Speech Enhancement Based on Auto Gain Control

Yoshifumi Nagata, Toyota Fujioka, and Masato Abe, *Member, IEEE*

***Abstract*—We propose a new method of speech enhancement based on auto gain control (AGC) using two channel inputs to deal with transient noises. Auto gain control is considered to be relatively ineffective for reducing noises that are superimposed on speech. Nevertheless, it offers advantages for addressing problems posed by musical noise and spectral distortion. This method combines two operations for obtaining accurate gain. One is spectral subtraction for two-channel input (2chSS); the other is self-offset of the noise with pre-whitening. This study also addresses a coherence based post-filter to reduce uncorrelated noise components among channels. The proposed method is evaluated in experiments across three noise conditions in which (i) impulsive noises, (ii) stationary car noise, and (iii) speech noise are present, respectively. Objective measures and spectrograms demonstrate marked improvements over other two-microphone based methods, but subjective preference tests reveal that the proposed method is less preferred than the equivalent of a nonprocessed signal in the case of stationary car noise (ii). The performance of the proposed method and the conventional 2chSS were even in the case of speech noise (iii). These results of subjective tests reflect some disadvantages of the AGC processing. Those drawbacks involve degradation of noise consistency in stationary noise conditions and residual noises in desired speech segments. Nevertheless, subjective tests in the case of noise (i) demonstrate that the proposed method is the most preferred among the methods compared here. The effectiveness of the proposed method is confirmed particularly for this noise condition.**

***Index Terms*—Auto gain control, directional microphone, spectral subtraction, speech enhancement, weighting function, Wiener gain.**

I. Introduction

SPECTRAL SUBTRACTION (SS) [1] is a widely used technique. However, ordinary SS can suppress only the averaged spectrum of a noise because it assumes the noise to be stationary. To overcome this limitation, a two-channel version of SS (2chSS) has been proposed [2] for reducing not only averaged noise, but instantaneous noise. This method introduces a blocking filter as used in the Griffiths-Jim generalized sidelobe canceler (GSC) [3], [4], which outputs instantaneous noise in the current frame while suppressing the desired speech. 2chSS compensates this noise signal to properly represent the noise power spectrum contained in the primary signal while ignoring the phase spectrum.

Reportedly, 2chSS is effective even in conditions with multiple noise sources. Nevertheless, it is difficult to deal with transient noises, such as impulsive noises, because transient noises provide insufficient duration to learn the compensation coefficients in most cases. Because transient noises are observed commonly and can be more perceptible than continuous noise, we believe that it is important to cope with those. Speech enhancement techniques based on an adaptive beamformer [3]–[6] are also unsuitable for suppressing transient noises because such methods have filters that must be produced through learning from the noises.

On the other hand, other authors have proposed microphone array post-filters which are the estimates of a Wiener filter obtained from multi channel input [7]–[12]. A coherence based filter [12]–[14], which has similar characteristics to the Wiener type filter, is also used as a post-filter. While ordinary post-filters are optimal for reducing uncorrelated noise, their performance in the presence of correlated noise is insufficient. To properly process both correlated and uncorrelated noise, spectral subtraction on the cross spectrum (SSCS) has been proposed [12]. That method estimates the noise cross spectrum in the noise period and subtracts the spectrum from that for the current input frame, assuming that the correlated noise is stationary. Its effectiveness in the presence of diffuse automobile noise was demonstrated in [12]. Utilization of *a priori* information of the spatial correlation function for compensating the cross spectrum has also been proposed for cases of diffuse noise fields [10]. In addition, a post-filter to deal with nonstationary noise [11] has been proposed. This method uses noise spectrum estimation with recursive averaging from multi-input signals. However, this averaging can make it difficult to deal with impulsive transient noises for the same reason as 2chSS. Performance in such a noise field where we intend to deal with is unknown.

This paper introduces a new method of two-channel speech enhancement that is based on auto gain control (AGC). An estimate of the Wiener gain calculated from the weighted cross spectrum is used for our method. Mere gain control of the waveform amplitude is considered to be less effective in eliminating noise which is superimposed on the speech segments of the desired signal. Nevertheless, gain control does not alter the shape of the spectrum in each frame; it hardly generates musical noise. In addition, because of the averaging along the frequency axis involved in the gain estimation, which is not useful in filter estimation, gain estimation can be improved even in the presence of transient noises. Better performance is possible if major noise components in the noise period, e.g., impulsive noises, are eliminated using the proper gain because noise is more perceptible in a noise-dominated period. To estimate that proper gain, we propose pre-processing based on 2chSS, followed by gain calculation and self-offset of the noise components by introducing a weighting function for whitening the noise spectrum to deal with transient noises.

The remainder of the paper is organized as follows. Section II describes a signal model of the two-channel system and a brief summary of the conventional two-channel spectral subtraction.

Manuscript received November 5, 2003; revised October 21, 2004. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Futoshi Asano.

The authors are with the Department of Computer and Information Science, Iwate University, Morioka, Japan 020-8551 (e-mail: nagata@cis.iwate-u.ac.jp; toy@cis.iwate-u.ac.jp).

Digital Object Identifier 10.1109/TSA.2005.854112

Section III introduces the principle of speech enhancement based on AGC along with the two weighting functions for estimating proper gain. Section IV describes implementation of the proposed speech enhancement method and experimental condition for evaluation. Section V describes experimental evaluation of the proposed method and subjective tests. Finally, Section VI summarizes the conclusions.

II. Two-Channel Speech Enhancement

A. Signal and Noise Models

We assume that two directional microphones are placed in a noisy environment to receive the identical desired signals, as shown in Fig. 1. Let the discrete time samples of the signals received at the microphones be

$$\begin{aligned} x(i) &= s(i) + n_x(i) \\ y(i) &= s(i) + n_y(i) \end{aligned} \tag{1}$$

where $x(i)$ and $y(i)$ denote the L- and R-channel microphone signals, $s(i)$ the desired signal, and $n_x(i)$ and $n_y(i)$ the noises received at respective microphones.

Subjecting the above samples to short-time discrete Fourier transform (DFT), we obtain

$$\begin{aligned} X_{n,k} &= S_{n,k} + N_{x,n,k} \\ Y_{n,k} &= S_{n,k} + N_{y,n,k} \end{aligned} \tag{2}$$

where $X_{n,k}$ and $Y_{n,k}$ denote the DFT of the $x(i)$ and $y(i)$ for the frame n and the k-th frequency bin; $S_{n,k}$ denotes that of $s(i)$, $N_{x,n,k}$ and $N_{y,n,k}$ denote those of $n_x(i)$ and $n_y(i)$, respectively.

We further assume that the received noise signals contain uncorrelated background noise and one broad-band interference arriving from angle θ. The interference can include continuous and transient noises, such that the ordinary adaptive beamformer is ineffective. Taking into account that the received interference signals differ in amplitude and phase in the above microphone arrangement, (2) becomes

$$\begin{aligned} X_{n,k} &= S_{n,k} + V_{n,k} + B_{x,n,k} \\ Y_{n,k} &= S_{n,k} + \alpha_{\theta,k} V_{n,k} e^{-j2\pi\tau k/K} + B_{y,n,k} \end{aligned} \tag{3}$$

where $B_{x,n,k}$ and $B_{y,n,k}$ are the DFTs of uncorrelated background noise, $V_{n,k}$ is the DFT of the interference, $\alpha_{\theta,k}$ is the relative amplitude of the interference normalized by that contained in the L-channel signal, τ is the time delay of the interference between channels, and K represents points of the DFT.

B. Two-Channel Spectral Subtraction

In [2], $X_{n,k}$ is used as a primary signal. Subtraction of the noise spectrum $N_{x,n,k} \equiv V_{x,n,k} + B_{x,n,k}$ contained in $X_{n,k}$ is performed with recursive estimation of the noise spectrum as

$$\nu_{n,k} = \frac{1}{M} \sum_{m=1}^{M} \frac{|X_{n-m,k} - Y_{n-m,k}|^2}{|\hat{N}_{x,n-m,k}|^2} \tag{4}$$

$$|\hat{N}_{x,n,k}|^2 = \frac{|X_{n,k} - Y_{n,k}|^2}{\nu_{n,k}} \tag{5}$$

$$|\hat{S}_{n,k}|^2 = |X_{n,k}|^2 - |\hat{N}_{x,n,k}|^2 \tag{6}$$

$$\arg(\hat{S}_{n,k}) = \arg(X_{n,k}) \tag{7}$$

where $\hat{S}_{n,k}$ denotes the estimate of the desired signal, $\hat{N}_{x,n,k}$ is the estimate of the noise spectrum $N_{x,n,k}$, M represents the number of frames for time averaging, and $\nu_{n,k}$ is the compensation coefficient.

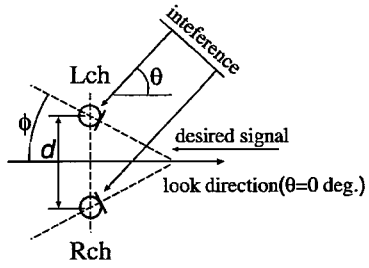


Fig. 1. Arrangement of directional microphones.

This method uses the differenced spectrum $X_{n,k} - Y_{n,k}$ to estimate the noise spectrum $N_{x,n,k}$. Then, the components of the interference signal contained in the differenced spectrum are expressed as follows:

$$V'_{n,k} = V_{n,k}(1 - \alpha_{\theta,k} e^{-j2\pi\tau k/K}). \tag{8}$$

If $\alpha_{\theta,k} = 1$, this operation can produce zero values on the interference spectrum. Because the distortion of the noise spectrum is critical to the performance of 2chSS, this algorithm requires compensation using the coefficient $\nu_{n,k}$.

III. Auto Gain Control for Speech Enhancement

A. Weighted Wiener Gain

Consider the case in which the average of the received signals $Z_{n,k} = (X_{n,k} + Y_{n,k})/2$ is multiplied by a scalar gain ρ_n for approximating the desired signal contained in $Z_{n,k}$ as follows:

$$\hat{S}_{n,k} = Z_{n,k}\, \rho_n. \tag{9}$$

Gain ρ_n can be obtained as a weighted least square solution to minimize the following cost function assuming that gain ρ_n and weighting function $\Psi_{n,k}$ are constant within the period of time averaging, as

$$J(\rho_n) = \sum_k \overline{|Z_{n,k}\, \rho_n - S_{n,k}|^2 \Psi_{n,k}} \tag{10}$$

$$= \frac{1}{2L+1} \sum_k \sum_{j=n-L}^{j<=n+L} |Z_{j,k}\, \rho_n - S_{j,k}|^2 \Psi_{n,k} \tag{11}$$

where $\overline{(\ )}$ denotes time averaging and $2L+1$ denotes the number of frames for time averaging. Therefore, the weighted version of the Wiener gain is obtained as

$$\rho_n = \frac{\sum_k G_{ss,n,k} \Psi_{n,k}}{\sum_k G_{zz,n,k} \Psi_{n,k}} \tag{12}$$

where

$$G_{ss,n,k} = \overline{|S_{n,k}|^2} = \frac{1}{2L+1} \sum_{j=n-L}^{j<=n+L} |S_{j,k}|^2 \tag{13}$$

and

$$G_{zz,n,k} = \overline{|Z_{n,k}|^2} = \frac{1}{2L+1} \sum_{j=n-L}^{j<=n+L} |Z_{j,k}|^2 \tag{14}$$

denote the power spectra of the desired signal and the primary signal, respectively.

If interference $V_{n,k}$ is not present, then $G_{ss,n,k}$ in (12) can be replaced by the cross spectrum, as

$$G_{xy,n,k} = \overline{X^*_{n,k} Y_{n,k}} = \frac{1}{2L+1} \sum_{j=n-L}^{j<=n+L} X^*_{n,k} Y_{n,k}. \quad (15)$$

Thereby, (12) becomes

$$\rho_n = \frac{\sum_k \mathrm{Re}(G_{xy,n,k}) \Psi_{n,k}}{\sum_k G_{zz,n,k} \Psi_{n,k}} \quad (16)$$

where Re() denotes the operation to take a real part of the complex number and $\Psi_{n,k}$ is assumed to have a real value. The imaginary part can be ignored because the desired signals are assumed to be identical among channels.

In the case where interference is present, the numerator of (16) is expressed as the following by substituting (3) into (16)

$$\sum_k \mathrm{Re}(G_{xy,n,k}) \Psi_{n,k} = \sum_k \mathrm{Rc}(\overline{|S_{n,k}|^2} + \alpha_{\theta,k} \overline{|V_{n,k}|^2} e^{-j2\pi\tau k/K}) \Psi_{n,k}. \quad (17)$$

If the weighting function $\Psi_{n,k}$ works to whiten the second term in (17) and the phase $2\pi\tau k/K$ is distributed uniformly within the range of $-\pi$ to π, the summation along frequency bin k reduces the summed power of the interference lower than that of the desired signal. We can closely simulate this condition in most cases by taking a sufficient inter-microphone distance. Because the spectrum of the interference can not be estimated directly from the observations, we choose the inverse of the power spectrum of the differenced signal

$$\Psi_{n,k} = \frac{1}{\overline{|X_{n,k} - Y_{n,k}|^2}} \quad (18)$$

as the weighting function for whitening. The desired signal is reduced by the differencing operation in (18) so as to be used as an approximation of the noise signal. Because the relative amplitude $\alpha_{\theta,k}$ depends on the microphone directivity, we can infer that $\alpha_{\theta,k}$ does not differ significantly depending on the frequency within the speech frequency band. Consequently, we ignore its effect.

The noise spectrum distortion that is attributable to the differencing operation is critical for 2chSS. It should be compensated in the process of 2chSS. On the other hand, the spectral zeros that are attributable to the differencing can be avoided using the microphone arrangement, as shown in Fig. 1, because $\alpha_{\theta,k} \neq 1$. In addition, the distortion does not directly affect the gain estimation because the averaging over speech frequency band can moderate that effect. Moreover, estimation of the compensation coefficient requires a sufficient observation time of noise so that the compensation can lead to a reduced effect on the transient noises. For that reason, we use $\Psi_{n,k}$ without compensation as the whitening function.

B. Weighting Function Based on Spectral Subtraction

Because reduction of the noise components contained in the cross spectrum can improve the accuracy of the gain, as seen in (17), we intend to reduce them before calculating the gain. For that purpose, we introduce the 2chSS based weighting function because 2chSS can deal with nonstationary noise conditions in cases with multiple noise sources.

To make it correspond to our setup, we replace the primary signal $X_{n,k}$ in the 2chSS with the averaged signal $Z_{n,k}$. Then we rewrite the process of estimating the desired speech amplitude in 2chSS (6), using the real function $\Phi^{(\mathrm{org})}_{n,k}$

$$|\hat{S}_{n,k}|^2 = |Z_{n,k}|^2 \Phi^{(\mathrm{org})}_{n,k} \quad (19)$$

$$\Phi^{(\mathrm{org})}_{n,k} = \frac{|Z_{n,k}|^2 - \frac{|X_{n,k} - Y_{n,k}|^2}{\nu_{n,k}}}{|Z_{n,k}|^2}. \quad (20)$$

Replacement of the primary signal has already been mentioned in [2].

In our case, taking into account that primary signal containing noise to be subtracted is the cross spectrum $G_{xy,n,k}$, we further replace the primary signal power $|Z_{n,k}|^2$ in (19) and (20) with $|G_{xy,n,k}|$ as

$$\tilde{G}_{ss,n,k} = |G_{xy,n,k}| - \frac{\gamma \overline{|X_{n,k} - Y_{n,k}|^2}}{\nu_{n,k}} \quad (21)$$

$$= |G_{xy,n,k}| \Phi_{n,k,\gamma} \quad (22)$$

$$\Phi_{n,k,\gamma} = \frac{|G_{xy,n,k}| - \frac{\gamma \overline{|X_{n,k} - Y_{n,k}|^2}}{\nu_{n,k}}}{|G_{xy,n,k}|} \quad (23)$$

where γ is a positive constant to control the strength of the subtraction, and $\Phi_{n,k,\gamma}$ is the weighting function to perform spectral subtraction. The short term power spectrum of the differenced signal in (20) is also replaced by the averaged power spectrum in the above expression.

Moreover, the estimation of the compensation coefficients $\nu_{n,k}$ is modified as follows:

$$\nu_{n,k} = \frac{D_{n,k}}{|Q_{xy,n,k}|} \quad (24)$$

$$D_{n,k} = \begin{cases} |X_{n,k} - Y_{n,k}|^2 \lambda + D_{n-1,k}(1-\lambda) & \text{(noise period)} \\ D_{n-1,k} & \text{(speech period)} \end{cases} \quad (25)$$

$$Q_{xy,n,k} = \begin{cases} X^*_{n,k} Y_{n,k} \lambda + Q_{xy,n-1,k}(1-\lambda) & \text{(noise period)} \\ Q_{xy,n-1,k} & \text{(speech period)} \end{cases} \quad (26)$$

where $D_{n,k}$ is the averaged differenced spectrum in the noise period, $Q_{xy,n,k}$ is the cross spectrum in the noise period, and λ represents the learning factor. We do not use recursive estimation as expressed by (4) and (5) because of possible leakage of the desired signal into the differenced signal. Detection of the noise period is described in Section IV-B

C. Combined Weighting Function

We propose calculation of the total gain by combining (16) with (23) as follows:

$$\rho'_n(\beta, \gamma) = \frac{\sum_k \mathrm{Re}(G_{xy,n,k}) \Psi^{\beta}_{n,k} \Phi_{n,k,\gamma}}{\sum_k G_{zz,n,k} \Psi^{\beta}_{n,k}} \quad (27)$$

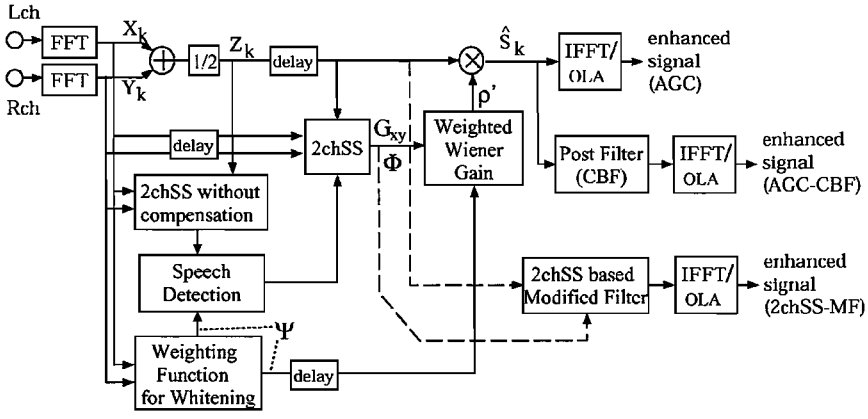


Fig. 2. Block diagram of the proposed system.

where β is a positive constant introduced to control strength of the whitening. If 2chSS works well, it seems that whitening by $\Psi_{n,k}$ is not necessary. Nevertheless, it is considered to be difficult to estimate the accurate compensation coefficients for $\Phi_{n,k,\gamma}$ in the case where impulsive disturbances arrive because the noise periods are very short. Whitening combined with noise reduction is considered to be effective to deal with such a case. Parameters β and γ are determined empirically because this is an ad hoc combination.

IV. Implementation and Experimental Setup

A. Processing System for Experiment

A block diagram of the proposed speech enhancement system with auto gain control is depicted in Fig. 2. This figure also contains a block, connected by dotted lines, for preliminary evaluation of 2chSS. First, the DFTs of the received signals are obtained via the fast Fourier transform (FFT). The DFT spectra are averaged and multiplied by the gain that was estimated according to (27). An enhanced spectrum of the desired speech based on auto gain control is obtained as

$$\hat{S}_{n,k} = \frac{X_{n,k} + Y_{n,k}}{2} \rho'_n(\beta, \gamma). \tag{28}$$

The waveform of the enhanced speech can be obtained via the inverse fast Fourier transform (IFFT) and standard overlap-add processing. In Fig. 2, this enhanced signal is denoted as "enhanced signal (AGC)".

In addition, post-filtering is mentioned to suppress uncorrelated noise components. An optimal post-filter derived from Wiener theory [15]–[17] is generally used for this purpose, but we chose the coherence based filter (CBF) here in anticipation of moderate performance. Thereby, we may prevent over-suppression caused by duplicated noise reduction of AGC and the post-filter as described below. Because the coherence estimate is always larger than or equal to the Wiener gain, CBF seems to provide moderate noise reduction performance compared to the optimum Wiener filter. The CBF is obtained as

$$F_{n,k} = \frac{|G_{xy,n,k}|}{\sqrt{G_{xx,n,k} G_{yy,n,k}}} \tag{29}$$

where $G_{xy,n,k}$ is the cross spectrum in (15), and $G_{xx,n,k}$ and $G_{yy,n,k}$ are the power spectra of input signals X and Y, respectively. We obtained $G_{xx,n,k}$ and $G_{yy,n,k}$ in the following manner:

$$G_{xx,n,k} = \overline{|X_{n,k}|^2} = \frac{1}{2L+1} \sum_{j=n-L}^{j<=n+L} |X_{j,k}|^2 \tag{30}$$

$$G_{yy,n,k} = \overline{|Y_{n,k}|^2} = \frac{1}{2L+1} \sum_{j=n-L}^{j<=n+L} |Y_{j,k}|^2. \tag{31}$$

Therefore, the enhanced spectrum with the post-filter (AGC-CBF) is obtained as the following:

$$\hat{S}_{n,k} = \frac{X_{n,k} + Y_{n,k}}{2} \rho'_n(\beta, \gamma) F_{n,k}. \tag{32}$$

The primary signal can be over-suppressed as stated above because the contribution of uncorrelated noise is decreased both in AGC and in CBF. Regarding the Wiener filter type post-filter, it has been reported that this type of post-filter can suffer from musical noise, particularly in cases where a directional noise source brings about zeros on the transfer function of the preceding beamformer [9] because such a situation forces the denominator of the Wiener filter estimate to be zero. In contrast, CBF inherently obviates such a problem. Moreover, the beamformer mentioned in this work is a 2ch half-sum type, which has no zero in the beamformer transfer function. Indeed, the effect of the duplication is shown to be negligible in the later section of evaluation. In the evaluation, to ensure the performance apart from AGC in our experimental setup, we calculate an enhanced signal using CBF alone. The CBF enhanced signal is given as

$$\hat{S}^{(\mathrm{CBF})}_{n,k} = \frac{X_{n,k} + Y_{n,k}}{2} F_{n,k}. \tag{33}$$

For implementing a 2chSS based weighting function, we have the option to choose not only $\Phi_{n,k,\gamma}$ (23), but $\Phi^{(\mathrm{org})}_{n,k}$ (20) to apply to AGC. We can decide which function is better for our method by evaluating the signal enhanced by each function as a post-filter that will be described in a later section. Because $\Phi_{n,k,\gamma}$ is not the estimate of a Wiener filter, we calculate enhanced signal with this filter as

$$\hat{S}^{(\mathrm{2chSS\text{-}MF})}_{n,k} = \frac{X_{n,k} + Y_{n,k}}{2} \frac{|G_{xy,n,k}|}{G_{zz,k}} \Phi_{n,k,\gamma} \tag{34}$$

TABLE I
SPEECH DETECTION ALGORITHM

Initialize $\eta_b = 0.03, \eta_v = 0.005$,

$\xi = 2.0,\ \kappa = 0.1$

b_0, v_0 : averaged values of the first 30 frames of the input.

$n = 0$

For each frame n do

if $|C_n - b_n| < \kappa$

$b_n = (1 - \eta_b) b_{n-1} + \eta_b C_n$

$v_n = (1 - \eta_v) v_{n-1} + \eta_v |C_n - b_n|^2$

else

$b_n = b_{n-1}$

$v_n = v_{n-1}$

endif

$h_n = b_n + \sqrt{v_n} \cdot \xi$

if $(C_n > h_n)$

speech is present

else

speech is not present

$n++$

end

whereas the original 2chSS is performed as

$$\hat{S}_{n,k}^{(\text{2chSS-ORG})} = \frac{X_{n,k} + Y_{n,k}}{2} \Phi_{n,k}^{(\text{org})}. \tag{35}$$

The above modified version of 2chSS is denoted as "2chSS-Modified filter (2chSS-MF)" in Fig. 2 and the original 2chSS is called "2chSS-ORG" hereafter.

B. Speech Detection

As described in Section III-B, detection of the noise period is required to estimate compensation coefficients for obtaining the 2chSS based weighting function. To determine the noise period for updating $D_{n,k}$ and $Q_{xy,n,k}$, we used the criterion

$$C_n = \rho_n'(\beta, \gamma)\big|_{\nu_{n,k}=1.0} \tag{36}$$

because C_n is a good approximation of the signal-to-signal + noise ratio. Moreover, it is obtainable without speech detection for calculating the compensation coefficients $\nu_{n,k}$. We used values of the constant parameters $(\beta, \gamma) = (2.0, 0.5)$ which are to be determined in the experiments described in Section V-C.

The detection algorithm is listed in Table I. In Table I: n is the frame number, b_n is the estimate of the bias of C_n (36) in the noise period, v_n is the estimate of the variance of C_n in the noise period, h_n is the threshold for detection, ξ is the constant parameter for setting h_n, η_b and η_v are the learning factors for estimating b_n and v_n respectively, and κ is a threshold for rough detection. This procedure updates the bias b_n and the variance v_n in the noise period determined by rough detection. Values of the above parameters were determined as shown in Table I to attain good performance through the three noise conditions that are described in Section IV-E.

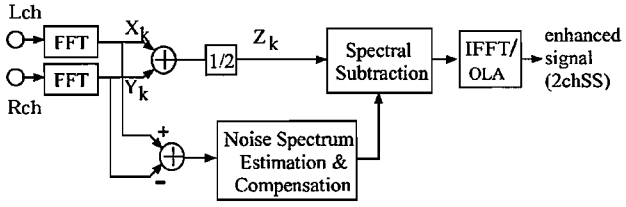


Fig. 3. Block diagrams of the two-channel spectral subtraction (2chSS-ORG).

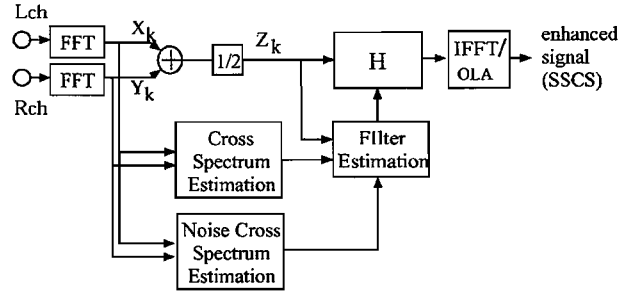


Fig. 4. Block diagrams of Wiener filtering with cross spectral subtraction.

C. Methods for Comparison

For comparison to conventional methods, we also present results that were obtained by 2chSS-ORG, SSCS and half-sums of the input signals. We regard SSCS as an improved version of a post-filter that can deal with correlated and uncorrelated noise environments.

1) 2chSS-ORG: A block diagram of the original 2chSS is depicted in Fig. 3. This diagram corresponds to the algorithm (4)–(6) in which $X_{n,k}$ replaces the averaged signal $Z_{n,k}$ to create correspondence to the proposed system for a comparison. The number of frames M for time averaging (4) is set to 5, as was done in [2].

2) SSCS: A block diagram of SSCS is shown in Fig. 4. As described in [12], we estimate the power spectrum of the desired signal $\hat{G}_{ss,n,k}$ and obtain a Wiener filter $H_{n,k}$ as

$$\hat{G}_{ss,n,k} = |G_{xy,n,k}| - |Q_{xy,n,k}| \tag{37}$$

$$H_{n,k} = \frac{\hat{G}_{ss,n,k}}{G_{zz,n,k}} \tag{38}$$

where $Q_{xy,n,k}$ is the noise cross spectrum updated by (26). Whereas the performance of SSCS was evaluated with a manually determined noise period in [12], we employ automatic detection with the detection criterion (36) and algorithm listed in Table I as used in our proposed system. The block of the noise cross spectrum estimation in Fig. 4 includes the speech detection procedure.

In the two methods above and in the proposed method, we commonly use the 256-point FFT and a Hanning window with

the 128-point frame shift. The number of frames for time averaging is set to 15 (0.18 s, $L = 7$) to perform time averaging for calculating the spectra. The learning factor λ in (25) and (26) is set to 0.1. The frequency range to estimate the gain for AGC and AGC-CBF is set between 260 Hz and 4000 Hz. These values were determined in preliminary experiments to obtain good performance through the three noise conditions that are described in Section IV-E.

D. Objective Measures

Next we examine evaluation of an enhanced signal which contains segments where speech is absent. For this purpose, we use log spectral distortion measure (LSD) [18] and overall SNR. To avoid $\log(0)$ in the calculation of LSD, -30 dB white noise is added both to enhanced speech and clean speech, as was done in [18]. The same series of white noise is used for the two signals. If some segments of the clean signal and the corresponding segments of the enhanced signal are both 0, this distortion measure becomes 0. We infer that distortion 0 is reasonable for such a case. Segmental SNR is not used because the treatment of such a silent period is difficult.

We used the average of the two channel clean signals $c(i) = (c_x(i) + c_y(i))/2$ as the clean signal to compute the above measures, where $c_x(i)$ and $c_y(i)$ are the clean signals of the L- and R-channels, respectively. Spectral analysis for LSD is implemented with Hamming windows of 512-sample length (46 ms) and a 128-sample frame update step.

The AGC and AGC-CBF enhanced signals become smaller than the original ones because the gain for noise reduction is smaller than 1.0. Next, we calculate the compensated SNR for all enhanced signals as the measure of overall SNR using a scaling factor μ, which is obtained by minimizing the cost function

$$I(\mu) = \overline{|\hat{s}(i) - c(i)\mu|^2} \tag{39}$$

where $\hat{s}(i)$ is the enhanced signal. Consequently, μ is easily obtained as

$$\mu = \frac{\overline{\hat{s}(i)c(i)}}{\overline{|c(i)|^2}} \tag{40}$$

and the compensated overall SNR becomes

$$\mathrm{SNR} = \frac{\overline{|c(i)|^2}\mu^2}{\overline{|\hat{s}(i) - c(i)\mu|^2}}. \tag{41}$$

E. Speech and Noise Data

We recorded two-channel clean speech and noises for evaluation. All recordings were done in an automobile, as illustrated in Fig. 5. Microphones were cardioid with an inter-microphone distance of 12 cm. We set the angle ϕ to about 20°. Speech and noises were recorded separately; the noisy speech of the desired SNR were generated in each experiment by adding noise to clean speech.

The desired speech comprised 100 Japanese city names uttered by a female speaker. The average duration of each word was 0.7 s; the average interval between the words was 1.0 s. The total length of the speech was about 170 s. The SNR of the input data was computed during the speech activity and averaged over the two channel signals as

$$\mathrm{SNR} = 10\log\left[\frac{\dfrac{\sum_{m=1}^{2}\sum_{j=1}^{100}\sum_{i=1}^{L_{m,j}} s_{m,j}^2(i)}{\sum_{m=1}^{2}\sum_{j=1}^{100} L_{m,j}}}{\dfrac{\sum_{m=1}^{2}\sum_{i=1}^{N_m} n_m^2(i)}{\sum_{m=1}^{2} N_m}}\right] \tag{42}$$

where m is the channel number, j is the data number of word speech, $L_{m,j}$ is the number of samples of the jth word in mth channel, $s_{m,j}$ is the jth word speech signal in mth channel, N_m is the number of samples of the noise in mth channel, and n_m represents the noise signal in the mth channel.

The recorded noises were attributable to the three conditions that are listed in Table II. Noise (i) was recorded near a construction site; it contained many impulsive noises of hammer impacts. Noise (ii) is a sound in a car moving at 70 km/h; it comprises road, wind, and engine noises. Noise (iii) contains speech radiated from loudspeakers and clicking sounds of car turn signals. The speed of the car was 15 km/h.

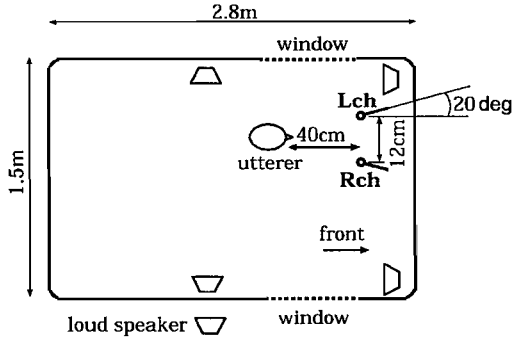


Fig. 5. Recording environment. The automobile was an off-road wagon type.

TABLE II
SPEECH AND NOISE DATA

Speech	100 Japanese city names
Utterer	1 female
Sampling	11 kHz (L/R)
Noise (i)	Noise from a construction site (the car was stopped and the front windows of the car were opened)
Noise (ii)	Noise in a car moving at 70 km/h (windows were opened halfway)
Noise (iii)	Speech from loudspeakers and clicks of blinkers in a car moving at 15 km/h (windows were closed)

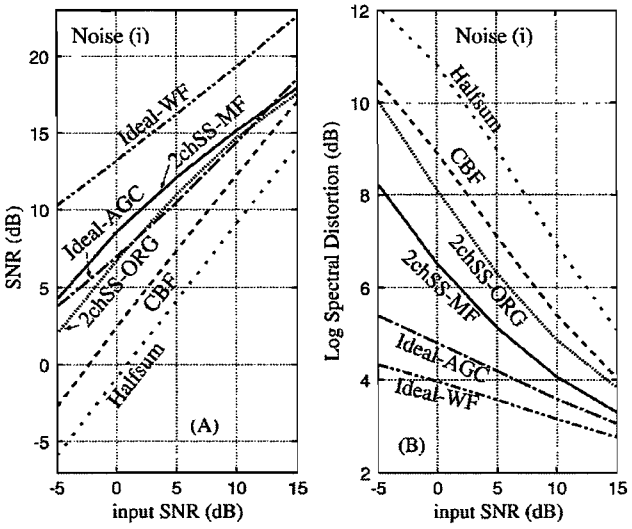


Fig. 6. Performance of CBF and 2chSS based post-filters in terms of (a) overall SNR and (b) LSD in the presence of noise (i).

Because noises (ii) and (iii) contain noises that are attributable to car motion, all data were high pass filtered using an FIR filter with a cutoff frequency of 150 Hz. This allows cancellation of a part of the noise without degrading the speech signal.

V. Evaluation

A. Evaluation of CBF and 2chSS Based Weighting Functions as a Post-Filter

We first evaluate the enhanced signal obtained using CBF (33), 2chSS-MF (34), and 2chSS-ORG (35) to assess each function independently. We calculated the 2chSS-MF enhanced signal with $\gamma = 0.5$. Thereby, the performance on overall SNR became nearly identical to that of 2chSS-ORG. Because we are mainly interested in the performance in situations where transient noises are present, we particularly show results using noise (i). Resultant performances in terms of overall SNR and LSD are shown in Fig. 6(A) and (B), respectively. This figure also contains the result of an ideal Wiener filter denoted as "Ideal-WF" and that of AGC processing with the ideal gain denoted as "Ideal-AGC." These figures show that the CBF performance is lower than both 2chSS based filters. Moreover, 2chSS-MF attains higher performance compared to 2chSS-ORG in term of LSD while these are comparable in terms of overall SNR. Performance degradation compared to the ideal Wiener filter is considered to be attributable to residual impulsive noises. The performance of Ideal-AGC is demonstrably lower than that of Ideal-WF, but it is still higher than that of 2chSS-MF in terms of LSD.

Next, we show spectrograms of the enhanced signals mentioned previously in Fig. 7. This figure also shows spectrograms of clean speech and half-sum signals. These signals were pre-emphasized to allow display of high frequency components. We can observe again that performance of CBF is weak; that of 2chSS-MF is better than 2chSS-ORG. Although

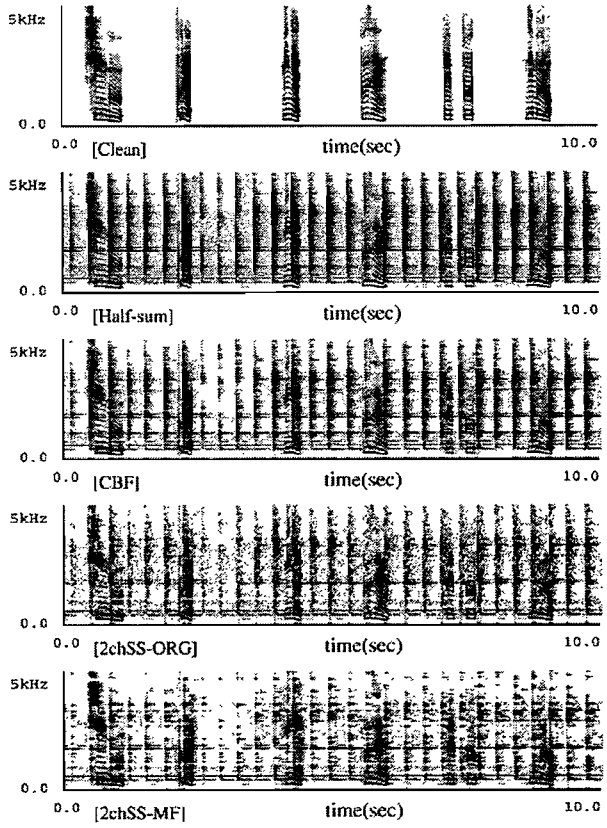


Fig. 7. Spectrograms of a clean signal, a half-sum signal, a CBF-enhanced signal, a conventional 2chSS enhanced signal, and a 2chSS-MF enhanced signal for noise (i). (SNR = 5 dB).

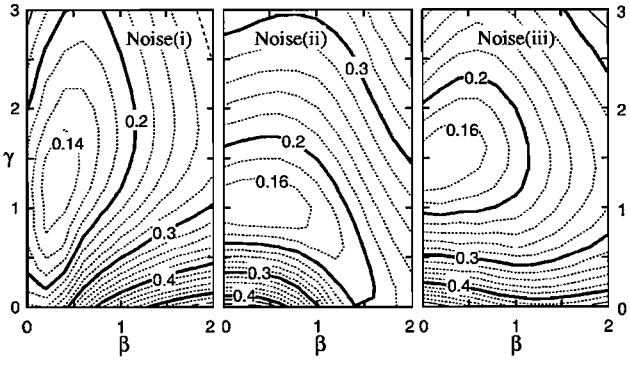


Fig. 8. Averaged gain error $E_g(\beta, \gamma)$. (input SNR = 5 dB).

2chSS-MF appears to provide good performance in terms of LSD, the 2chSS-MF enhanced signal still contains residual impulsive noises that sound like musical noise.

These results confirm that 2chSS-MF is more suitable than 2chSS-ORG for application to the AGC because a smaller residual noise can be expected. In addition, it is suggested that the effect of CBF is sufficiently small to permit duplicated noise reduction by both AGC and a post-filter. Although 2chSS-MF seems to be promising as a post-filter, further investigation about this point is beyond the scope of this study.

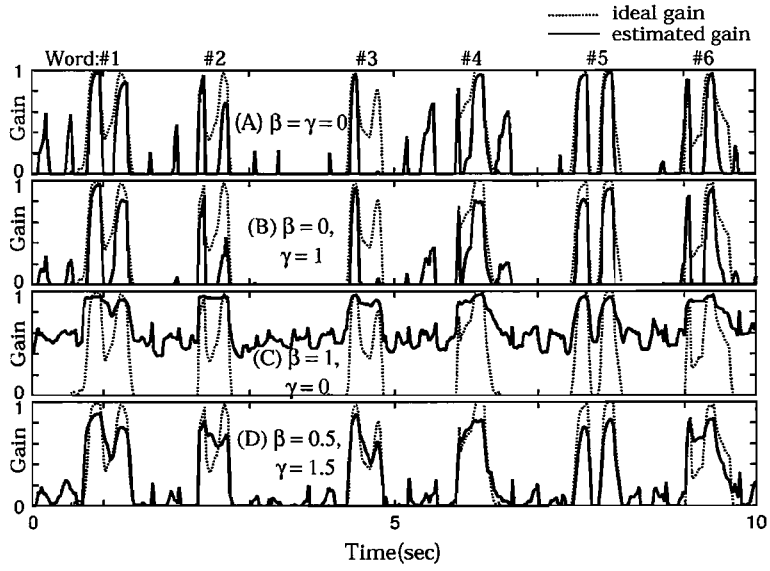


Fig. 9. Estimated gain versus time in the presence of noise (i). (SNR = 5 dB).

B. *Effect of Weighting Functions on the Gain Estimation*

1) Gain Error Dependency on the Weighting Functions: First, to elucidate the relation between parameters of weighting functions and the gain error, we calculated the average difference between the estimated gain and the ideal one by changing the parameters β and γ in (27) as follows:

$$E_g(\beta,\gamma) = \sqrt{|\rho_n'(\beta,\gamma) - \rho_n^{(\mathrm{idcal})}|^2}. \tag{43}$$

The ideal gain was obtained using clean speech; averaging was performed over the entire speech signal. Fig. 8 was obtained by changing β from 0 to 2 and γ from 0 to 3. This figure contains three results, which correspond to noises (i), (ii), and (iii).

These figures show that the minimum gain errors in this parameter range are obtained when $(\beta,\gamma) = (0.35, 1.3)$ for the case of noise (i), (0.5, 1.0) for the case of noise (ii), and (0.35, 1.5) for the case of noise (iii). Whereas an almost minimal gain error can be achieved without the whitening function ($\beta = 0$) in the cases of noise (ii) and noise (iii), both weighting functions are required to attain almost minimal error in the cases of noise (i). This result indicates the necessity of both weighting functions, and that whitening is particularly effective in the presence of impulsive noises.

2) Temporal Change of Gain: Next, we compare the temporal change of gain over frames using different parameter values of (β,γ) to confirm the roles of the two weighting functions. Noise (i) was used for the calculation. The SNR was set to 5 dB. Fig. 9(A) – (D) show the resultant curves of the gain when parameters $(\beta,\gamma) = (0,0),(0,1),(1,0)$, and $(0.5,1.5)$ were used, respectively. Positions of six word speech signals are shown at the top of Fig. 9(A) as $\#1,\#2,\ldots,\#6$. The estimated gains are plotted in solid curves and the ideal gain is plotted in dotted curves.

Fig. 9(B) shows the result when only the 2chSS-based weighting function was enabled $[(\beta,\gamma) = (0,1)]$. This result resembles Fig. 9(A), which was obtained when weighting functions were both disabled ($(\beta,\gamma) = (0,0)$): it is the ordinary estimate of the Wiener gain. These figures show that both curves of the estimated gain differ markedly from the ideal one. Serious eliminations of gain are observed in the periods of

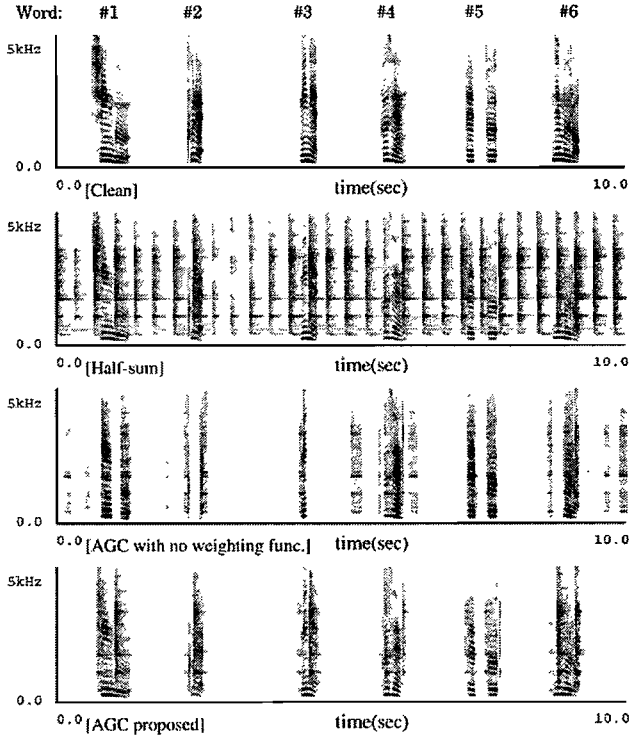


Fig. 10. Spectrograms of the clean signal, half-sum signal, and AGC enhanced signal obtained with $(\beta,\gamma) = (0.0)$, and AGC enhanced signal obtained with $(\beta,\gamma) = (0.5, 1.5)$ in the presence of noise (i) (construction site noise) (Input SNR = 5 dB).

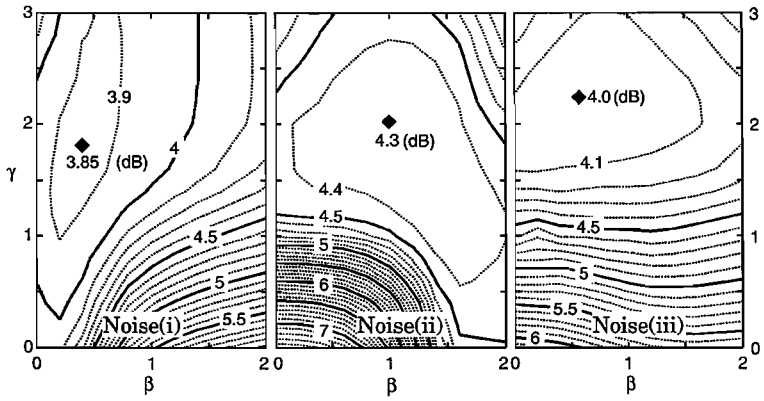


Fig. 11. Log spectral distortion (LSD) obtained from AGC-CBF enhanced signal versus parameters β and γ. ◆: the point at which the minimum LSD is obtained. (input SNR = 5 dB).

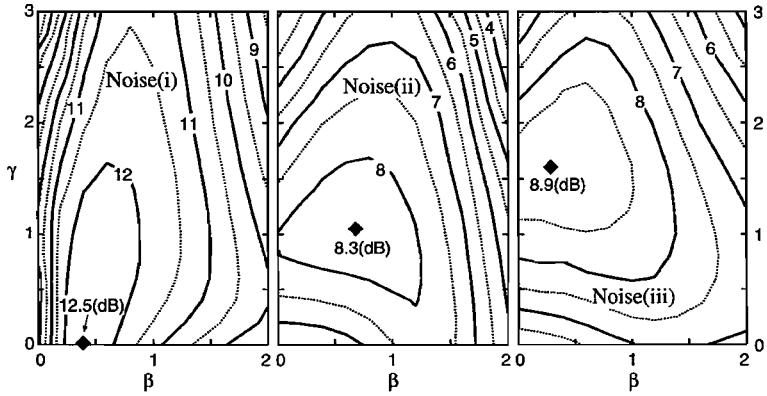


Fig. 12. Overall SNR obtained from AGC-CBF enhanced signal versus parameters β and γ. ◆: the point at which the maximum SNR is obtained. (input SNR = 5 dB).

speech e.g., #2 and #3 that are led by the overlap of impulsive noises and the speech signal. These eliminations tend to grow as γ increases. For that reason, we confirmed that 2ch-SS alone hardly achieves good estimation of the gain in the presence of impulsive noises. This result agrees with the result of the averaged gain error described in the previous section. On the other hand, Fig. 9(C) is obtained when only the whitening function is enabled: $((\beta, \gamma) = (1, 0))$ shows no eliminations in the speech period. Instead, the gain in the background is raised. That increase of background gain is attributable to whitening, which enlarges small noise components while reducing large noise components. Fig. 9(D), which was obtained when both weighting functions were enabled $((\beta, \gamma) = (0.5, 1.5))$, shows reduction of both the increase of the background gain and eliminations of gain in the speech periods.

Moreover, to ensure the relation between the above results of gain estimation and the enhanced signals, we show spectrograms of the signals enhanced by AGC with no weighting function $(\beta, \gamma) = (0, 0)$ and AGC with both weighting functions $(\beta, \gamma) = (0.5, 1.5)$. These spectrograms are shown in Fig. 10 together with spectrograms of clean speech and the half-sum signal. Almost identical portions of the signals were used in the gain estimation above.

We again observe that AGC with no weighting function provides not only some eliminations of speech segment, but also residual noise during the noise period. In contrast, AGC with both weighting functions greatly reduces such eliminations and residuals. These results confirm the effectiveness on the enhanced signal of the weighting functions and the necessity of both functions.

C. Parameters of the Weighting Functions

The proposed gain for speech enhancement (27) has fixed parameters β and γ. To determine the values that are suitable for the three noise conditions, we calculated LSD of the enhanced speech changing β from 0 to 2 and γ from 0 to 3. The input SNR was set to 5 dB. Results obtained from AGC-CBF enhanced signals are shown in Fig. 11.

The values of the parameters at which the minimum LSDs are obtained are $(\beta, \gamma) = (0.4, 1.8)$ in noise condition (i), (1.0, 2.0) in noise condition (ii), and (0.6, 2.2) in noise condition (iii). We can observe that the difference of the LSD is relatively small around the above point in each condition. Consequently, we choose a mean value of the above values to set $(\beta, \gamma) = (0.5, 2.0)$ for AGC and AGC-CBF through all the noise conditions.

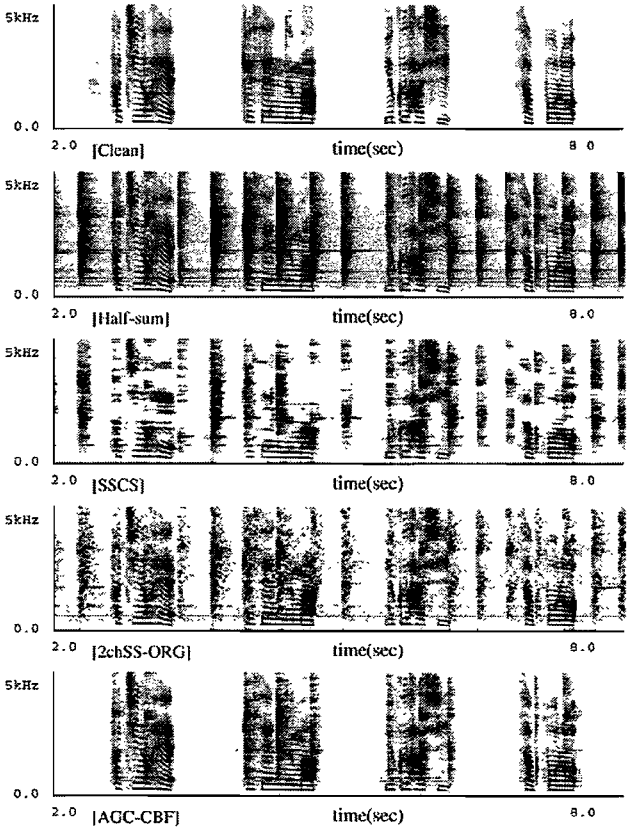


Fig. 13. Spectrograms of the clean signal, half-sum signal, SSCS enhanced signal, 2chSS-ORG enhanced signal, and AGC-CBF enhanced signal obtained in noise condition (i) (construction site noise) (Input SNR = 3 dB).

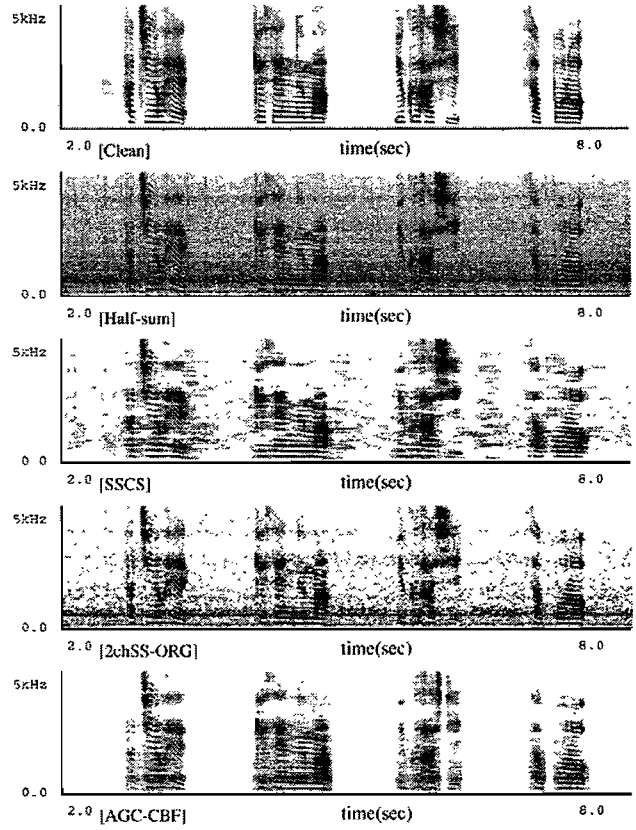


Fig. 14. Spectrograms of the clean signal, half-sum signal, SSCS enhanced signal, 2chSS-ORG enhanced signal, and AGC-CBF enhanced signal obtained in noise condition (ii) (car noise) (Input SNR = 3 dB).

Furthermore, we show results of the overall SNR in Fig. 12. In Fig. 12(i), we observe that the maximum SNR is obtained at $(\beta, \gamma) = (0.4, 0.0)$. The SNR decreases rapidly as β varies from 0.4 to 0.0. This result indicates that the whitening can improve SNR in the presence of the impulsive noises because $\beta = 0.0$ indicates that no whitening is performed.

D. Comparison With Conventional Methods in Spectrograms

Figs. 13–15 show spectrograms of the clean signal, the half-sum signal, the SSCS enhanced signal, the 2chSS-ORG enhanced signal, and the AGC-CBF enhanced signal. Results of the AGC enhanced signals are not presented because they differ only slightly from those of AGC-CBF. Results of the half-sum signals are shown instead of those of the noisy signals. The spoken words for the displayed spectrograms were"Hachinohe", "Kesennuma", "Yukuhashi", and "Sapporo"; the input SNR was set to 3 dB. All signals were pre-emphasized to enable display of high frequency components.

In the case of noise (i) (Fig. 13), residuals of the impulsive noises are observed in the spectrograms of 2chSS-ORG and SSCS. The elimination of the frequency components of speech are observed in the SSCS results. On the other hand, results for the AGC-CBF show that the impulsive noises between the desired speech words become almost indistinguishable even though the superimposed impulsive noises on the speech segments remain, as seen near 4.5 s.

In the case of noise (ii) (Fig. 14), we can observe that moderate noise reduction is achieved by 2chSS-ORG and SSCS. However, 2chSS-ORG has large residual noise components of low frequency. On the contrary, results of the AGC-CBF show that noise between the speech words is reduced, whereas the speech frequency components are almost entirely preserved.

In the case of noise (iii) (Fig. 15), we can observe that the results of 2chSS-ORG and SSCS have large residual noises. The SSCS result shows the elimination of the frequency components of the desired speech. Unlike these two methods, the AGC-CBF result shows that the noises are adequately reduced in the intervals of the desired speech words while the spectrum of the desired speech remains almost unchanged.

Ultimately, these spectrograms demonstrate that, whereas 2chSS-ORG and SSCS attain insufficient noise reduction for nonstationary noises, the proposed enhancer AGC-CBF reduces noise in the noise period markedly. It also preserve the desired speech components.

Informal listening tests showed that noises were reduced moderately by 2chSS-ORG and SSCS both in the period of speech and noise in the conditions (ii) and (iii), whereas the impulsive noises almost all remained in condition (i). The timbre

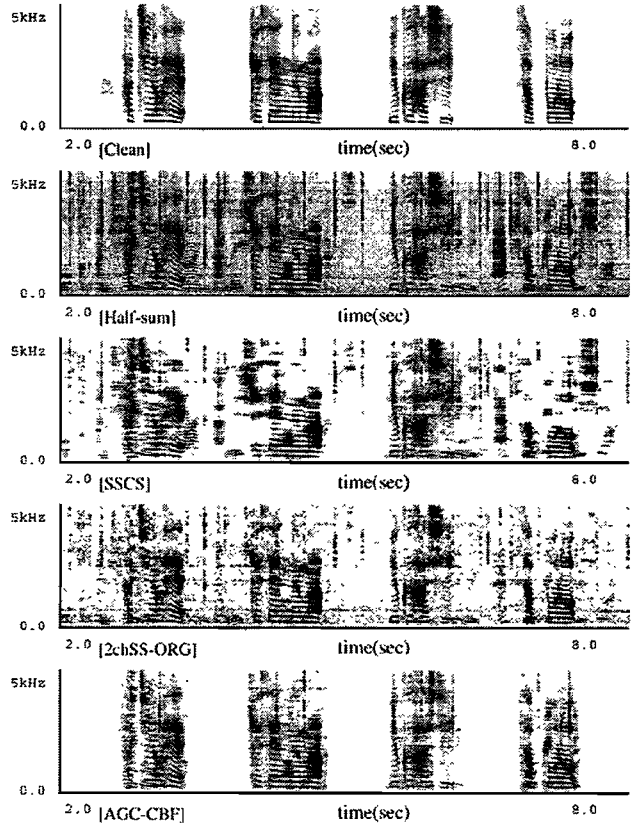


Fig. 15. Spectrograms of the clean signal, half-sum signal, SSCS enhanced signal, 2chSS-ORG enhanced signal, and AGC-CBF enhanced signal obtained in noise condition (iii) (radio and car noise) (Input SNR = 3 dB).

of the residual noises of 2chSS-ORG and SSCS enhanced signals were altered from the original ones. Noisy musical noises were heard in the presence of noise (ii) from the 2chSS-ORG enhanced signals. In contrast, the noise-only periods were almost silent both in cases of AGC and AGC-CBF; the timbre of the desired speech remained intact in these methods. However, the impulsive noises superimposed on the speech segments remained as noisy in cases AGC and AGC-CBF even though these methods decreased these impulses' amplitudes. This suggests that additional filters to remove such noise are needed in periods where both impulsive noise and speech arrive simultaneously.

E. Comparison in Objective Measures

First, we show the results evaluated from the input signals from which intervals of the speech words are excluded. Figs. 16 and 17 show the results in terms of LSD and SNR, respectively. The input SNR was varied from −5 dB to 15 dB in 5 dB steps. Those results are given separately for each noise condition.

As seen from the figures of noise (i) and noise (ii) in Fig. 16, 2chSS-ORG attained high SNR compared to other methods in the noise condition (i); SSCS attained high SNR compared to other methods in the noise condition (ii). However, the difference on LSD measure is not pronounced, as shown in Fig. 17; the LSD of AGC and AGC-CBF are comparable to other methods. AGC-CBF is slightly better than AGC in all the cases. These are the anticipated results because AGC does not promise to eliminate noise in the speech period.

Next, we show results obtained from whole signals. Figs. 18 and 19 show the results in terms of overall SNR and LSD, respectively. The input SNR was varied from −5 dB to 15 dB in 5 dB steps. We can observe that the AGC and AGC-CBF outperform the other methods in all the noise conditions in terms of LSD, whereas overall SNR of SSCS in the case of noise (ii) is the highest, as seen in Fig. 18. The AGC-CBF improved LSD compared to AGC by about 0.3 dB at the input SNR = 0 dB when evaluated using whole signals. These results suggest that the spectral distortion using CBF can be ignored. It rather lowers the distortion.

TABLE III
RESULT OF PREFERENCE TEST

Noise	SNR	Compared with half-sum	Compared with 2chSS-ORG	Compared with SSCS
Noise(i)	5 dB	72%	68%	82%
	-5 dB	75%	73%	95%
Noise(ii)	5 dB	50%	67%	68%
	-5 dB	32%	55%	65%
Noise(iii)	5 dB	73%	51%	67%
	-5 dB	65%	48%	77%

F. Subjective Listening Test

Finally, we show results of a subjective evaluation test. We used a preference test algorithm similar to those used in [19], [20]. The number of subjects were 20. All of them were engineering students whose ages ranged from 18 to 22. None of them had speech processing area problems. The speech material consisted of three consecutive words among the 100 Japanese city names, which were spoken by two female and two male speakers. Two series of the three words "Hachinohe, Kesennuma, Yukuhashi" and "Yokote, Toride, Warabi" were used. Consequently, at each stage of the test, each subject was presented eight pairs of signals. The above speech signals and noises were added with SNRs of −5 dB and 5 dB. Subsequently, they were processed by the three methods of AGC-CBF, 2chSS-ORG, SSCS, and half-sum. In this test, the half-sum signal was regarded to be almost equivalent to the nonprocessed signal. Subjects were presented pairs of signals which each comprised one AGC-CBF enhanced signal and one signal that was enhanced by one of the other three methods. Subjects were asked to choose one of the two signals. Table III summarizes the results.

As shown in this table, in the case of noise (ii), the proposed method, AGC-CBF, is less preferred to the half-sum method.

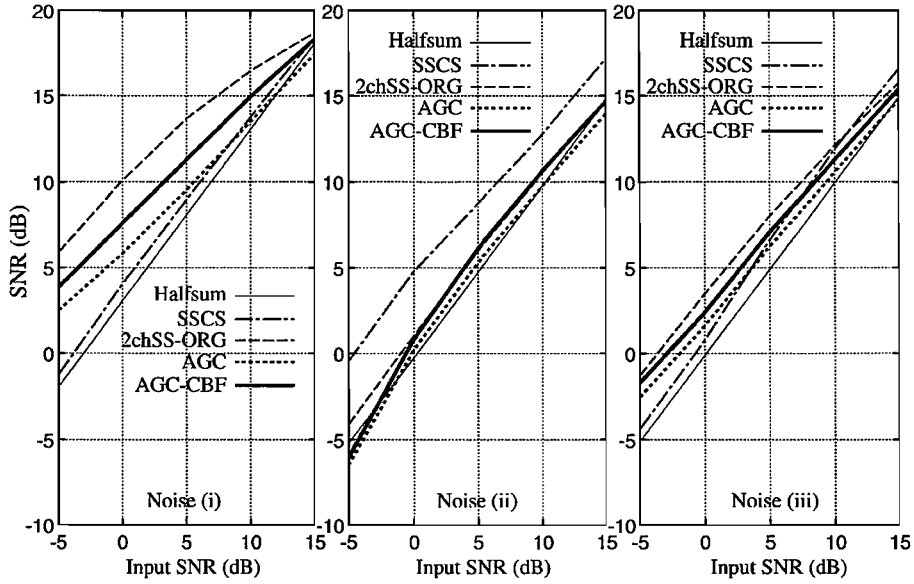


Fig. 16. Comparative performance for the speech period in terms of overall SNR.

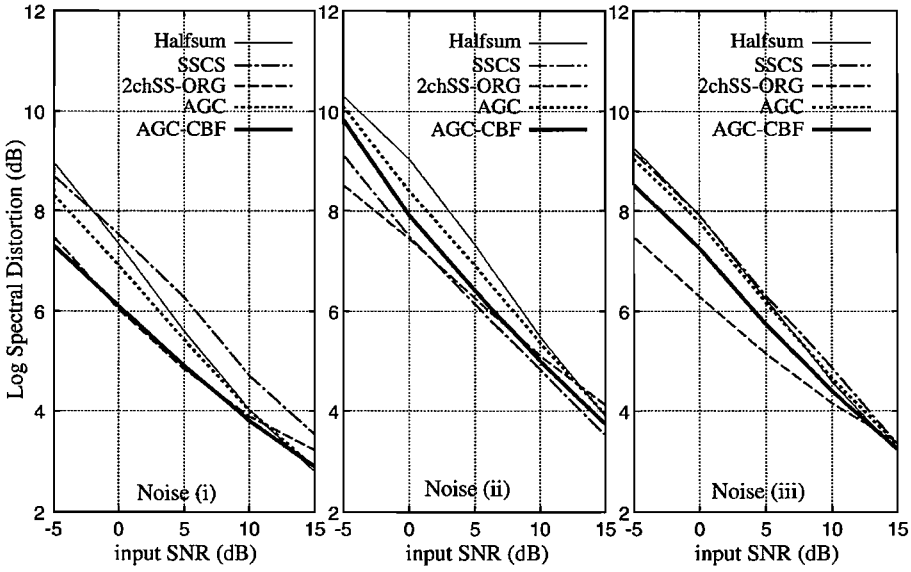


Fig. 17. Comparative performance for the speech period in terms of log spectral distortion (LSD).

Superiority in comparison with 2chSS-ORG is not obvious when SNR $= -5$ dB. This result indicates that because noise (ii) is stationary car noise, the noise consistency is markedly degraded by AGC processing, particularly at low SNR. Performance of 2chSS-ORG was also degraded as a result of musical noise in this noise case. Therefore, AGC-CBF is slightly preferred to 2chSS-ORG when compared with 2chSS-ORG. In the case of noise (iii), whereas superiority to the half-sum signal is obvious, performance of AGC-CBF is almost identical to that of 2chSS-ORG. Because noise (iii) is nonstationary speech noise, the loss of noise consistency seems to be less perceptible than that of stationary noise. Nevertheless, this overlap might be more annoying than that of 2chSS-ORG enhanced signal because AGC does not reduce overlapped speech noise on the desired speech. Unlike above two noise cases, in the case of noise (i), AGC-CBF is the most preferred among all methods in this comparison: noise (i) is impulsive noise and lost consistency of noise was less perceptive in this case. In addition, the proposed method was preferred to SSCS in all noise cases.

The absolute removal of noise is not advantageous to maintain naturalness of a processed signal. Some residual noise left over in the output signal is preferred. In the proposed method, residual noise can be controlled easily by forcing the gain to be larger than a given minimum value, e.g., 0.1. Furthermore, this setup generates no musical noise. We can expect to be able to improve AGC performance in listening tests by controlling residual noise over the processed signal, but noise control should

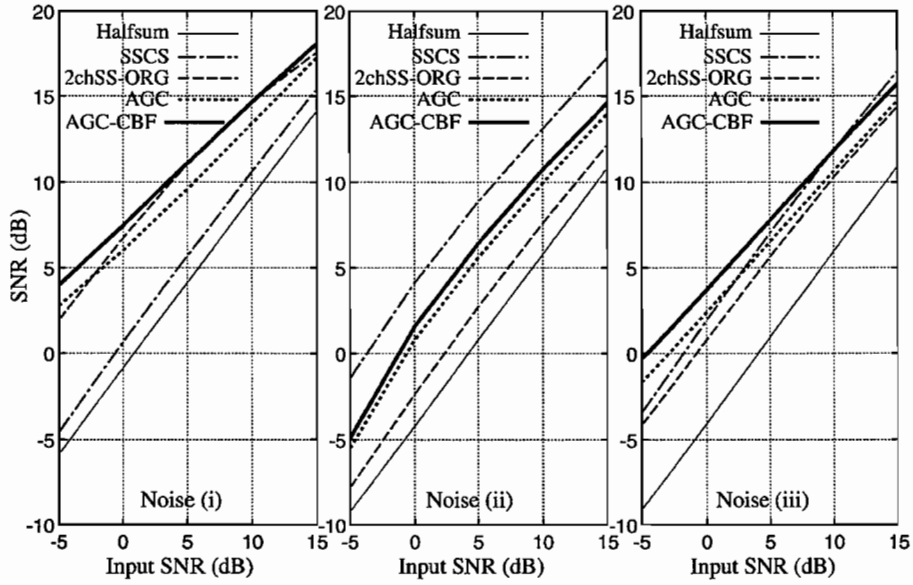


Fig. 18. Comparative performance for the whole signal in terms of overall SNR.

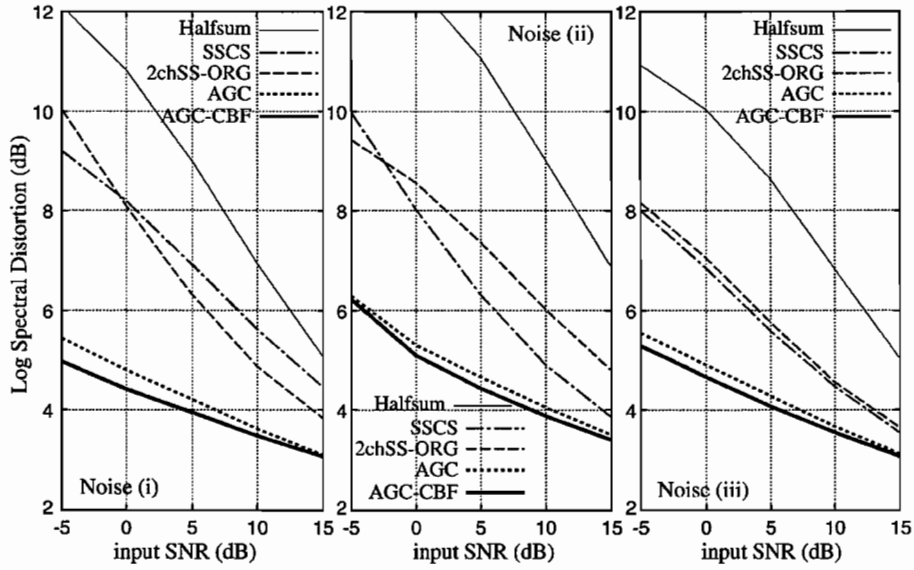


Fig. 19. Comparative performance for the whole signal in terms of log spectral distortion (LSD).

be mentioned for all methods for comparison to maintain a fair comparison. Such noise control preparation requires intensive investigation that could not be adequately addressed in this limited study.

VI. Conclusion

This study introduced a new method of two-channel speech enhancement based on auto gain control. Unlike other methods, the proposed method preserves the spectrum of the desired speech and hardly generates musical noise. Particularly, the proposed method can deal with transient noise better than most speech enhancement methods. Objective measures and spectrograms demonstrated significant improvements over other two-channel based speech enhancement methods in the three noise conditions. On the other hand, subjective preference tests revealed that the proposed method is less preferred to a half-sum signal in the case of stationary car noise. These results suggest that AGC can affect the consistency of stationary noise markedly, engendering an unnatural quality of the processed signal. Improvement of this problem can be accomplished by controlling residual noise to maintain a natural quality of the signal. Nevertheless, these preference tests supported the superiority of the proposed method over other methods in the presence of impulsive noises.

REFERENCES

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

[2] H. Y. Kim, F. Asano, Y. Suzuki, and T. Sone, "Speech enhancement based on short-time spectral amplitude estimation with two-channel beamformer," *IEICE Trans. Fund.*, vol. E79-A, no. 12, pp. 2151–2158, Dec. 1996.

[3] L. J. Griffiths and C. W. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Trans. Antennas Propagat.*, vol. AP-30, pp. 27–34, 1982.

[4] J. E. Greenberg and P. M. Zurek, "Evaluation of an adaptive beamforming method for hearing aids," *J. Acoust. Soc. Amer.*, vol. 91, no. 3, pp. 1662–1676, Mar. 1992.

[5] Y. Kaneda and J. Ohga, "Adaptive microphone-array system for noise reduction," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP–34, pp. 1391–1400, Dec. 1986.

[6] O. L. Frost III, "An algorithm for linearly constrained adaptive array processing," *Proc. IEEE*, vol. 60, pp. 926–935, Aug. 1972.

[7] R. Zelinski, "A microphone array with adaptive post-filtering for noise reduction in reverberant rooms," in *Proc. ICASSP'88*, 1988, pp. 2578–2581.

[8] S. Fischer and K. D. Kammeyer, "Broadband beamforming with adaptive postfiltering for speech acquisition in noisy environments," in *Proc. ICASSP'97*, 1997, pp. 359–362.

[9] C. Marro, Y. Mahieux, and K. U. Simmer, "Analysis of noise reduction and dereverberation techniques based on microphone arrays with postfiltering," *IEEE Trans. Speech Audio Process.*, vol. 6, no. 3, pp. 240–259, May 1998.

[10] I. A. McCowan and H. Bourlard, "Microphone array post-filter for diffuse noise field," in *Proc. ICASSP'2002*, 2002, pp. I-905–I-908.

[11] I. Cohen and B. Berdugo, "Microphone array post-filtering for nonstationary noise suppression," in *Proc. ICASSP'2002*, 2002, pp. 901–904.

[12] R. Le Bouquin-Jeannès, A. A. Azirani, and G. Faucon, "Enhancement of speech degraded by coherent and incoherent noise using a cross-spectral estimator," *IEEE Trans. Speech Audio Process.*, vol. 5, no. 5, pp. 484–487, Sep. 1997.

[13] R. Le Bouquin-Jeannès and G. Faucon, "Using the coherence function for noise reduction," *Proc. Inst. Elect. Eng.*, vol. 139, no. 3, pp. 276–280, Jun. 1992.

[14] Y. Kaneda and M. Tohyama, "Noise suppression signal processing using 2-point received signals," *Electron. Commun. Jpn.*, vol. 67-A, pp. 19–28, 1984.

[15] *Extrapolation, Interpolation and Smoothing of Stationary Time Series*, N. Wiener, Ed., MIT, Cambridge, MA, 1964.

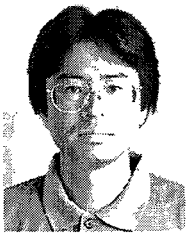
[16] H. W. Bode and C. E. Shannon, "A simplified derivation of linear least square smoothing and prediction theory," *Proc. IRE*, vol. 38, pp. 417–425, Apr. 1950.

[17] *Digital Signal Processing*, J. G. Proakis and D. G. Manolakis, Eds., Prentice-Hall, Englewood Cliffs, NJ, 1996.

[18] U. Mittal and N. Phamdo, "Signal/noise KLT based approach for enhancing speech degraded by colored noise," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 2, pp. 159–167, Mar. 2000.

[19] Y. Ephraim and H. L. Van Trees, "A signal subspace approach for speech enhancement," *IEEE Trans. Speech Audio Process.*, vol. 3, no. 4, pp. 251–266, Jul. 1995.

[20] A. Rezayee and S. Gazor, "An adaptive KLT approach for speech enhancement," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 2, pp. 87–95, Feb. 2001.



Yoshifumi Nagata received the B.E. degree in electronics in 1984, and the M.E. and Dr.Eng. degrees in information science in 1987 and 1990, respectively, all from Tohoku University, Sendai, Japan.

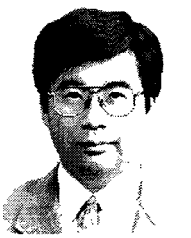
In 1990, he joined the Research and Development Center, Toshiba Corporation, where he has been engaged in research and development of speech processing systems. Since 1997, he has been an Associate Professor, Iwate University, Morioka, Japan. His interests include multimedia human interface and speech signal processing.

Dr. Nagata is a member of Acoustical Society of Japan and the Information Processing Society of Japan.

Toyota Fujioka was born in Akita, Japan, on August 21, 1969. He received the B.E. and M.E. degrees in electrical and electronic engineering from the Mining Collage, Akita University, Japan, in 1992 and 1994, respectively, and the Ph.D. degree in electrical and communication engineering from Tohoku University, Sendai, Japan, in 1997.

He is currently a Research Associate in the Department of Computer and Information Science Faculty of Engineering, Iwate University. His research interests include parallel computer and data compression.

Dr. Fujioka is a member of the Information Processing Society of Japan.



Masato Abe (M'84) received the B.E, M.E., and Ph.D. degrees in electrical engineering from Tohoku University, Sendai, Japan, in 1976, 1978, and 1981, respectively.

From 1981 to 1989, he was a Research Associate with the Research Center for Applied Information Sciences, Tohoku University. From 1989 to 1996, he was an Associate Professor in the Department of Information Science, Iwate University, Morioka, Japan. His research interests include digital signal processing for acoustics and computer architecture.

Dr. Abe is a member of the Acoustical Society of America, the Acoustical Society of Japan, the Institute of Noise Control Engineering of Japan, the Information Processing Society of Japan, the Institute of Electronics, Information and Communication Engineers, the Association for Computing Machinery, and the Japan Society of Mechanical Engineers.