

Adaptation Control for Stereophonic Acoustic Echo Canceller without Pre-Processing

前処理を用いないステレオエコーキャンセラの学習制御

Akihiro HIRANO Kenji Nakayama Daisuke SOMEDA Masahiko TANAKA
平野 晃宏 中山 謙二 染田大輔 田中正彦

Faculty of Engineering, Kanazawa University
金沢大学 工学部

E-mail: hirano@t.kanazawa-u.ac.jp, nakayama@t.kanazawa-u.ac.jp

ABSTRACT

This paper proposes an adaptation control algorithm for a stereophonic acoustic echo canceller without pre-processing which can identify the correct echo-paths. A convergence detection and an adaptive step-size are introduced into the partial update algorithm which have a unique solution by dividing the filter coefficients. The modification amount of the filter coefficients detects the convergence and also determines the step-size. By using the adaptive step-size, the time-averaging of the filter coefficients required for the conventional algorithm is not necessary. Faster convergence is achieved by convergence detection.

あらまし

前処理を用いずにエコーパスを同定できるステレオエコーキャンセラのための学習制御アルゴリズムを提案する。フィルタ係数を分割することで解が一意に定まる分割学習方式において、フィルタ係数収束の自動判定と可変ステップサイズを用いる。フィルタ係数の修正量が収束したら係数が収束したと判定する。また、フィルタ係数の修正量に応じたステップサイズを使用する。可変ステップサイズの導入により、従来法で必要だったフィルタ係数の時間平均が不要になる。係数収束の自動判定により、収束速度を高速化できる。

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 (SAECs) [1–4] have been studied.

SAEC's have a fundamental problem in which their filter coefficients cannot have a unique solution [1]. Though SAEC's with pre-processing [2, 3] are good candidates for solving this problem, audible sound distortion caused by the pre-processing arises.

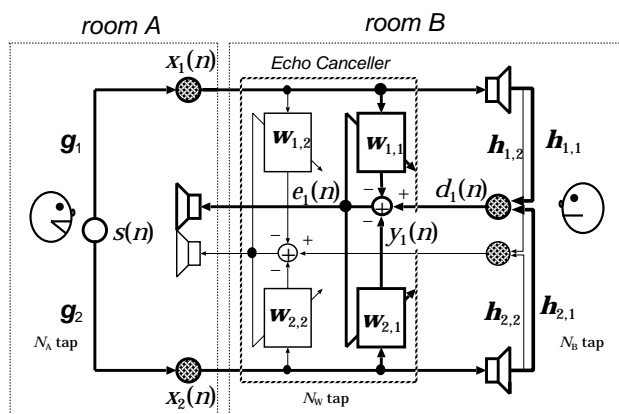


Figure 1: Teleconferencing using SAEC

A stereophonic acoustic echo canceller without pre-processing [5] has been proposed. Though this SAEC can identify the correct echo-paths without sound distortion, its adaptation control and convergence speed makes it difficult to apply this SAEC for real systems.

This paper proposes an adaptation control algorithm for a stereophonic acoustic echo canceller without pre-processing. Section 2 reviews the SAEC and its fundamental problem, followed by the SAEC without pre-processing. The improved adaptation control is proposed in Section 4. 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 \mathbf{g}_i from the talker to the i -th microphone. $\mathbf{x}_i(n)$ passes an echo path $\mathbf{h}_{i,j}$ from the i -th loudspeaker to the j -th microphone and become an echo $d_j(n)$. Similarly, adaptive filters $\mathbf{w}_{i,j}(n)$ generates an echo replica $y_j(n)$. $\mathbf{w}_{i,j}(n)$ 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 $\mathbf{w}_{i,j}(n) = \mathbf{h}_{i,j}$.

Further analyses show that SAEC's may have unique and optimum solution when $N_W < N_A$ is satisfied [6, 7]. For echo cancellation performance, $N_B < N_W$ is preferable. 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 by Partial Updates [5]

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 $\mathbf{w}_{i,j}(n)$ is divided into two sub-vectors $\mathbf{w}_{i,j,f}(n)$ and $\mathbf{w}_{i,j,b}(n)$ show by

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

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

In the first stage, $\mathbf{w}_{i,j,f}(n)$ is updated while $\mathbf{w}_{i,j,b}(n)$ is fixed. This stage is repeated until $\mathbf{w}_{i,j,f}(n)$ converges. As the second stage, $\mathbf{w}_{i,j,b}(n)$ is updated while $\mathbf{w}_{i,j,f}(n)$ is fixed. This stage is also repeated until $\mathbf{w}_{i,j,b}(n)$ 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

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

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

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

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

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

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

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

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

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

$$\mathbf{h}_{i,j,b} = [h_{i,j,N_W/2}, \dots, h_{i,j,N_W}]^T \quad (9)$$

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

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

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

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

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

$\mathbf{s}_f(n)$, $\mathbf{s}_b(n - N_W)$ is defined by

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

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

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

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

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

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

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

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

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

is derived.

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

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

where

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

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

$p = f \text{ or } b$

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

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

$$\mathbf{Q}_f = \mathbf{G}_f \mathbf{R}_f \mathbf{G}_f^T \quad (25)$$

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

$$\mathbf{Q}_b = \mathbf{G}_b \mathbf{R}_b \mathbf{G}_b^T \quad (27)$$

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

$$\frac{\partial e_f^2(n)}{\partial \overline{\mathbf{d}_f(n)}} = 2\mathbf{Q}_f \overline{\mathbf{d}_f(n)} + 2\mathbf{Q}_{fb} \mathbf{d}_b(0) = 0, \quad (28)$$

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

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

Therefore, the filter coefficients will converge on

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

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

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

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

where

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

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

Solving (31) and (32) results in

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

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

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

3.3 Problems on Adaptation Control

In the conventional algorithm, 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.

A long switching interval is used because a time interval toward convergence is not known a priori. A long switching interval causes slow convergence. To avoid an influence of a random walk around the convergence value on the performance, an time averaging of the filter coefficients is used. Otherwise, a small step-size should be used. The averaging increases the computational cost while a small step-size causes slow convergence.

4 Proposed Adaptation Control

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

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

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

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

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

Table 1: Simulation Conditions

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

Table 2: Adaptation Control Methods

Name	Step-size	Interval	Averaging
Proposed	Adaptive	Adaptive	No
Conventional	Fixed	Fixed	Yes
Reference 1	Adaptive	Adaptive	Yes
Reference 2	Fixed	Adaptive	Yes
Reference 3	Fixed	Adaptive	No

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 $D_{max} = D(m)$.
7. Goto 3.

5 Computer Simulations

Simulations have been carried out to show the performance of the proposed algorithm. Table 1 depicts the simulation conditions. All room impulse responses, g_i and $h_{i,j}$, 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 [8] is used. Adaptation control methods are compared by Tab. 2. To show the effect of the adaptive step-size and averaging, several combinations are compared in addition to the proposed method.

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

$$NCE(n) = \frac{\sum_{i=1}^2 \|w_{i,j}(n) - h_{i,j}\|^2}{\sum_{i=1}^2 \|h_{i,j}\|^2}. \quad (39)$$

The proposed algorithm converges to -20dB of NCE almost 10 times faster than the conventional algorithm. Since the convergence characteristics of the proposed algorithm

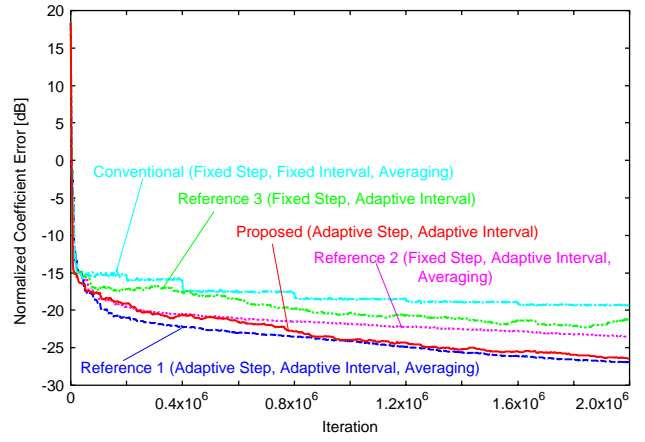


Figure 2: Normalized Coefficient Error

is almost same as “Reference 1,” the averaging is not required. Comparison of the proposed algorithm with “Reference 2” and “Reference 3” shows the advantage of the adaptive step-size. Without the adaptive step-size, the NCE becomes large even if the averaging is used.

6 Conclusions

This paper proposes an adaptation control algorithm for a stereophonic acoustic echo canceller without pre-processing which can identify the correct echo-paths. A convergence detection and an adaptive step-size based on the modification amount of the filter coefficients are introduced. Simulation results show the faster convergence and smaller coefficient error.

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] M. Ali, “Stereophonic acoustic echo cancellation system using time-varying all-path filtering for signal decorrelation,” *Proc. of ICASSP’98*, pp. 3689–3692, May 1998.
- [4] 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.

- [5] Masahiko Tanaka, Akihiro Hirano, and Kenji Nakayama, "Echo path identification without pre-processing in stereophonic acoustic echo canceller (in japanese)," *Proceedings of 15th DSP Symposium*, pp. 541–545, Nov. 2000.
- [6] 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.
- [7] Masahiko Tanaka, Akihiro Hirano, and Kenji Nakayama, "On uniqueness of filter coefficients in stereophonic acoustic echo canceller (in japanese)," *Proceedings of 14th DSP Symposium*, pp. 31–36, Nov. 1999.
- [8] J. Nagumo and A. Noda, "A learning method for system identification," *IEEE Trans. AC*, vol. 12, no. 3, pp. 282–287, Mar 1967.