Convergence Characteristics of a Multi-Channel Echo Canceller with Strongly Cross-Correlated Input Signals - Analytical Results -

相互相関を持つ入力信号に対する多チャンネルエコーキャンセラの収束特性

Akihiro Hirano 平野晃宏
Akihiko Sugiyama 杉山昭彦

C&C Systems Research Laboratories, NEC Corporation
日本電気株式会社 C&Cシステム研究所

ABSTRACT

An Analysis of convergence characteristics for a multi-channel echo canceller with cross-correlated input signals is presented. The gradient with respect to the filter coefficients of a conventional multi-channel echo canceller contain disturbing components. These components behave like additive noises when the input signals have strong cross-correlation. With an example of strongly cross-correlated signals, it is analytically shown that the convergence speed is slower than that for uncorrelated signals and that the filter coefficients do not converge to their optimum values. A new multi-channel echo canceller whose filter coefficients converge faster to their optimum values is derived. Simulation results show the validity of the analysis and the performance of the new echo canceller.

1. INTRODUCTION

Echo cancellers are used for a wide range of applications, such as TV conference systems, to reduce echoes. Conventional TV conference systems do not make us feel as if the other party were at the same table. This is because monaural audio does not help identify the talker position. For multi-channel audio, which is essential to a realistic TV conference system, a multi-channel echo canceller is necessary.

Two types of multi-channel echo cancellers based on either linear combination or cascade connection have been proposed[1]. In [1], cross-correlated input signals commonly encountered in multi-channel applications are not considered. However, cross-correlation in multi-channel signals is more critical for convergence of echo cancelers.

In this paper, convergence characteristics for a multi-channel echo canceller with cross-correlated input signals is analyzed. An analysis with cross-correlated input signals reveals two problems caused by disturbance in coefficient adaptation. A new echo canceller free from these problems is then derived. Finally, simulation results are presented to show the validity of the analysis and the performance of the new echo canceller.

2. MULTI-CHANNEL ECHO CANCELLER BASED ON LINEAR COMBINATION

Let us concentrate on the multi-channel echo canceller based on linear combination because it is superior to that based on cascade connection[1]. It consists of M^2 adaptive filters for M^2 echo paths from M loudspeakers to M microphones for an M channel case. Each adaptive filter estimates the impulse response of the corresponding echo path. Figure 1 shows a two-channel case.

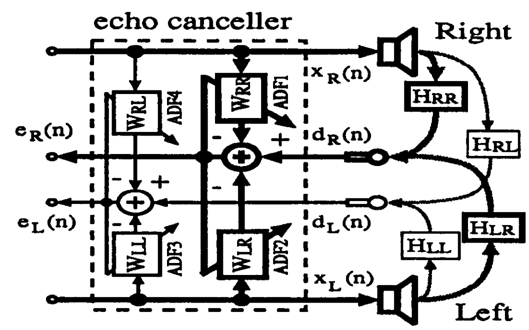


Figure 1. Two-channel echo canceller based on linear combination.

In the following discussions, for simplicity and symmetry, only the right channel echo in a two-channel case will be considered. The signal paths and the echo paths for the right channel are shown with bold lines in Figure 1. Obtained results on a two-channel case could easily be extended to a multi-channel case. The length of the echo paths and the number of the adaptive filter taps are both assumed to be N.

The echo is the sum of two signals which propagate from the left and the right loudspeakers to the microphone. The echo replica is generated by adding the outputs of the two adaptive filters ADF1 and ADF2. Let us define the echo $d_R(n)$ and the echo replica $\hat{d}_R(n)$ at time n as

$$d_R(n) = \boldsymbol{H}_{RR}^T \boldsymbol{X}_R(n) + \boldsymbol{H}_{LR}^T \boldsymbol{X}_L(n). \quad (1)$$

$$\hat{d}_R(n) = \boldsymbol{W}_{RR}^T(n) \boldsymbol{X}_R(n) - \boldsymbol{W}_{LR}^T(n) \boldsymbol{X}_L(n). \quad (2)$$

$\boldsymbol{X}_R(n)$ and $\boldsymbol{X}_L(n)$ are the input signal vectors which contain the input signals $x_R(n)$ and $x_L(n)$ for the right and the left channel.

$$\boldsymbol{X}_R(n) = \left[x_R(n) \cdots x_R(n-N+1)\right]^T \quad (3)$$

$$\boldsymbol{X}_L(n) = \left[x_L(n) \cdots x_L(n-N+1)\right]^T \quad (4)$$

$\boldsymbol{H}_{RR}$ and $\boldsymbol{H}_{LR}$ are the impulse response vectors of the echo paths from the right and the left loudspeaker to the right microphone.

$$\boldsymbol{H}_{RR} = \left[h_{RR,0} \cdots h_{RR,N-1}\right]^T \quad (5)$$

$$\boldsymbol{H}_{LR} = \left[h_{LR,0} \cdots h_{LR,N-1}\right]^T \quad (6)$$

$\boldsymbol{W}_{RR}(n)$ and $\boldsymbol{W}_{LR}(n)$ are the filter coefficient vectors of the adaptive filters ADF1 and ADF2 which estimate $\boldsymbol{H}_{RR}$ and $\boldsymbol{H}_{LR}$.

$$\boldsymbol{W}_{RR}(n) = \left[w_{RR,0}(n) \cdots w_{RR,N-1}(n)\right]^T \quad (7)$$

$$\boldsymbol{W}_{LR}(n) = \left[w_{LR,0}(n) \cdots w_{LR,N-1}(n)\right]^T \quad (8)$$

$\boldsymbol{A}^T$ denotes transpose of a matrix $\boldsymbol{A}$. The residual echo $e_R(n)$ is given by

$$e_R(n) = d_R(n) - \hat{d}_R(n). \quad (9)$$

The filter coefficient vectors $\boldsymbol{W}_{RR}(n)$ and $\boldsymbol{W}_{LR}(n)$ are adapted so that the residual echo $e_R(n)$ is minimized.

3. DISTURBING TERMS IN FILTER COEFFICIENT ADAPTATION

Assuming the LMS algorithm[2], the filter coefficient adaptation is described by

$$\boldsymbol{W}_{RR}(n+1) = \boldsymbol{W}_{RR}(n) + \mu e_R(n) \boldsymbol{X}_R(n) \quad (10)$$

$$\boldsymbol{W}_{LR}(n+1) = \boldsymbol{W}_{LR}(n) + \mu e_R(n) \boldsymbol{X}_L(n) \quad (11)$$

where a positive constant μ is the adaptation step size. By substituting (1), (2) and (9) into (10) and (11),

$$\begin{aligned} \boldsymbol{W}_{RR}(n+1) &= \boldsymbol{W}_{RR}(n) \\ &+ \mu \left\{ \boldsymbol{H}_{RR} - \boldsymbol{W}_{RR}(n) \right\}^T \boldsymbol{X}_R(n) \boldsymbol{X}_R(n) \\ &+ \mu \left\{ \boldsymbol{H}_{LR} - \boldsymbol{W}_{LR}(n) \right\}^T \boldsymbol{X}_L(n) \boldsymbol{X}_R(n) \quad (12) \end{aligned}$$

$$\begin{aligned} \boldsymbol{W}_{LR}(n+1) &= \boldsymbol{W}_{LR}(n) \\ &+ \mu \left\{ \boldsymbol{H}_{LR} - \boldsymbol{W}_{LR}(n) \right\}^T \boldsymbol{X}_L(n) \boldsymbol{X}_L(n) \\ &+ \mu \left\{ \boldsymbol{H}_{RR} - \boldsymbol{W}_{RR}(n) \right\}^T \boldsymbol{X}_R(n) \boldsymbol{X}_L(n) \quad (13) \end{aligned}$$

are obtained. In order to evaluate this adaptation, it is compared with that of a monaural echo canceller shown in Figure 2.

The filter coefficient adaptation by the LMS algorithm for the monaural echo canceller is

$$\begin{aligned} \boldsymbol{W}(n+1) &= \boldsymbol{W}(n) + \mu e(n) \boldsymbol{X}(n) \\ &= \boldsymbol{W}(n) \\ &+ \mu \left\{ \boldsymbol{H} - \boldsymbol{W}(n) \right\}^T \boldsymbol{X}(n) \boldsymbol{X}(n) \quad (14) \end{aligned}$$

where $\boldsymbol{X}(n)$, $\boldsymbol{H}$, $\boldsymbol{W}(n)$ and $e(n)$ are the input signal vector, the inpulse response vector of the echo path, and the filter coefficient vector of the adaptive filter which estimates $\boldsymbol{H}$, and the residual echo, respectively.

The filter coefficient adaptation for the monaural and the two-channel case are geometrically shown in Figure 3. In the monaural case, the filter coefficient vector $\boldsymbol{W}(n)$ is updated by a component of the error vector

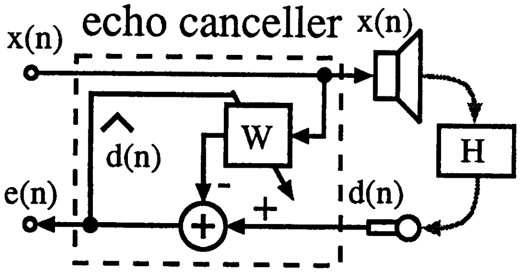
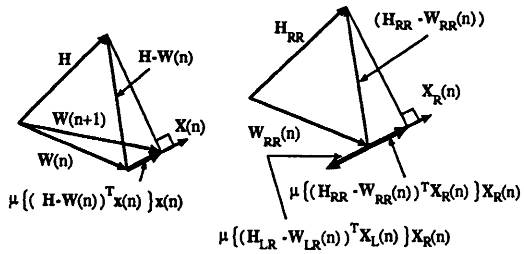


Figure 2. Monaural echo canceller.



(a) monaural (b) 2-channel

Figure 3. Comparison of filter coefficient adaptation.

$W(n)-H$ which is parallel to $X(n)$ as shown in Figure 3 (a). In the two-channel case in Figure 3 (b), the first terms of (12) and (13) update the filter coefficient vectors $W_{RR}(n)$ and $W_{LR}(n)$ in the same manner as the monaural case. However, the second terms disturb convergence of $W_{RR}(n)$ and $W_{LR}(n)$ because their direction might be either the same as or the opposite to the first term at random. The disturbance of the second terms will be examined in the following section.

4. DISTURBANCE BY THE CROSS-CORRELATION

By calculating an ensemble average, (12) and (13) become

$$E[W_{RR}(n+1)] = E[W_{RR}(n)] + \mu R_{RR}\left\{H_{RR} - E[W_{RR}(n)]\right\} + \mu R_{RL}\left\{H_{LR} - E[W_{LR}(n)]\right\} \quad (15)$$

$$E[W_{LR}(n+1)] = E[W_{LR}(n)] + \mu R_{LL}\left\{H_{LR} - E[W_{LR}(n)]\right\} + \mu R_{RL}^{T}\left\{H_{RR} - E[W_{RR}(n)]\right\} \quad (16)$$

where $E[A]$ denotes an ensemble average of A. R_{RR} and R_{LL} are autocorrelation matrices of the input signals, and R_{RL} is a cross-correlation matrix defined by

$$R_{RR} = E[X_R(n)X_R^{T}(n)] \quad (17)$$

$$R_{LL} = E[X_L(n)X_L^{T}(n)] \quad (18)$$

$$R_{RL} = E[X_R(n)X_L^{T}(n)]. \quad (19)$$

The coefficient adaptation in the two-channel case (15) and (16) are compared with the monaural case (14).

By calculating an ensemble average, (14) becomes

$$E[W(n+1)] = E[W(n)] + \mu R\left\{H - E[W(n)]\right\} \quad (20)$$

where R is an autocorrelation matrix of the input signal vector $X(n)$ defined by

$$R = E[X(n)X^{T}(n)]. \quad (21)$$

The first terms of (15) and (16) update $W_{RR}(n)$ and $W_{LR}(n)$ toward their optimum values H_{RR} and H_{LR} in the same manner as the monaural case. On the other hand, the second terms may update $W_{RR}(n)$ and $W_{LR}(n)$ in a wrong direction. If $x_R(n)$ and $x_L(n)$ are not cross-correlated, i.e. $R_{RL}=0$, the second terms of (15) and (16) are equal to 0, thus, the filter coefficients converge to their optimum values like the monaural case. However, if $x_R(n)$ and $x_L(n)$ are cross-correlated, the second terms disturb the convergence of $W_{RR}(n)$ and $W_{LR}(n)$ as additive noises.

5. AN EXAMPLE OF CROSS-CORRELATED INPUT SIGNALS

As an example of strongly cross-correlated signals, a stereo TV conference system shown in Figure 4 with a single talker in room A is considered. It could be assumed without loss of generality that the talker is closer to the right microphone than to the left. The results can easily be modified for the opposite case.

The input signals $x_R(n)$ and $x_L(n)$ are almost the same except for the differences in

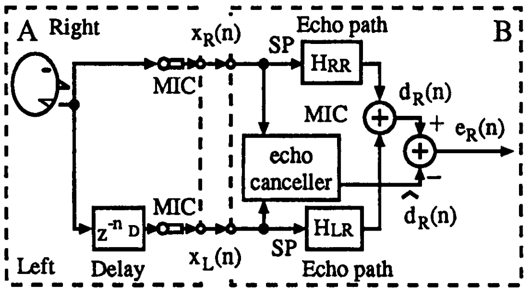


Figure 4. Two-channel TV conference system.

the delay and the power. Neglecting the difference of the power between the input signals since they are generally small, the relation between the input signals $x_R(n)$ and $x_L(n)$ is

$$x_L(n) = x_R(n - n_D) \qquad (22)$$

where a non-negative integer n_D is the relative delay between the input signals and is determined by the talker and the microphone positions.

By substituting (1) - (4) and (22) into (9), the residual echo $e_R(n)$ is given by

$$e_R(n) = \sum_{i=0}^{n_D-1} \left\{ h_{RR,i} - w_{RR,i}(n) \right\} x_R(n-i)$$
$$+ \sum_{i=n_D}^{N-1} \left\{ h_{RR,i} + h_{LR,i-n_D} - w_{RR,i}(n) - w_{LR,i-n_D}(n) \right\} x_R(n-i)$$
$$+ \sum_{i=N-n_D}^{N-1} \left\{ h_{LR,i} - w_{LR,i}(n) \right\} x_R(N-n_D-i). \qquad (23)$$

The condition that the echo $d_R(n)$ is completely canceled is determined by

$$w_{RR,i}(n) = h_{RR,i} \quad \text{for } i = 0, \cdots, n_D - 1 \qquad (24.1)$$
$$w_{LR,i}(n) = h_{LR,i} \quad \text{for } i = N - n_D, \cdots, N-1 \qquad (24.2)$$
$$w_{RR,i}(n) + w_{LR,i-n_D}(n) = h_{RR,i} + h_{LR,i-n_D} \quad \text{for } i = n_D, \cdots, N-1. \qquad (24.3)$$

From (24), the filter coefficients $w_{RR,0}(n), \cdots, w_{RR,n_D-1}(n)$ and $w_{LR,N-n_D}(n), \cdots, w_{LR,N-1}(n)$ have unique solutions. However, $w_{RR,n_D}(n), \cdots, w_{RR,N-1}(n)$ and $w_{LR,0}(n), \cdots, w_{LR,N-n_D-1}(n)$ do not have unique solutions and only the sum of two coefficients $w_{RR,i}(n) + w_{LR,i-n_D}(n)$ could be determined.

By substituting (23) into (10) and (11) and taking an ensemble average with an assumption of input-signal whiteness, the filter coefficient adaptation bocomes

$$E[w_{RR,i}(n+1)] = E[w_{RR,i}(n)] + \mu\sigma_X{}^2 \left\{ h_{RR,i} - E[w_{RR,i}(n)] \right\} \qquad (25.1)$$
$$E[w_{LR,N-n_D+i}(n+1)] = E[w_{LR,N-n_D+i}(n)] + \mu\sigma_X{}^2 \left\{ h_{LR,N-n_D+i} - E[w_{LR,N-n_D+i}(n)] \right\}$$
$$\text{for } i = 0, 1, \cdots, n_D - 1 \qquad (25.2)$$

and

$$E[w_{RR,i}(n+1)] = E[w_{RR,i}(n)] + \mu\sigma_X{}^2 \left\{ h_{RR,i} - E[w_{RR,i}(n)] \right\} + \mu\sigma_X{}^2 \left\{ h_{LR,i} - E[w_{LR,i}(n)] \right\} \qquad (26.1)$$
$$E[w_{LR,i-n_D}(n+1)] = E[w_{LR,i-n_D}(n)] + \mu\sigma_X{}^2 \left\{ h_{LR,i-n_D} - E[w_{LR,i-n_D}(n)] \right\} + \mu\sigma_X{}^2 \left\{ h_{RR,i} - E[w_{RR,i}(n)] \right\}$$
$$\text{for } i = n_D, n_D + 1, \cdots, N-1. \qquad (26.2)$$

$\sigma_X{}^2$ is the variance of the input signal $x_R(n)$. From (25.1) and (25.2), which are the same as the monaural LMS algorithm[2], $w_{RR,0}(n), \cdots, w_{RR,n_D-1}(n)$ and $w_{LR,N-n_D}(n), \cdots, w_{LR,N-1}(n)$ converge to their optimum values if and only if

$$0 \le \mu \le \frac{2}{\sigma_X{}^2}. \qquad (27)$$

By introducing new variables

$$w_1(n) = E[w_{RR,i}(n) - h_{RR,i}] + E[w_{LR,i-n_D}(n) - h_{LR,i-n_D}] \qquad (28)$$
$$w_2(n) = E[w_{RR,i}(n) - h_{RR,i}] - E[w_{LR,i-n_D}(n) - h_{LR,i-n_D}], \qquad (29)$$

(26.1) and (26.2) are reduced to a pair of new recursive equations with a single variable $w_1(n)$ or $w_2(n)$. These simple equations lead to the solutions to $w_{RR,n_D}(n), \cdots, w_{RR,N-1}(n)$ and $w_{LR,0}(n), \cdots, w_{LR,N-n_D-1}(n)$ as

$$E[w_{RR,i}(n)] = h_{RR,i} + \frac{1}{2} \left\{ (1 - 2\mu\sigma_X{}^2)^n + 1 \right\} \times \left\{ E[w_{RR,i}(0)] - h_{RR,i} \right\}$$
$$+ \frac{1}{2} \left\{ (1 - 2\mu\sigma_X{}^2)^n - 1 \right\} \times \left\{ E[w_{LR,i-n_D}(0)] - h_{LR,i-n_D} \right\} \qquad (30)$$

and

$$E[w_{LR,i-n_D}(n)] = h_{LR,i-n_D} + \frac{1}{2}\left\{(1-2\mu\sigma_X{}^2)^n + 1\right\} \times \left\{E[w_{LR,i-n_D}(0)] - h_{LR,i-n_D}\right\} + \frac{1}{2}\left\{(1-2\mu\sigma_X{}^2)^n - 1\right\} \times \left\{E[w_{RR,i}(0)] - h_{RR,i}\right\} \quad (31)$$

where $w_{RR,i}(0)$ and $w_{LR,i}(0)$ are the initial values of $w_{RR,i}(n)$ and $w_{LR,i}(n)$. (30) and (31) converge to

$$E[w_{RR,i}(\infty)] = h_{RR,i} + \frac{1}{2}\left\{E[w_{RR,i}(0)] - h_{RR,i}\right\} - \frac{1}{2}\left\{E[w_{LR,i-n_D}(0)] - h_{LR,i-n_D}\right\} \quad (32)$$

and

$$E[w_{LR,i-n_D}(\infty)] = h_{LR,i-n_D} + \frac{1}{2}\left\{E[w_{LR,i-n_D}(0)] - h_{LR,i-n_D}\right\} - \frac{1}{2}\left\{E[w_{RR,i}(0)] - h_{RR,i}\right\} \quad (33)$$

if and only if

$$0 < \mu < \frac{1}{\sigma_X{}^2}. \quad (34)$$

Two important things are observed from (32), (33) and (34); the final coefficient values and the convergence speed. (32) and (33) show that the filter coefficients $w_{RR,i}(n)$ and $w_{LR,i-n_D}(n)$ $(i = 1, \cdots, n_D - 1)$ $w_{RR,n_D}(n), \cdots, w_{RR,N-1}(n)$ and $w_{LR,0}(n), \cdots, w_{LR,N-n_D-1}(n)$ do not converge to their optimum values $h_{RR,i}$ and $h_{LR,i-n_D}$. The final values depend on the talker position because (32) and (33) contain n_D. Therefore, a move or a change of the talker has the same influence on the echo cancellation as a change of the echo path. (34) shows that convergence speed is slower than the monaural case because the adaptation step size μ is limited to the half range of (27).

6. A NEW MULTI CHANNEL ECHO CANCELLER

A new multi-channel echo canceller can be realized with a single adaptive filter per channel[4]. The new structure is based on an observation that in Figure 4, the signal path from the talker in room A to the microphone in room B, which consists of two paths connected in parallel, is equivalent to a single path. Therefore, the same echo replica as in Figure 1 can be generated by a single adaptive filter. A two-channel case of the new echo canceller is shown in Figure 5. The input signal which is ahead in phase is determined by a detector (DET) and used by all adaptive filters as their input signals[4].

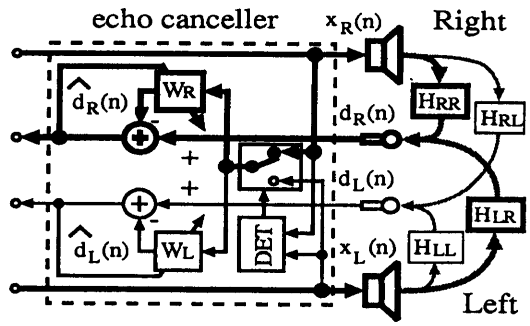


Figure 5. New two-channel echo canceller.

7. SIMULATION RESULTS

Simulations have been carried out in a two-channel and a single-talk case. Only the echo in the right channel is canceled because of the symmetry. 20-tap FIR filters whose impulse response is exponentially decreased are used as the echo paths. 40-tap FIR adaptive filters equipped with the normalized LMS algorithm[3] with the adaptation step size $\mu = 0.8$ are used for both the conventional and the new echo cancellers. The signal-to-noise ratio (SNR), which is defined by the ratio of the echo power to the additive noise power, is chosen as 40dB.

Figure 6 compares the convergence characteristics for cross-correlated and uncorrelated input signals. The cross-correlated signals are generated by adding two independent white Gaussian noises with different amplitude and delay for each channel. The uncorrelated signals and the additive noise to the echo are also white Gaussian noises. The convergence speed for cross-correlated signals is about 1/4 of that for the uncorrelated signals. The

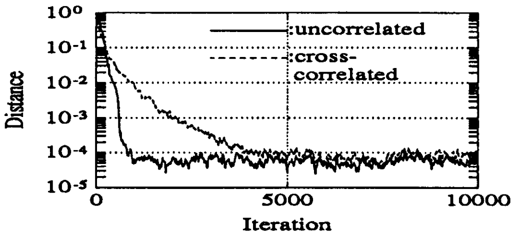


Figure 6. Convergence characteristics for cross-correlated and uncorrelated signals.

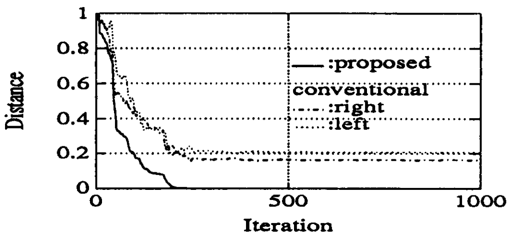


Figure 7. Distance from the optimum value.

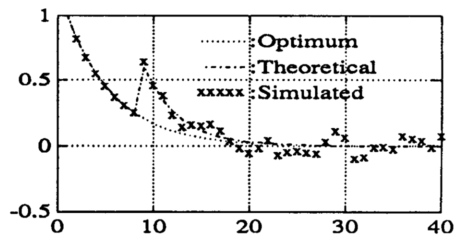


Figure 8. Coefficients after 1000 iterations.

distance $|W_{RR}(n) - H_{RR}|$ after convergence for cross-correlated signals is larger than that for the uncorrelated signals.

Figure 7 and 8 shows the convergence characteristics of the conventional and the new echo canceller with a single talker. Two independent white Gaussian noises are used as the input signal and the additive noise to the echo. The input signal in the left channel is delayed by 8 samples from the right channel. Figure 7 shows that the new echo canceller converges twice as fast as the conventional one, which agrees with (27) and (34). The filter coefficients of the new echo canceller after 1000 iterations are nearly equal to their optimum values. Figure 8 illustrates the filter coefficients $W_{RR}(1000)$ of the conventional echo canceller. They converge to those given by (32) rather than their optimum values H_{RR}.

8. CONCLUSION

Analysis of convergence characteristics for a multi-channel echo canceller with cross-correlated input signals have been presented. The gradient with respect to the filter coefficients of the conventional multi-channel echo canceller contain disturbing components. These components behave like additive noises when the input signals have strong cross-correlation. With an example of strongly cross-correlated signals, it has been analytically shown that the convergence speed is slower than that for the uncorrelated signals. The filter coefficients do not converge to their optimum values either. A new multi-channel echo canceller, which converges twice as fast as the conventional one for cross-correlated signals, has been derived. Its filter coefficients converge to their optimum values while those of the conventional echo canceller do not.

ACKNOWLEDGMENT

The authors wish to thank Dr. Takao Nishitani, the manager of Terminal Systems Research Laboratory, C&C Systems Research Laboratories, NEC Corporation, for his helpful comments and continuous encouragement.

REFERENCES

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

[2] B. Widrow and S. D. Stearns:"Adaptive Signal Processing" ,Chapter 6, pp. 99-116, Prentice-Hall,NJ,1985.

[3] G. C. Goodwin et al., "Adaptive Filtering, Prediction and Control," Englewood Cliffs, NJ: Prentice-Hall Info. Syst. Sci. Ser., 1985.

[4] A. Hirano and A. Sugiyama, "A New Multi-Channel Echo Canceller with a Single Adaptive Filter per Channel," National Convention Record of IEICE, pp. 1-202, Mar. 1991 (in Japanese).