SPEECH NOISE REDUCTION SYSTEM BASED ON FREQUENCY DOMAIN ALE USING MODIFIED DFT PAIR

[1] *Isao Nakanishi,* [2] *Takenori Asakura,* [2] *Yoshio Itoh and* [2] *Yutaka Fukui*

[1]isao@rstu.jp

[1] Tottori University, Faculty of Regional Sciences, 4-101 Koyama-Minami, Tottori, 680-8551 Japan

[2] Tottori University, Faculty of Engineering, 4-101 Koyama-Minami, Tottori, 680-8552 Japan

ABSTRACT

This paper presents the speech noise reduction system based on frequency domain adaptive line enhancer using the modified DFT (MDFT) pair. The adaptive line enhancer (ALE) is effective to extract sinusoidal signals blurred by broadband noise and utilizes only one microphone; therefore, it is suitable for realization of speech noise reduction in the portable electronic devices. In the ALE, an input signal is generated by delaying a desired signal by the de-correlation parameter, which makes the noise in the input signal de-correlated with that in the desired one. In this paper, we propose to set the de-correlation parameter in the frequency domain signal and adjust it to an optimal value. Moreover, we introduce the window function for suppressing the spectral leakage in the MDFT. The performance of the proposed noise reduction system is confirmed through computer simulations.

1. INTRODUCTION

The speech noise reduction technique has attracted attention as speech communication systems are widely used in our daily life. There have been proposed several methods: noise canceller, microphone array system, spectral subtraction method and so on. The noise canceller and the microphone array system require multiple microphones [1]; therefore, they are not cost effective and not suitable for miniaturizing portable systems. On the other hand, the spectral subtraction requires only one microphone, which is used to detect noise spectrum during speech pause. In other words, the spectral subtraction is considered as the time-sharing method of one microphone for noisy speech and only noise. However, such time-sharing of microphone is disadvantageous to non-stationary environment.

As another noise reduction method using one microphone, the speech noise reduction system based on the Adaptive Line Enhancer (ALE) has been proposed [2]. The ALE is effective for extracting sinusoidal signals blurred by broadband noise [1]. Also, the ALE is based on the Adaptive Digital Filter (ADF), which is advantageous to non-stationary environment. However, it is well known that the convergence speed of the ADF is degraded when an input signal of the ADF is such a colored signal as the speech. It results in degradation of noise reduction performance.

In order to solve such a problem, we have proposed to introduce the frequency domain adaptive filter (FDAF) into the ALE [3-5]. The FDAF always achieves faster convergence than the time domain ADF even if the input signal is colored. Especially, we utilized the modified discrete Fourier transform (MDFT) pair [6] in the FDAF. In the ALE, an input signal for the ADF is generated by delaying a desired signal. The time delay is constant and called the de-correlation parameter, which makes the noise in the desired signal de-correlated with that in the input signal. On the other hand, our proposed frequency domain ALE enables to set the de-correlation parameter in frequency domain. In this paper, we propose to set the de-correlation parameter in frequency domain signal and adjust it to an optimal value. In addition, we introduce the window function into the MDFT for avoiding the spectral leakage.

2. SPEECH NOISE REDUCTION SYSTEM BASED ON FREQUENCY DOMAIN ALE

Figure 1 illustrates the proposed speech noise reduction system based on the frequency domain ALE with frequency domain de-correlation parameters. Noisy speech as the input signal x_i is decomposed into frequency domain signals $X_{k,i}$ by using the MDFT [6]. The MDFT is obtained by simplifying the original DFT and defined as

$$X_{k,i} = \sum_{n=0}^{N-1} x_{i-n} \cos(2\pi nk / N) \quad (k = 0,1,2\cdot\cdot, N/2-1) \tag{1}$$

where N is the number of samples for the DFT analysis and assumed to be even hereafter. The inverse MDFT (MIDFT) is defined as

$$x_i = \frac{X_{0,i}}{N} + \frac{2}{N}\sum_{k=1}^{N/2-1} X_{k,i} \tag{2}$$

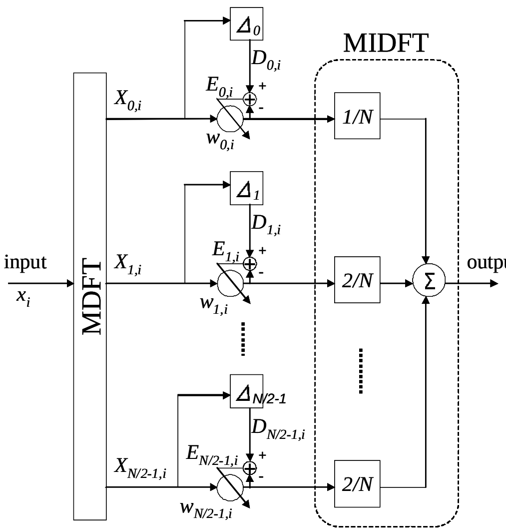


Figure 1. *Speech noise reduction system based on frequency domain ALE.*

The MDFT pair requires only real-value operations and the MIDFT is achieved by summing all MDFT results. When the input signal consists of multiple harmonic frequency signals, the MDFT decomposes it into each harmonic signal keeping the phase difference. Therefore, the input signal can be reconstructed by summing such decomposed frequency signals through the MIDFT.

$D_{k,i}$ is a desired frequency domain signal which is generated by delaying $X_{k,}$ by Δ_k, which is the frequency domain de-correlation parameter. An adaptive weight $w_{k,i}$ is multiplied to each MDFT output $X_{k,}$ and updated to reduce the error $E_{k,i}$ between the weighted MDFT output and the desired one.

$$E_{k,i} = D_{k,i} - w_{k,i} X_{k,i} \tag{3}$$

The normalized step size algorithm is used for updating the adaptive weight [3].

$$w_{k,i+1} = w_{k,i} + 2\mu_{k,i} E_{k,i} X_{k,i} \tag{4}$$

$$\mu_{k,i} = 0.5 / \left(X_k^{perk}\right)^2 \tag{5}$$

where $\mu_{k,i}$ is the normalized step size and X_k^{perk} is the maximum value of each MDFT output. Finally, adapted MDFT outputs are summed in the MIDFT and then the noise-reduced speech signal is reconstructed. Adaptive signal processing is achieved by adjusting the amplitude of the MDFT output; therefore, the phase spectrum is not adjusted. The phase information of the input signal is used in the output signal.

As shown in Fig.1, the frequency domain ALE enables to set the de-correlation parameters in frequency domain. Moreover, if they are adjusted independently, the noise reduction performance may be improved. Since the noise reduction system based on the ALE reduces noise by utilizing the difference of correlation between the speech and the noise, we examine the frequency domain de-correlation parameters by using autocorrelation. In the following, we explain by giving examples.

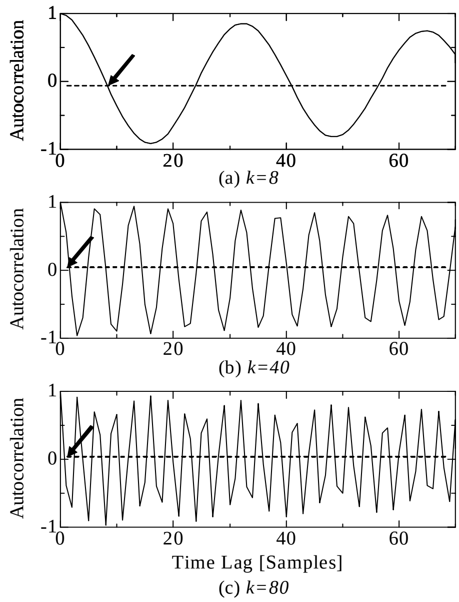


Figure 2. *Auto-correlations of frequency domain signal in a white noise.*

3. FREQUENCY DOMAIN DE-CORRELATION PARAMETER

Firstly, we examine the effective setting for noise reduction. We used the white noise of variance 0.04^2 sampled by 8 kHz with 16 bit resolution. Also, we set N for the DFT to 256 samples (32ms). Figures 2 (a), (b) and (c) show the autocorrelations of frequency domain signal at k=8 (250Hz), 40 (1250Hz), 80 (2500Hz), respectively. In general, the white noise has no correlation except at the time lag of 0. However, the frequency signals of the white noise have correlation and their characteristics depend on the period of frequency signal because they are periodic in frequency domain. Concretely, each auto-correlation has local maximum value at the time lag of integral multiple of N/k and then it becomes zero at the time lag of $N/(k\times 4)$, which makes the frequency signal de-correlated. As a result, the optimal setting of the frequency domain de-correlation parameter for noise reduction can be preliminary determined based on N as

$$\Delta_k = \left\langle \frac{N}{k \times 4} \right\rangle \tag{6}$$

where < > is a constant which adjusts Δ_k to be integer and it must be odd number. For example, when N is 256, the period of the frequency domain signal at k=40 is N/k=6.4. Then, its autocorrelation also has period of 6.4. The first zero-crossing point of the auto-correlation is

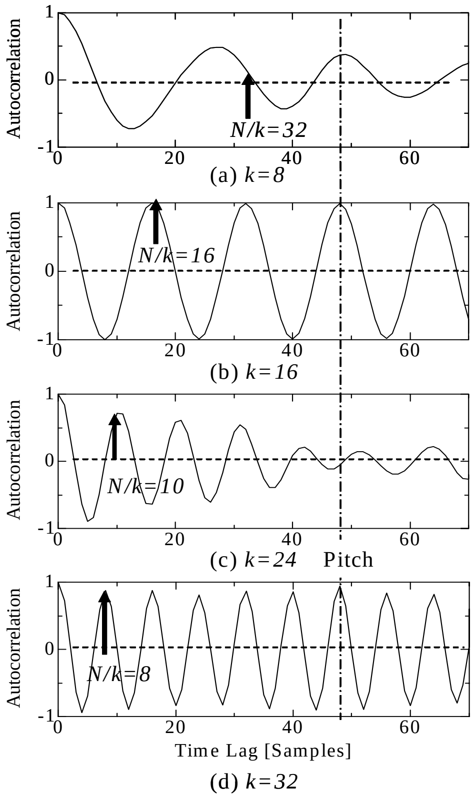


Figure 3. *Auto-correlations of frequency domain signal in a speech.*

calculated by $N/(k\times 4)$ as a black arrow in Fig.2 (b) and then we obtain Δ_k=1.6. However, Δ_k must be integer, so that it is multiplied by 5 and we obtain Δ_k=8.
Next, we examine the frequency domain de-correlation parameter which is effective for speech enhancement. Figure 3 shows the auto-correlations of frequency domain signal in a speech. (a), (b), (c) and (d) are at k=8 (250Hz), 16 (500Hz), 24 (750Hz) and 32 (1000Hz), respectively. The pitch of the used speech was 48 samples (166.7Hz); therefore, frequencies at (b) 500Hz and (d) 1000Hz correspond to those of the harmonics of the speech. It is obvious that auto-correlation has a maximum at the time lag of 48 samples which corresponds to the pitch. The de-correlation parameter as the time lag with maximal auto-correlation enhances the frequency domain signal. From the viewpoint of speech enhancement, the frequency domain de-correlation parameter should be set equal to the pitch.

$$\Delta_k = pitch \tag{7}$$

As a result, it is confirmed that the optimal setting of the frequency domain de-correlation parameter for noise suppression is different from that for speech enhancement. Therefore, it is necessary to change the frequency domain de-correlation parameter, which is adjusted by Eq. (6) in speech pause and by Eq. (7) in speech existent period. This requires the voice activity detector (VAD).

4. INTRODUCTION OF WINDOW FUNCTION

On the other hand, the frequencies at (a) 250Hz and (c) 750Hz in Fig.3 correspond to those of non-harmonics. Therefore, these frequency domain signals must be regarded as of noise and their auto-correlations must have equivalent characteristics with those of noise. However, the auto-correlations in Fig.3 (a) and (c) do not have local maximum at the time lag of integral multiple of N/k as shown in Fig.2. It is due to the spectral leakage.
It is well known that the spectrum side-lobe causes the leakage when the period of the truncation function for the DFT analysis is not equal to the fundamental period of the input (pitch). This phenomenon brings the MDFT output with multiple frequencies. However, the proposed adjusting method of Δ_k demonstrates its ability most effectively when the frequency domain signal is purely a sinusoidal wave.
The window function is effective for coping with such a problem. However, it was not easy to introduce the window function into the MDFT. Let the window function be $W(n)$, the windowed input signal $x_i W(n)$ is processed through the MDFT-MIDFT; therefore, $x_i W(n)$ is reconstructed in the MIDFT. However, $W(n)$ is less than 1 in general, so that we never obtain the original signal of x_i. To solve this problem, we propose to modify the MDFT as given by

$$X_{k,i} = \sum_{n=0}^{N-1} x_{i-n} \cos\left(2\pi\left(n-\frac{N}{2}\right)k/N\right)\times W(n) \tag{8}$$

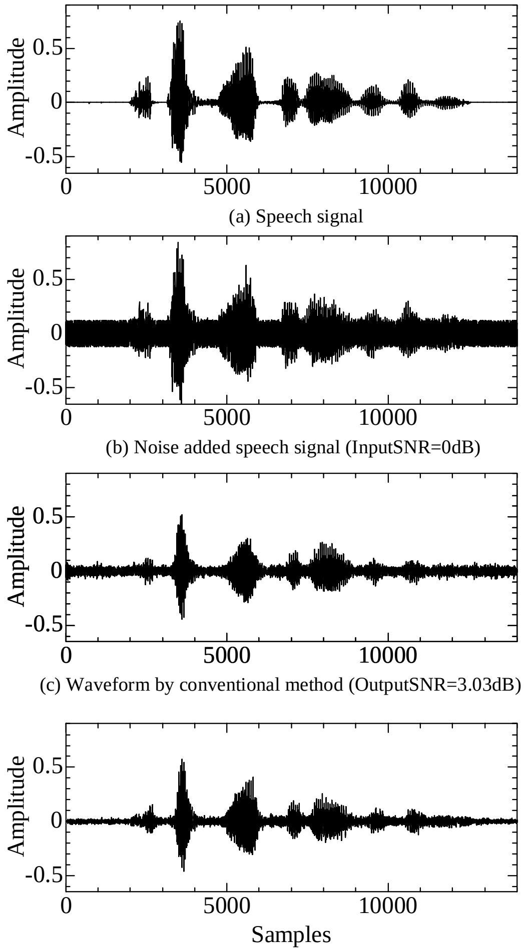
where $\cos(2\pi nk/N)$ in Eq.(1) is shifted by $N/2$. This equation is rewritten as

$$X_{k,i} = \sum_{n'=-N/2}^{N/2-1} x_{i-(n'+N/2)} \cos(2\pi n' k/N)\times W(n) \tag{9}$$

where $n'=n-N/2$. When $n'=0$, that is, $n=N/2$, $W(N/2)$ corresponds to the center of the window function. In general, $W(N/2)$=1; therefore, it can be escaped to reconstruct decreased input signal.
By using the modified MDFT, we can perfectly reconstruct the input signal; however, it is necessary to know that the phase of the MDFT output is delayed by $N/2$ and it results in the system output delay.

5. SIMULATION RESULTS

In order to verify the effectiveness of the proposed methods, we carried out the simulations. The proposed method requires the real-time pitch detection to adjust the frequency domain de-correlation parameter using Eq.(7). While many pitch detection methods have been



(a) Speech signal

(b) Noise added speech signal (InputSNR=0dB)

(c) Waveform by conventional method (OutputSNR=3.03dB)

(d) Waveform by proposed method (OutputSNR=5.99dB)

Figure 4 . *Examples of waveform.*

proposed [7], we adopted a simple detection method based on the auto-correlation in this simulation. The auto-correlation was calculated by using passed 300 sample data sample by sample. The detected period of autocorrelation was regarded as the pitch. Also, the proposed noise reduction system needs the VAD [8]. However, the speech existent or speech pause period was assumed to be known in this simulation.

Original clean speech signal was of a male and sampled by 8 kHz with 16 bit resolution. Noisy speech signal was generated by adding a white noise to the clean speech signal. Figure 4 (a) and (b) show the waveform of the clean speech signal and that of the noisy speech signal, respectively. Input SNR was 0 dB.

For reference, the result by the conventional system with a time domain de-correlation parameter is shown in (c). N for the MDFT was 64 (If N=128, the performance was degraded). The de-correlation parameter in the time domain was adjusted to the pitch in speech existent period and to be constant 64 in speech pause. In this case, output SNR was 3.03 dB.

The result by the proposed noise reduction system with the frequency domain de-correlation parameter and the window function is shown in Fig.4 (d). N for the MDFT was 128. The Hamming window was adopted in this simulation. The output SNR was 5.99 dB. Comparing (d) with (c), the noise reduction performance was improved by 2.96 dB. It is confirmed as less residual noise in the waveform. In subjective evaluation, we also confirmed that the proposed noise reduction technique was effective especially in speech existent period.

6. CONCLUSION

We proposed the speech noise reduction system based on the frequency domain ALE. In addition, we proposed the optimal setting of the frequency domain de-correlation parameter and the modified MDFT to introduce the window function. The noise reduction performance was confirmed through the simulations using a speech signal.

In this paper, we assumed that the speech existent and speech pause periods were known. Some VAD is needed in practical applications.

7. REFERENCES

[1] S. Haykin, Introduction to Adaptive Filters, Macmillan publishing Company, New York, 1984.

[2] R. Sambur, "Adaptive Noise Canceling for Speech Signals," *IEEE Trans. Acoust., Speech, Signal Processing*, vol.ASSP-26, no.5, pp.113-120, Oct. 1978.

[3] I. Nakanishi, Y. Hamahashi, Y. Itoh, and Y. Fukui, "A new structure of frequency domain adaptive filter with composite algorithm," *IEICE Trans. Fundamental*, vol.E81-A, no.4, pp.649-655, Apr. 1998.

[4] I. Nakanishi, Y. Itoh, and Y. Fukui, "Accelerated frequency domain adaptive filter using modified DFT pair and its application to noise canceling," *Proc. of the 2000 IEEE International Symposium on Circuits and Systems (ISCAS2000)*, Geneva, Switzerland, vol.IV, pp.361-364, May 2000.

[5] I. Nakanishi, Y. Itoh, and Y. Fukui, "Noise reduction system based on frequency domain adaptive filter using modified DFT pair," *Proc. of the 2001 IEEE International Symposium on Circuits and Systems (ISCAS2001)*, Sydney, Australia, vol.II, pp.737-740, May 2000.

[6] S. Yoneda, I. Nakanishi, I. Sasaki and A. Ogihara, "Switched-capacitor DFT and IDFT Circuit," *Int. J. Electronics*, vol.67, no.6, pp.839-851, 1989.

[7] B. Gold and N. Morgan, Speech and audio signal processing, John Wiley & Sons Inc., New York, 2000.

[8] G. M. Davis, Noise reduction in speech applications, CRC Press, Florida, 2002.