# STEREO PROJECTION ECHO CANCELLER WITH TRUE ECHO PATH ESTIMATION

*Suehiro Shimauchi and Shoji Makino*

NTT Human Interface Laboratories
3-9-11, Midori-cho, Musashino-shi, Tokyo, 180 Japan

**ABSTRACT**

The effect that cross-correlation between stereo signals has on a stereo echo canceller is studied and a new stereo projection echo canceller is proposed that can identify the true echo path impulse responses. This echo canceller accelerates filter coefficient error convergence by adding the variations in the cross-correlation to stereo signals or by utilizing the fact that the cross-correlation between stereo signals varies slightly in actual teleconferencing situations. Computer simulations demonstrate that this echo canceller can reduce the filter coefficient error much faster than a conventional stereo normalized least mean squares (NLMS) echo canceller.

## 1. INTRODUCTION

A stereo teleconferencing system provides a more realistic teleconferencing presence than a monaural system. It lets listeners use spatial information to help distinguish who is speaking. However, a practical stereo echo canceller has not yet been developed. The most significant obstacle to stereo echo cancellation using a conventional linear combiner structure is that the adaptive filters converge very slowly[1]. This is due to misconvergence caused by the cross-correlation between stereo signals[2][3]. The performance of a structure based on preliminary decorrelation of stereo signals and that of a conventional structure with frequency domain least mean squares (LMS) are reported to be somewhat unsatisfactory[2]. A compact stereo echo canceller[3], which has a single adaptive filter for each channel, cannot structurally identify the true echo path impulse responses.

In this paper, we examine the influence of the cross-correlation between stereo signals on stereo echo canceller performance. We point out that, even if the conventional linear combiner structure is used, the variation in the cross-correlation between stereo signals can be effectively used to estimate the true echo path impulse responses, but it is slow. We thus propose a new stereo projection echo canceller. The structure of this echo canceller includes a function block for adding the slight variations in the cross-correlation to the stereo signals. The stereo projection algorithm emphasizes these variations and accelerates the filter coefficient error convergence. This echo canceller also exploits the fact that cross-correlation between stereo signals varies slightly in actual teleconferencing situations. Computer simulations demonstrate that this echo canceller can reduce the filter coefficient error much faster than a conventional stereo normalized least mean squares (NLMS) echo canceller.

## 2. STRUCTURE OF CONVENTIONAL STEREO ECHO CANCELLER

Conventional stereo echo cancellation is achieved by linearly combining stereo signals (Fig. 1)[1]. Input signal vectors $\boldsymbol{x}_1(k)$ and $\boldsymbol{x}_2(k)$ and filter coefficient vectors $\hat{\boldsymbol{h}}_1(k)$ and $\hat{\boldsymbol{h}}_2(k)$ are combined as $\boldsymbol{x}(k) = [\boldsymbol{x}_1^T(k), \boldsymbol{x}_2^T(k)]^T$ and $\hat{\boldsymbol{h}}(k) = [\hat{\boldsymbol{h}}_1^T(k), \hat{\boldsymbol{h}}_2^T(k)]^T$. The combined filter coefficient vector $\hat{\boldsymbol{h}}(k)$ is updated by an NLMS algorithm. Thus, stereo echo cancellation is achieved by linearly combining two monaural echo cancellers. Note that input signal vectors $\boldsymbol{x}_1(k)$ and $\boldsymbol{x}_2(k)$ have a strong cross-correlation.

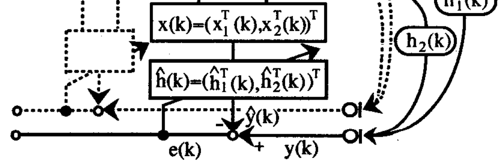


Fig. 1. Configuration of a conventional stereo echo canceller.



0-7803-2431-5/95 $4.00 © 1995 IEEE

## 3. PROBLEM IN STEREO ECHO CANCELLATION

The most significant problem in stereo echo cancellation is that the filter coefficient does not converge to the true echo path impulse responses. If input signals $\boldsymbol{x}_1(k)$ and $\boldsymbol{x}_2(k)$ have a constant cross-correlation, the steady-state solution $[\hat{\boldsymbol{h}}_1^T(k), \hat{\boldsymbol{h}}_2^T(k)]$ that satisfies $y(k) = \hat{y}(k)$, where

$$y(k) = \boldsymbol{h}^T(k)\boldsymbol{x}(k) = \boldsymbol{h}_1^T(k)\boldsymbol{x}_1(k) + \boldsymbol{h}_2^T(k)\boldsymbol{x}_2(k) \quad (1)$$

and

$$\hat{y}(k) = \hat{\boldsymbol{h}}^T(k)\boldsymbol{x}(k) = \hat{\boldsymbol{h}}_1^T(k)\boldsymbol{x}_1(k) + \hat{\boldsymbol{h}}_2^T(k)\boldsymbol{x}_2(k), \quad (2)$$

is not unique. By minimizing the error between echo $y(k)$ and echo replica $\hat{y}(k)$, the steady-state solution $[\hat{\boldsymbol{h}}_1^T(k), \hat{\boldsymbol{h}}_2^T(k)]$ converges to the point in subspace $\boldsymbol{H}_x$ nearest the initial point, where $\boldsymbol{H}_x$ is uniquely determined by the cross-correlation between $\boldsymbol{x}_1(k)$ and $\boldsymbol{x}_2(k)$. This does not mean that $\hat{\boldsymbol{h}}(k)$ equals $\boldsymbol{h}(k)$.

For example, in the simplest case, if input stereo signals $\boldsymbol{x}_1(k)$ and $\boldsymbol{x}_2(k)$ are denoted as

$$\boldsymbol{x}_1(k) = a_1\boldsymbol{s}(k) \quad \text{and} \quad \boldsymbol{x}_2(k) = a_2\boldsymbol{s}(k), \quad (3)$$

where $a_1$ and $a_2$ are scalar constant values and $\boldsymbol{s}(k)$ is a source signal vector, and if the initial value $\hat{\boldsymbol{h}}(0)$ is set to the zero vector, subspace $\boldsymbol{H}_x$ corresponds to a line for convenience' sake, as shown in Fig. 2, and filter coefficients $\hat{\boldsymbol{h}}_1(k)$ and $\hat{\boldsymbol{h}}_2(k)$ converge to

$$\hat{\boldsymbol{h}}_{1a}(k) = \frac{a_1^2}{a_1^2+a_2^2}(\boldsymbol{h}_1(k) + \frac{a_2}{a_1}\boldsymbol{h}_2(k)) \neq \boldsymbol{h}_1(k) \quad (4)$$

$$\hat{\boldsymbol{h}}_{2a}(k) = \frac{a_2^2}{a_1^2+a_2^2}(\frac{a_1}{a_2}\boldsymbol{h}_1(k) + \boldsymbol{h}_2(k)) \neq \boldsymbol{h}_2(k). \quad (5)$$

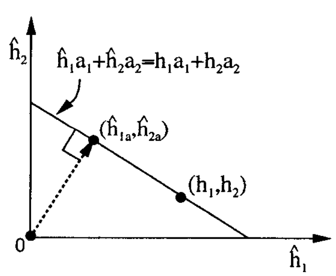


Fig. 2. Effect of cross-correlation on the steady-state solution.

## 4. NEW STEREO ECHO CANCELLER

### 4.1. Effect of variation in cross-correlation between stereo signals

Consider the case where the cross-correlation between input signals $\boldsymbol{x}_1(k)$ and $\boldsymbol{x}_2(k)$ varies, *e.g.*, $a_1$ and $a_2$ in Eq. (3) vary respectively to $b_1$ and $b_2$ (Fig. 3). First, $[\hat{\boldsymbol{h}}_1^T(k), \hat{\boldsymbol{h}}_2^T(k)]$ converges to $[\hat{\boldsymbol{h}}_{1a}^T(k), \hat{\boldsymbol{h}}_{2a}^T(k)]$, as shown in Eqs. (4) and (5). Then it converges to $[\hat{\boldsymbol{h}}_{1b}^T(k), \hat{\boldsymbol{h}}_{2b}^T(k)]$, the point nearest the "initial" point $[\hat{\boldsymbol{h}}_{1a}^T(k), \hat{\boldsymbol{h}}_{2a}^T(k)]$. Consequently, the norm of filter coefficient error vector $\boldsymbol{e}_b$ becomes smaller than the norm of $\boldsymbol{e}_a$. The coefficient error between $\boldsymbol{h}(k)$ and $\hat{\boldsymbol{h}}(k)$ becomes smaller with every variation in the cross-correlation between the stereo signals.

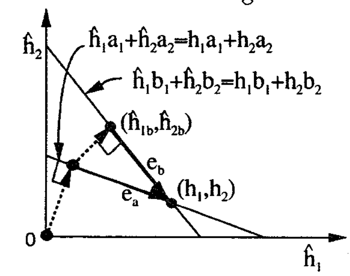


Fig. 3. Effect of variation in the cross-correlation between stereo signals.

### 4.2. Addition of variation in cross-correlation to signals

The analysis in Sec. **4.1.** suggests that the variations in the cross-correlation between stereo signals can be used to achieve filter coefficient error convergence. A new our proposed stereo echo canceller structure is shown in Fig. 4. The function block for adding the variations in the cross-correlation to stereo signals is necessary when receiving fixed cross-correlation stereo signals, *e.g.*, those generated by a mixing machine. There are many ways of achieving this function block, *e.g.*, modulations, filterings, or noise additions. Any of these methods are acceptable as long as their influence is inaudible.

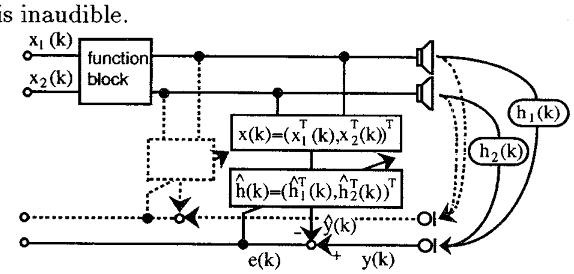


Fig. 4. Configuration of a new stereo echo canceller.

### 4.3. Emphasis of variation in cross-correlation

The variation in the cross-correlation between stereo signals can be used to estimate the true echo path impulse responses; the cross-correlation will always vary slightly due to the function block described in Sec. **4.2.**. The next step is how to emphasize these variations and use them to accelerate the filter coefficient error convergence.

If the cross-correlation is constant until time $k-1$, combined stereo signal vectors exist in the subspace determined by the cross-correlation, *i.e.*, $\boldsymbol{x}(k-1), \boldsymbol{x}(k-2), \ldots \in \boldsymbol{S}$, as shown in Fig. 5. If the cross-correlation then varies at time $k$, the combined stereo signal vector $\boldsymbol{x}(k)$ exists in a subspace different from $\boldsymbol{S}$, *i.e.*, $\boldsymbol{x}(k) \in \boldsymbol{S}'$. We can thus treat $\boldsymbol{x}(k)$ as the sum of two orthogonal components, that is,

$$\boldsymbol{x}(k) = \boldsymbol{v}(k) + \bar{\boldsymbol{v}}(k) \quad (\boldsymbol{v}(k) \in \boldsymbol{S},\ \bar{\boldsymbol{v}}(k) \perp \boldsymbol{S}), \qquad (6)$$

where $\bar{\boldsymbol{v}}(k)$ is the new component used for convergence and $\boldsymbol{v}(k)$ is the redundant component. Thus, if the direction of adjustment vector $\Delta\hat{\boldsymbol{h}}(k)$ is made the same as that of $\bar{\boldsymbol{v}}(k)$ by removing $\boldsymbol{v}(k)$ from $\boldsymbol{x}(k)$, the adjustment of $\hat{\boldsymbol{h}}(k)$ is achieved with no redundancy; the variation emphasis in the cross-correlation is achieved by deriving $\bar{\boldsymbol{v}}(k)$, and filter coefficient vector $\hat{\boldsymbol{h}}(k)$ is adjusted in the direction of the "emphasized" vector $\bar{\boldsymbol{v}}(k)$. This discussion also applies to the case where the cross-correlation varies slightly until time $k-1$. In this case, subspace $\boldsymbol{S}$ becomes slightly wider; however, the vector $\bar{\boldsymbol{v}}(k)$ still becomes an "emphasized" vector orthogonal to "wide" subspace $\boldsymbol{S}$.

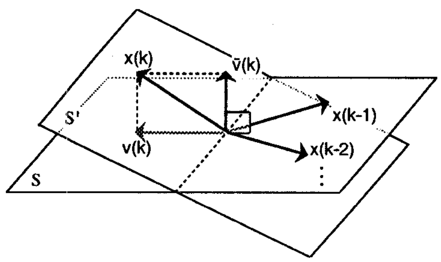


Fig. 5. Geometric interpretation of subspace determined by cross-correlation.

Since $\boldsymbol{v}(k)$ is correlated with $\boldsymbol{x}(k-1), \boldsymbol{x}(k-2), \ldots$, it can be removed from $\boldsymbol{x}(k)$ by the decorrelation; for example, by using the recursive least squares (RLS) algorithm or the projection algorithm. The RLS algorithm removes the redundant component completely, but its computation is very complex. On the other hand, a projection algorithm of order p removes p major redundant components at small computational cost[4].

### 4.4. Stereo projection algorithm

We use the projection algorithm, *i.e.*, the stereo projection algorithm[5], in our canceller as follows.

A $p$-th order projection algorithm updates the filter coefficient vector $\hat{\boldsymbol{h}}(k)$ as

$$\hat{\boldsymbol{h}}(k+1) = \hat{\boldsymbol{h}}(k) + \alpha\Delta\hat{\boldsymbol{h}}(k). \qquad (7)$$

$$\Delta\hat{\boldsymbol{h}}(k) = \beta_1(k)\boldsymbol{x}(k) + \beta_2(k)\boldsymbol{x}(k-1) + \cdots + \beta_p(k)\boldsymbol{x}(k-p+1), \qquad (8)$$

where $\alpha$ is scalar step size $(0 < \alpha < 2)$ and $\beta_1(k), \beta_2(k), \ldots, \beta_p(k)$ are determined so that $\hat{\boldsymbol{h}}(k+1)$ satisfies the following equations when $\alpha = 1$ [6].

$$\begin{aligned} \hat{\boldsymbol{h}}^T(k+1)\boldsymbol{x}(k) &= y(k) \\ \hat{\boldsymbol{h}}^T(k+1)\boldsymbol{x}(k-1) &= y(k-1) \\ &\vdots \\ \hat{\boldsymbol{h}}^T(k+1)\boldsymbol{x}(k-p+1) &= y(k-p+1). \end{aligned} \qquad (9)$$

Consequently, $\Delta\hat{\boldsymbol{h}}(k)$ becomes a decorrelated component of $\boldsymbol{x}(k)$ by subtracting the correlated components of $\boldsymbol{x}(k-1), \boldsymbol{x}(k-2), \ldots, \boldsymbol{x}(k-p+1)$ from $\boldsymbol{x}(k)$.

In Eq. (6), $\boldsymbol{v}(k)$ is the only component of $\boldsymbol{x}(k)$ from which the correlated components of $\boldsymbol{x}(k-1), \boldsymbol{x}(k-2), \ldots, \boldsymbol{x}(k-p+1)$ are removed because $\bar{\boldsymbol{v}}(k)$ is orthogonal to $\boldsymbol{x}(k-1), \boldsymbol{x}(k-2), \ldots, \boldsymbol{x}(k-p+1)$. If order p is adequately determined, almost all the components are removed from $\boldsymbol{v}(k)$ and the direction of the adjustment vector $\Delta\hat{\boldsymbol{h}}(k)$ becomes the same as that of $\bar{\boldsymbol{v}}(k)$.

## 5. COMPUTER SIMULATIONS

In our computer simulations, we use 500 taps in each $\hat{\boldsymbol{h}}_1(k)$ and $\hat{\boldsymbol{h}}_2(k)$ filter. The sampling frequency was 8 kHz. The true echo path impulse responses were measured in a conference room with a reverberation time of 150 ms.

### 5.1. Performance of conventional stereo echo canceller

The performance of a conventional stereo NLMS echo canceller was evaluated with input stereo signals that had a strong cross-correlation. We used three types of stereo signal pairs:

(a) **Fixed cross-correlation signals** were made from a monaural speech signal by convoluting two different impulse responses in the computer.

(b) **Modulated signals** were made from fixed cross-correlation signals by 5% amplitude modulation with white noise that had a different phase for each channel.

(c) **Recorded signals** were speech signals recorded with two microphones in the conference room while the speaker remained in one place.

With all of these input signals, the echo return loss enhancement (ERLE) reached at least 20 dB (Fig. 6). On the other hand, as shown in Fig. 7, the coefficient error convergence leveled off after a few dB modification. This is due to incorrectly estimating the true echo path impulse responses. Note that the coefficient errors with the modulated signals and recorded signals decreased slightly; the cross-correlation varied slightly. This suggests that the function block described in Sec.

4.2. is not always necessary in cases where the cross-correlation between stereo signals itself varies.

### 5.2. Performance of new stereo projection echo canceller

The performance of our proposed stereo projection echo canceller was evaluated for the projection order $p = 1, 2, 4, 8$. Setting $p = 1$ is equivalent to using a conventional NLMS algorithm. The input stereo signals were speech signals made by two people speaking alternately; these were recorded with two microphones in the conference room. The speakers were in different parts of the room and did not move their bodies or heads.

As the projection order $p$ increased, the canceller became more effective in preventing the residual echo increase caused by speaker alternation, as shown in Fig. 8, and filter coefficient error convergence was achieved more quickly, as shown in Fig. 9.

## 6. CONCLUSIONS

We proposed a new stereo projection echo canceller that identifies true echo path impulse responses by analyzing the variations in cross-correlation between the stereo signals. Computer simulations showed that it can expedite filter coefficient error convergence much faster than a conventional stereo NLMS echo canceller.

## REFERENCES


[1] M. M. Sondhi and D. R. Morgan, "Acoustic Echo Cancellation for Stereophonic Teleconferencing," *1991 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* (1991).

[2] Y. Mahieux, A. Gilloire, and F. Khalil, "Annulation D'echo en Teleconference Stereophonique," *EUSIPCO 93* (1993) (in French).

[3] A. Hirano, A. Sugiyama, Y. Arasawa, and N. Kawayachi, "DSP Implementation and Performance Evaluation of A Compact Stereo Echo Canceller," *ICASSP94*, pp. 245-248 (1994).

[4] M. Tanaka, Y. Kaneda, S. Makino, and J. Kojima, "Fast Projection Algorithm and Its Step Size Control," *ICASSP95* (to be published in 1995).

[5] S. Shimauchi and S. Makino, "Stereo projection echo canceller based on the variations in the cross-correlation between stereo signals," *Proc. Autumn Meet. Acoust. Soc. Jpn.*, pp. 655-656 (1994) (in Japanese).

[6] S. Makino and Y. Kaneda, "Exponentially Weighted Step-size Projection Algorithm for Acoustic Echo Cancellers," *IEICE Trans.*, Vol. E75-A, No. 11, pp. 1500-1508 (Nov. 1992).


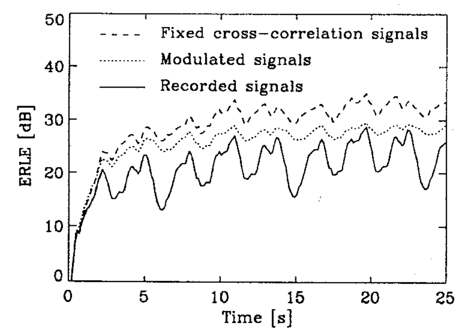


Fig. 6. ERLE with conventional NLMS algorithm.

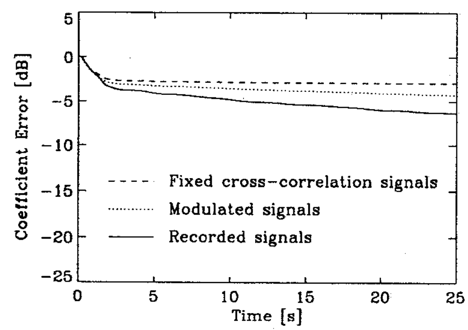


Fig. 7. Coefficient error convergence with conventional NLMS algorithm.

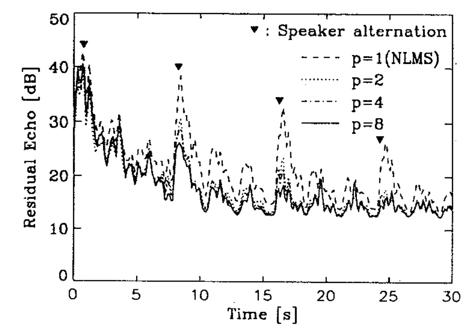


Fig. 8. Residual echo suppression with stereo projection algorithm.

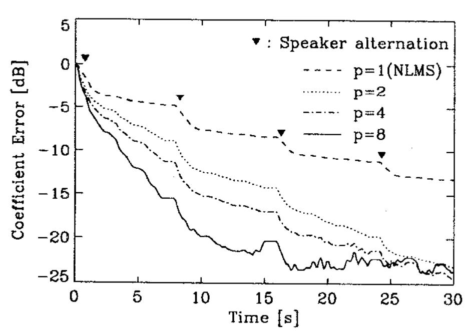


Fig. 9. Coefficient error convergence with stereo projection algorithm.