CONVERGENCE ANALYSIS
OF STEREOPHONIC ECHO CANCELLER WITH PRE-PROCESSING
– RELATION BETWEEN PRE-PROCESSING AND CONVERGENCE –

Akihiro HIRANO   Kenji NAKAYAMA   Kazunobu WATANABE

Faculty of Engineering, Kanazawa University
40-20, Kodatsuno 2-Chome, Kanazawa 920-8667, Japan
E-mail: hirano@t.kanazawa-u.ac.jp

ABSTRACT

This paper presents convergence characteristics of stereophonic echo cancellers with pre-processing. The convergence analysis of the averaged tap-weights show that the convergence characteristics depends on the relation between the impulse response in the far-end room and the changes of the pre-processing filters. Examining the uniqueness of the solution in the frequency domain leads us to the same relation. Computer simulation results show the validity of these analyses.

1. INTRODUCTION

Echo cancellers are used to reduce echoes in a wide range of applications, such as TV conference systems and hands-free telephones. To realistic TV conferencing, multi-channel audio, at least stereophonic, is essential. For stereophonic teleconferencing, stereophonic acoustic echo cancellers have been studied [1-9].

In stereophonic echo cancellers, an uniqueness problem is one of the most serious problems [2-4]. There are infinite number of solutions for tap-weights of adaptive filters. Strong cross-correlation between input signals causes incorrect identification of the echo paths. To overcome this problem, improved echo cancellation algorithms have been proposed [5-9]. Some of these algorithms introduce a pre-processing which artificially varies the cross-correlation [5, 6]. The others incorporate independent components into input signals [7-9]. However, the convergence of these algorithm has not been analyzed.

This paper investigates convergence characteristics of stereophonic echo cancellers with pre-processing. The convergence analysis of the averaged tap-weights show that the convergence characteristics depends on the relation between the impulse response in the far-end room and the changes of the pre-processing filters. Examining the uniqueness of the solution in the frequency domain leads us to the same relation. Computer simulations validate the analyses.

2. STEREOPHONIC ACOUSTIC ECHO CANCELLER WITH PRE-PROCESSING

Figure 1 depicts an teleconferencing using stereophonic echo canceller with pre-processing [5, 6]. This echo canceller consists of four adaptive filters corresponding to four echo paths from two loudspeakers to two microphones. Each adaptive filter estimates the corresponding echo path. Pre-processing units $\mathbf{F}_{1,k}$
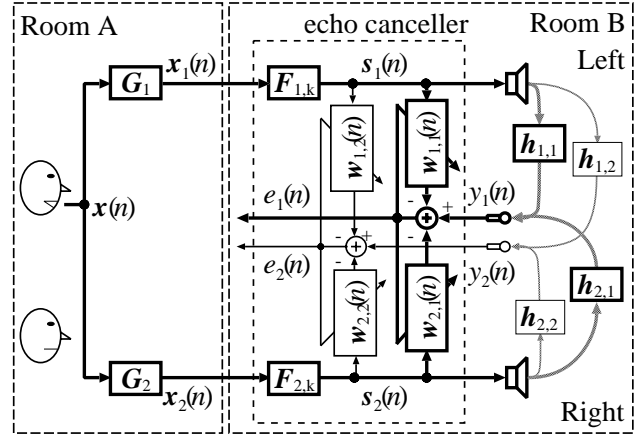

Fig. 1. Teleconferencing using
stereophonic echo canceller with pre-processing.

and $\mathbf{F}_{2,k}$, which are time-varying filters, are introduced for correct echo path identification.

The far-end signal vector $\mathbf{x}_j(n)$ in the j-th channel at time index n is generated from a white signal vector $\mathbf{x}(n)$ by

$$\mathbf{x}_j(n) = \mathbf{G}_j\mathbf{x}(n). \tag{1}$$

$\mathbf{G}_j$ is a matrix consists of the far-end room impulse response $g_{j,k}$ defined by

$$\mathbf{G}_j = \begin{bmatrix} \mathbf{g}_{j,1} & \mathbf{g}_{j,2} & \cdots & \mathbf{g}_{j,N_f+N_h-1} \end{bmatrix}^T \tag{2}$$

$$\mathbf{g}_{j,k} = \begin{bmatrix} \mathbf{0}_{1,k-1} & g_{j,0} & g_{j,1} & \cdots & g_{j,N_g-1} & \mathbf{0}_{1,N_f+N_h-k-1} \end{bmatrix}^T \tag{3}$$

where N_g is the length of the far-end room impulse response, N_f is the number of taps of the pre-processing filters, N_g is the number of taps of the adaptive filters, $\mathbf{0}_{i,j}$ is an $i \times j$ zero matrix, $[\cdot]^T$ denotes the transpose of a matrix $[\cdot]$. For colored signals, $g_{j,k}$ contains both a coloring filter and the room impulse response. The pre-processed signal vector $\mathbf{s}_j(n)$ is calculated by

$$\mathbf{s}_j(n) = \mathbf{F}_{j,k}\mathbf{x}_j(n) = \mathbf{F}_{j,k}\mathbf{G}_j\mathbf{x}(n)$$

where $\mathbf{F}_{j,k}$ contains the coefficients of the j-th pre-processing filter in the k-th state given by

$$\mathbf{F}_{j,k} = \begin{bmatrix} \mathbf{f}_{j,k,1} & \mathbf{f}_{j,k,2} & \cdots & \mathbf{f}_{j,k,N_h} \end{bmatrix}^T \tag{4}$$

$$\mathbf{f}_{j,k,l} = \begin{bmatrix} \mathbf{0}_{1,l-1} & f_{j,k,0} & f_{j,k,1} & \cdots & f_{j,k,N_f-1} & \mathbf{0}_{1,N_h-l} \end{bmatrix}^T. \qquad (5)$$

For a recursive pre-processing filters, an FIR approximation can be used. In order to overcome the uniqueness problem, multiple $\mathbf{F}_{j,k}$'s are used. Each $\mathbf{F}_{j,k}$ will be selected in the order of k, or in a random order.

The echo $y_i(n)$ in the i-th channel is generated by

$$y_i(n) = \sum_{j=1}^{2} \mathbf{s}_j^T(n)\mathbf{h}_{j,i} \qquad (6)$$

where $\mathbf{h}_{j,i}$ is the impulse response vector of the echo path from the j-th loudspeaker to the i-th microphone. The echo replica $\hat{y}_i(n)$ is calculated by

$$\hat{y}_i(n) = \sum_{j=1}^{2} \mathbf{s}_j^T(n)\mathbf{w}_{j,i}(n) \qquad (7)$$

where $\mathbf{w}_{j,i}(n)$ is the tap-weight vector of the adaptive filter which estimates $\mathbf{h}_{j,i}$. The residual echo $e_i(n)$ is given by

$$e_i(n) = y_i(n) - \hat{y}_i(n)$$

$$= \mathbf{x}^T(n) \sum_{i=1}^{2} \mathbf{G}_j^T \mathbf{F}_{j,k}^T (\mathbf{h}_{j,i} - \mathbf{w}_{j,i}(n)). \qquad (8)$$

Assuming an LMS (Least Mean Squares) algorithm, the tap-weight vector is updated by

$$\mathbf{w}_{j,i}(n+1) = \mathbf{w}_{j,i}(n) + \mu \mathbf{x}_j(n) e_i(n) \qquad (9)$$

where the step size μ is a positive constant.

3. CONVERGENCE OF AVERAGED TAP-WEIGHTS

Ensemble average of the tap-weight error defined by

$$\mathbf{m}_{j,i}(n) = E[\mathbf{h}_{j,i} - \mathbf{w}_{j,i}(n)] \qquad (10)$$

will be analyzed. The simultaneous difference equation for $\mathbf{m}_{j,i}(n)$ are

$$\mathbf{m}_{1,i}(n+1) = (\mathbf{I} - \mu \mathbf{F}_{1,k} \mathbf{G}_1 \mathbf{G}_1^T \mathbf{F}_{1,k}^T) \mathbf{m}_{1,i}(n)$$
$$-\mu \mathbf{F}_{1,k} \mathbf{G}_1 \mathbf{G}_2^T \mathbf{F}_{2,k}^T \mathbf{m}_{2,i}(n) \qquad (11)$$

$$\mathbf{m}_{2,i}(n+1) = (\mathbf{I} - \mu \mathbf{F}_{2,k} \mathbf{G}_2 \mathbf{G}_2^T \mathbf{F}_{2,k}^T) \mathbf{m}_{1,i}(n)$$
$$-\mu \mathbf{F}_{2,k} \mathbf{G}_2 \mathbf{G}_1^T \mathbf{F}_{1,k}^T \mathbf{m}_{1,i}(n) \qquad (12)$$

where $\mathbf{I}$ is an unit matrix. These equations can be re-written as

$$\mathbf{M}(n+1) = \left(\mathbf{I} - \mu \mathbf{F}_k \mathbf{G} \mathbf{G}^T \mathbf{F}_k^T\right) \mathbf{M}(n) \qquad (13)$$

where

$$\mathbf{M}(n) = \begin{bmatrix} \mathbf{m}_{1,i}(n) \\ \mathbf{m}_{2,i}(n) \end{bmatrix}, \quad \mathbf{G} = \begin{bmatrix} \mathbf{G}_1 \\ \mathbf{G}_2 \end{bmatrix}, \quad \mathbf{F}_k = \begin{bmatrix} \mathbf{F}_{1,k} & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_{2,k} \end{bmatrix}. \quad (14)$$

Let us assume that there are two pre-processing filters $\mathbf{F}_1$ and $\mathbf{F}_2$ and that they are periodically switched; using $\mathbf{F}_1$ for L iterations and then $\mathbf{F}_2$ for L iterations. Extension of this results to more general case is strait forward. The averaged tap-weight error matrix $\mathbf{M}(n)$ becomes

$$\mathbf{M}(2(m+1)L) =$$

$$\left(\mathbf{I} - \mu \mathbf{F}_2 \mathbf{G} \mathbf{G}^T \mathbf{F}_2^T\right)^L \left(\mathbf{I} - \mu \mathbf{F}_1 \mathbf{G} \mathbf{G}^T \mathbf{F}_1^T\right)^L \mathbf{M}(2mL) \qquad (15)$$

where m is an integer. By introducing the difference of two filters

$$\Delta \mathbf{F} = \mathbf{F}_2 - \mathbf{F}_1, \qquad (16)$$

$\mathbf{M}(2(m+1)L)$ is derived as

$$\mathbf{M}(2(m+1)L) = \left(\mathbf{I} - \mu(\mathbf{F}_1 + \Delta \mathbf{F})\mathbf{G}\mathbf{G}^T(\mathbf{F}_1^T + \Delta \mathbf{F}^T)\right)^L$$

$$\cdot \left(\mathbf{I} - \mu \mathbf{F}_1 \mathbf{G}\mathbf{G}^T \mathbf{F}_1^T\right)^L \mathbf{M}(2mL)$$

$$= \left\{\left(\mathbf{I} - \mu \mathbf{F}_1 \mathbf{G}\mathbf{G}^T \mathbf{F}_1^T\right)\right.$$

$$\left. -\mu\left(\Delta \mathbf{F}\mathbf{G}\mathbf{G}^T \mathbf{F}_1^T + \mathbf{F}_1 \mathbf{G}\mathbf{G}^T \Delta \mathbf{F}^T + \Delta \mathbf{F}\mathbf{G}\mathbf{G}^T \Delta \mathbf{F}^T\right)\right\}^L$$

$$\cdot \left(\mathbf{I} - \mu \mathbf{F}_1 \mathbf{G}\mathbf{G}^T \mathbf{F}_1^T\right)^L \mathbf{M}(2mL). \qquad (17)$$

Using a matrix $\mathbf{D}$ defined by

$$\mathbf{D} = -\mu\left(\Delta \mathbf{F}\mathbf{G}\mathbf{G}^T \mathbf{F}_1^T + \mathbf{F}_1 \mathbf{G}\mathbf{G}^T \Delta \mathbf{F}^T + \Delta \mathbf{F}\mathbf{G}\mathbf{G}^T \Delta \mathbf{F}^T\right), \qquad (18)$$

$\mathbf{M}(2(m+1)L)$ is derived as

$$\mathbf{M}(2(m+1)L) =$$

$$\sum_{i=0}^{L} C(i,L) \left(\mathbf{I} - \mu \mathbf{F}_1 \mathbf{G}\mathbf{G}^T \mathbf{F}_1^T\right)^{L+i} \mathbf{D}^{L-i} \mathbf{M}(2mL) \qquad (19)$$

where $C(i,L)$ is number of combinations selecting i from L.

If $\mathbf{D} = \mathbf{0}$, i.e., the pre-processor is a fixed filter, the right hand side of (19) becomes $(\mathbf{I} - \mu \mathbf{F}_1 \mathbf{G}\mathbf{G}^T \mathbf{F}_1^T)^{2L} \mathbf{M}(2mL)$. This corresponds to the convergence of the traditional stereophonic echo canceller [1]. Because of the uniqueness problem, $(\mathbf{I} - \mu \mathbf{F}_1 \mathbf{G}\mathbf{G}^T \mathbf{F}_1^T)^i$ do not converge to zero. Some of the eigenvalues of $\mathbf{F}_1 \mathbf{G}\mathbf{G}^T \mathbf{F}_1^T$ are zero and components corresponding to zero eigenvalues do not converge to the optimum values[4].

For convergence of the tap-weights to the echo paths, $\mathbf{D}$ should not be zero. Non-zero $\mathbf{D}$ will change the minimum eigenvalue from zero to non-zero. Larger $\mathbf{D}$ may results in smaller eigenvalue spread, and therefore, faster convergence will be achieved.

In $\mathbf{D}$, $\Delta \mathbf{F}\mathbf{G}$ denotes a combined characteristics of $\mathbf{G}$ and $\mathbf{F}_1 - \mathbf{F}_2$, i.e., filtering signals by $\mathbf{G}$ and then by $\mathbf{F}_1 - \mathbf{F}_2$. Thus, relation between $\mathbf{G}$ and $\mathbf{F}_1 - \mathbf{F}_2$ has a large influence on the convergence of the echo canceller. If some frequency components are removed by either $\mathbf{G}$ or $\mathbf{F}_1 - \mathbf{F}_2$, the frequency response of the adaptive filter corresponding to these frequencies is not determined.

Since users cannot modify the far-end room characteristics $\mathbf{G}$, $\mathbf{F}_1 - \mathbf{F}_2$ should be so designed as not to reduce any frequency components which $\mathbf{G}$ passes. Otherwise, the convergence speed becomes slow. Similarly, large $\Delta \mathbf{F}\mathbf{G}\mathbf{G}^T \mathbf{F}_1^T$ requires that the pass-band of $\mathbf{F}_1$ should cover that of $\mathbf{G}$.

4. ANALYSIS IN FREQUENCY DOMAIN

The influence of the relation between the far-end room characteristics and the pre-processing filters on the convergence will further be analyzed in the frequency domain. For k-th pre-

processing filter, the residual echo $E_{i,k}(z)$ in the i-th channel is calculated by

$$E_{i,k}(z) = \sum_{j=1}^{2}\left(H_{j,i}(z) - W_{j,i}(z)\right)F_{j,k}(z)G_j(z)X(z) \quad (20)$$

where $G_j(z)$ is the far-end room transfer function, $F_{j,k}(z)$ is the pre-processing filter, $H_{j,i}(z)$ is the echo path, $W_{j,i}(z)$ is the adaptive filter, $X(z)$ is the signal source. For a simple case where $F_{1,1}(z) = F_{1,2}(z) = 1$, solving $E_{i,1}(z) = E_{i,2}(z) = 0$ results in

$$\left(H_{1,i}(z) - W_{1,i}(z)\right)\left(F_{2,2}(z) - F_{2,1}(z)\right)G_1(z)X(z) = 0. \quad (21)$$

In order to have an optimum solution $W_{1,i}(z) = H_{1,i}(z)$, $(F_{2,2}(z) - F_{2,1}(z))G_1(z)X(z)$ should not be zero. Therefore, the pass-band of $F_{2,2}(z) - F_{2,1}(z)$ should cover all the pass-band of $G_1(z)$. Similarly, the pass-band of $F_{2,2}(z) - F_{2,1}(z)$ should also cover all the pass-band of $H_{1,i}(z) - W_{1,i}(z)$.

For more general case, the convergence condition becomes

$$\left(F_{1,1}(z)F_{2,2}(z) - F_{1,2}(z)F_{2,1}(z)\right)G_j(z) \neq 0. \quad (22)$$

Introducing

$$F_{j,2}(z) = F_{j,1}(z)\Delta F_j(z) \quad (23)$$

leads to

$$F_{1,1}(z)F_{2,1}(z)(\Delta F_2(z) - \Delta F_1(z))G_j(z) \neq 0. \quad (24)$$

This suggests that the pass-bands of $F_{1,1}(z)$, $F_{2,1}(z)$ and $\Delta F_2(z) - \Delta F_1(z)$ should cover the pass-band of $G_j(z)$ for good convergence. These results from frequency domain analysis agree with the convergence analysis of the averaged tap-weights shown in previous section.

5. COMPUTER SIMULATIONS

Simulations have been carried out to examine the relation between the characteristics of the pre-processor and that of the room impulse response. 20-th order Butterworth filters are used as the transfer functions in the far-end room $G_j(z)$'s and the echo paths $H_{j,i}(z)$'s. Both high-pass filters (HPF's) and low-pass filters (LPF's) are used. Example of the transfer functions are shown in Fig. 2.

For simplicity, $F_{1,1}(z) = F_{2,1}(z) = F_{2,2}(z) = 1$ are assumed. Thus, $F_{1,2}(z) = F(z)$ is periodically inserted and then removed in the first channel. As a pre-processing filter, four $F(z)$'s are used:

(1) $F(z) = z^{-1}$. $1 - F(z)$ is HPF. This can be considered as a simplified version of [5].

(2) $F(z) = \dfrac{a + z^{-1}}{1 + az^{-1}}$. $1 - F(z)$ is a HPF. This can be considered as a simplified version of [6].

(3) $F(z) = -z^{-1}$. $1 - F(z)$ is a LPF.

(4) $F(z) = \dfrac{(1 - a) - z^{-1}}{1 + az^{-1}}$. $1 - F(z)$ is an all-pass filter (APF).

Figure 3 demonstrates the transfer function $1 - F(z)$ for each $F(z)$. 64-tap adaptive FIR filters with the normalized LMS (least mean squares) algorithm, which is identical to the LMS algorithm for stationary inputs, are used. The step size is $\mu = 0.5$.
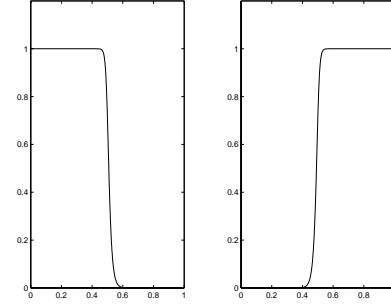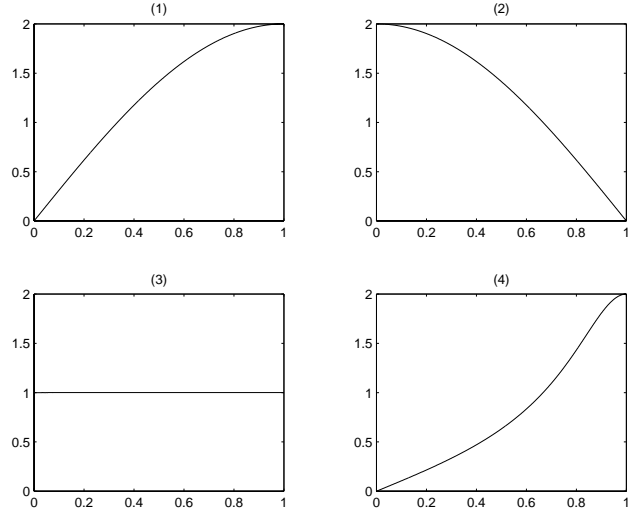

Fig. 2. Amplitude response of echo path.

Fig. 3 Amplitude response of $1 - F(z)$.

Figure 4 (1) depicts the convergence characteristics of the tap-weight error vector norm for $F(z) = z^{-1}$. A curve with "LP" label denotes the characteristics when both the echo path and the far-end room transfer functions are LPF's, and vice versa. As expected by the analyses, convergence speed for the LPF case is slower than that for the HPF case. For the LPF case, $(1 - F(z))G_j(z)$ becomes too small for a low frequency range. The convergence becomes slow for the low frequency range. The results in Fig. 4 (2) is similar to that in (1) because $1 - F(z)$ is also a LPF.

The opposite situation occurs when $1 - F(z)$ is a LPF. As shown in Fig. 4 (3), the convergence is faster when the echo path and the far-end room transfer functions are LPF's. Figure 4 (4) shows that the convergence speed is almost independent of the $G_j(z)$ and $H_{j,i}(z)$ if $1 - F(z)$ is an APF. This is because the amplitude of $(1 - F(z))G_j(z)$ is same as that of $G_j(z)$.
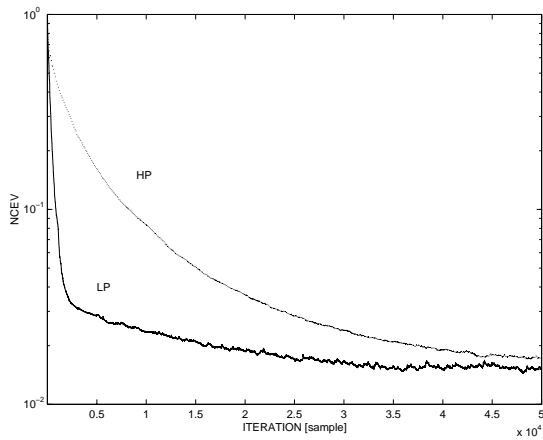
6. CONCLUSION

The convergence characteristics of the stereophonic echo cancellers with pre-processing have been examined. The convergence analysis of the averaged tap-weights show that the
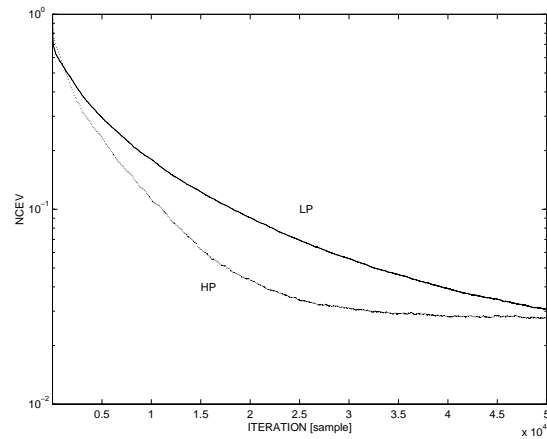
(1) $F(z) = z^{-1}$

(2) $F(z) = -z^{-1}$

Fig. 4 Convergence of tap-weight error vector norm.
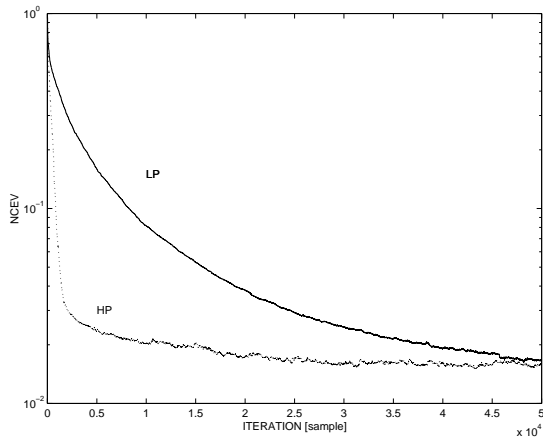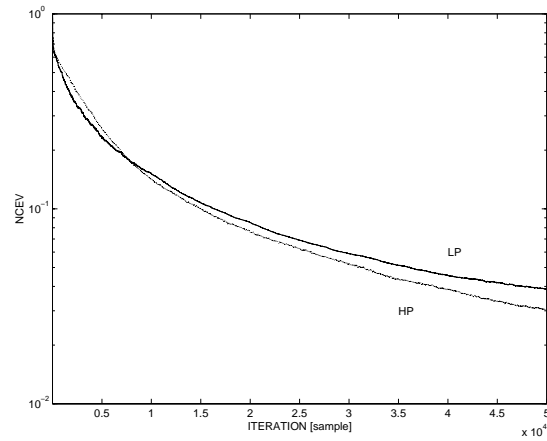
(3) $1 - F(z)$ is all-pass filter.

(4) $F(z)$ is all-pass filter.

Fig. 4 Convergence of tap-weight error vector norm (Conti.).

convergence characteristics depends on the relation between the impulse response in the far-end room and the changes of the pre-processing filters. The analysis in the frequency domain results in a same relation. Computer simulation results show the validity of these analyses.

REFERENCES

[1] T. Fujii et al., "A Note on Multi-Channel Echo Cancellers," Technical Reports of IEICE on CS, pp. 7-14, January 1984 (in Japanese).

[2] A. Hirano et al., "Convergence Characteristics of a Multi-Channel Echo Canceller with Strongly Cross-correlated Input Signals − Analytical Results −," Proc. of 6th DSP Symposium, pp. 144-149, November 1991.

[3] M. M. Sondhi et al., "Stereophonic Acoustic Echo Cancellation − An Overview of the Fundamental Problem," IEEE SP Letters, vol. 2, no. 8, pp. 148-151, August 1995.

[4] A. Hirano et al., "Convergence Analysis of a Stereophonic Acoustic Echo Canceller Part I: Convergence Characteristics of Tap Weights," Proc. of 11th DSP Symposium, pp. 569-574, November 1996.

[5] Y. Joncour et al., "A Stereo Echo Canceler with Pre-Processing for Correct Echo Path Identification," Proc. of ICASSP '98, pp. 3677-3680, May 1998.

[6] M. Ali, "Stereophonic Acoustic Echo Cancellation System Using Time-Varying All-Pass Filtering for Signal Decorrelation," Proc. of ICASSP '98, pp. 3689-3692, May 1998.

[7] J. Benesty et al., "Stereophonic Acoustic Echo Cancellation Using Nonlinear Transformations and Comb Filtering," Proc. of ICASSP '98, pp. 3637-3676, May 1998.

[8] A. Gilloire et al., "Using Auditory Properties to Improve The Behaviour of Stereophonic Acoustic Echo Cancellers," Proc. of ICASSP '98, pp. 3681-3684, May 1998.

[9] S. Shimauchi et al., "New Configuration for A Stereo Echo Canceller with Nonlinear Pre-Processing," Proc. of ICASSP '98, pp. 3685-3688, May 1998.