Speech Noise Reduction System Based on Frequency Domain ALE Using Windowed Modified DFT Pair

Isao NAKANISHI[†a)], *Member*, Yuudai NAGATA[††], *Student Member*, Takenori ASAKURA[††*],
Yoshio ITOH[††], *Members*, *and* Yutaka FUKUI[††], *Fellow*

SUMMARY   The speech noise reduction system based on the frequency domain adaptive line enhancer using a windowed modified DFT (MDFT) pair is presented. The adaptive line enhancer (ALE) is effective for extracting sinusoidal signals blurred by a broadband noise. In addition, it utilizes only one microphone. Therefore, it is suitable for the realization of speech noise reduction in portable electronic devices. In the ALE, an input signal is generated by delaying a desired signal using the decorrelation parameter, which makes the noise in the input signal decorrelated with that in the desired one. In the present paper, we propose to set decorrelation parameters in the frequency domain and adjust them to optimal values according to the relationship between speech and noise. Such frequency domain decorrelation parameters enable the reduction of the computational complexity of the proposed system. Also, we introduce the window function into MDFT for suppressing spectral leakage. The performance of the proposed noise reduction system is examined through computer simulations.
key words: speech processing, noise reduction, adaptive line enhancer, frequency domain, decorrelation parameter, modified DFT pair, window function

1.   Introduction

Speech noise reduction techniques have attracted much attention as speech communication systems are widely used in our daily life. Several methods have been proposed: the use of a noise canceller, the use of a microphone array system and the spectral subtraction method.

The noise canceller and microphone array system require multiple microphones [1]; therefore, they are not cost-effective or suitable for miniaturizing portable systems. On the other hand, the spectral subtraction method requires only one microphone, which is used to estimate a noise spectrum during a speech pause [2]. In other words, the spectral subtraction method is considered as a time-sharing method of one microphone for extracting both a noisy speech and a noise. However, such time sharing of a microphone is disadvantageous to a non-stationary environment.

As another noise reduction method that uses one microphone, the speech noise reduction system based on the adaptive line enhancer (ALE) has been proposed [3]. The ALE is effective for extracting sinusoidal signals blurred by

a broadband noise [1], and is based on the adaptive digital filter (ADF), which is advantageous for a nonstationary environment. However, the convergence speed of the ADF is degraded when the input signal of the ADF is a colored signal such as speech. This results in the degradation of noise reduction performance.

To solve the above problem, we have introduced the frequency domain adaptive filter (FDAF) into the ALE [4]–[6]. The FDAF always achieves a faster convergence than the time domain ADF even if the input signal is colored [7]. In particular, we utilized the modified discrete Fourier transform (MDFT) pair [8] in the FDAF.

In the ALE, the input signal of the ADF is generated by delaying a desired signal. The time delay is constant and called the decorrelation parameter, which makes the noise in the desired signal decorrelated with that in the input signal. On the other hand, the frequency domain ALE enables the setting of the decorrelation parameter in the frequency domain [9]. In the present paper, the authors propose to set the decorrelation parameter in the frequency domain and adjust it according to the relationship between speech and noise. In addition, such a frequency domain ALE reduces enables decreasing the number of MDFT operations.

The present paper is organized as follows: In Sect. 2, the speech noise reduction system based on the frequency domain adaptive line enhancer is explained. In particular, the MDFT is remodified for introducing the window function. Next, the optimal setting of frequency domain decorrelation parameters is introduced in Sect. 3. In Sect. 4, the effectiveness of the proposed system is examined through several simulations. Finally, in Sect. 5, the conclusions are presented.

2.   Speech Noise Reduction System Based on Frequency Domain ALE

2.1   Conventional System

The fundamental structure of the proposed speech noise reduction system is based on the adaptive line enhancer (ALE). Figure 1 shows a block diagram of the ALE. The input signal x_i is generated by delaying the desired signal d_i by the decorrelation parameter Δ, which makes the noise in the desired signal decorrelated with that in the input signal. As a result, the noise is reduced while correlative sinusoidal waves are emphasized in the ALE.
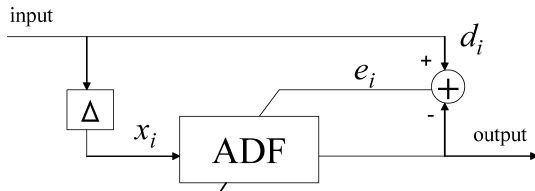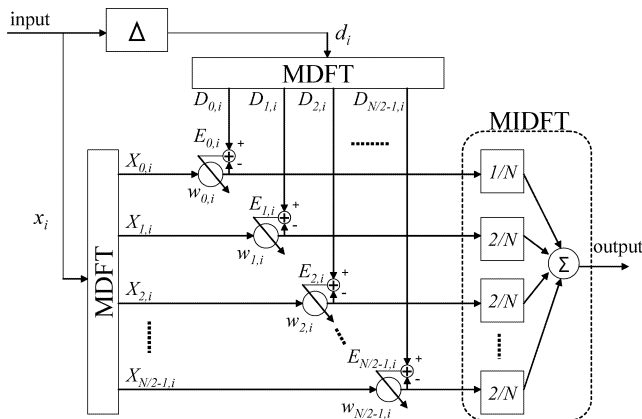
Fig. 1 Block diagram of ALE.

Fig. 2 Speech noise reduction system based on frequency domain ALE.

However, when the input signal of the adaptive digital filter (ADF) is colored such as speech, the convergence speed of the ADF is decreased. This degrades the noise reduction performance of the ALE.

To cope with the above problem, we have proposed to adopt the frequency domain adaptive filter (FDAF) instead of the ADF [4]–[6]. Such a speech noise reduction system based on the frequency domain ALE is shown in Fig. 2. The FDAF always achieves faster convergence than the ADF even if the input signal is colored [7].

Moreover, the proposed FDAF adopts the modified DFT (MDFT) pair [8]. The MDFT is obtained by simplifying the original DFT and defined as

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

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

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

An MDFT pair requires only real-value operations and the MIDFT is achieved by summing the MDFT outputs. The MDFT decomposes an input signal into harmonic signals while maintaining phase differences; therefore, adaptive signal processing can be simply realized by adjusting the amplitude of the MDFT output signal.

By using the MDFT, the input and desired signals are simultaneously decomposed into the harmonic signals $X_{k,i}$
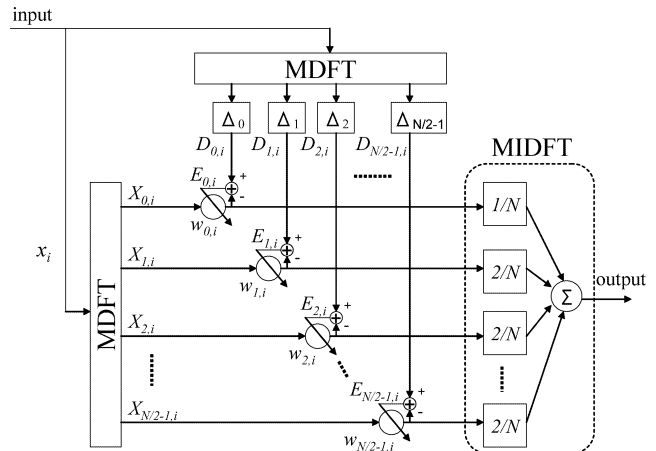

Fig. 3 Speech noise reduction system based on frequency domain ALE with frequency domain decorrelation parameters.

and $D_{k,i}$, respectively. The adaptive weight $w_{k,i}$ is multiplied by each $X_{k,i}$ and updated to reduce the error $E_{k,i}$ between $X_{k,i}$ and $D_{k,i}$:

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

For updating adaptive weight, a normalized step size algorithm is used because it is essential for achieving fast convergence in the FDAF [7]. Thus,

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

$$\mu_{k,i} = \frac{0.5}{|X_{k,i}|_p^2}, \tag{5}$$

where $\mu_{k,i}$ is a normalized step size, and $|X_{k,i}|_p$ is the maximum of each MDFT output.

Finally, adapted MDFT outputs are summed in the MIDFT and then a noise-reduced speech signal is reconstructed. Phase information on the input signal is also used in the output signal.

2.2 Frequency Domain Decorrelation Parameter

As shown in Figs. 1 and 2, the decorrelation parameter is generally inserted in the time domain. However, it can be moved to the frequency domain [9]. Such a structure is illustrated in Fig. 3 where $\Delta_k (k = 0, 1, \cdots, N/2 - 1)$ are the frequency domain decorrelation parameters that can be set independently. If the frequency domain decorrelation parameter is individually adjusted in each MDFT output, it is expected to improve noise reduction performance.

Figure 3 is revised to Fig. 4. Note that only one MDFT is required in this structure whereas two MDFTs are required in the conventional system in Fig. 2. Such a reduction in computational complexity becomes possible only after the introduction of the frequency domain decorrelation parameter.

2.3 Introduction of Window Function into MDFT

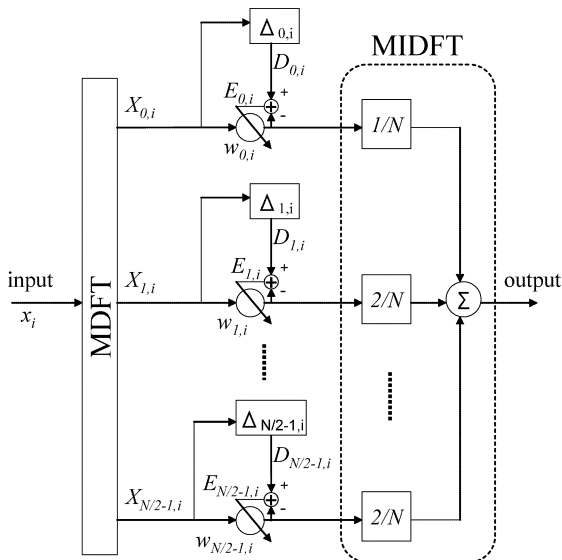The spectrum sidelobe causes *leakage* when the period of
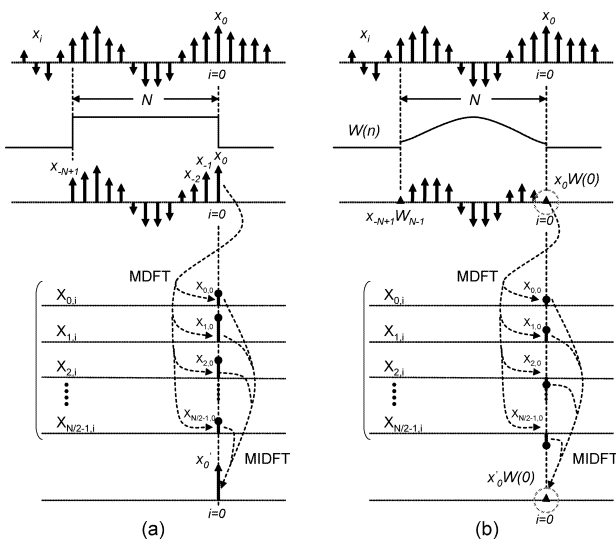
Fig. 4 Proposed speech noise reduction system.

Fig. 5 Influence of window function on MDFT.

Fig. 6 Relationships between cosine function and window function: (a) for MDFT case and (b) for modified MDFT case.

the truncation function for DFT analysis is not equal to the fundamental period of the analyzed signal. This phenomenon causes each MDFT output to have mixed neighbor frequencies. As described later, it is assumed that frequency domain signals are sinusoidal in the adjustment of the frequency domain decorrelation parameter. If the frequency domain signals are not purely sinusoidal, noise reduction performance is degraded.

For reducing such spectrum leakage, the window function is effective; however, there are two problems for introducing the window function into MDFT. One is the well-known problem that introducing the window function leads to low-pass filtering and the other is the problem inherent in the MDFT.

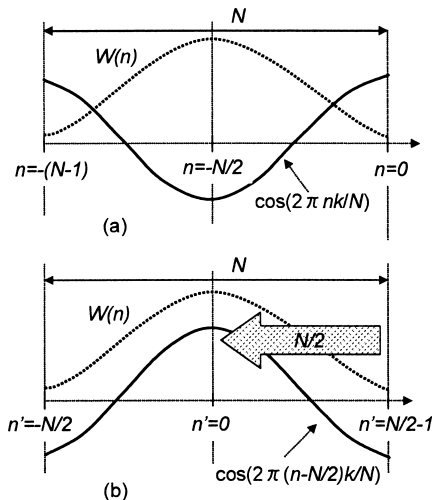Figure 5 shows MDFT-MIDFT processing. In the case of nonwindowed (a), at $i = 0$, the current input x_0 and past

data: total N sampled data in an input signal is processed through the MDFT (Eq. (1)) and then current data of frequency domain signals are obtained as the MDFT output $X_{k,0}$. Next, the frequency domain signals are recomposed in the MIDFT (Eq. (2)) and then the current input is perfectly reconstructed as x_0'.

On the other hand, in the windowed case (b), where $W(n)$ represents the window function, windowed N sample data are processed in the MDFT and then the windowed input $x_0'W(0)$ is recomposed in the MIDFT. As a result, the current input signal x_0 is not reconstructed through the MDFT-MIDFT if the window function is simply introduced into the MDFT.

To enable the introduction of the window function, we propose to modify the MDFT again as given by

$$X_{k,i} = \sum_{n=0}^{N-1} x_{i-n} \cos(2\pi(n - N/2)k/N), \qquad (6)$$

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

$$X_{k,i} = \sum_{n'=-N/2}^{N/2-1} x_{i-(n'+N/2)} \cos(2\pi n'k/N), \qquad (7)$$

where $n' = n - N/2$.

Figure 6 shows the relationships between the cosine function and the window function: (a) for the MDFT case and (b) for the above-modified MDFT case. When $n' = 0$, that is, $n = N/2$, $W(N/2)$ corresponds to the center of the window function. In general window functions, $W(N/2) = 1$; therefore, the reconstructed input signal $x_{i-N/2}' \, W(N/2)$ becomes equal to the current original input signal $x_{i-N/2}$ and so we can avoid decreasing the original input signal by multiplying the window function.

In addition, this modification simultaneously overcomes the problem of low-pass filtering. It is certain that the window function influences MDFT outputs as spectra;

however, the original input signal is perfectly reconstructed and not filtered.

By using the modified MDFT, the problems with the window function can be solved; however, it is necessary to know that the phase of the modified MDFT output is delayed by $N/2$, which results in system output delay. It corresponds to 16 ms when the sampling rate is 8 kHz and $N = 256$. The effectiveness of introducing the window function into the MDFT was reported in detail in Ref. [10].

3. Optimal Setting of Frequency Domain Decorrelation Parameters

Since noise reduction systems based on the ALE reduce noise by utilizing the difference in correlation between speech and noise, we examine frequency domain decorrelation parameters using the autocorrelation

$$\phi(\tau) = \left(\sum_{l=0}^{L} s_l \cdot s_{l+\tau} \right) \Big/ \sum_{l=0}^{L} s_l^2, \qquad (8)$$

where τ is a time lag index and L is the number of all sampled data in the signal s_l.

3.1 Frequency Domain Decorrelation Parameters for Noise Suppression

Firstly, the effective setting for noise suppression is examined by giving an example. A white noise of variance 0.04^2 sampled at 8 kHz and a 16 bit resolution was used, and N for DFT analysis was 256 samples (32 ms). A Hamming window was used as the window function. Figure 7 shows the autocorrelation of frequency domain signals of the white noise at $k = 8, 9, 10, 20$. In general, the white noise has no correlation except at the time lag of 0. However, its frequency domain signals have correlation because they are periodic in the frequency domain. Moreover, their characteristics depend on the period of the frequency domain signal. Concretely, each autocorrelation has local maxima at the time lag of integral multiple of N/k; thus, it becomes zero at the time lag of $N/(k \times 4)$.

If the time lag causes a zero correlation between two signals, it is effective for their decorrelation. As a result, the optimal setting of the frequency domain decorrelation parameter for noise suppression can be determined based on N as

$$\Delta_k = \left\langle \frac{N}{k \times 4} \right\rangle, \qquad (9)$$

where $\langle \ \rangle$ expresses processing to an integer. For instance, when $N = 256$ and $k = 10$, $N/(k \times 4) = 1.6$ is multiplied by 5 and then we obtain $\Delta_k = 8$.

3.2 Frequency Domain Decorrelation Parameters for Speech Enhancement
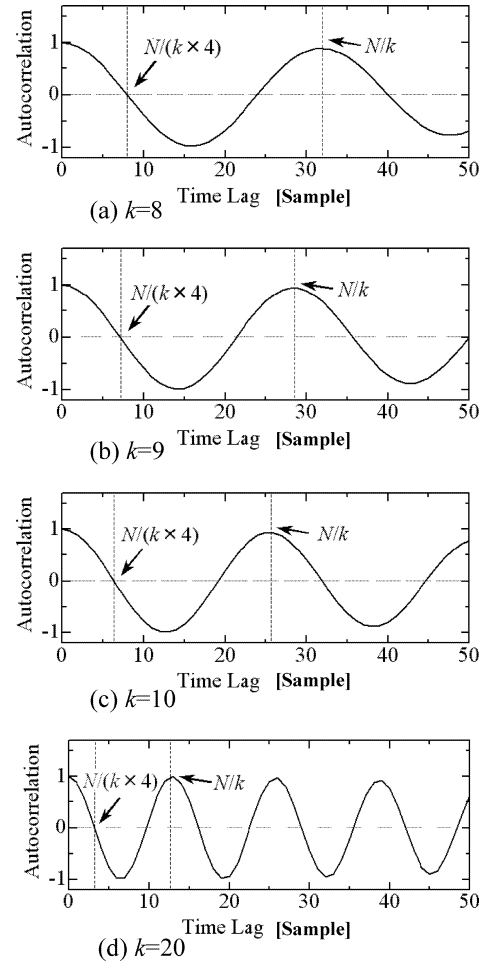
Next, the frequency domain decorrelation parameter, which

Fig. 7 Autocorrelations of frequency domain signals in white noise.

is effective for speech enhancement, is examined. Part of a male speech sampled at 8 kHz and a 16 bit resolution was used as an example. The pitch of the speech was 64 samples (125.0 Hz).

Figure 8 shows the autocorrelations of frequency domain signals in the above speech; (a), (b), (c) and (d) are at $k = 8$ (250.0 Hz), 9 (281.3 Hz), 10 (312.5 Hz) and 40 (1250.0 Hz), respectively. From these results, it is confirmed that the autocorrelation becomes maximum almost at the time lag of N/k. It is easy to understand this phenomenon because the period of each frequency signal in MDFT is N/k.

For reference, the spectrum of the speech used is shown in Fig. 9. This spectrum was obtained by averaging 256 sampled data in each MDFT power. There are spectral peaks at the integral multiple of pitch frequency since speeches generally have a harmonic structure consisting of a fundamental frequency signal and its harmonic signals.

The time lag where the autocorrelation of a signal becomes maximum is effective for enhancing the signal. As a result, when k is identical to a harmonic frequency as (a) or (d) in Fig. 8, the frequency domain decorrelation parameter must be set equal to the time lag where its autocorrelation
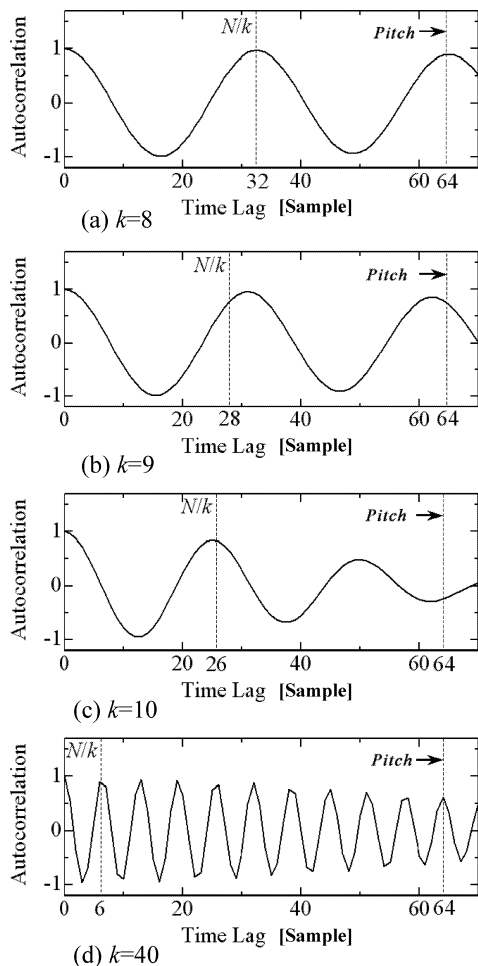
(a) *k*=8

(b) *k*=9

(c) *k*=10

(d) *k*=40

Fig. 8 Autocorrelations of frequency domain signals in speech.
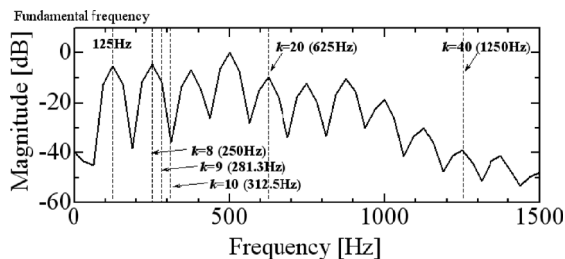

Fig. 9 Spectrum of speech used.

becomes maximum. Thus,

$$\Delta_k = \frac{N}{k}. \tag{10}$$

However, the above setting is not always effective. In general, the speech is a narrow-band signal while the noise is a wide-band signal. When a noise element is more dominant than a speech element in a frequency domain signal, enhancing the frequency domain signal results in increased noise.

To solve such a trade-off problem, the frequency domain decorrelation parameter is set equal to the pitch when the speech element is not dominant. Let

$$\Delta_k = Pitch, \tag{11}$$

where *Pitch* expresses the pitch period. As shown in Fig. 8, the autocorrelation of a speech harmonic also has a local maximum at the time lag, which is equal to the pitch. On the other hand, autocorrelation decreases as time lag increases. Therefore, the frequency domain decorrelation parameter, which is larger than N/k is expected to suppress the noise element compared with that in Eq. (10), while enhancing the frequency domain signal.

3.3 Adjusting Frequency Domain Decorrelation Parameters

As mentioned above, the effective setting of the frequency domain decorrelation parameter for noise suppression is different from that for speech enhancement. In the present paper, we propose to adjust the decorrelation parameter as follows.

In a speech pause, the frequency domain decorrelation parameter is set to suppress the noise using Eq. (9). In a speech-existent period, the decorrelation parameter is set using Eq. (10) or Eq. (11) for speech harmonics according to the dominance of the speech element in each frequency domain signal. For nonharmonics, the decorrelation parameter is set to suppress the noise using Eq. (9). Consequently, it becomes important to detect voice activity and speech dominancy. Voice activity detection is described later. In the following paragrapgh, how to determine whether a speech element is dominant is explained.

First of all, pitch, that is, the fundamental frequency (f_f) of the speech, is detected. The frequency of a harmonic (f_h) can be easily calculated since it corresponds to the integral multiple of fundamental frequency. However, how to deal with its neighbor frequencies is a problem. By introducing the window function, the spectrum resolution of the mainlobe is degraded, so that neighbor frequency domain signals also contain considerable speech elements. In the present paper, a neighbor frequency domain signal is regarded as a harmonic when its frequency f_k satisfies

$$(f_h - f_f \times \alpha) \le f_k \le (f_h + f_f \times \alpha), \tag{12}$$

where $f_k = k \times f_s/N$ when the sampling rate is f_s. α is a coefficient for neighbors on harmonics and is less than 1.

Moreover, even if a frequency domain signal corresponds to a harmonic, it is important to switch the setting of the decorrelation parameter according to the dominance of the speech element in the frequency domain signal. In the present paper, such dominance is determined by dividing the frequency domain into four regions. In general, speech elements are maldistributed in the low-frequency range. Additionally, it is mentioned in Ref. [11] that the spectrum of Japanese vowels is represented by roughly two "mountains." Therefore, the distribution of speech elements is modeled as in Fig. 10. Assuming that the noise is a broad band and its spectrum is described as a dotted line, the speech element is regarded as dominant in Domain 1 and inversely the noise

element is dominant in Domain 3. Domains 2 and 4 correspond to their intermediate regions. As a result, harmonic signals in Domain 1 are processed using Eq. (10) to enhance speech elements. On the other hand, they are processed using Eq. (9) in Domain 3 to suppress noise elements. In Domains 2 and 4, a compromise setting of speech enhancement and noise suppression is applied using Eq. (11).

Finally, the flowchart of the proposed adjustment method for the frequency domain decorrelation parameter is shown in Fig. 11.

3.4 Detections of Pitch and Speech Existence

The proposed adjustment method requires the detection of pitch and speech existence. Several detection methods have been proposed [12], [13]. However, our main aim in this study is not to propose such detection methods but to con-

firm the effectiveness of the frequency domain decorrelation parameter. In this study, we adopt simple methods for the detection of pitch and speech existence.

Firstly, the pitch detection method is explained. The speech can be assumed to be stationary during 20–40 ms [12]. It corresponds to a 256 sample period at an 8 kHz sampling rate. By calculating the autocorrelation of the speech using the current and past 255 data sample by sample, pitch can be detected as the time lag where the autocorrelation becomes maximum. Moreover, the tolerance range of the detection is set to reduce misdetection. The pitch frequency is normally distributed in the logarithmic frequency domain, and its average and standard deviation are respectively 125.0 Hz and 20.5 Hz for the male voice, and they are 250.0 Hz and 41.0 Hz for the female voice [12]. Assuming an occurrence probability of 99 %, the pitch ranges are 84.5–184.8 Hz for the male voice and 169.1–369.7 Hz for the female voice. As a result, the pitch frequency range is assumed to be 84.5–369.7 Hz, which corresponds to 2.7–11.8 ms. If the calculated pitch period is out of this range, pitch is not updated.

Next, for detecting speech existence, we use the speech and noise level detectors proposed in Ref. [14]. The description of the detectors is omitted for lack of space. It is described in detail in Ref. [14]. Based on the above assumption, it can be considered that principal speech elements are contained in the MDFT output at $k = 1$ (250.0 Hz) when $N = 32$ at an 8 kHz sampling rate. Firstly, input signal-to-noise ratio (SNR) is estimated using such an MDFT output. Concretely, when the output of the speech level detector is DS_i and that of the noise level detector is DN_i, input SNR
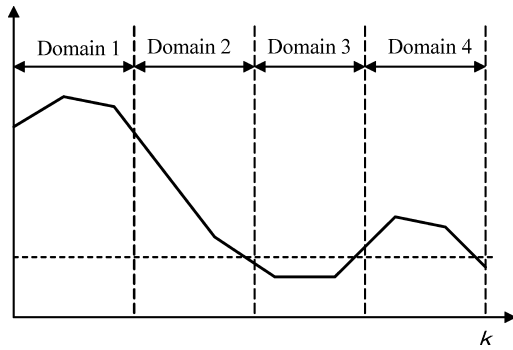

Fig. 10   Modeled distribution of speech elements.
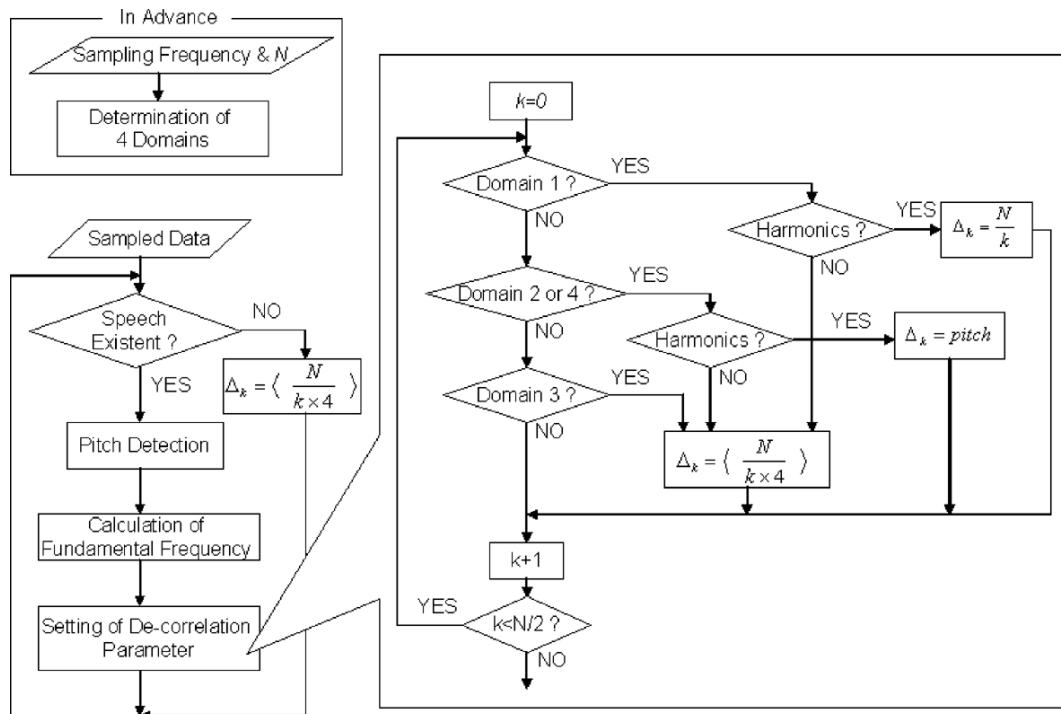

Fig. 11   Proposed adjustment method of frequency domain decorrelation parameter.

is estimated by

$$\text{Estimated input SNR} = 10\log_{10}\frac{DS_i}{DN_i}. \qquad (13)$$

Secondly, the estimated input SNR is compared with the threshold β. If it is larger than the threshold, then the speech is considered to be existent.

4. Simulations

To verify the effectiveness of the proposed methods, the authors carried out simulations. An original clean speech signal was /watashiwasoreonozomu/ of a male and sampled at 8 kHz and a 16 bit resolution. A noisy speech signal was generated by adding a white noise of variance 0.04^2 to the original speech signal. It was assumed that Domain 1 was $0 \leq k \leq 24$, Domain 2 was $24 < k \leq 48$, Domain 3 was $48 < k \leq 80$, and Domain 4 was $80 < k$. The coefficient

for neighbors on harmonics α was 0.25 and the threshold for the detection of speech existence β was 12.5 dB. These were obtained by trial and error in preliminary experiments. N was 256 and a Hamming window was used.

In Fig. 12, (a) and (b) show the waveform of the original speech signal and that of the noisy speech signal, respectively. The input SNR was 0.0 dB. For reference, the result obtained using a time domain decorrelation parameter is shown as (c). The decorrelation parameter was adjusted to the pitch in speech-existent periods and to be constant (64) in speech pauses. In this case, the output SNR was 5.42 dB. The result obtained using the proposed noise reduction system with the frequency domain decorrelation parameter is shown as (d). The output SNR was 7.81 dB. Comparing (d) with (c), noise reduction performance was improved by 2.39 dB. This is also confirmed as less residual noise in waveforms. Their spectrograms are shown in Fig. 13. It can be observed that the residual noise obtained using the proposed system is less than that obtained using
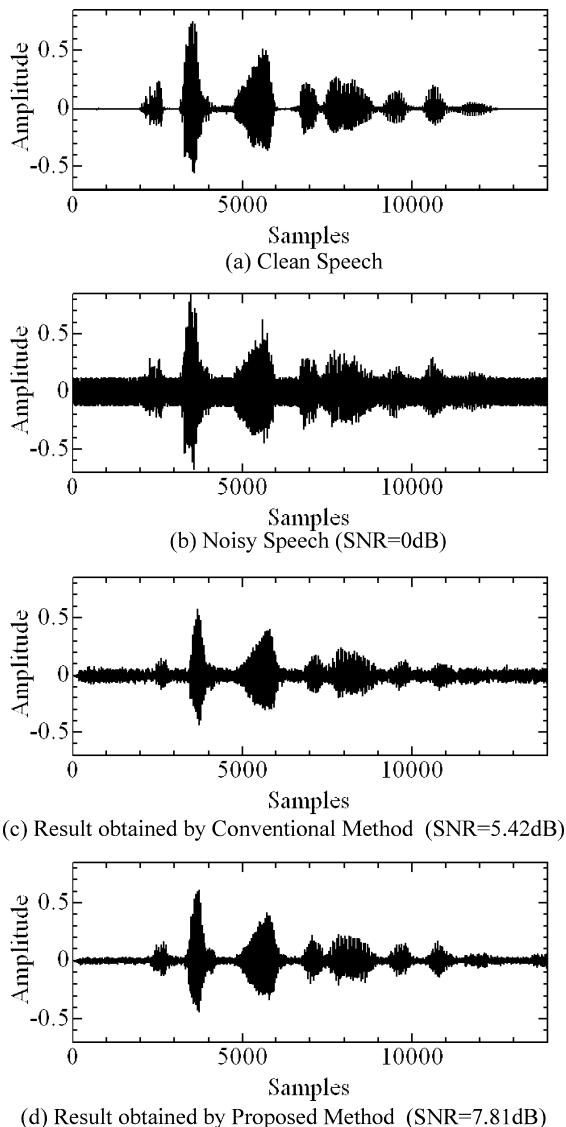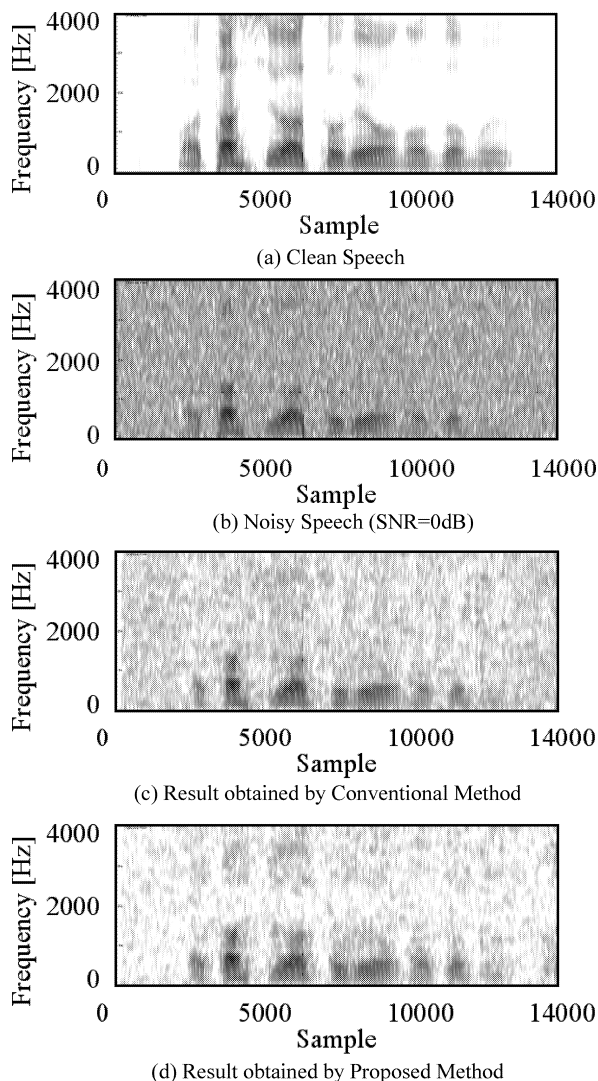

(a) Clean Speech

(b) Noisy Speech (SNR=0dB)

(c) Result obtained by Conventional Method (SNR=5.42dB)

(d) Result obtained by Proposed Method (SNR=7.81dB)

Fig. 12   Waveforms in stationary case.

(a) Clean Speech

(b) Noisy Speech (SNR=0dB)

(c) Result obtained by Conventional Method

(d) Result obtained by Proposed Method

Fig. 13   Spectrograms.

957
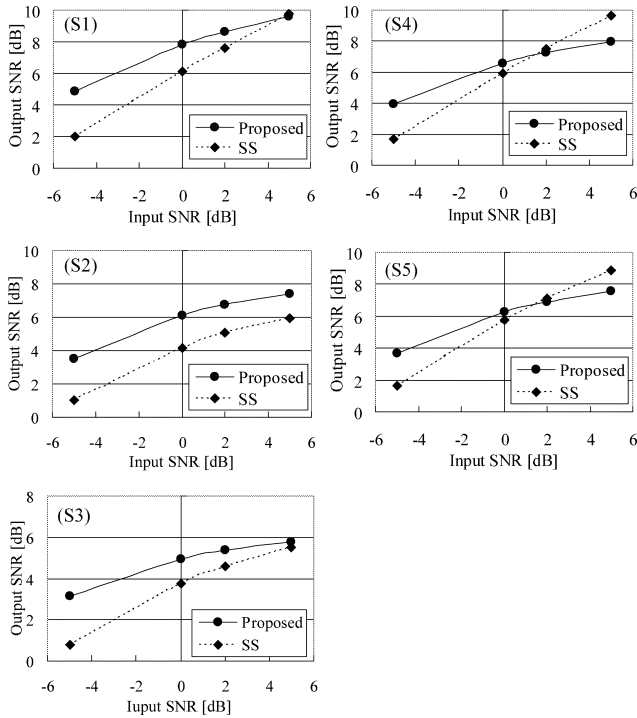
Fig. 14 Output SNR vs. Input SNR in stationary case.

(a) Speech: s4

(b) Noisy Speech (SNR=5.0dB)

(c) Result obtained by proposed method (SNR=8.0dB)

(d) Result obtained by modified SS method (SNR=9.7dB)

Fig. 15 Waveforms in comparison with modified SS system.

the time domain decorrelation parameter and speech elements are extracted relatively clearer using the system.

Next, the comparisons of the proposed system with a modified spectral subtraction (SS) system [15] were performed using the following speech signals.

- s1:/watashiwasoreonozomu/
- s2:/soredejidainonagaregamienai/
- s3:/nazekoredakenokanegaugoitanoka/
- s4:/tanosekaidemosorewaonajidatoomoundesu/
- s5:/hyousyoudaidenoegaowatotemokireideshita/

The SS is well known as a speech noise reduction system in the frequency domain and it requires only one microphone similarly to the proposed system. Other conditions of simulation were identical to those in the previous simulation.

Results are summarized in Fig. 14. It is clear that the proposed system is more effective than the modified SS system when the input SNR is less than approximately 2 dB. The processing of noise reduction in the frequency domain is common to both systems but speech elements can be enhanced based on their correlations in the proposed system even if noise elements are dominant. In the SS system, if the speech element is buried in the noise element, it cannot be extracted by subtraction.

Waveform examples are shown in Fig. 15: (a) original speech signal: s4; (b) noise-added speech; and (c) and (d) are obtained using the proposed system and the modified SS system, respectively. The input SNR was 5.0 dB. The residual noise of the proposed method is less than that of the modified SS system, whereas the output SNR of the proposed system becomes lower than that of the modified
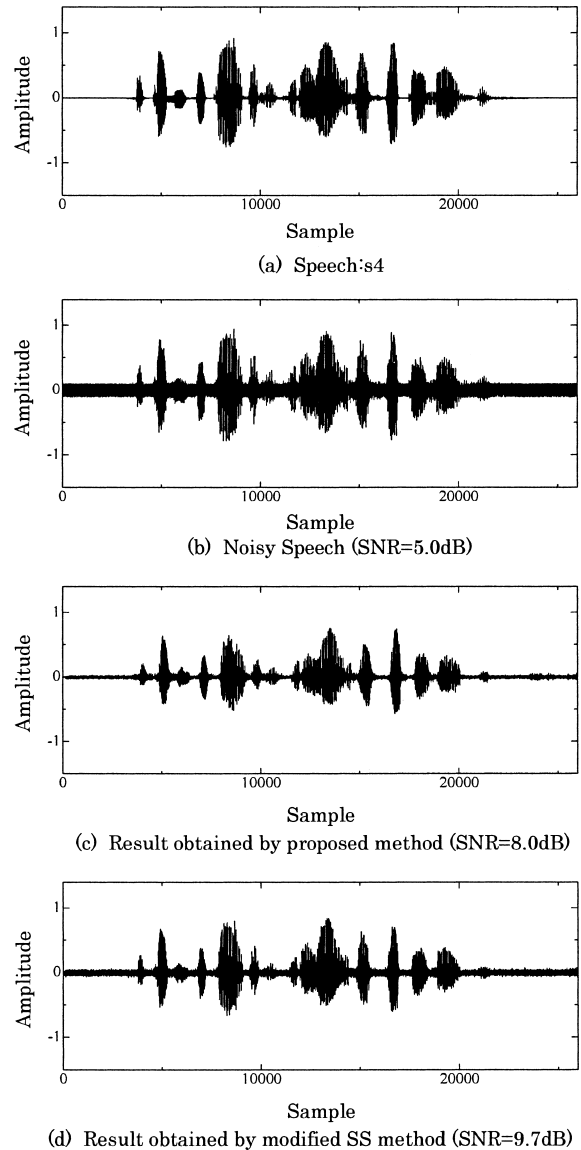
SS system. The reason is that even speech elements are regarded as noises and then suppressed in Domain 3. For improving the performance of the proposed system, it is a future work to adjust the processing in four frequency domains according to the estimated input SNR.

Next, comparisons using a nonstationary noise were performed. As the nonstationary noise, the authors used a noise generated by a car's motion. Other conditions were identical to those in the previous simulations. Figure 16 shows difference between output SNR and input SNR. Compared with that in the stationary case, the degradation of the performance in the proposed system is mainly due to the misdetection of speech existence. The original SS system [2] is ineffective in reduceing nonstationary noise since a noise spectrum is preliminarily estimated during a speech pause and is then subtracted from the noise speech spectrum in the following speech-existent period. The modified
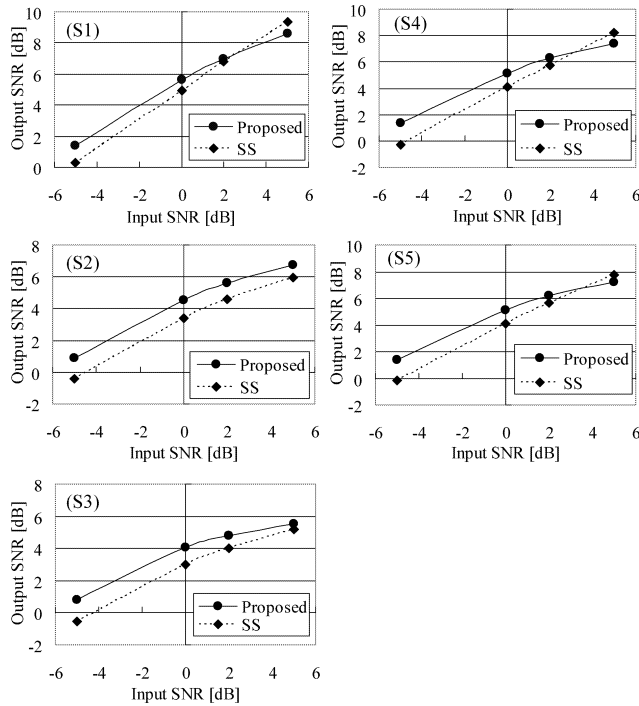
Fig. 16 Output SNR vs. Input SNR in nonstationary case.

(a) Speech: s5

(b) Noisy Speech (SNR=-5.0dB)

(c) Result obtained by proposed method (SNR=1.4dB)

(d) Result obtained by modified SS method (SNR=-0.2dB)

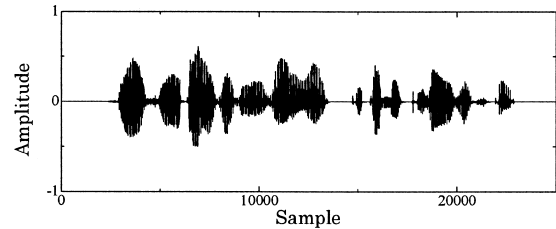**Fig. 17** Waveforms in nonstationary case.

SS system is designed not to require such preliminary estimation but it assumes that the noise is relatively stationary compared with the speech [15]. In these simulations, such an assumption was not made, resulting in the degradation of noise reduction performance.

For example, waveforms are shown in Fig. 17 where (a) and (b) are the waveforms of the original speech: s5 and the noisy speech, and (b) and (c) are those obtained by the proposed system and the modified SS system, respectively. The input SNR was −5.0 dB. It is confirmed that the proposed system is also effective in this nonstationary case.
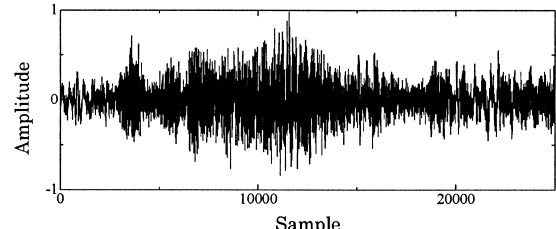
5. Conclusion

The speech noise reduction system based on the frequency domain ALE was proposed. The proposed structure required only one MDFT pair. In addition, the optimal setting of frequency domain decorrelation parameters was proposed. They were adjusted according to the existence of speech, the harmonic structure of speech, and the dominancy of speech elements in each frequency domain signal. It was confirmed that the proposed noise reduction system is effective through several simulations.
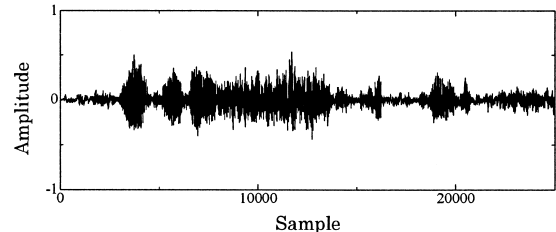
The detections of pitch and speech existence are important in the proposed system. In the present study, we adopted simple methods of such detections to confirm the effectiveness of the proposed frequency domain decorrelation parameter. As the performance of those detections is increased, that of the proposed noise reduction system is improved. More accurate and robust methods of detection must be further studied.
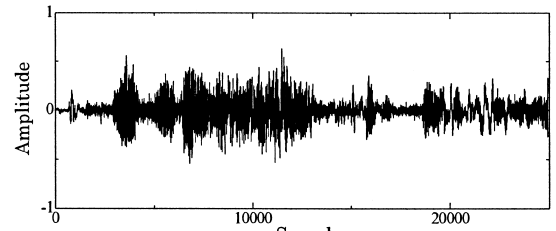
Moreover, the dominance of speech elements was determined by dividing the frequency domain into four regions. This method is simple and easy to realize but non-adjustable. Therefore, damage to the original speech signals was remarkable, particularly under a high-SNR condition. It is a future work to adjust such determination depending on the SNR condition.

References

[1] B. Widrow and S.D. Stearns, Adaptive Signal Processing, Prentice-

Hall, N.J., 1985.

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

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

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

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

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

[7] D.F. Marshall, W.K. Jenkins, and J.J. Murphy, "The use of orthogonal transforms for improving performance of adaptive filters," IEEE Trans. Circuits Syst., vol.36, no.4, pp.474–484, April 1989.

[8] S. Yoneda, I. Nakanishi, I. Sasaki, and A. Ogihara, "Switched-capacitor DFT and IDFT circuit," Int. J. Electron., vol.67, no.6, pp.839–851, Dec. 1989.

[9] I. Nakanishi, T. Asakura, Y. Itoh, and Y. Fukui, "Frequency domain decorrelation parameter in speech noise reduction system based on frequency domain adaptive line enhancer," Proc. 2004 47th IEEE Midwest Symposium on Circuits and Systems (MWSCAS2004), vol.2, pp.13–16, Hiroshima, Japan, July 2004.

[10] I. Nakanishi, T. Asakura, Y. Itoh, and Y. Fukui, "Speech noise reduction using frequency domain ALE—Introduction of window function for performance improvement (in Japanese)," IEICE Technical Report, SIP2004-155, March 2005.

[11] T. Miura, eds., Hearing and Speech, Coronasha, Tokyo, 1980.

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

[13] F. Basbug, S. Nandkumar, and K. Swaminathan, "Noise reduction for IS-136 systems," Proc. International Workshop on Acoustic Echo and Noise Control (IWAENC1999), pp.56–59, Sept. 1999.

[14] I. Nakanishi, Y. Itoh, Y. Fukui, and K. Fujii, "Noise reduction system using modified DFT pair," Proc. 2001 IEEE International Symposium on Circuits and Systems (ISCAS2001), vol.II, pp.9–12, Sydney, Australia, May 2001.

[15] A. Kouda, T. Usagawa, and M. Ebata, "A new spectral subtraction method using the power change for noise spectrum estimation (in Japanese)," J. Acoust. Soc. Jpn., vol.58, no.8, pp.493–500, Aug. 2002.

Yuudai Nagata is now a Master's student in the Department of Electrical and Electronic Engineering, Tottori University, Japan. His research interests are in speech noise reduction.

Takenori Asakura received his B.E. and M.E. degrees in Electrical Electronic Engineering from Tottori University, Japan in 2003 and 2005, respectively. In 2005, he joined the NTT DoCoMo Cyugoku.

Yoshio Itoh received his B.E. degree in Electronics Engineering from Osaka Institute of Technology, M.E. and Dr.E. degrees in Electrical Engineering from Osaka Prefecture University in 1979, 1981, and 1991, respectively. He is now a professor in the Faculty of Engineering, Tottori University, Japan. His research interests are in digital signal processing and digital communication systems. From 2000 to 2004, he was an associate editor of Trans. on Fundamentals of IEICE. He is a member of the IEEE and the Acoustical Society of Japan.

Yutaka Fukui received his B.E., M.E., and Dr.E. degrees in Electrical Engineering from Osaka Prefecture University, Japan in 1964, 1966, and 1974, respectively. He is currently a professor in the Faculty of Engineering, Tottori University, Japan. His research interests are in analog and digital signal processing. He is a member of the IEEE and the Institute of Electrical Engineers of Japan.

Isao Nakanishi received his B.E., M.E., and Dr.E. degrees in Electrical Engineering from Osaka Prefecture University, Japan in 1984, 1986, and 1997, respectively. He is now an assistant professor in the Faculty of Regional Sciences, Tottori University, Japan. His research interests are in digital signal processing and biometrics. He is a member of the IEEE and the Information Processing Society of Japan.