NEW CONFIGURATION FOR A STEREO ECHO CANCELLER WITH NONLINEAR PRE-PROCESSING

Suehiro Shimauchi, Yoichi Haneda, Shoji Makino, and Yutaka Kaneda

NTT Human Interface Laboratories
3-9-11, Midori-cho, Musashino-shi, Tokyo, 180 Japan
sima@splab.hil.ntt.co.jp, http://www.hil.ntt.co.jp/AcousticEnv/onjo-e.html

ABSTRACT

A new configuration for a stereo echo canceller with nonlinear pre-processing is proposed. The pre-processor which adds uncorrelated components to the original received stereo signals improves the adaptive filter convergence even in the conventional configuration. However, because of the inaudibility restriction, the pre-processed signals still include a large amount of the original stereo signals which are often highly cross-correlated. Therefore, the improvement is limited. To overcome this, our new stereo echo canceller includes exclusive adaptive filters whose inputs are the uncorrelated signals generated in the pre-processor. These exclusive adaptive filters converge to true solutions without suffering from cross-correlation between the original stereo signals. This is demonstrated through computer simulation results.

1. INTRODUCTION

A stereo teleconferencing system provides greater presence than a monaural system. However, such a system requires stereo echo cancellers to suppress the acoustic echoes which may otherwise cause howling. One of the most significant problems with the stereo echo canceller is that the adaptive filters often misconverge or, if not, convergence speeds are very slow because of the cross-correlation between stereo signals[1, 2, 9]. Recently, there has been interest in applying pre-processed stereo signals to both the adaptive filter inputs and the loudspeaker inputs, which may help the adaptive filters converge to their true solutions[9]. In fact, practical pre-processing methods have been proposed whose distortion audibilities are very low for their improvement of the adaptive filter convergence[4, 5]. However, the pre-processor itself may be inherently limited in terms of improving the adaptive filter convergence because of the inaudibility restriction. On the other hand, no adaptive algorithm actively takes into account how the input signals are pre-processed.

In this paper, we propose a new configuration for a stereo echo canceller with nonlinear pre-processing. It includes separate adaptive filters for both the original signals and the additive nonlinear signals, both of which are extracted before being combined as the pre-processed signals. If the additive signals are generated to be uncorrelated to the original ones, the adaptive filters for the additive signals converge to the true solutions without suffering from cross-correlation between the original signals.

2. CONVENTIONAL STEREO ECHO CANCELLER CONFIGURATION

A conventional stereo echo canceller is shown in Fig. 1. Although one more microphone is actually necessary, we omit showing it because the process is symmetrical and independent. The function block is a pre-processor for adding the variation in cross-correlation between the stereo signals[9]. Practical methods for pre-processing have been proposed[4, 5]. Following the description in Ref. [4], the pre-processed signals are represented as $x_1'(k) = x_1(k) + \alpha_1 f_1[x_1(k)]$ and $x_2'(k) = x_2(k) + \alpha_2 f_2[x_2(k)]$, where $x_1(k)$ and $x_2(k)$ are the signals received from the far end, α_1 and α_2 are tuning parameters, and f_1 and f_2 are generally nonlinear functions. The echo $y(k)$ picked up by the microphone is modeled as

$$y(k) = \boldsymbol{x}'^T_1(k)\boldsymbol{h}_1(k) + \boldsymbol{x}'^T_2(k)\boldsymbol{h}_2(k), \tag{1}$$

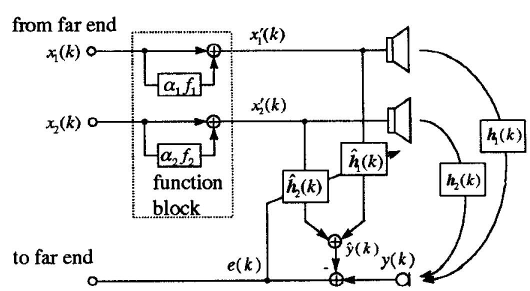

Figure 1: Conventional configuration of a stereo echo canceller with pre-processing.

0-7803-4428-6/98 $10.00 © 1998 IEEE

where $\boldsymbol{x}_1'(k) = [x_1'(k), x_1'(k-1), \ldots, x_1'(k-L+1)]^T$, $\boldsymbol{x}_2'(k) = [x_2'(k), x_2'(k-1), \ldots, x_2'(k-L+1)]^T$, $\boldsymbol{h}_1(k)$ and $\boldsymbol{h}_2(k)$ are the echo path impulse response coefficient vectors, L is the length of $\boldsymbol{h}_1(k)$ and $\boldsymbol{h}_2(k)$, k is a discrete time index, and ^{T} indicates a transpose. The adaptive filters $\hat{\boldsymbol{h}}_1(k)$ and $\hat{\boldsymbol{h}}_2(k)$ should identify the true echo paths $\boldsymbol{h}_1(k)$ and $\boldsymbol{h}_2(k)$ and cancel the echo $y(k)$ generating the echo replica $\hat{y}(k)$:

$$\hat{y}(k) = \boldsymbol{x}_1'^T(k)\hat{\boldsymbol{h}}_1(k) + \boldsymbol{x}_2'^T(k)\hat{\boldsymbol{h}}_2(k). \tag{2}$$

The two-channel versions of the normalized least mean squares (NLMS) algorithm[5], the (affine) projection algorithm[3, 9, 10], and the recursive least squares (RLS) algorithm[2, 4] are mainly used to update $\hat{\boldsymbol{h}}_1(k)$ and $\hat{\boldsymbol{h}}_2(k)$.

The pre-processor improves the adaptive filter convergences. However, because of the inaudibility restriction, the pre-processed signals $x_1'(k)$ and $x_2'(k)$ still include a large amount of the original signals $x_1(k)$ and $x_2(k)$ which are often highly cross-correlated. Therefore, the signals added in the pre-processor cannot effectively contribute to improving the convergences of the adaptive filters in the conventional configuration.

3. NEW STEREO ECHO CANCELLER CONFIGURATION

3.1. Four-adaptive-filter structure

We propose a new configuration for the stereo echo canceller, in which exclusive adaptive filters are arranged for the additive signals (Fig. 2).

The pre-processed signal vectors $\boldsymbol{x}_1'(k)$ and $\boldsymbol{x}_2'(k)$ are denoted as

$$\boldsymbol{x}_1'(k) = \boldsymbol{x}_1(k) + \alpha_1 \bar{\boldsymbol{x}}_1(k), \tag{3}$$
$$\boldsymbol{x}_2'(k) = \boldsymbol{x}_2(k) + \alpha_2 \bar{\boldsymbol{x}}_2(k), \tag{4}$$

where

$$\boldsymbol{x}_1(k) = [x_1(k), x_1(k-1), \ldots, x_1(k-L+1)], \tag{5}$$
$$\boldsymbol{x}_2(k) = [x_2(k), x_2(k-1), \ldots, x_2(k-L+1)], \tag{6}$$
$$\bar{\boldsymbol{x}}_1(k) = [f_1[x_1(k)], f_1[x_1(k-1)], \ldots, f_1[x_1(k-L+1)]], \tag{7}$$
$$\bar{\boldsymbol{x}}_2(k) = [f_2[x_2(k)], f_2[x_2(k-1)], \ldots, f_2[x_2(k-L+1)]]. \tag{8}$$

Substituting Eqs. (3) and (4) to Eq. (1), we obtain

$$y(k) = \boldsymbol{x}_1^T(k)\boldsymbol{h}_1(k) + \boldsymbol{x}_2^T(k)\boldsymbol{h}_2(k) + \alpha_1 \bar{\boldsymbol{x}}_1^T(k)\boldsymbol{h}_1(k) + \alpha_2 \bar{\boldsymbol{x}}_2^T(k)\boldsymbol{h}_2(k). \tag{9}$$

Then, by introducing the exclusive adaptive filters $\bar{\boldsymbol{h}}_1(k)$ and $\bar{\boldsymbol{h}}_2(k)$, the echo replica $\hat{y}(k)$ is generated as

$$\hat{y}(k) = \boldsymbol{x}_1^T(k)\hat{\boldsymbol{h}}_1(k) + \boldsymbol{x}_2^T(k)\hat{\boldsymbol{h}}_2(k) + \alpha_1 \bar{\boldsymbol{x}}_1^T(k)\bar{\boldsymbol{h}}_1(k) + \alpha_2 \bar{\boldsymbol{x}}_2^T(k)\bar{\boldsymbol{h}}_2(k). \tag{10}$$

Even though $\boldsymbol{x}_1(k)$ and $\boldsymbol{x}_2(k)$ are highly correlated, $\bar{\boldsymbol{h}}_1(k)$ and $\bar{\boldsymbol{h}}_2(k)$ should converge to $\boldsymbol{h}_1(k)$ and $\boldsymbol{h}_2(k)$ respectively, if $\bar{\boldsymbol{x}}_1(k)$ and $\bar{\boldsymbol{x}}_2(k)$ are uncorrelated to each other, and

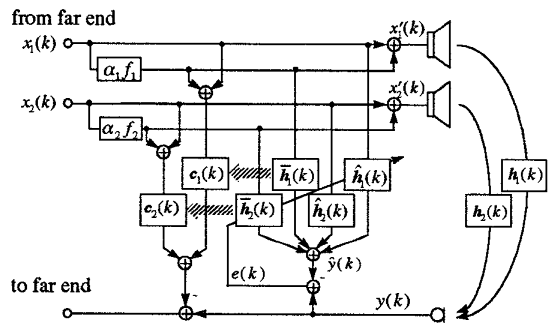

Figure 2: A new configuration for the stereo echo canceller.

also to both $\boldsymbol{x}_1(k)$ and $\boldsymbol{x}_2(k)$. Then, as shown in Fig. 2, the echo cancellation is actually achieved by using the echo cancellation filters $\boldsymbol{c}_1(k)$ and $\boldsymbol{c}_2(k)$ whose coefficients are transferred from $\bar{\boldsymbol{h}}_1(k)$ and $\bar{\boldsymbol{h}}_2(k)$ respectively only when $\bar{\boldsymbol{h}}_1(k)$ and $\bar{\boldsymbol{h}}_2(k)$ are judged to be in good condition. This can be regarded as a kind of "duo-filter control system[8]" which has the advantage of robustness during double-talk.

3.2. Desirable pre-processing

From **3.1**, f_1 and f_2 in the pre-processor should satisfy the following conditions:

$$r_{f1f2}(n) = E\,[f_1[x_1(k)]f_2[x_2(k-n)]] \approx 0, \tag{11}$$
$$r_{f1x1}(n) = E\,[f_1[x_1(k)]x_1(k-n)] \approx 0, \tag{12}$$
$$r_{f2x2}(n) = E\,[f_2[x_2(k)]x_2(k-n)] \approx 0, \tag{13}$$
$$r_{f1x2}(n) = E\,[f_1[x_1(k)]x_2(k-n)] \approx 0, \tag{14}$$
$$r_{f2x1}(n) = E\,[f_2[x_2(k)]x_1(k-n)] \approx 0, \tag{15}$$

for every discrete time index n, where $E[\cdot]$ denotes the expectation.

We consider choosing the absolute f_a as f_1:

$$f_1[x(k)] = f_a[x(k)] = |x(k)|. \tag{16}$$

This satisfies Eq. (12), if $E[x^2(k)]|_{x(k)>0} \approx E[x^2(k)]|_{x(k)<0}$. We note that

$$f_a[x(k)] = 2f_b[x(k)] - x(k), \tag{17}$$

where the function f_b is given in Ref. [4] as

$$f_b[x(k)] = \begin{cases} x & \text{if } x(k) \geq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{18}$$

Thus, the psychoacoustical influence of the function f_a on $x(k)+\alpha f_a[x(k)]$ is similar to that of f_b, whose psychoacoustical advantage is supported in Ref. [4]. However, $f_1 = f_a$ satisfies Eq. (12), while $f_1 = f_b$ does not. This is significant for our configuration. Next, we have to choose f_2 to satisfy Eq. (11). A possible example is

$$f_2[x(k)] = \sigma|x(k)|, \tag{19}$$

where $\sigma = 1$ or -1, with the sign being changed every time both of the following conditions are satisfied: $x(k-1)x(k) \leq 0$ and $x(k) \geq 0$. Then f_2 is psychoacoustically similar to f_1, but Eqs. (11) and (13) are satisfied. The above chosen functions f_1 and f_2 also satisfy Eqs. (14) and (15) in usual situations.

It should be remarked that the desirable inputs of $\bar{h}_1(k)$ and $\bar{h}_2(k)$ are not necessarily signals such that formulated as $\alpha_1 f_1[x_1(k)]$ and $\alpha_2 f_2[x_2(k)]$. In general, those signals which are uncorrelated to each other and also to both of the original received stereo signals are required. In this sense, while the configuration proposed in Ref. [6] is similar to ours, its exclusive adaptive filter inputs $\varepsilon_1(k)$ and $\varepsilon_2(k)$, which are generated to be uncorrelated to $x_1(k)$ and $x_2(k)$ respectively, are still correlated to $x_2(k)$ and $x_1(k)$ respectively. Therefore, this can be a reason why the configuration[6] does not achieve a satisfactory result.

3.3. Algorithm modification

Using f_1 and f_2 described in Eqs. (16) and (19), and setting $\alpha_1 = \alpha_2 = \alpha$, we consider the case where the four adaptive filters $\hat{h}_1(k)$, $\hat{h}_2(k)$, $\bar{h}_1(k)$, and $\bar{h}_2(k)$ are updated using the four-channel NLMS algorithm:

$$\begin{bmatrix} \hat{\boldsymbol{h}}_1(k+1) \\ \hat{\boldsymbol{h}}_2(k+1) \\ \bar{\boldsymbol{h}}_1(k+1) \\ \bar{\boldsymbol{h}}_2(k+1) \end{bmatrix} = \begin{bmatrix} \hat{\boldsymbol{h}}_1(k) \\ \hat{\boldsymbol{h}}_2(k) \\ \bar{\boldsymbol{h}}_1(k) \\ \bar{\boldsymbol{h}}_2(k) \end{bmatrix} + \frac{\mu e(k)}{(1+\alpha^2)(||\boldsymbol{x}_1(k)||^2+||\boldsymbol{x}_2(k)||^2)} \begin{bmatrix} \boldsymbol{x}_1(k) \\ \boldsymbol{x}_2(k) \\ \alpha\bar{\boldsymbol{x}}_1(k) \\ \alpha\bar{\boldsymbol{x}}_2(k) \end{bmatrix}, \tag{20}$$

where μ is a stepsize parameter and $e(k) = y(k) - \hat{y}(k)$. The characteristics of the adaptive filter convergences can be analyzed using the covariance matrix $\boldsymbol{R}'$:

$$\boldsymbol{R}' = \frac{1}{1+\alpha^2} \begin{bmatrix} \boldsymbol{R} & \boldsymbol{0} \\ \boldsymbol{0} & \alpha^2 \bar{\boldsymbol{R}} \end{bmatrix}, \tag{21}$$

where

$$\boldsymbol{R} = E\left[\frac{1}{||\boldsymbol{x}_1(k)||^2+||\boldsymbol{x}_2(k)||^2} \begin{bmatrix} \boldsymbol{x}_1(k)\boldsymbol{x}_1^T(k) & \boldsymbol{x}_1(k)\boldsymbol{x}_2^T(k) \\ \boldsymbol{x}_2(k)\boldsymbol{x}_1^T(k) & \boldsymbol{x}_2(k)\boldsymbol{x}_2^T(k) \end{bmatrix}\right], \tag{22}$$

$$\bar{\boldsymbol{R}} = E\left[\frac{1}{||\boldsymbol{x}_1(k)||^2+||\boldsymbol{x}_2(k)||^2} \begin{bmatrix} \bar{\boldsymbol{x}}_1(k)\bar{\boldsymbol{x}}_1^T(k) & \boldsymbol{0} \\ \boldsymbol{0} & \bar{\boldsymbol{x}}_2(k)\bar{\boldsymbol{x}}_2^T(k) \end{bmatrix}\right], \tag{23}$$

and Eqs. (11) - (15) are assumed to be satisfied. Equation (21) indicates that the steady-state solutions of $\bar{h}_1(k)$ and $\bar{h}_2(k)$ are independent of other channel signals, while those of $\hat{h}_1(k)$ and $\hat{h}_2(k)$ are affected by the cross-correlation between $\boldsymbol{x}_1(k)$ and $\boldsymbol{x}_2(k)$. However, the convergence speed of $\bar{h}_1(k)$ and $\bar{h}_2(k)$ may still be slow, because the maximum eigenvalue of $\alpha^2\bar{\boldsymbol{R}}$ is approximately α^2 times smaller than that of $\boldsymbol{R}$, where α is smaller than 1, e.g. $\alpha = 0.2$. A technique to improve such a slow convergence was proposed in Ref. [7]. It is based on a power-normalization of each channel signal. Before applying it to our case, we note that it guarantees the adaptive filters will converge to their true solutions only if the inter-channel independence is held. Thus, the update equations are modified as

$$\begin{bmatrix} \hat{\boldsymbol{h}}_1(k+1) \\ \hat{\boldsymbol{h}}_2(k+1) \end{bmatrix} = \begin{bmatrix} \hat{\boldsymbol{h}}_1(k) \\ \hat{\boldsymbol{h}}_2(k) \end{bmatrix} + \frac{\mu e(k)}{2(||\boldsymbol{x}_1(k)||^2+||\boldsymbol{x}_2(k)||^2)} \begin{bmatrix} \boldsymbol{x}_1(k) \\ \boldsymbol{x}_2(k) \end{bmatrix}, \tag{24}$$

$$\begin{bmatrix} \bar{\boldsymbol{h}}_1(k+1) \\ \bar{\boldsymbol{h}}_2(k+1) \end{bmatrix} = \begin{bmatrix} \bar{\boldsymbol{h}}_1(k) \\ \bar{\boldsymbol{h}}_2(k) \end{bmatrix} + \frac{\mu\beta e(k)}{2\alpha^2(||\boldsymbol{x}_1(k)||^2+||\boldsymbol{x}_2(k)||^2)} \begin{bmatrix} \alpha\bar{\boldsymbol{x}}_1(k) \\ \alpha\bar{\boldsymbol{x}}_2(k) \end{bmatrix}, \tag{25}$$

where β is a parameter for relaxation. In this case, the covariance matrix $\boldsymbol{R}''$ becomes

$$\boldsymbol{R}'' = \frac{1}{2} \begin{bmatrix} \boldsymbol{R} & \boldsymbol{0} \\ \boldsymbol{0} & \beta\bar{\boldsymbol{R}} \end{bmatrix}. \tag{26}$$

Since β can be much larger than α^2, the convergence speeds of $\bar{h}_1(k)$ and $\bar{h}_2(k)$ can be improved.

4. COMPUTER SIMULATIONS

4.1. Independence of processed signals

We confirm how well f_1 and f_2 described in Eqs. (16) and (19) satisfy the conditions mentioned in **3.2**, by calculating the 80000-sample-time averaged cross-correlation functions $r'_{f1f2}(n)$, $r'_{f1x1}(n)$, and $r'_{f2x2}(n)$. The signals $x_1(k)$ and $x_2(k)$ were made from one-channel signal $x(k)$ by convolving 2000 samples each of two different impulse responses measured in our conference room at an 8 kHz sampling rate. White Gaussian noise and Japanese male speech at the 8 kHz sampling rate were used for the signal $x(k)$. Table 1 shows the maximum amplitude of each normalized cross-correlation function for $n = -2000$ to 2000. Comparing $r'_{x1x2}(n)$, which corresponds to the cross-correlation function between $x_1(k)$ and $x_2(k)$, we found that $r'_{f1f2}(n)$, $r'_{f1x1}(n)$, and $r'_{f2x2}(n)$ were sufficient small.

Table 1: Maximum amplitude of normalized cross-correlation functions ($-2000 \leq n \leq 2000$).

	White Gaussian	Japanese speech
$r'_{f1f2}(n)$	0.010	0.030
$r'_{f1x1}(n)$	0.007	0.035
$r'_{f2x2}(n)$	0.016	0.045
$r'_{x1x2}(n)$	0.752	0.787

4.2. Convergence of adaptive filters

We show here the convergence characteristic of each adaptive filter updated by using Eqs. (24) and (25). The true echo path impulse responses $\boldsymbol{h}_1(k)$ and $\boldsymbol{h}_2(k)$ were measured in a conference room at the 8 kHz sampling rate and truncated at 1000 samples. The functions f_1 and f_2 were chosen as Eqs. (16) and (19). The input stereo signals were those mentioned in **4.1**. Each of the adaptive filters $\hat{h}_1(k)$, $\hat{h}_2(k)$, $\bar{h}_1(k)$, and $\bar{h}_2(k)$ had 1000 taps. The ambient noise was added to the echo $y(k)$ to achieve a 30 dB SNR. The parameters were $\alpha = 0.2$, $\beta = 0.7$, and $\mu = 0.5$. Then, the adaptive filter coefficient error convergences were evaluated for inputs of the white Gaussian noise (Fig. 3) and the Japanese male speech (Fig. 4). In both figures, the errors between $[h_1^T(k), h_2^T(k)]^T$ and (a) $[\hat{h}_1^T(k), \hat{h}_2^T(k)]^T$ (proposed), (b) $[\bar{h}_1^T(k), \bar{h}_2^T(k)]^T$ (proposed), and (c) $[\hat{h}_1^T(k), \hat{h}_2^T(k)]^T$ (conventional (Fig. 1) with NLMS) are compared. Note that, for the speech inputs, the coefficient errors were calculated after weighting by the averaged speech signal spectrum. These results show that the exclusive adaptive filters $\bar{h}_1(k)$ and $\bar{h}_2(k)$ quickly converge to the true solutions.

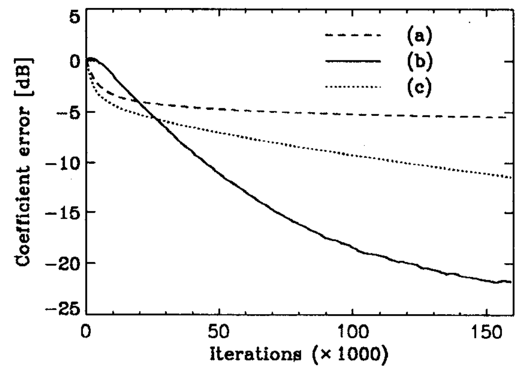

Figure 3: Coefficient error for the white Gaussian noise.

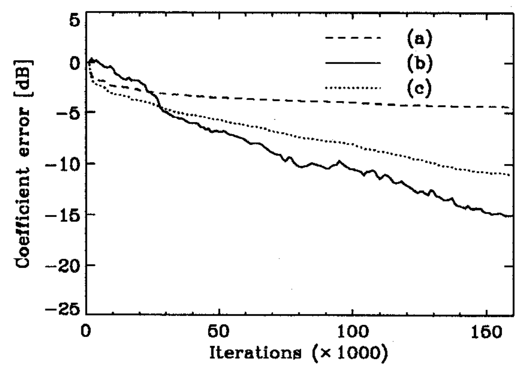

Figure 4: Coefficient error for the Japanese male speech. (calculated after weighting by averaged speech signal spectrum.)

5. CONCLUSION

We have proposed a new configuration for a stereo echo canceller with nonlinear pre-processing. In this configuration, the nonlinearly processed signals are input to the exclusive adaptive filters arranged for them. By choosing nonlinear functions in the pre-processor to generate uncorrelated signals from input stereo signals, the exclusive adaptive filters converge to the true solutions without suffering from cross-correlation between the original stereo signals.

6. ACKNOWLEDGMENTS

We would like to thank Mr. Y. Nishino, Mr. M. Tanaka, Mr. A. Nakagawa, and Mr. K. Suzuki for fruitful discussions and many suggestions.

7. REFERENCES

[1] M. M. Sondhi, D. R. Morgan, and J. L. Hall, "Stereophonic Acoustic Echo Cancellation -An Overview of the Fundamental Problem," *IEEE Signal Processing Letters*, vol. 2, no. 8, pp. 148-151 (1995).

[2] J. Benesty, F. Amand, A. Gilloire, and Y. Grenier, "Adaptive Filtering Algorithms for Stereophonic Acoustic Echo Cancellation," *Proc. ICASSP95*, vol. 5, pp. 3099-3102 (1995).

[3] J. Benesty, P. Duhamel, and Y. Grenier, "A Multi-Channel Affine Projection Algorithm with Applications to Multi-Channel Acoustic Echo Cancellation," *IEEE Signal Processing Letters*, vol. 3, no. 2, pp. 35-37 (1996).

[4] J. Benesty, D. R. Morgan, and M. M. Sondhi, "A Better Understanding and an Improved Solution to the Problems of Stereophonic Acoustic Echo Cancellation," *Proc. ICASSP97*, vol. 1, pp. 303-306 (1997).

[5] Y. Joncour and A. Sugiyama, "A New Stereo Echo Canceller with Pre-Processing of One Input Signal: I. Strict Identification of the Echo Paths," *Proc. 1997 Engineering Sciences Society Conf. of IEICE*, A-4-10 (1997).

[6] Y. Mahieux, A. Gilloire, and F. Khalil, "Annulation D'echo en Teleconference Stereophonique," *Proc. Quatorzieme Colloque GRETSI* (1993) (in French).

[7] T. Fujii and S. Shimada, "Linear-Combined Multi-Channel Adaptive Digital Filters," *Trans. IEICE*, vol. J69-A, no. 10, pp. 1226-1233 (1986) (in Japanese).

[8] Y. Haneda, S. Makino, J. Kojima, and S. Shimauchi, "Implementation and Evaluation of an Acoustic Echo Canceller Using Duo-Filter Control System," *Proc. EUSIPCO96*, vol. 2, pp. 1115-1118 (1996).

[9] S. Shimauchi and S. Makino, "Stereo Projection Echo Canceller with True Echo Path Estimation," *Proc. ICASSP95*, vol. 5, pp. 3059-3062 (1995).

[10] S. Makino, K. Strauss, S. Shimauchi, Y. Haneda, and A. Nakagawa, "Subband Stereo Echo Canceller Using the Projection Algorithm with Fast Convergence to the True Echo Path," *Proc. ICASSP97*, vol. 1, pp. 299-302 (1997).