A Noise Reduction Method with Linear Prediction Using Periodicity of Voiced Speech

Naoto Sasaoka[1] Arata Kawamura[1] Kensaku Fujii[2] Yoshio Itoh[1] and Yutaka Fukui[1]

[1] Faculty of Engineering, Tottori University
4-101 Koyama-minami, Tottori 680-8552, Japan
Tel: +81-857-31-5698, Fax: +81-857-31-5698
e-mail : sasaoka@dacom2.ele.tottori-u.ac.jp

[2] Faculty of Engineering, Himeji Institute of Technology
2167 Shosha, Himeji 671-2201, Japan

Abstract: A noise reduction technique to reduce background noise in corrupted voice is proposed. The proposed method is based on linear prediction and takes advantages of periodicity of voiced speech. A voiced sound is regarded as a periodic stationary signal in short time interval. Therefore, the current voice signal is correlated with the voice signal delayed by a pitch period. A linear predictor can estimate only the current signal correlated with the delayed signal. Therefore, the enhanced voice can be obtained as output of the linear predictor. Simulation results show that the proposed method is able to reduce the background noise.

1. Introduction

A large variety of noise reduction techniques have been proposed for reducing noise in corrupted voice, with microphone array [1], spectral subtraction [2] and so on. Assuming that many noise sources exist, the microphone array cannot avoid increasing the number of microphones. On the other hand, the spectral subtraction is known as the noise reduction method which uses one microphone. However, musical tones arise from residual noise, and the processing delay occurs. Additionally, the spectral subtraction method requires advanced estimation of a noise spectrum, therefore, it is difficult to reduce non-stationary noise whose characteristics change with time. In order to improve these problems, we have proposed noise reduction method based on linear prediction [3],[4]. The conventional noise reduction method [3] effectively performs with white noise. However, the method does not perform very well with colored noise, because the linear predictor estimates not only a voice component but also a colored noise component.

In this paper, we propose a noise reduction technique, which uses the linear prediction exploiting periodicity of voiced speech, to reduce background noise.

A voiced sound is regarded as a periodic stationary signal in short time interval. The proposed method takes into consideration periodicity of the voiced sound signal. A linear predictor can estimate only the current signal correlated with the delayed signal. The current voice signal is correlated with the voice signal delayed by a pitch period and the current noise signal is not correlated with the noise signal delayed by a pitch period. Therefore, the enhanced voice can be obtained as output of the linear predictor. Since the proposed noise reduction method uses signal delayed by a pitch period, the pitch period detector is required. We propose the pitch period detection method as well.

This paper is organized as follows. In section 2, we describe the principle of noise reduction with linear prediction using periodicity of voiced speech. In section 3, we show the experimental results of the proposed method. In section 4, we conclude the paper.

2. Noise Reduction Method

2.1 Linear prediction using periodicity

The pitch period is known as about 8ms on the average for a male and about 4ms on the average for a female [5]. 64 samples and 32 samples are the average pitch period for a male and a female by the signal sampled by 8kHz, respectively. This paper takes the male voice sampled by 8kHz as an object of voice.

The corrupted voice signal $x(n)$ is represented as

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

where $v(n)$ is the clean voice signal and $\xi(n)$ is the noise signal. The signal delayed by a pitch period of the voice signal is represented as

$$x(n-\tau) = v(n-\tau) + \xi(n-\tau) \tag{2}$$

where τ is a pitch period of the voice signal. Therefore, $v(n-\tau)$ is correlated with $v(n)$ and $\xi(n-\tau)$ is not correlated with $\xi(n)$.

Figure 1 shows a noise reduction system using periodicity, which is a transversal type linear predictor. The linear predictor output $y(n)$ is giving by

$$y(n) = \sum_{k=-L}^{L} h_{\tau+k} \cdot x(n-\tau-k) \tag{3}$$

where $h_{\tau+k}(n)$ are tap coefficents whose size is $2L+1$. The linear predictor can estimate only the current signal correlated with the delayed signal. Therefore, the voice $v(n)$ can be predicted and the noise $\xi(n)$ can be reduced.

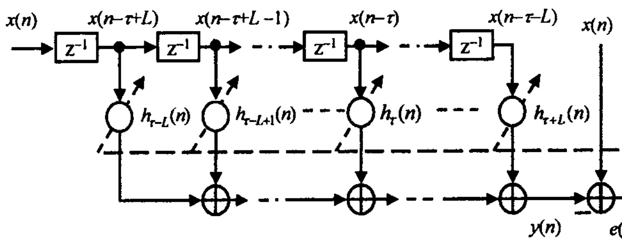


Fig. 1 Linear predictor using periodicity

2.2 Proposed noise reduction system

A proposed noise reduction system is illustrated in Fig. 2, where $x(n)$ is the corrupted voice, $y(n)$ is the output of the proposed noise reduction system. LP_0 is a transversal type linear predictor, which has tap inputs $x(n-40), x(n-41), \ldots, x(n-90)$. Parallel linear predictor LP_p ($40 \le p \le 90$) has tap inputs $x(n-p+L), x(n-p+L-1), \ldots, x(n-p-L)$ in parallel with LP_0.

First, LP_0 estimates the pitch period. The pitch period detection method is described in next section. Next, the output of parallel linear predictor having the pitch period position $x(n-\tau)$ at the center of tap inputs is chosen from $y^{(40)}(n) \sim y^{(90)}(n)$. Therefore, the voice $v(n)$ can be enhanced and the noise $\xi(n)$ can be reduced. For instance, LP_0 detects the pitch period p. The output $y^{(p)}(n)$ of parallel linear predictor LP_p is chosen and $y^{(p)}(n)$ is established as the voice enhanced from the corrupted voice $x(n)$.

However, the enhanced voice signal $y^{(p)}(n)$ excludes consonant information, which is like the unvoiced sound not having the pitch period. In order to compensate consonant information, the proposed system multiplies the voice component which is not predicted in LP_p $(x(n)-y^{(p)}(n))$ by K $(0 \le K \le 1)$ and adds it to the LP_p output $y^{(p)}(n)$.

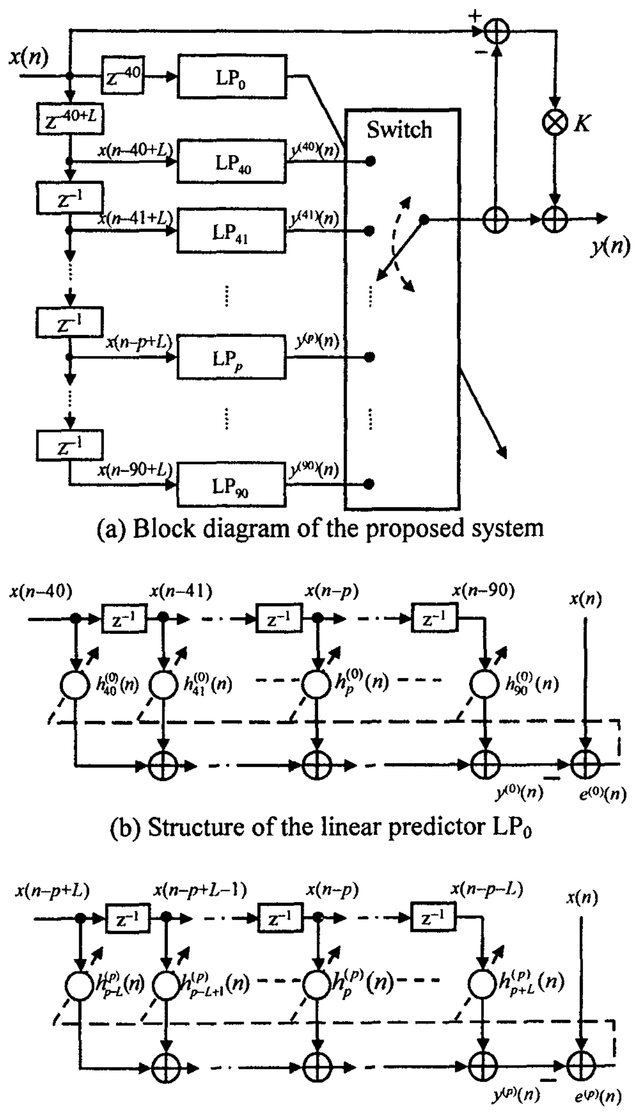
2.3 Pitch period detector

The pitch period detector is required, since the proposed noise reduction method uses the signal delayed by a pitch period. The pitch period detector using the linear predictor is proposed. This method takes into consideration auto-correlation of periodic signal. Since the average male pitch is 64 samples by the signal sampled by 8kHz, we assume that the male pitch period ranges from 40 samples to 90 samples. Therefore, a transversal type linear predictor in Fig. 2(b) has the tap inputs $x(n-40), x(n-41), \ldots, x(n-90)$. The correlation of the current signal $x(n)$ with $x(n-\tau)$ is higher than that of $x(n)$ with $x(n-m)$ ($m=40,41,\ldots,90$, $m \ne \tau$). We assume that a value of the tap coefficient increases with correlation thus we detect $h^{(0)}{}_p(n)$ giving maximum. $h^{(0)}{}_p(n)$ is given by

$$h_p^{(0)}(n) = \mathrm{MAX}\left\{h_{40}^{(0)}(n), h_{41}^{(0)}(n), \ldots, h_{90}^{(0)}(n)\right\} \quad (4)$$

$$(40 \le p \le 90)$$

where p represents the position of the tap coefficient giving maximum. Thus p is established as the pitch period. LP_0 estimates the pitch period like this.



(a) Block diagram of the proposed system

(b) Structure of the linear predictor LP_0

(c) Structure of the parallel linear predictor LP_p
(p=40,41,42,...,90)

Fig. 2 Noise reduction system

3. Simulation Results

The performance of the proposed noise reduction system was evaluated. As the adaptive algorithm, we use the LMS (Least Mean Square) algorithm [6] in this simulation. The LMS algorithm of LP_0 and LP_p are given as follows:

(LMS algorithm for LP_0)

$$\mathbf{h}^{(0)}(n+1) = \mathbf{h}^{(0)}(n) + \mu_0 \cdot \mathbf{x}^{(0)}(n) \cdot e^{(0)}(n) \quad (5)$$

$$\mathbf{h}^{(0)}(n) = \left[h_{40}^{(0)}(n), h_{41}^{(0)}(n), \ldots, h_{90}^{(0)}(n)\right]^T \quad (6)$$

$$\mathbf{x}^{(0)}(n) = \left[x(n-40), x(n-41), \ldots, x(n-90)\right]^T \quad (7)$$

(LMS algorithm for LP_p)

$$\mathbf{h}^{(p)}(n+1) = \mathbf{h}^{(p)}(n) + \mu_p \cdot \mathbf{x}^{(p)}(n) \cdot e^{(p)}(n) \quad (8)$$

$$\mathbf{h}^{(p)}(n) = \left[h_{p-L}^{(p)}(n), h_{p-L+1}^{(p)}(n), \ldots, h_{p+L}^{(p)}(n)\right]^T \quad (9)$$

$$\mathbf{x}^{(p)}(n) = \left[x(n-p+L), x(n-p+L-1), \ldots, x(n-p-L)\right]^T \quad (10)$$

$$(p=40,41,\ldots,90)$$

where μ_0 and μ_p are the step size of adaptation and T represents a transposition of vector. In this simulation, μ_0 and μ_p were set to 0.1 and 0.2, respectively.

To evaluate the noise reduction ability, the SNR (Signal to Noise Ratio) was used. The SNR_{in} and SNR_{out}

represent the input SNR and output SNR, respectively. These indexes are define as follows:

$$\mathrm{SNR}_{in} = 10\log_{10}\frac{\sum_{j=1}^{N} v^2(j)}{\sum_{j=1}^{N} \xi^2(j)} \tag{11}$$

$$\mathrm{SNR}_{out} = 10\log_{10}\frac{\sum_{j=1}^{N} y_v^2(j)}{\sum_{j=1}^{N} y_\xi^2(j)} \tag{12}$$

where N is the number of samples, and $y_v(j)$ and $y_\xi(j)$ are components of the voice signal and the noise signal included in the output $y(j)$, respectively.

All sound data prepared in simulations were sampled by 8kHz with 16bit resolution. The voice signal spoken by a male was used. The input signals were generated by artificially adding colored noise or tunnel noise of an expressway to voice. The colored noise whose spectrum has a peak at 1kHz was generated by passing white noise through the second order IIR (Infinite Impulse response) filter which is shown in Fig. 3. As the parameter, the tap size($2L+1$) of LP_p was 21(L=10) and the voice compensator K was set to 0.2.

3.1 Noise reduction

The ability of noise reduction was evaluated. The input signals were generated by the colored noise or the tunnel noise added to the voice signal. The number of samples N was 40,000. Figure 4 is the results of colored noise reduction. From the results, we see that the SNR is improved by 8.0dB in a SNR_{in}=0dB environment. Figure 5 and 6 are respectively the results of tunnel noise reduction and the waveforms. These results show that the SNR is improved by 6.7dB in a 0dB environment. As can be seen, the proposed noise reduction method reduced the background noise effectively.

3.2 Pitch period detection

The proposed pitch period detector was tested under noisy condition. The input signal was generated by the tunnel noise added to the voiced sound signal. The pitch period of voiced sound /a/ and /i/ are about 63 samples and about 53 samples, respectively. The number of samples N was 6,500. Figure 7 shows the results of pitch period detection in SNR_{in}=0dB environment. From simulation results, the proposed method detects the pitch period accurately. These results indicate that the proposed pitch period detection method can estimate the pitch period exactly.

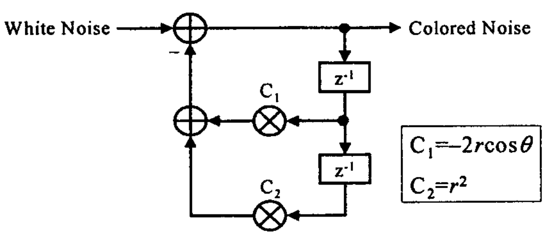
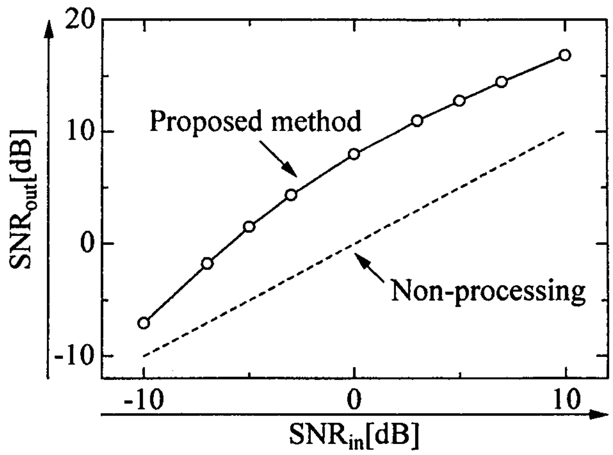
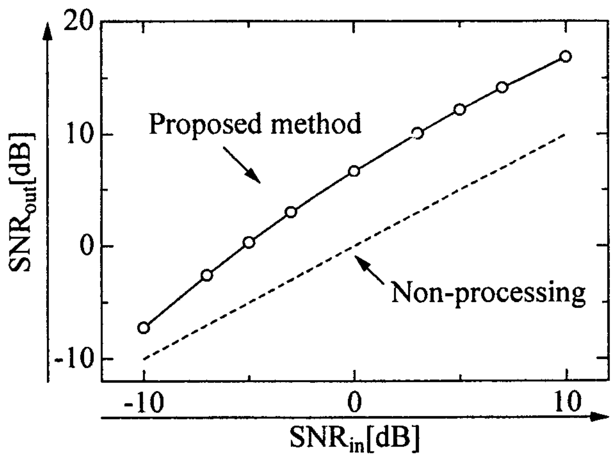


Fig. 3 Second order IIR filter (θ=π/4, r=0.8)



SNR_{in}[dB]	SNR_{out}[dB]
-10.0	-7.0
-7.0	-1.8
-5.0	1.5
-3.0	4.3
0.0	8.0
3.0	11.0
5.0	12.8
7.0	14.5
10.0	16.8

Fig. 4 Simulation results of the colored noise reduction



SNR_{in}[dB]	SNR_{out}[dB]
-10.0	-7.2
-7.0	-2.6
-5.0	0.3
-3.0	3.0
0.0	6.7
3.0	10.1
5.0	12.2
7.0	14.1
10.0	16.8

Fig. 5 Simulation results of the tunnel noise reduction

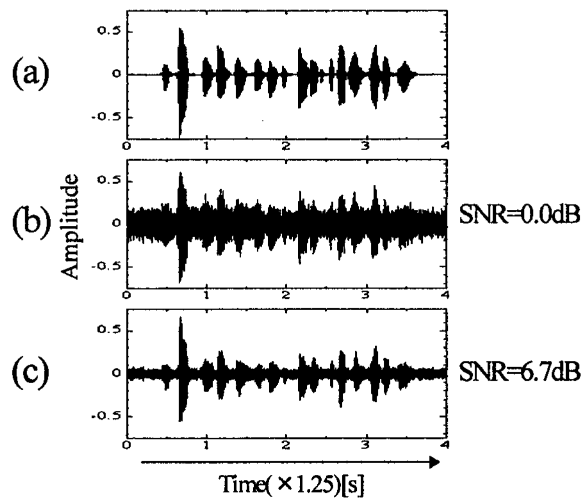


Fig. 6 Waveforms of the tunnel noise reduction result
(a) Original voice signal
(b) Voice corrupted with tunnel noise at 0dB SNR_{in}
(c) Enhanced voice

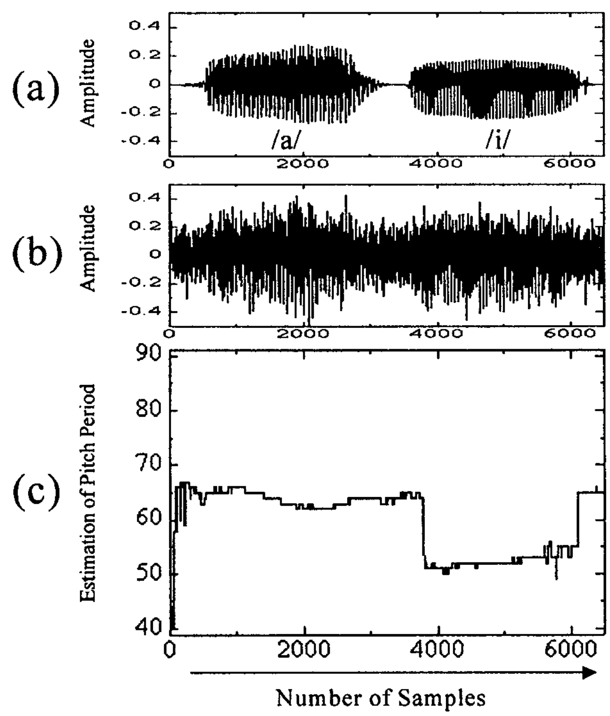


Fig. 7 Results of pitch period detection
(a) Original voiced sound signal
(b) Voiced sound signal corrupted with tunnel noise at 0dB SNR_{in}
(c) Estination of pitch period

4. Conclusions

A noise reduction technique has been proposed for reducing the background noise. The proposed method uses the linear prediction exploiting periodicity of voiced speech. This method takes into consideration the autocorrelation of the periodic signal. From the results of the tunnel noise reduction, the proposed method improved SNR by about 6.7dB in a 0 dB environment. A pitch period detector using linear prediction was also proposed. We verified that the proposed pitch period detector has potential for detecting the pitch period accurately. From the simulation results, it was shown that the proposed method performs in reducing the background noise effectively. In our future work, we will explore a method of extracting the unvoiced sound effectively.

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, pp.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 and Y. Fukui, "A New Noise Redction Method Based on Linear Prediction," *Proc. ITC-CSCC2000*, Vol. 1, pp.260-263, July 2000.

[4] A. Kawamura, K. Fujii, Y. Itoh and Y. Fukui, "A New Noise Reduction Method Using Estimated Noise spectrum," *IEICE Trans. Fundamentals*, Vol. E85-A, no. 4, pp.784-789, Apr. 2002.

[5] S. Furui, *Digital Speech Processing*, Tokai University Press, Tokyo, 1985, In Japanese.

[6] S. Haykin, *Introduction to Adaptive Filiters*, Macmillan Publishing Company, New york, 1984.