PAPER Special Section on Digital Signal Processing

Alternative Learning Algorithm for Stereophonic Acoustic Echo Canceller without Pre-Processing

Akihiro HIRANO[†a)], **Kenji NAKAYAMA**[†], ***Members***, **Daisuke SOMEDA**[†*], ***and*** **Masahiko TANAKA**[†**], ***Nonmembers***

SUMMARY This paper proposes an alternative learning algorithm for a stereophonic acoustic echo canceller without pre-processing which can identify the correct echo-paths. By dividing the filter coefficients into the former/latter parts and updating them alternatively, conditions both for unique solution and for perfect echo cancellation are satisfied. The learning for each part is switched from one part to the other when that part converges. Convergence analysis clarifies the condition for correct echo-path identification. For fast and stable convergence, a convergence detection and an adaptive step-size are introduced. The modification amount of the filter coefficients determines the convergence state and the step-size. Computer simulations show 10 dB smaller filter coefficient error than those of the conventional algorithms without pre-processing.

key words: *stereophonic acoustic echo canceller, pre-processing*

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 (SAEC's) [1]–[13] have been studied.

SAEC's have a fundamental problem in which their filter coefficients cannot have a unique solution [1]–[5]. Though SAEC's with pre-processing [6]–[10] are good candidates for solving this problem, audible sound distortion caused by the pre-processing arises. An SAEC without pre-processing, XM-NLMS algorithm [11], has also been proposed. Though the XM-NLMS converges faster than a standard SAEC [1], its convergence at the optimum coefficient is not confirmed.

This paper proposes a stereophonic acoustic echo canceller without pre-processing. Section 2 reviews the SAEC and its fundamental problem. An SAEC without pre-processing, its convergence analysis and adaptation control are presented in Sect. 3. Computer simulation results show the performance of the proposed algorithm.

2. Uniqueness Condition for Stereophonic Acoustic Echo Canceller

Figure 1 shows a teleconferencing using an SAEC. 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.

The far-end signal $\boldsymbol{x}_i(n)$ in the i-th channel at time index n is generated from a talker speech $s(n)$ by passing room A impulse response $\boldsymbol{g}_i$ from the talker to the i-th microphone. $\boldsymbol{x}_i(n)$ passes an echo path $\boldsymbol{h}_{i,j}$ from the i-th loudspeaker to the j-th microphone and become an echo $d_j(n)$. Similarly, adaptive filters $\boldsymbol{w}_{i,j}(n)$ generates an echo replica $y_j(n)$. $\boldsymbol{w}_{i,j}(n)$ is so updated as to reduce the residual echo $e_j(n)$

SAEC's have a fundamental problem in which their filter coefficients cannot have a unique solution [1]. SAEC's may have infinite number of solutions other than the optimum solution $\boldsymbol{w}_{i,j}(n) = \boldsymbol{h}_{i,j}$.

Further analyses show that SAEC's may have a unique and optimum solution when the number of taps N_W for SAEC and the impulse response length N_A in room A satisfy $N_W < N_A$ [5], [10]. For echo cancellation performance, $N_B \le N_W$ is preferable where N_B is the impulse response length in room B. Therefore, if $N_B \le N_W < N_A$, SAEC in room B achieves both perfect echo cancellation and optimum solution. Such a condition, however, cannot be satisfied for SAEC's in both room A and B.

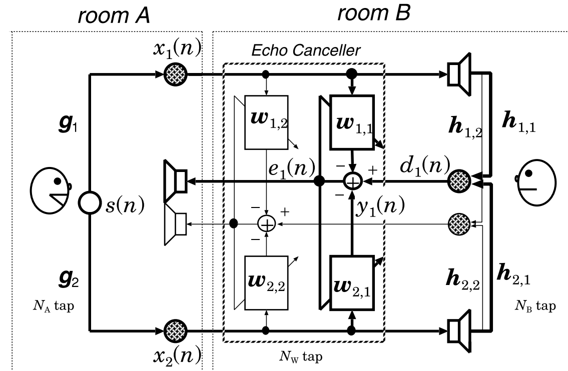


Fig. 1 Teleconferencing using SAEC.

Manuscript received November 28, 2003.
Manuscript revised March 1, 2004.
Final manuscript received April 15, 2004.

[†]The authors are with the Division of Electrical Engineering and Computer Science, Graduate School of Natural Science and Technology, Kanazawa University, Kanazawa-shi, 920-8667 Japan.
[*]Presently, with the Designg Department No. 4, Electronics Engineering Division, Electronics Business Group, Tokairtika Co., Ltd.
[**]Presently, with the Second Department, Financial Systems Division, Hitachi Software Engineering Co., Ltd.
a) E-mail: hirano@t.kanazawa-u.ac.jp

3. Correct Echo-Path Identification without Pre-Processing

3.1 Algorithm

In order to satisfy the uniqueness condition for both SAEC's in room A and room B, the number of taps for SAEC N_W is so chosen as to satisfy $N_W/2 < N_A \leq N_W$ and $N_W/2 < N_B \leq N_W$. If the size of both rooms are similar, which is usual case, such N_W may exist. In adaptation, $N_W/2$ taps are updated at a time; thus the effective number of taps for SAEC $N_W/2$ is smaller than the impulse response length in the far-end room N_A. To avoid the performance degradation caused by the tap shortage, another $N_W/2$ taps will also update at the other time.

The filter coefficient vector $\boldsymbol{w}_{i,j}(n)$ is divided into two sub-vectors $\boldsymbol{w}_{i,j,f}(n)$ and $\boldsymbol{w}_{i,j,b}(n)$ shown by

$$\boldsymbol{w}_{i,j,f}(n) = [w_{i,j,0}(n), \cdots, w_{i,j,N_W/2-1}(n)]^T \tag{1}$$

$$\boldsymbol{w}_{i,j,b}(n) = [w_{i,j,N_W/2}(n), \cdots, w_{i,j,N_W-1}(n)]^T. \tag{2}$$

In the first stage, $\boldsymbol{w}_{i,j,f}(n)$ is updated while $\boldsymbol{w}_{i,j,b}(n)$ is fixed. This stage is repeated until $\boldsymbol{w}_{i,j,f}(n)$ converges. As the second stage, $\boldsymbol{w}_{i,j,b}(n)$ is updated while $\boldsymbol{w}_{i,j,f}(n)$ is fixed. This stage is also repeated until $\boldsymbol{w}_{i,j,b}(n)$ converges. These two stages are repeated one after another.

3.2 Convergence Analysis

Convergence of the averaged filter coefficients has been analyzed. First, convergence of $\boldsymbol{w}_{i,j,f}(n)$ when $\boldsymbol{w}_{i,j,b}(n)$ is fixed as an initial value $\boldsymbol{w}_{i,j,b}(0)$ is analyzed. Convergence of $\boldsymbol{w}_{i,j,b}(n)$ can be derived in the same manner as $\boldsymbol{w}_{i,j,b}(n)$.

The far-end signal on i-th channel $x_i(n)$ is derived as

$$x_i(n) = \boldsymbol{g}_i^T \boldsymbol{s}(n) \tag{3}$$

where the talker speech vector $\boldsymbol{s}(n)$ and the impulse response vector $\boldsymbol{g}_i$ are defined by

$$\boldsymbol{g}_i = [g_{i,0}, g_{i,1}, \cdots, g_{i,N_A-1}]^T \tag{4}$$

$$\boldsymbol{s}(n) = [s(n), \cdots, s(n-N_A+1)]^T. \tag{5}$$

$[]^T$ denotes the transpose of []. The echo $d_j(n)$ and the echo replica $y_j(n)$ are calculated as

$$d_j(n) = \sum_{i=1}^{2}\{\boldsymbol{h}_{i,j,f}^T \boldsymbol{x}_{i,f}(n) + \boldsymbol{h}_{i,j,b}^T \boldsymbol{x}_{i,b}(n)\} \tag{6}$$

$$y_j(n) = \sum_{i=1}^{2}\{\boldsymbol{w}_{i,j,f}^T(n) \boldsymbol{x}_{i,f}(n) + \boldsymbol{w}_{i,j,b}^T(0) \boldsymbol{x}_{i,b}(n)\}. \tag{7}$$

$\boldsymbol{h}_{i,j,f}$, $\boldsymbol{h}_{i,j,b}$, $\boldsymbol{x}_{i,f}(n)$ and $\boldsymbol{x}_{i,b}(n)$, are defined as

$$\boldsymbol{h}_{i,j,f} = [h_{i,j,0}, \cdots, h_{i,j,N_W/2-1}]^T \tag{8}$$

$$\boldsymbol{h}_{i,j,b} = [h_{i,j,N_W/2}, \cdots, h_{i,j,N_W}]^T \tag{9}$$

$$\boldsymbol{x}_{i,f}(n) = [x_i(n), \cdots, x_i(n-N_W/2+1)]^T \tag{10}$$

$$\boldsymbol{x}_{i,b}(n) = [x_i(n-N_W/2), \cdots, x_i(n-N_W)]^T, \tag{11}$$

which are sub-vectors of $\boldsymbol{h}_{i,j}$ and $\boldsymbol{x}_j(n)$.

By using (3), the residual echo $e_j(n)$ is calculated by

$$\begin{aligned} e_j(n) = & \sum_{i=1}^{2}\{\boldsymbol{h}_{i,j,f} - \boldsymbol{w}_{i,j,f}(n)\}^T \boldsymbol{G}_{i,f} \boldsymbol{s}_f(n) \\ & + \sum_{i=1}^{2}\{\boldsymbol{h}_{i,j,b} - \boldsymbol{w}_{i,j,b}(0)\}^T \boldsymbol{G}_{i,b} \boldsymbol{s}_b(n). \end{aligned} \tag{12}$$

$\boldsymbol{s}_f(n), \boldsymbol{s}_b(n-N_W)$ are defined by

$$\boldsymbol{s}_{i,f}(n) = [s(n), \cdots, s(n-N_W/2-N_A+1)] \tag{13}$$

$$\boldsymbol{s}_{i,b}(n) = [s(n-N_W/2), \cdots, s(n-N_W-N_A+1)]. \tag{14}$$

$\boldsymbol{G}_{i,p}$ is a matrix defined by (15), which contains $\boldsymbol{g}_i$ and performs convolution between $\boldsymbol{s}_{i,p}(n)$ and $\boldsymbol{g}_{i,p}$, where p is either b or f. By introducing vectors and matrices defined by

$$\boldsymbol{d}_f(n) = \begin{bmatrix} \boldsymbol{h}_{1,j,f} - \boldsymbol{w}_{1,j,f}(n) \\ \boldsymbol{h}_{2,j,f} - \boldsymbol{w}_{2,j,f}(n) \end{bmatrix} \tag{16}$$

$$\boldsymbol{d}_b(n) = \begin{bmatrix} \boldsymbol{h}_{1,j,b} - \boldsymbol{w}_{1,j,b}(n) \\ \boldsymbol{h}_{2,j,b} - \boldsymbol{w}_{2,j,b}(n) \end{bmatrix} \tag{17}$$

$$\boldsymbol{G}_f = \begin{bmatrix} \boldsymbol{G}_{1,f} \\ \boldsymbol{G}_{2,f} \end{bmatrix} \tag{18}$$

$$\boldsymbol{G}_b = \begin{bmatrix} \boldsymbol{G}_{1,b} \\ \boldsymbol{G}_{2,b} \end{bmatrix}, \tag{19}$$

simplified result for $e_j(n)$, i.e.,

$$e_j(n) = \boldsymbol{d}_f^T(n) \boldsymbol{G}_f \boldsymbol{s}_f(n) + \boldsymbol{d}_b^T(0) \boldsymbol{G}_b \boldsymbol{s}_b(n) \tag{20}$$

is derived.

Taking an ensemble average of $e_j(n)$ leads us to

$$\begin{aligned} E[e_j^2(n)] &= E[e_j(n) e_j^T(n)] \\ &= \overline{\boldsymbol{d}_f^T(n)} \boldsymbol{Q}_f \overline{\boldsymbol{d}_f(n)} + \boldsymbol{d}_b^T(0) \boldsymbol{Q}_b \boldsymbol{d}_b(0) \\ &\quad + 2\overline{\boldsymbol{d}_f^T(n)} \boldsymbol{Q}_{fb} \boldsymbol{d}_b(0) \end{aligned} \tag{21}$$

where

$$\boldsymbol{R}_f = \overline{\boldsymbol{s}_f(n) \boldsymbol{s}_f^T(n)} \tag{22}$$

$$\boldsymbol{R}_{fb} = \overline{\boldsymbol{s}_f(n) \boldsymbol{s}_b^T(n)} \tag{23}$$

$$\boldsymbol{R}_b = \overline{\boldsymbol{s}_b(n) \boldsymbol{s}_b^T(n)} \tag{24}$$

$$\boldsymbol{Q}_f = \boldsymbol{G}_f \boldsymbol{R}_f \boldsymbol{G}_f^T \tag{25}$$

$$\boldsymbol{Q}_{fb} = \boldsymbol{G}_f \boldsymbol{R}_{fb} \boldsymbol{G}_b^T \tag{26}$$

$$\boldsymbol{Q}_b = \boldsymbol{G}_b \boldsymbol{R}_b \boldsymbol{G}_b^T \tag{27}$$

and $\overline{\boldsymbol{d}_f(n)}$ denotes an average of $\boldsymbol{d}_f(n)$. From

$$\frac{\partial\, E[e_j^2(n)]}{\partial \overline{\boldsymbol{d}_f(n)}} = 2\boldsymbol{Q}_f \overline{\boldsymbol{d}_f(n)} + 2\boldsymbol{Q}_{fb} \boldsymbol{d}_b(0) = 0, \tag{28}$$

the averaged filter coefficient error $\overline{\boldsymbol{d}_f^*}$ which minimizes

$$
\boldsymbol{G}_{i,p} = \underbrace{\begin{bmatrix} g_{i,p,0} & g_{i,p,1} & \cdots & g_{i,p,N_A-1} & 0 & \cdots & \cdots & 0 \\ 0 & g_{i,p,0} & g_{i,p,1} & \cdots & g_{i,p,N_A-1} & \ddots & \vdots & \vdots \\ \vdots & 0 & \ddots & \ddots & \cdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \cdots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & g_{i,p,0} & g_{i,p,1} & \cdots & g_{i,p,N_A-1} \end{bmatrix}}_{N_W/2+N_A-1} \left.\vphantom{\begin{bmatrix} 0\\0\\0\\0\\0 \end{bmatrix}}\right\} N_W/2 \tag{15}
$$

$p = f$ or b

$E[e_j^2(n)]$ is determined as

$$\overline{\boldsymbol{d}_f^*} = -\boldsymbol{Q}_f^{-1}\boldsymbol{Q}_{fb}\boldsymbol{d}_b(0). \tag{29}$$

Therefore, the filter coefficients will converge on

$$\begin{bmatrix} \overline{\boldsymbol{w}_{1,j,f}(\infty)} \\ \overline{\boldsymbol{w}_{2,j,f}(\infty)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{h}_{1,j,f} \\ \boldsymbol{h}_{2,j,f} \end{bmatrix} + \boldsymbol{Q}_f^{-1}\boldsymbol{Q}_{fb}\begin{bmatrix} \boldsymbol{h}_{1,j,b} - \boldsymbol{w}_{1,j,b}(0) \\ \boldsymbol{h}_{2,j,b} - \boldsymbol{w}_{2,j,b}(0) \end{bmatrix}. \tag{30}$$

The first term in the right hand side is the optimum solution. The second term is the error caused by the tap shortage. Results for $\boldsymbol{w}_{i,j,b}(n)$ becomes

$$\overline{\boldsymbol{d}_b^*} = -\boldsymbol{Q}_b^{-1}\boldsymbol{Q}_{fb}^T\boldsymbol{d}_f(0) \tag{31}$$

which can be derived by the same manner as for $\boldsymbol{w}_{i,j,f}(n)$.

By updating $\boldsymbol{w}_{i,j,f}(n)$ until convergence, the coefficient error $\boldsymbol{d}_f^{(1)}$ for $\boldsymbol{w}_{i,j,f}(n)$ converges at

$$\boldsymbol{d}_f^{(1)} = -\boldsymbol{K}_f\boldsymbol{d}_b^{(0)} \tag{32}$$

where $\boldsymbol{K}_f$ is defined by

$$\boldsymbol{K}_f = \boldsymbol{Q}_f^{-1}\boldsymbol{Q}_{fb}. \tag{33}$$

Then updating $\boldsymbol{w}_{i,j,b}(n)$ with the initial coefficient error $\boldsymbol{d}_f^{(1)}$ for $\boldsymbol{w}_{i,j,f}(n)$, the coefficient error $\boldsymbol{d}_b^{(1)}$ for $\boldsymbol{w}_{i,j,b}(n)$ converges at

$$\boldsymbol{d}_b^{(1)} = -\boldsymbol{K}_b\boldsymbol{d}_f^{(1)} = \boldsymbol{K}_b\boldsymbol{K}_f\boldsymbol{d}_b^{(0)} \tag{34}$$

where $\boldsymbol{K}_b$ is defined by

$$\boldsymbol{K}_b = \boldsymbol{Q}_b^{-1}\boldsymbol{Q}_{fb}^T. \tag{35}$$

By repeating these processes once more, $\boldsymbol{d}_f^{(2)}$ and $\boldsymbol{d}_b^{(2)}$ are calculated as

$$\boldsymbol{d}_f^{(2)} = -\boldsymbol{K}_f\boldsymbol{d}_b^{(1)} = -\boldsymbol{K}_f\boldsymbol{K}_b\boldsymbol{K}_f\boldsymbol{d}_b^{(0)} \tag{36}$$

$$\boldsymbol{d}_b^{(2)} = -\boldsymbol{K}_b\boldsymbol{d}_f^{(2)} = (\boldsymbol{K}_b\boldsymbol{K}_f)^2\boldsymbol{d}_b^{(0)}. \tag{37}$$

Finally,

$$\boldsymbol{d}_f^{(m)} = -\boldsymbol{K}_f(\boldsymbol{K}_b\boldsymbol{K}_f)^{m-1}\boldsymbol{d}_b^{(0)} \tag{38}$$

$$\boldsymbol{d}_b^{(m)} = (\boldsymbol{K}_b\boldsymbol{K}_f)^m\boldsymbol{d}_b^{(0)} \tag{39}$$

are obtained.

The conditions in which the filter coefficients converge at the optimum value are the followings.

1. The inverse matrices $\boldsymbol{Q}_f^{-1}$ and $\boldsymbol{Q}_b^{-1}$ exist.
2. The maximum absolute eigenvalue of $\boldsymbol{K}_b\boldsymbol{K}_f$ is less than 1.

The first condition is similar to the uniqueness condition discussed in [5], [10]. To satisfy this, the matrices $\boldsymbol{R}_f$ and $\boldsymbol{R}_b$ should have their inverse. Also, $\boldsymbol{G}_f$ and $\boldsymbol{G}_b$ should have pseudo-inverse. $\boldsymbol{R}_f^{-1}$ and $\boldsymbol{R}_b^{-1}$ always exist because $\boldsymbol{R}_f$ and $\boldsymbol{R}_b$ are the auto-correlation matrices.

A neccesary condition for existence of pseudo-inverse is $N_W/2 < N_A$, which is assumed in this algorithm. The neccesary and sufficient condition is $rank\boldsymbol{G}_f = rank\boldsymbol{G}_b = N_W/2$. This condition might hold if $\boldsymbol{g}_1$ and $\boldsymbol{g}_2$ sufficiently differ.

In order to examine the condition on eigenvalue of $\boldsymbol{K}_b\boldsymbol{K}_f$, let up rewite the matrices $\boldsymbol{Q}_f$, $\boldsymbol{Q}_b$ and $\boldsymbol{Q}_{fb}$. By introducing vectors

$$\boldsymbol{x}_f(n) = \begin{bmatrix} \boldsymbol{x}_{1,f} \\ \boldsymbol{x}_{2,f} \end{bmatrix} \tag{40}$$

$$\boldsymbol{x}_b(n) = \begin{bmatrix} \boldsymbol{x}_{1,b} \\ \boldsymbol{x}_{2,b} \end{bmatrix}, \tag{41}$$

$\boldsymbol{Q}_f$, $\boldsymbol{Q}_b$ and $\boldsymbol{Q}_{fb}$ becomes

$$\boldsymbol{Q}_f = \boldsymbol{G}_f\boldsymbol{R}_f\boldsymbol{G}_f^T = E[\boldsymbol{x}_f(n)\boldsymbol{x}_f^T(n)] \tag{42}$$

$$\boldsymbol{Q}_b = \boldsymbol{G}_b\boldsymbol{R}_b\boldsymbol{G}_b^T = E[\boldsymbol{x}_b(n)\boldsymbol{x}_b^T(n)] \tag{43}$$

$$\boldsymbol{Q}_{fb} = \boldsymbol{G}_f\boldsymbol{R}_{fb}\boldsymbol{G}_b^T = E[\boldsymbol{x}_f(n)\boldsymbol{x}_b^T(n)]. \tag{44}$$

The matrices $\boldsymbol{Q}_f$ and $\boldsymbol{Q}_b$ are the auto-correlation of $\boldsymbol{x}_f(n)$ and $\boldsymbol{x}_b(n)$, respectively. Similarly, $\boldsymbol{Q}_{fb}$ is the cross-correlation between $\boldsymbol{x}_f(n)$ and $\boldsymbol{x}_b(n)$. Note that these matrices include inter-channel correlation.

Using (42), (43) and (44), $\boldsymbol{K}_b\boldsymbol{K}_f$ becomes

$$\begin{aligned} \boldsymbol{K}_b\boldsymbol{K}_f &= \boldsymbol{Q}_b^{-1}\boldsymbol{Q}_{fb}\boldsymbol{Q}_f^{-1}\boldsymbol{Q}_{fb} \\ &= E[\boldsymbol{x}_b(n)\boldsymbol{x}_b^T(n)]^{-1}E[\boldsymbol{x}_f(n)\boldsymbol{x}_b^T(n)] \\ &\quad \cdot E[\boldsymbol{x}_f(n)\boldsymbol{x}_f^T(n)]^{-1}E[\boldsymbol{x}_f(n)\boldsymbol{x}_b^T(n)]. \end{aligned} \tag{45}$$

Thus, $\boldsymbol{K}_b\boldsymbol{K}_f$ is square of the cross-correlation $\boldsymbol{Q}_{fb}$ multiplied by the inverse of the auto-correlation $\boldsymbol{Q}_f$ and $\boldsymbol{Q}_b$. Intuitively, the eigenvalues of $\boldsymbol{K}_b\boldsymbol{K}_f$ might not be so large bacause the auto-correlation is larger than the cross-correlation.

3.3 Adaptation Control

In this approach, the adaptation control is important. To identify the echo path, each stage should be terminated when the filter coefficients converge. This causes two problems: selection of the switching interval between two stages and a random walk around the convergence value.

An adaptive step-size and a convergence detection are introduced for fast convergence with a small computational cost. The adaptive step-size and the convergence detection are carried out based on the coefficient modification amount defined by

$$D_j(m) = \frac{\sum_{i=1}^{2} ||\boldsymbol{w}_{i,j,p}(mK) - \boldsymbol{w}_{i,j,p}((m-1)K)||^2}{\sum_{i=1}^{2} ||\boldsymbol{w}_{i,j,p}(mK)||^2} \tag{46}$$

where i and j are the channel index for the loudspeaker and the microphone in the near-end room, respectively. p is either f or b. To avoid the increase of the computational cost, (46) is calculated once in a K iterations. Coefficient adaptation is stopped when (46) is calculated.

Figure 2 demonstrates an example of coefficient modification amount $D_j(m)$. $D_j(m)$ decreases until the filter coefficients converge. After convergence, $D_j(m)$ walks up and down caused by a random walk around the convergence value. Therefore, the filter coefficients are considered to be converged if $D_j(m-1) < D_j(m)$ is satisfied. The step-size is controlled by

$$\mu_j(m) = \mu_{max} f\left(\frac{D_j(m)}{D_{j,max}}\right) \tag{47}$$

where $D_{j,max}$ is the maximum value of $D_j(m)$ in the same stage. Usually, $D_j(1)$ is used as the $D_{j,max}$. $\mu_j(n)$ is used within $mK < n < (m+1)K$. $f(x)$ is a monotonically increasing function of x. An example is

$$\mu_j(m) = \mu_{max}\left(\frac{D_j(m)}{D_{j,max}}\right)^{\alpha} \tag{48}$$

where α is a constant. The step-size $\mu_j(m)$ in (48) is used for both $\boldsymbol{w}_{i,j,f}$ and $\boldsymbol{w}_{i,j,b}$.

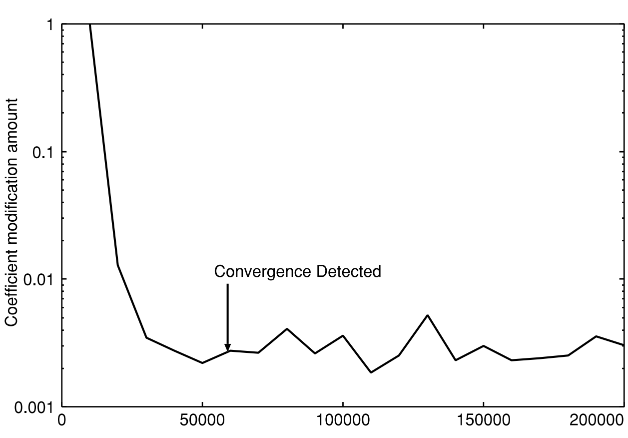


Fig. 2 An example of coefficient modification amount.

An overview of the adaptation control is as follows:

1. Update filter coefficients with $\mu_j(0) = \mu_{max}$ for the first K iterations.
2. Calculate $D_j(1)$. $D_{j,max} = D_j(1)$.
3. Update filter coefficients with $\mu_j(m)$ controlled by (48) for the next K iterations.
4. Calculate $D_j(m)$.
5. If $D_j(m-1) < D_j(m)$, then proceed to the next stage.
6. If $D_{j,max} < D_j(m)$, then $D_{j,max} = D_j(m)$.
7. Goto 3.

4. Computer Simulations

Simulations have been carried out to show the performance of the proposed algorithm. The simulation conditions and parameters in Table 1 are used, except for specially noted cases. Far-end room impulse responses $\boldsymbol{g}_i$ are 60-tap FIR filters ($N_A = 60$) while those for near-end room $\boldsymbol{h}_{i,j}$ are 64-tap FIR filters ($N_B = 64$). In this case, SAEC's do not have an unique solution. Adaptive filters are 64-tap FIR filters ($N_W = 64$). In this case, the condition for echo-path identification $N_W/2 < N_A \leq N_B$ is satisfied. White Gaussian signals are used as a talker signal and an additive noise. The echo-to-noise ratio (ENR) is 60 dB.

As an adaptation algorithm, Normalized Least Mean Squares (NLMS) algorithm [14] is used. The proposed algorithm is compared with the standard SAEC [1] and the XM-NLMS algorithm [11]. For XM-NLMS, smaller step-size is also used because of the stability. As a performance measure, the normalized coefficient error (NCE) and the echo return loss enhancement (ERLE) defined by

$$NCE(n) = \frac{\sum_{i=1}^{2} ||\boldsymbol{w}_{i,j}(n) - \boldsymbol{h}_{i,j}||^2}{\sum_{i=1}^{2} ||\boldsymbol{h}_{i,j}||^2} \tag{49}$$

$$ERLE(n) = \frac{\sum_{k=0}^{9999} d_j^2(n-k)}{\sum_{k=0}^{9999} e_j^2(n-k)} \tag{50}$$

are used. The ERLE is time-averaged over 10000 samples.

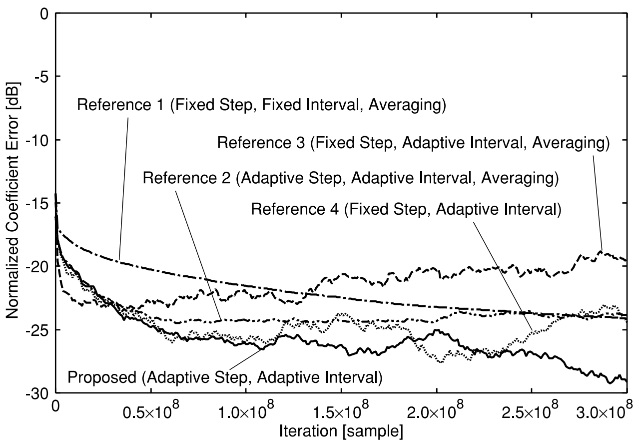
To show the effect of the adaptive step-size and adaptive interval, several combinations are compared in addition

Table 1 Simulation conditions.

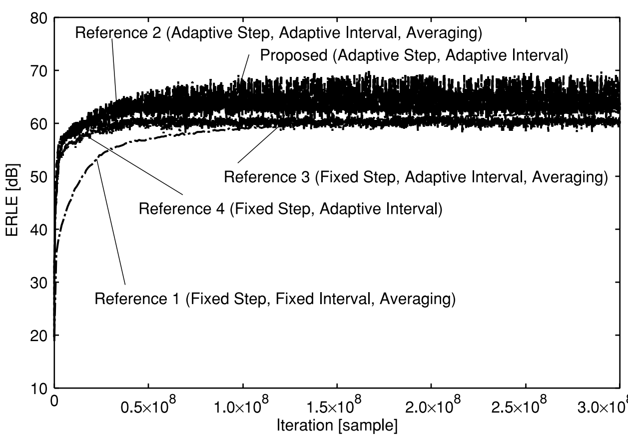
Parameters	
N_A	60
N_B, N_W	64
Fixed switching interval	20000
Averaging interval	10000
K	4000
Adaptation algorithm	NLMS
Fixed step-size	1.0
Variable step-size	$\mu_{max}\left(\frac{D_j(m)}{D_{j,max}}\right)^{1/4}$
μ_{max}	1.0
Far-end talker signal $s(n)$	White Gaussian
Additive noise	White Gaussian, ENR=60 dB

Table 2 Adaptation control methods.

Name	Step-size	Interval	Averaging	K
Proposed	Adaptive	Adaptive	No	4000
Reference 1	Fixed	Fixed	Yes	5000
Reference 2	Adaptive	Adaptive	Yes	5000
Reference 3	Fixed	Adaptive	Yes	5000
Reference 4	Fixed	Adaptive	No	4000



(a) NCE

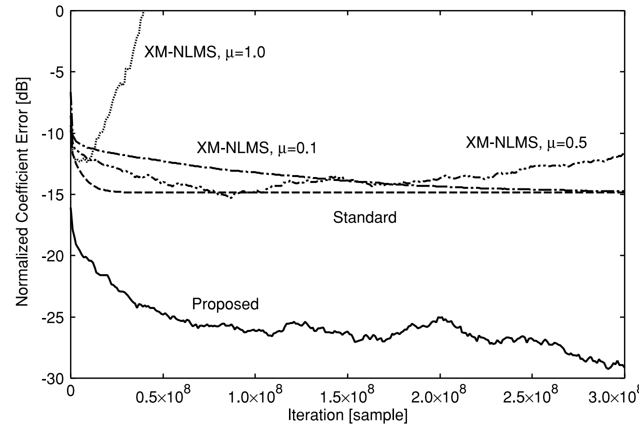


(b) ERLE

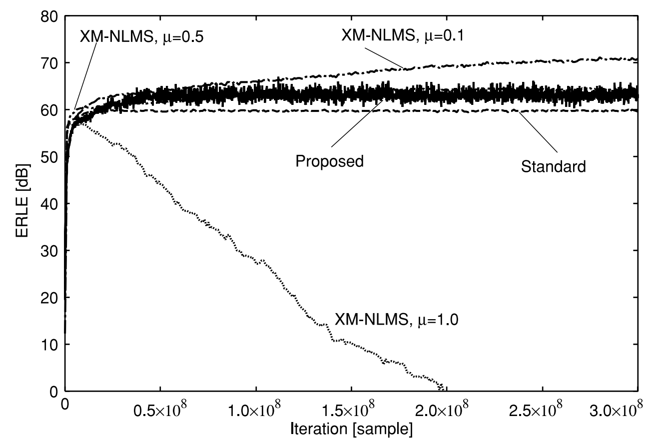
Fig. 3 NCE and ERLE for some adaptation control methods.

to the proposed method. Adaptation control methods are shown in Table 2. In some combinations, time-averaging of the filter coefficients is introduced in order to supress an influence of a random walk around the convergence value.

The NCE is depicted in Fig. 3(a). The proposed algorithm converges to −20 dB of the NCE almost 10 times faster than "Reference 1" with fixed step-size and fixed interval. Since the convergence characteristics of the proposed algorithm is better than "Reference 2," the averaging is not required. Comparison of the proposed algorithm with "Reference 3" and "Reference 4" shows the advantage of the adaptive step-size. Without the adaptive step-size, the NCE becomes large even if the averaging is used. Though the NCE for the proposed method reaches −16 dB within 10000 samples, its convergence to the final value requires very long time.



(a) NCE



(b) ERLE

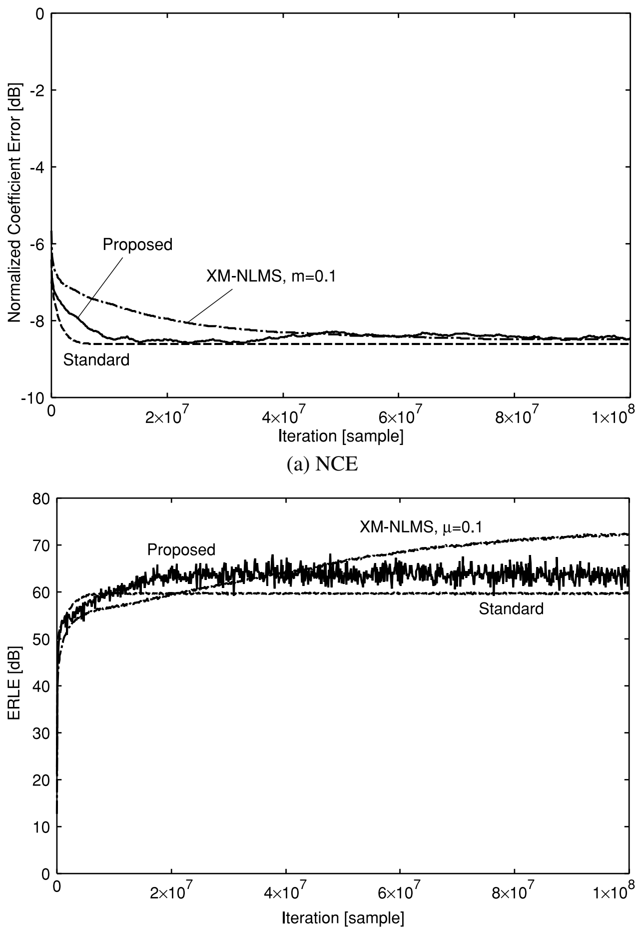
Fig. 4 NCE and ERLE for some algorithms witout pre-processing.

The ERLE is compared by 3(b). Higher ERLE is achieved by the proposed and "Reference 2" algorithms with the adaptive step-size, compared with those with the fixed step-size. "Reference 1" with fixed interval converges slowest.

Figure 4(a) compares the NCE for some adaptation algorithms without pre-processing. The proposed algorithm achieves −28 dB of the NCE which is almost 10 dB smaller than the standard SAEC. The XM-NLMS failed to converge at the optimum value for larger step-size. Though the XM-NLMS with $\mu = 0.1$ converges, the NCE is almost same as that for the standard SAEC. The convergence speed of XM-NLMS with $\mu = 0.1$ is slow because of a small step-size.

Figure 4(b) shows the ERLE for some adaptation algorithms without pre-processing. The proposed algorithm, the standard SAEC, and the XM-NLMS with $\mu = 0.5$ achieve almost the same ERLE, almost 60 dB. The ELRE for the XM-NLMS with $\mu = 0.1$ is slightly larger because of the smaller step-size.

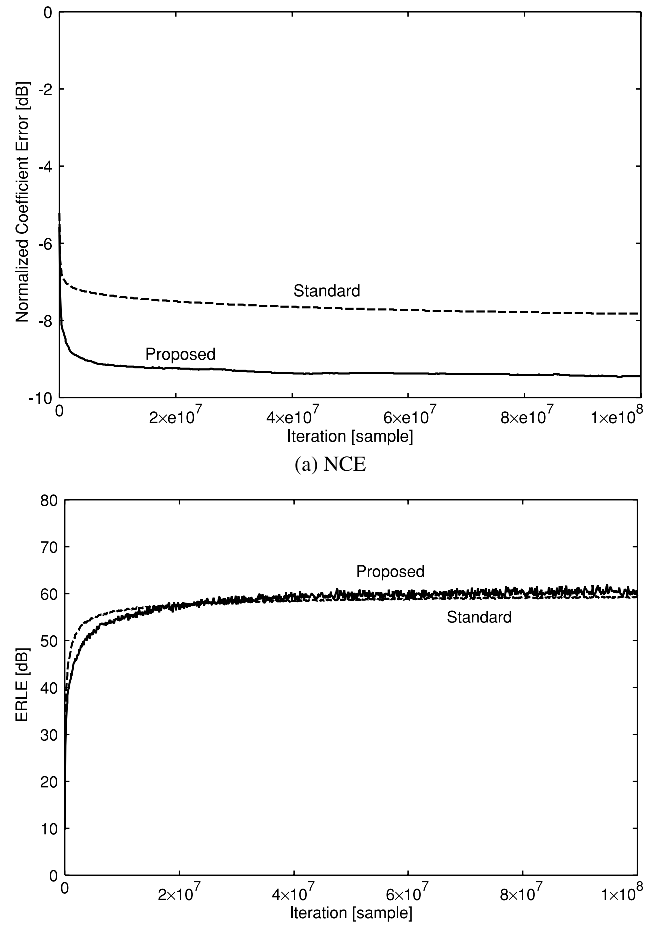
Figure 5 compares the performance when the correct echo-path identification condition $N_W/2 < N_A \leq N_B$ is not satisfied. N_A is selected as 30, which is less than $N_W/2 = 32$. In this case, neither $\boldsymbol{Q}_f^{-1}$ nor $\boldsymbol{Q}_b^{-1}$ exists. The NCE for the proposed algorithm is almost same as those for the standard



(a) NCE

(b) ERLE

Fig. 5 NCE and ERLE when correct identification condition is not satisfied.



(a) NCE

(b) ERLE

Fig. 6 NCE and ERLE for longer impulse response.

SAEC and the XM-NLMS. The ERLE for the proposed algorithm is almost same as that for the standard SAEC. The ELRE for the XM-NLMS is slightly larger because of the smaller step-size. These results show that the performance of the proposed algorithm is comparable to those for the conventional algorithms even when the correct echo-path identification condition is not satisfied.

Simulations for a longer impulse response case have also been carried out. $N_A = 800$, $N_B = N_W = 1024$ are used. For the proposed algorithm, $K = 2000$ is used. The NCE and the ERLE are shown in Fig. 6. The proposed algorithm improves the NCE, while almost the same ERLE is achieved.

5. Conclusions

This paper proposes an alternative learning algorithm for a stereophonic acoustic echo canceller without pre-processing which can identify the correct echo-paths. Convergence analysis clarifies the condition for correct echo-path identification. A convergence detection and an adaptive step-size based on the modification amount of the filter coefficients are also introduced. Simulation results show 10 dB smaller coefficient error than those of the conventional algorithms without pre-processing. Detailed analyses on the convergence condition, performance evaluation for more noisy case such as double-talk, and for real speech signals should be future study.

References

[1] M.M. Sondhi and D.R. Morgan, "Stereophonic acoustic echo cancellation—An overview of the fundamental problem," IEEE Signal Process. Lett., vol.2, no.8, pp.148–151, Aug. 1995.

[2] A. Hirano and A. Sugiyama, "Convergence characteristics of a multi-channel echo canceller with strongly cross-correlated input signals—Analytical results," Proc. 6th DSP Symposium, pp.144–149, Nov. 1991.

[3] A. Hirano and S. Koike, "Convergence analysis of a stereophonic acoustic echo canceller part 1: Convergence characteristics of tap weights," Proc. 11th DSP Symposium, pp.521–526, Nov. 1996.

[4] A. Hirano, "Convergence analysis of a multi-channel acoustic echo canceller," Proc. 12th DSP Symposium, pp.521–526, Nov. 1997.

[5] M. Tanaka, A. Hirano, and K. Nakayama, "On uniqueness of filter coefficients in stereophonic acoustic echo canceller," Proc. 14th DSP Symposium, pp.31–36, Nov. 1999.

[6] A. Sugiyama, Y. Joncour, and A. Hirano, "A stereo echo canceler with correct echo-path identification based on an input-sliding technique," IEEE Trans. Signal Process., vol.49, no.11, pp.2577–2587, Nov. 2001.

[7] M. Ali, "Stereophonic acoustic echo cancellation system using time-varying all-path filtering for signal decorrelation," Proc. ICASSP'98, pp.3689–3692, May 1998.

[8] A. Hirano, K. Nakayama, and K. Watanabe, "Convergence analysis of stereophonic echo canceller with pre-processing—Relation between pre-processing and convergence," Proc. ICASSP'99, pp.861–864, March 1999.

[9] A. Hirano, K. Nakayama, and K. Takebe, "Stereophonic acoustic echo canceller with pre-processing—Second-order pre-processing filter and its convergence," Proc. 8th IWAENC, pp.63–66, Sept. 2003.

[10] J. Benesty, D. Morgan, and M. Sondhi, "A better understanding and an improved solution to the specific problems of stereophonic acoustic echo cancellation," IEEE Trans. Speech Audio Process., vol.6, no.2, pp.156–165, March 1998.

[11] A.W.H. Khong and P. Naylor, "The use of partial update schemes to reduce inter-channel coherence in adaptive stereophonic acoustic echo cancellation," Proc. IWAENC 2003, pp.59–62, Sept. 2003.

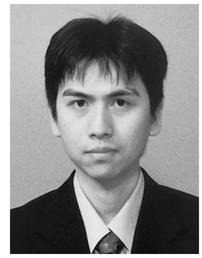
[12] A. Hirano and A. Sugiyama, "A compact multi-channel echo canceller with a single adaptive filter per channel," Proc. IEEE 1992 ISCAS, pp.1922–1925, May 1992.

[13] A. Hirano, A. Sugiyama, Y. Arasawa, and N. Kawayachi, "Dsp implementation and performance evaluation of a compact stereo echo canceller," Proc. 1994 IEEE ICASSP, vol.2, pp.249–252, April 1994.

[14] J. Nagumo and A. Noda, "A learning method for system identification," IEEE Trans. Autom. Control, vol.12, no.3, pp.282–287, March 1967.

Akihiro Hirano received the B.Eng., M.Eng. and Dr. Eng. degrees from Kanazawa University, Kanazawa, Japan in 1987, 1989 and 2000, respectively. He was involved in NEC Corporation, Kawasaski, Japan from 1989 to 1998, where he had been a Research Engineer in Research and Development Group. He joined Faculty of Engineering, Kanazawa University in 1998. He is currently an Assistant Professor in Division of Electrical Engineering and Computer Science, Graduate School of Natural Science and Technology, Kanazawa University. He has been engaged in researches on adaptive signal processing and neural networks. He was awarded the 1995 Academic Encouragement Award by IEICE. Dr. Hirano is a member of IEEE.

Kenji Nakayama received the B.E. and Dr. degrees in electronics engineering from Tokyo Institute of Technology (TIT), Tokyo, Japan, in 1971 and 1983, respectively. From 1971 to 1972 he was engaged in the research on classical network theory in TIT. He was involved in NEC Corporation from 1972 to 1988, where his research subjects were filter design methodology and digital signal processing. He joined the Department of Electrical and Computer Engineering at Kanazawa University, in Aug. 1998. He is currently a Professor of Graduate School of Natural Science and Technology. His current research interests include adaptive signal processing and neural networks. He is a senior member of IEEE and a member of INNS.



Daisuke Someda entered Department of Electrical and Computer Engineering, Faculty of Engineering, Knazawa University in 1999. He belonged to the adaptative systems laboratory in 2002, and studied a stereo echo canceller. After graduation, he joined Tokairtika Co., Ltd. and engaged in design of car electronics devices.

Masahiko Tanaka received the B.Eng. and M.Eng. degrees from Kanazawa University, Kanazawa, Japan in 1999, 2001, respectively. He has been engaged in researches on adaptive signal processing. He was involved in Hitachi SoftWare Engineering Corporation, Japan from 2000, where he had been a System Engineer in Second Department, Financial Systems Division.