereby, the prediction error signal has not cross-correlation to the voice. The background noise can be reconstructed from the prediction error signal by using the adaptive digital filter (ADF) which is used as one for the system identification. The coefficients of the ADF will converge such that the error signal which is obtained by subtracting the output of the ADF from a voice mixed with noise, is minimized. Then, the ADF estimates only the background noise since that the input of the ADF (the prediction error signal) has not cross-correlation to the voice. Noise reduction is achieved by subtracting the output of the ADF from the voice mixed with noise.

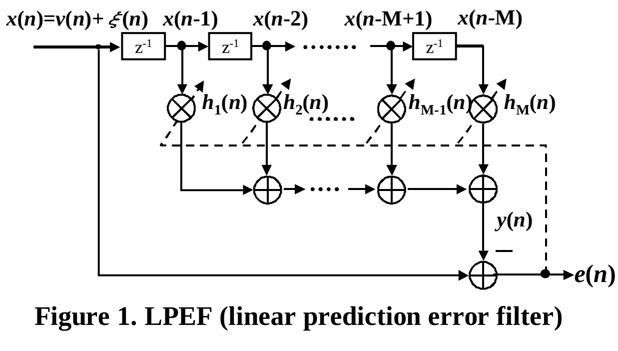


Figure 1. LPEF (linear prediction error filter)

2. PRINCIPLE OF THE PROPOSED NOISE REDUCTION METHOD

In this section, we describe the principle of the proposed noise reduction method based on the linear prediction and the system identification.

Figure 1 shows a transversal type LPEF. In this Figure, $x(n)$ is a voice $v(n)$ mixed with background noise $\xi(n)$, and the prediction error signal $e(n)$ whose mean square value is minimized, is defined as $e(n)=x(n)-y(n)$ where $y(n)$ is a predicting signal. $y(n)$ is given by

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

where $h_k(n)$ (k=1,2,…,M) are the tap coefficients [4].

The coefficients of the LPEF will converge such that the prediction error signal $e(n)$ becomes white. However, it is difficult to whiten a voice signal, since that a voice signal is a non-stationary signal. If a short time average value of the error signal $e(n)$ is used for adaptation of the coefficients in the LPEF (i.e., using large step size for adaptation), then the LPEF estimates the input signal $x(n)$ which is a voice mixed with noise with high tracking ability at the expense of roughly estimation. On the other hand, if a long time average value of the $e(n)$ is used for adaptation of the LPEF's coefficients (i.e., using small step size for adaptation), the LPEF estimates the input signal $x(n)$ with high fidelity at the expense of poor tracking ability.

In order to improve this trade-off, the LPEF which has large step size for adaptation of the tap coefficients corresponding to the parts considered as pitch period of the voice, and which has small step size for adaptation of other tap coefficients, is proposed. This LPEF estimates the voice signal having pitch frequency with high tracking ability, and the other signal with high fidelity. Thereby, the signal whitening by the LPEF is efficiently performed.

Next, we consider the reconstruction of the background noise from the prediction error signal. If the background noise generation process is represented in terms of a certain transformation from a white noise as shown in Figure 2, then we can consider the system identification model as shown in Figures 3(a),(b). In these Figures, $H_\xi(z)$ represents the spectrum of the noise, it is an imaginary system, $H_{ADF}(z)$ is the transfer function of the ADF, $\hat{\xi}(n)$ and $e_{ADF}(n)=\hat{v}(n)$ represent the reconstructed noise and the extracted voice, respectively, and $h'_0(n),h'_1(n),...,h'_L(n)$ are the tap coefficients of the ADF. The ADF can not estimate the $v(n)$ due to the $e(n)$ which has not cross-correlation to the voice signal $v(n)$, then the output of the ADF represents only the background noise.

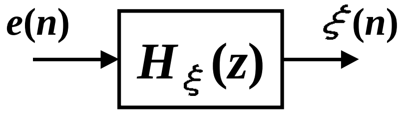
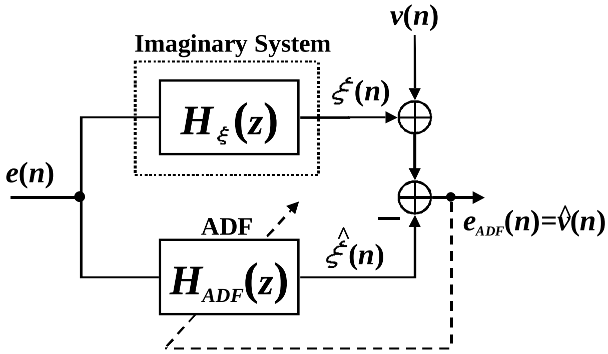
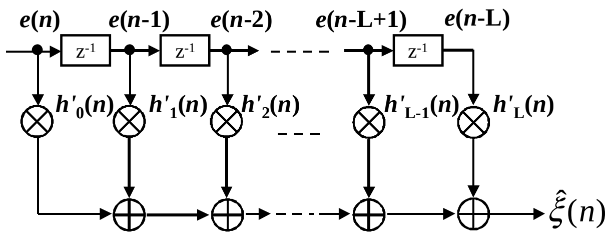


Figure 2. Imaginary noise generation process



(a) Block diagram of system identification model



(b) Structure of the ADF (adaptive digital filter)

Figure 3. System identification model

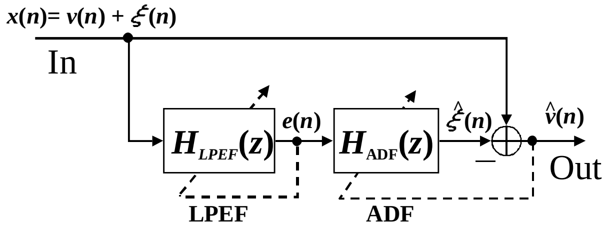


Figure 4. Noise reduction system

We shall now incorporate the signal whitening process by using the LPEF with the noise reconstruction process by using the ADF for noise reduction. The proposed noise reduction system is shown in Figure 4, where the $H_{LPEF}(z)$ represents the transfer function of the LPEF. The noise reduction is achieved by subtracting the reconstructed noise from the input signal $x(n)$ which is a voice mixed with noise.

3. SIMULATION RESULTS

We performed computer simulation using the proposed noise reduction system as shown in Figure 4. The adaptive algorithms used for updating the coefficients of the LPEF and the ADF are the LMS (least mean square) and the NLMS (normalized least mean square) algorithms, respectively. The LMS and the NLMS algorithms are given by following equations [4].

(LMS algorithm for the LPEF)

$$\boldsymbol{h}(n+1) = \boldsymbol{h}(n) + \mu \cdot \boldsymbol{x}(n) \cdot e(n) \tag{2}$$

$$\mathbf{h}(n) = [h_1(n), \ldots, h_M(n)]^T \tag{3}$$

$$\mathbf{x}(n) = [x(n-1), \ldots, x(n-M)]^T \tag{4}$$

(NLMS algorithm for the ADF)

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

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

$$\mathbf{e}(n) = [e(n), e(n-1), \ldots, e(n-L)]^T \tag{7}$$

where μ and μ' are the step size of adaptation and $\|\cdot\|$ represents the norm.

All sound data prepared in the simulations are sampled at 8kHz with 16bit resolution. The input signals to the noise reduction system are generated by artificially adding white noise, or tunnel noise of an expressway to a voice, respectively. To evaluate the noise reduction ability, the SNR_{in} and the SNR_{out} is used. These factors are defined by

$$SNR_{in} = 10 \log_{10} \frac{\sum_{j=1}^{N} v^2(j)}{\sum_{j=1}^{N} \xi^2(j)} \tag{8}$$

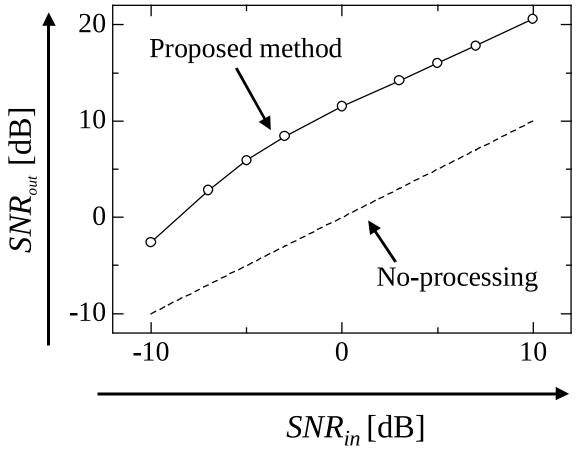
$$SNR_{out} = 10 \log_{10} \frac{\sum_{j=1}^{N} y_v^2(j)}{\sum_{j=1}^{N} y_\xi^2(j)} \tag{9}$$

where N is the number of samples and $y_v(j)$ and $y_\xi(j)$ are the component of the voice and the noise involved in the extracted voice signal $\hat{v}(j)$, respectively.

The parameters used in this simulations are shown in table 1, where it is assumed that the pitch period of the

Table 1 Parameters used in the simulation of Fig.5.

LPEF	The number of tap coefficients	256
	Step size of adaptation for $h_{40}(n)$～$h_{90}(n)$	0.1
	Step size of adaptation for $h_1(n)$~$h_{39}(n)$, $h_{91}(n)$~$h_{256}(n)$	0.01
ADF	The number of tap coefficients	256
	Step size of adaptation	0.001

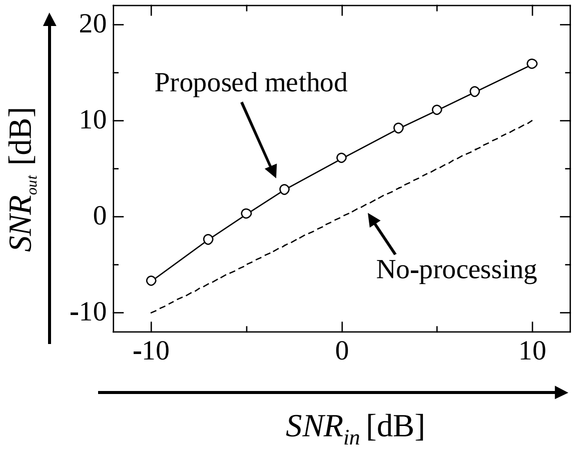


SNR_{in} [dB]	SNR_{out} [dB]
-10.0	-2.6
-7.0	2.8
-5.0	5.9
-3.0	8.4
0.0	11.5
3.0	14.2
5.0	16.0
7.0	17.8
10.0	20.6

Figure 5. Simulation result of white noise reduction

voice certainly exists between the tap coefficients $h_{40}(n)$～$h_{90}(n)$.

Figures 5 and 6 show the results of white and tunnel noise reduction, respectively, and Figure 7 shows the waveforms of the tunnel noise reduction result. From these results, it can be seen that there is an improvement in the SNR (signal to noise ratio), and this proposed noise reduction method is also available under the practically environment.



SNR_{in} [dB]	SNR_{out} [dB]
-10	-6.7
-7	-2.4
-5	0.3
-3	2.8
0	6.1
3	9.2
5	11.1
7	13
10	15.9

Figure 6. Simulation Result of tunnel noise reduction

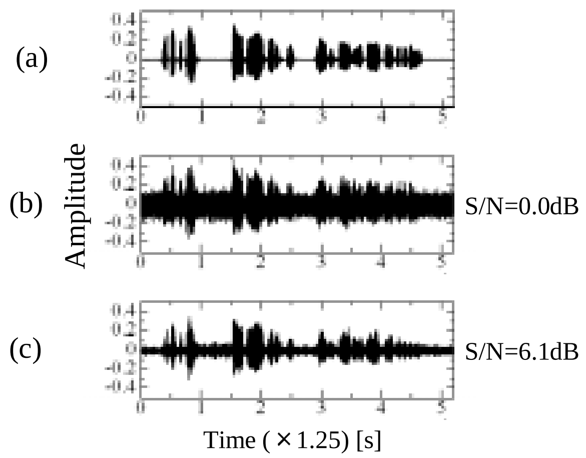


Figure 7. Waveforms of the tunnel noise reduction result (a) Original voice. (b) Nosy voice (SNR_{in}=0dB). (c) Extracted voice.

4. CONCLUSION

We have proposed a new noise reduction method using linear prediction and system identification. From the computer results, it was observed that there was improvement of SNR in the extracted voice signal, and this proposed noise reduction method is also available under the practically environment. Further researches with this proposed noise reduction method involve an improvement of the tracking ability to the non-stationary noise like a tunnel noise, a reduction of the residual noise and a performance evaluation in a test product.

REFERENCES

[1] J. L. Flanagan, J.D. Johnston, R. Zahn, and G. W. Elko. “Computer-steered microphone arrays for sound transaction in large rooms,” J. Acoust. Soc. Am., 78, pages 1508-1518, Nov. 1985.

[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, April 1979.

[3] A.Kawamura, K.Fujii, Y.Itoh, Y.Fukui, “A New Noise Reduction Method Based on Linear Prediction, ” Proc. ITC-CSCC2000, Vol. 1, pp.260-263, July 2000.

[4] S.Haykin. Adaptive Filter Theory, 3rd ed. Englewood Cliffs, NJ: Prentice-Hall, 1996.