STEREOPHONIC ACOUSTIC ECHO CANCELLER WITHOUT PRE-PROCESSING

Akihiro HIRANO   *Kenji NAKAYAMA*   *Daisuke SOMEDA*   *Masahiko TANAKA*

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

ABSTRACT

This paper proposes a stereophonic acoustic echo canceller without pre-processing which can identify the correct echo-paths. By dividing the filter coefficients into two portions and update one part at a time, the filter coefficient have an unique solution. Convergence analysis clarifies the condition for correct echo-path identification. For fast convergence and stable adaptation, a convergence detection and an adaptive step-size are also introduced. The modification amount of the filter coefficients detects the convergence and also determines the step-size. Computer simulations show 10dB smaller coefficient error than those of the conventional algorithms.

Fig. 1. Teleconferencing using SAEC

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) [1–3] have been studied.

SAEC's have a fundamental problem in which their filter coefficients cannot have an unique solution [1]. Though SAEC's with pre-processing [2] are good candidates for solving this problem, audible sound distortion caused by the pre-processing arises. An SAEC without pre-processing, XM-NLMS algorithm [4], 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 Section 3. Computer simulation results show the performance of the proposed algorithm.
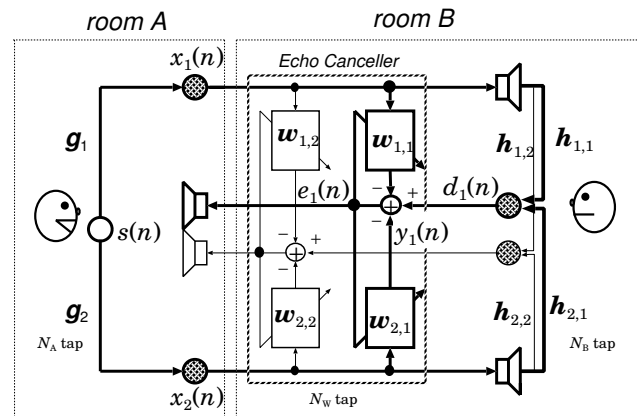
2. STEREOPHONIC ACOUSTIC ECHO CANCELLER AND UNIQUENESS PROBLEM

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 $x_i(n)$ in the i-th channel at time index n is generated from a talker speech $s(n)$ by passing room A impulse response g_i from the talker to the i-th microphone. $x_i(n)$ passes an echo path $h_{i,j}$ from the i-th loudspeaker to the j-th microphone and become an echo $d_j(n)$. Similarly, adaptive filters $w_{i,j}(n)$ generates an echo replica $y_j(n)$. $w_{i,j}(n)$ 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 an unique solution [1]. SAEC's may have infinite number of solutions other than the optimum solution $w_{i,j}(n) = h_{i,j}$.

Further analyses show that SAEC's may have 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, 6]. For echo cancellation performance,

$N_B < N_W$ is preferable where N_B is the impulse response length in room B. Therefore, if $N_B < N_W < N_A$, 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.

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 $N_W/2 < N_A < N_W$ and $N_W/2 < N_B < N_W$. 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 $\boldsymbol{w}_{i,j,b}(n)$ show by

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

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

The superscript ^{T} denotes the transpose of a matrix or a vector. 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. The far-end signal on i-th channel $x_i(n)$ is derived as

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

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

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

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

The echo $d_j(n)$ and the echo replica $y_j(n)$ is calculated as

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

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

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

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

$$\boldsymbol{h}_{i,j,b} = [h_{i,j,N_W/2}, \cdots, h_{i,j,N_W-1}]^T \quad (9)$$

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

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

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

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

$\boldsymbol{s}_f(n), \boldsymbol{s}_b(n)$ is defined by

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

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

$\boldsymbol{G}_i$ is a matrix defined by (15), which contains $\boldsymbol{g}_i$ and performs convolution between $\boldsymbol{s}_i(n)$ and $\boldsymbol{g}_i$. By introducing vectors and matrices defined by

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

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

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

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

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

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

is derived.

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

$$E[e_i^2(n)] = E[e_i(n)e_i^T(n)]$$
$$= \overline{\boldsymbol{d}_f^T(n)}\boldsymbol{Q}_f\overline{\boldsymbol{d}_f(n)} + \boldsymbol{d}_b^T(0)\boldsymbol{Q}_b\boldsymbol{d}_b(0)$$
$$+ 2\overline{\boldsymbol{d}_f^T(n)}\boldsymbol{Q}_{fb}\boldsymbol{d}_b(0) \quad (21)$$

where

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

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

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

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

$$\underbrace{\hphantom{g_{i,p,0} \quad g_{i,p,1} \quad \cdots \quad g_{i,p,N_A-1} \quad 0 \quad \cdots \quad \cdots \quad 0}}_{N_W/2+N_A-1}$$

$$p = f \ or \ b$$

$$\boldsymbol{Q}_f = \boldsymbol{G}_f \boldsymbol{R}_f \boldsymbol{G}_f^T \qquad (25)$$

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

$$\boldsymbol{Q}_b = \boldsymbol{G}_b \boldsymbol{R}_b \boldsymbol{G}_b^T \qquad (27)$$

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

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

the averaged filter coefficient error $\overline{\boldsymbol{d}_f^*}$ which minimizes $E[e_j^2(n)]$ is determined as

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

Therefore, the filter coefficients will converge on

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

The first term in the right hand side is the optimum solution. The second term is the error caused by the tap shortage.

By repeating updates of $\boldsymbol{w}_{i,j,f}(n)$ and $\boldsymbol{w}_{i,j,b}(n)$, the coefficient error vectors $\boldsymbol{d}_f^{(m)}$ and $\boldsymbol{d}_b^{(m)}$ after m-th iteration becomes

$$\boldsymbol{d}_f^{(m+1)} = \boldsymbol{K}_f \boldsymbol{d}_b^{(m)} \qquad (31)$$

$$\boldsymbol{d}_b^{(m+1)} = \boldsymbol{K}_b \boldsymbol{d}_f^{(m)} \qquad (32)$$

where

$$\boldsymbol{K}_f = \boldsymbol{Q}_f^{-1}\boldsymbol{Q}_{fb} \qquad (33)$$

$$\boldsymbol{K}_b = \boldsymbol{Q}_b^{-1}\boldsymbol{Q}_{fb}^T. \qquad (34)$$

Solving (31) and (32) results in

$$\boldsymbol{d}_f^{(m)} = -\boldsymbol{K}_f(\boldsymbol{K}_b\boldsymbol{K}_f)^{(m-1)}\boldsymbol{d}_b^{(0)} \qquad (35)$$

$$\boldsymbol{d}_b^{(m)} = (\boldsymbol{I} - (\boldsymbol{K}_b\boldsymbol{K}_f)^m)\boldsymbol{d}_b^{(0)}. \qquad (36)$$

If the maximum absolute eigenvalue of $\boldsymbol{K}_b\boldsymbol{K}_f$ is less than 1, the filter coefficients converge at the optimum value.

3.3. Adaptation Control

In this approach, the adaptation control is a difficult problem. For converge at the optimum value, 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

$$D(m) = \frac{\sum_{i=1}^2 \|\boldsymbol{w}_{i,j,p}(mK) - \boldsymbol{w}_{i,j,p}((m-1)K)\|^2}{\sum_{i=1}^2 \|\boldsymbol{w}_{i,j,p}(mK)\|^2} \qquad (37)$$

where p is either f or b. To avoid the increase of the computational cost, (37) is calculated once in a K iterations. Coefficient adaptation is stopped when (37) is calculated.

The filter coefficients are considered to be converged if $D(m-1) < D(m)$ is satisfied. The step-size is controlled by

$$\mu(m) = \mu_{max}\left(\frac{D(m)}{D_{max}}\right)^{1/4} \qquad (38)$$

where D_{max} is a maximum value of $D(m)$ in a same stage. Usually, $D(1)$ is used as a D_{max}. $\mu(n)$ is used within $mK < n < (m+1)K$.

The overview of the adaptation control is as follows:

1. Update filter coefficients with $\mu(0) = \mu_{max}$ for first K iterations.

2. calculate $D(1)$. $D_{max} = D(1)$.

3. Update filter coefficients with $\mu(m)$ by (38) for next K iterations.

4. calculate $D(m)$.

5. If $D(m-1) < D(m)$, then proceed to the next stage.

6. If $D_{max} < D(m)$, then $D_{max} = D(m)$.

Table 1. Simulation Conditions

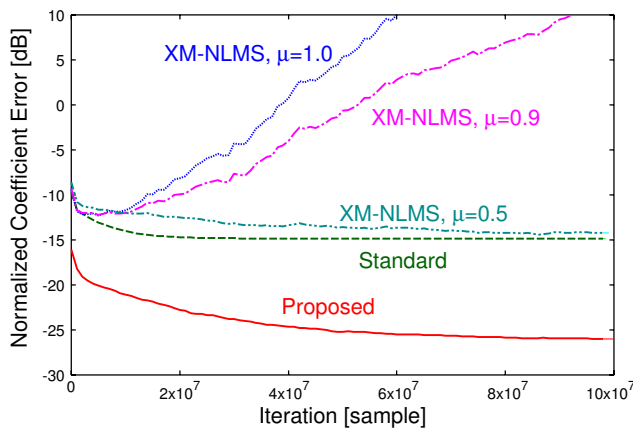| Parameters	
N_A	60
N_B, N_W	64
Fixed switching interval	20000
Averaging interval	10000
K	5000
Adaptation algorithm	NLMS
Fixed step-size μ	1.0
μ_{max}	1.0
Far-end talker signal $s(n)$	White Gaussian
Additive noise	None

Fig. 2. Normalized Coefficient Error

7. Goto 3.

4. COMPUTER SIMULATIONS

Simulations have been carried out to show the performance of the proposed algorithm. Table 1 depicts the simulation conditions. Far-end room impulse responses $\boldsymbol{g}_i$ are 60-tap FIR filters while those for near-end room $\boldsymbol{h}_{i,j}$ are 64-tap FIR filters. In this case, SAEC's do not have an unique solution. Adaptive filters are 64-tap FIR filters. As an adaptation algorithm, Normalized Least Mean Squares (NLMS) algorithm [7] is used. The proposed algorithm is compared with the standard SAEC [1] and the XM-NLMS algorithm [4].

Figure 2 compares the normalized coefficient error (NCE) defined by

$$NCE(n) = \frac{\sum_{i=1}^{2} \|\boldsymbol{w}_{i,j}(n) - \boldsymbol{h}_{i,j}\|^2}{\sum_{i=1}^{2} \|\boldsymbol{h}_{i,j}\|^2}. \qquad (39)$$

The proposed algorithm achieves -26dB of NCE which is almost 10dB smaller than the standard SAEC. The XM-

NLMS failed to converge for a large step-size μ. Though the XM-NLMS converges with $\mu = 0.5$, its convergence speed is slower than that of the standard SAEC.

5. CONCLUSIONS

This paper proposes 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 10dB smaller coefficient error.

6. REFERENCES

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

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

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

[4] 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. of IWAENC 2003*, pp. 59–62, 9 2003.

[5] 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 and Audio*, vol. 6, no. 2, pp. 156–165, Mar. 1998.

[6] M. Tanaka, A. Hirano, and K. Nakayama, "On uniqueness of filter coefficients in stereophonic acoustic echo canceller," *Proceedings of 14th DSP Symposium*, pp. 31–36, Nov. 1999 (In Japanese).

[7] J. Nagumo and A. Noda, "A learning method for system identification," *IEEE Trans. AC*, vol. 12, no. 3, pp. 282–287, Mar 1967.