ADAPTIVE FILTERING ALGORITHM ENHANCING DECORRELATED ADDITIVE SIGNALS FOR STEREO ECHO CANCELLATION

Satoru Emura   Yoichi Haneda*   Shoji Makino***

*NTT Cyber Space Laboratories   **NTT Communication Science Laboratories

3-9-11, Midori-cho, Musashino-shi, Tokyo 180-8585, Japan   2-4, Hikaridai, Seika-cho, Kyoto, 619-0237, Japan

emura.satoru@lab.ntt.co.jp   haneda.yoichi@lab.ntt.co.jp   maki@cslab.kecl.ntt.co.jp

ABSTRACT

The strong inter-channel cross correlation between stereo received signals creates a problem of non-unique solutions for stereo echo cancellation. To avoid this problem, the addition of inaudible non-linearly filtered received signals to themselves was proposed. However, the convergence of misalignment is still slow even when the affine projection algorithm is used, since strong inter-channel cross correlation remains in the adjustment vector. We propose an enhanced affine projection algorithm in which the adjustment vector is well-decorrelated by enhancing non-linearly filtered received signals. Computer simulation demonstrated that the enhanced affine projection algorithm can decrease misalignment much faster than the conventional affine projection algorithm.

1. INTRODUCTION

In a stereo teleconferencing system, a stereo acoustic echo cancellation is required to reduce echoes that result from coupling between loudspeakers and microphones. Since stereo received signals usually come from the individual speaker via two acoustic paths, inter-channel cross correlation is strong. This causes the problem of non-unique solutions to the normal equation for estimating unknown echo paths. Misalignment between the estimated and actual impulse responses usually remains even if echoes are cancelled [1].

The addition of non-linearly filtered received signals to the original received signals [2], [3] has been proposed to avoid this non-uniqueness problem. Such a non-linear filtering approach must have no audible effects and must have no effect on stereo perception.

Since the amount of non-linearity is restricted, the convergence of misalignment is slow when the NLMS algorithm is used. The affine projection algorithm [2] and the combination of inter-channel decorrelation and affine projection [4] have been proposed to improve the convergence speed. However, strong inter-channel cross correlation still remains in the adjustment vector. The improvement in convergence speed is thus insufficient [5].

We propose an enhanced affine projection algorithm that improves the convergence of misalignment with the adjustment vector decorrelated by enhancing non-linearly filtered received signals.

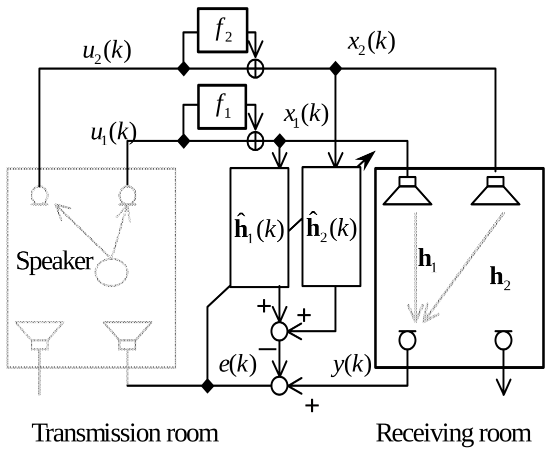

Fig. 1 Diagram of system for stereo echo cancellation.

2. CONFIGURATION FOR STEREO ECHO CANCELLATION

Figure 1 shows a diagram of a typical system for stereo echo cancellation. Stereo signals $u_1(k)$ and $u_2(k)$, picked up by two microphones in the transmission room, are transformed as

$$x_1(k) = u_1(k) + f_1[u_1(k)], \quad x_2(k) = u_2(k) + f_2[u_2(k)], \tag{1}$$

where $f_1[\bullet]$ and $f_2[\bullet]$ are non-linear functions used to avoid the non-uniqueness problem [3]. We hereafter refer to $f_1[u_1(k)]$ and $f_2[u_2(k)]$ as additive signals.

The echo signal $y(k)$ is expressed as

$$y(k) = \mathbf{h}^T \mathbf{x}(k), \tag{2}$$

where $\mathbf{h} = [\mathbf{h}_1^T \ \mathbf{h}_2^T]^T$ is the concatenation of true impulse response vectors $\mathbf{h}_1$ and $\mathbf{h}_2$, and $\mathbf{x}(k)$ is the input signal vector defined as

$$\mathbf{x}(k) = \begin{bmatrix} \mathbf{x}_1(k) \\ \mathbf{x}_2(k) \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1(k) \\ \mathbf{u}_2(k) \end{bmatrix} + \begin{bmatrix} \mathbf{u}_{f1}(k) \\ \mathbf{u}_{f2}(k) \end{bmatrix},$$

$$\mathbf{u}_i(k) = \left[u_i(k) \ \cdots \ u_i(k-L+1) \right]^T \ (i = 1, 2),$$

$$\mathbf{u}_{fi}(k) = \left[f_i[u_i(k)] \ \cdots \ f_i[u_i(k-L+1)] \right]^T \ (i = 1, 2),$$

where L is the length of impulse response vectors. The error $e(k)$ is expressed as

$$e(k) = y(k) - \hat{\mathbf{h}}^T(k)\mathbf{x}(k), \tag{3}$$

where $\hat{\mathbf{h}}(k) = [\hat{\mathbf{h}}_1^T(k) \ \hat{\mathbf{h}}_2^T(k)]^T$ is the concatenation of adaptive-filter-coefficient vector $\hat{\mathbf{h}}_1(k)$ and $\hat{\mathbf{h}}_2(k)$.

Shimauchi et al. [2] proposed updating of the adaptive filter coefficients by the *p*-th order affine projection algorithm as

$$\begin{aligned}\hat{\mathbf{h}}(k+1) &= \hat{\mathbf{h}}(k) + \mu\,\Delta\hat{\mathbf{h}}(k),\\ \Delta\hat{\mathbf{h}}(k) &= \mathbf{X}(k)[\mathbf{X}^T(k)\mathbf{X}(k)+\delta\,\mathbf{I}]^{-1}\mathbf{e}(k),\\ \mathbf{e}(k) &= [\,y(k)\ \cdots\ y(k-p+1)]^T - \mathbf{X}^T(k)\hat{\mathbf{h}}(k),\\ \mathbf{X}(k) &= [\,\mathbf{x}(k)\ \cdots\ \mathbf{x}(k-p+1)\,],\end{aligned} \tag{4}$$

where μ is the step-size and δ is a small positive constant. We hereafter refer to $\Delta\hat{\mathbf{h}}(k)$ as the adjustment vector.

The adjustment vector can be treated as the sum of two components, a component that is effective for convergence and a redundant component. The *p*-th order affine projection algorithm tries to reduce the redundant component by removing the correlation between the consecutive input signal vectors $\mathbf{x}(k),\cdots,\mathbf{x}(k-p+1)$. The achievement of faster convergence by using a more decorrelated adjustment vector with a higher-order projection has been reported [2].

Benesty et al. [4] pointed out that, after the above straightforward affine projection, the adjustment vector still contains the redundant component caused by inter-channel cross correlation, that is, the correlation between $\mathbf{x}_1(k)$ and $\mathbf{x}_2(k)$ in input signal vector $\mathbf{x}(k)$. They proposed the introduction of inter-channel decorrelation in the affine projection algorithm. They added the constraints that the adjustment weight of each channel must be orthogonal to the input signal vectors of the other channels. However, for the projection order p = 2, these constraints more than triple the amount of computation required, as compared with the NLMS algorithm.

These multi-channel adaptive algorithms use a linear combination of input signal vectors with high inter-channel cross correlation as the adjustment vector. Even after non-linear preprocessing, the inter-channel cross correlation remains in the adjustment vector when the projection order is low. This limits the convergence speed.

3. NEW ADAPTIVE ALGORITHM

We propose a new method of obtaining a well-decorrelated adjustment vector that has a low level of computational complexity. It is obvious that the additive signals $f_1[u_1(k)]$ and $f_2[u_2(k)]$ in Fig. 1 should be designed to have a much lower inter-channel cross correlation than the original input signals $x_1(k)$ and $x_2(k)$ [5]. Increasing the contribution of these decorrelated additive signals in the adjustment vector should improve the convergence speed. This is our basic idea.

In order to achieve this, we propose that the adjustment vector be obtained as a linear combination of vectors $\mathbf{z}(k),\cdots,\mathbf{z}(k-p+1)$, where

$$\mathbf{z}(k) = a\begin{bmatrix}\mathbf{u}_1(k)\\ \mathbf{u}_2(k)\end{bmatrix} + \begin{bmatrix}\mathbf{u}_{f1}(k)\\ \mathbf{u}_{f2}(k)\end{bmatrix}. \tag{5}$$

Here, a $(0<a<1)$ is a pre-determined attenuation factor. The enhanced affine projection algorithm determines the weights of the linear combination of $\mathbf{z}(k),\cdots,\mathbf{z}(k-p+1)$ by forcing the filter coefficient vector $\hat{\mathbf{h}}(k)+\Delta\hat{\mathbf{h}}(k)$ to satisfy the last p input-output relationships, as in the conventional affine projection algorithm. As a result, the proposed algorithm is represented by

$$\begin{aligned}\hat{\mathbf{h}}(k+1) &= \hat{\mathbf{h}}(k) + \mu\,\Delta\hat{\mathbf{h}}(k),\\ \Delta\hat{\mathbf{h}}(k) &= \mathbf{Z}(k)[\mathbf{X}^T(k)\mathbf{Z}(k)+\delta\,\mathbf{I}]^{-1}\mathbf{e}(k),\\ \mathbf{e}(k) &= [\,y(k)\ \cdots\ y(k-p+1)]^T - \mathbf{X}^T(k)\hat{\mathbf{h}}(k),\\ \mathbf{X}(k) &= [\,\mathbf{x}(k)\ \cdots\ \mathbf{x}(k-p+1)\,],\\ \mathbf{Z}(k) &= [\,\mathbf{z}(k)\ \cdots\ \mathbf{z}(k-p+1)\,],\end{aligned} \tag{6}$$

where μ is the step-size and δ is a small positive constant. The sound reproduced by loudspeakers after the proposed algorithm has been applied is the same as for the conventional affine projection algorithm.

The amount of computation is almost the same for the proposed algorithm as for the affine projection algorithm because the vector $\mathbf{z}(k)$ can be generated from signal samples $a\,u_i(k)+f_i[u_i(k)]\ (i=1,2)$ and there is no need for the vector calculation of (5) for each sample. Note that, with the projection order p = 2, the amount of computation required to apply the proposed algorithm can be reduced to be almost the same as is required to apply the NLMS algorithm. This is achieved by the technique used in the fast affine projection algorithm [6].

4 . CONVERGENCE ANALYSIS

The first-order statistical behavior of the LMS algorithm [7] shows that the convergence speed of the coefficient error $\hat{\mathbf{h}}(k)-\mathbf{h}$ is primarily limited by the eigenvalue spread of the correlation matrix of input signals, that is, the ratio of the largest eigenvalue to the smallest eigenvalues. A broad spread of eigenvalues is known to make the convergence slow. In this section, we show that the same analysis can explain the improvement achieved by the proposed algorithm for the projection order $p=1$.

To make the analysis easier, we use white-noise signals that satisfy

$$E\left[\begin{bmatrix}\mathbf{u}_{f1}(k)\\ \mathbf{u}_{f2}(k)\end{bmatrix}\begin{bmatrix}\mathbf{u}_{f1}^T(k) & \mathbf{u}_{f2}^T(k)\end{bmatrix}\right] = q\,\mathbf{I}, \tag{7}$$

$$E\left[\mathbf{u}_{f1}(k)\mathbf{u}_1^T(k)\right] = \mathbf{0},$$
$$E\left[\mathbf{u}_{f2}(k)\mathbf{u}_2^T(k)\right] = \mathbf{0}$$

as the additive signals.

We may rewrite the proposed algorithm in terms of the coefficient error $\hat{\mathbf{h}}(k) - \mathbf{h}$ by substituting

$$e(k) = \mathbf{x}^T(k)\mathbf{h} - \mathbf{x}^T(k)\hat{\mathbf{h}}(k) \tag{8}$$

into the update equation (6) and by subtracting $\mathbf{h}$ from both sides of (6). Taking the mathematical expectations yields the following sample-by-sample transition formula for the mean coefficient error [8],

$$\begin{aligned} &E[\hat{\mathbf{h}}(k+1) - \mathbf{h}] = E[\mathbf{I} - \mu\ \mathbf{P}(k)]E[\hat{\mathbf{h}}(k) - \mathbf{h}] \\ &\mathbf{P}(k) = \frac{\mathbf{z}(k)\mathbf{x}^T(k)}{\mathbf{x}^T(k)\mathbf{z}(k)}, \end{aligned} \tag{9}$$

where $\mathbf{I}$ is a $2L \times 2L$ identity matrix. The correlation matrix of stereo input signal is diagonalized by using orthogonal matrix $\mathbf{C}$ as

$$\begin{aligned} &\mathbf{C}^T E\left[\begin{bmatrix}\mathbf{u}_1(k)\\ \mathbf{u}_2(k)\end{bmatrix}\left[\mathbf{u}_1^T(k)\ \ \mathbf{u}_2^T(k)\right]\right]\mathbf{C} = \begin{bmatrix}\lambda_{2L} & & 0\\ & \ddots & \\ 0 & & \lambda_1\end{bmatrix}, \\ &\lambda_{2L} \geq \cdots \geq \lambda_1 \geq 0, \\ &\mathbf{C}^T\mathbf{C} = \mathbf{C}\mathbf{C}^T = \mathbf{I}, \end{aligned} \tag{10}$$

So the matrix $E[\mathbf{P}(k)]$ can be diagonalized as

$$\begin{aligned} &\mathbf{C}^T E[\mathbf{P}(k)]\,\mathbf{C} = \frac{1}{\gamma}\left\{a\begin{bmatrix}\lambda_{2L} & & 0\\ & \ddots & \\ 0 & & \lambda_1\end{bmatrix} + \begin{bmatrix}q & & 0\\ & \ddots & \\ 0 & & q\end{bmatrix}\right\} \\ &\gamma = a\,E\left[\mathbf{u}_1^T(k)\mathbf{u}_1(k) + \mathbf{u}_2^T(k)\mathbf{u}_2(k)\right] \\ &\qquad + E[\mathbf{u}_{f1}^T(k)\mathbf{u}_{f1}(k) + \mathbf{u}_{f2}^T(k)\mathbf{u}_{f2}(k)]. \end{aligned} \tag{11}$$

When stereo input signals are correlated, the smallest eigenvalue λ_1 is almost 0 compared with the largest eigenvalue λ_{2L}, and the eigenvalue spread of $E[\mathbf{P}(k)]$ can be approximated by $(a\lambda_{2L} + q)/q$. This relationship shows that the eigenvalue spread can be made narrower by setting the attenuation factor a $(0 < a < 1)$ small. This tendency is reflected in the matrix $E[\mathbf{I} - \mu\,\mathbf{P}(k)]$, and improves the convergence speed of the coefficient error. Finding the optimum attenuation factor is a problem for future work.

This discussion of convergence is not specific to the case of two input channels, but is applicable to any number of input channels.

5. SIMULATION

We confirmed the validity of the proposed algorithm by computer simulation. The received signals were obtained by convolving an input speech signal and transmission paths measured in a conference room with a reverberation time of 200 ms. The true echo paths were also measured in a conference room and were truncated to 700 taps. The sampling frequency was 8 kHz. A white- noise signal with 40dB SNR was added to $y(k)$ as ambient noise. The adaptive filter had 600 taps on each channel. Step-size μ was set to 0.5. We used

$$f_1[u] = \frac{\alpha\ (u + |u|)}{2},\ f_2[u] = \frac{\alpha\ (u - |u|)}{2} \tag{12}$$

with $\alpha = 0.26$ as non-linear functions [9].

We evaluated the performance of the adaptive filter in terms of the misalignment defined by $\|\mathbf{h} - \hat{\mathbf{h}}'(k)\|^2 / \|\mathbf{h}\|^2$, where $\hat{\mathbf{h}}'(k)$ was the filter coefficient vector of the adaptive filter expanded to the same size as $\mathbf{h}$ by padding with zeros.

Figure 2 shows the behavior of misalignment for the conventional NLMS algorithm (the first-order affine projection algorithm, p=1) (A) without and (B) with additive signals, and for (C) the proposed enhanced affine projection algorithm with p=1 and attenuation factor a=0.06. In (A), the misalignment of the filter coefficient vector decreased rapidly in the first 1s, but soon saturated at about -4.5 dB. When non-linear transformation was applied (B), the misalignment continued to decrease slowly and reached around -5.2 dB at t = 10 s. This could be improved to –7.9 dB by the proposed enhanced affine projection algorithm (C).

Figure 3 shows the behavior of misalignment for the second-order affine projection algorithm (p=2) (A) without and (B) with additive signals, (C) the second-order affine projection algorithm with inter-channel decorrelation [4], and (D) the proposed enhanced affine projection algorithm

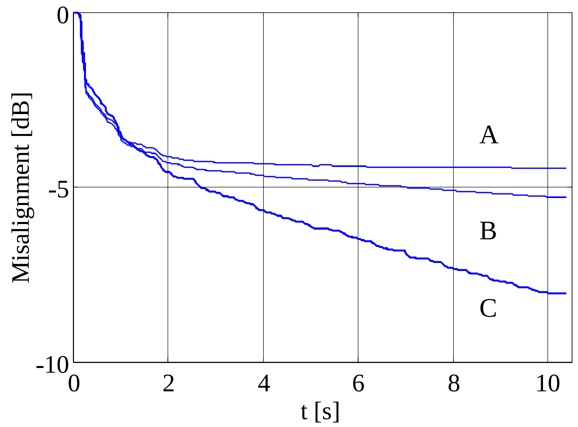

Fig. 2 Behavior of misalignment for the conventional NLMS algorithm (the first-order affine projection algorithm) (A) without and (B) with additive signals, and (C) the proposed first-order enhanced affine projection algorithm.

with p=2 and attenuation factor a=0.06.

In (A), the misalignment of the filter coefficient vector decreased rapidly in the first 1s, but soon saturated at about –4.5 dB. When non-linear transformation was applied (B), the misalignment continued to decrease slowly and reached around –7.0 dB at t = 10 s. When the second-order affine projection was combined with inter-channel decorrelation, the misalignment reached around –7.7 dB at t = 10 s. This could be improved to –14.2 dB by the proposed enhanced affine projection algorithm (D).

Figure 4 shows the behavior of misalignment when the stereo received signals were speech signals made by simulating two people speaking alternately.

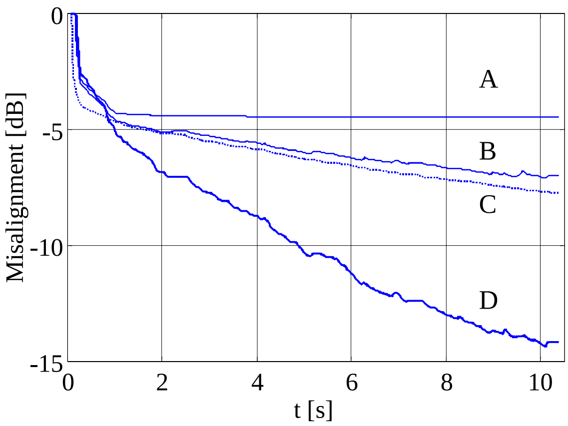

Fig. 3 Behavior of misalignment for the conventional second-order affine projection algorithm (A) without and (B) with additive signals, (C) the second-order affine projection algorithm with inter-channel decorrelation, and (D) the proposed second-order enhanced affine projection algorithm.

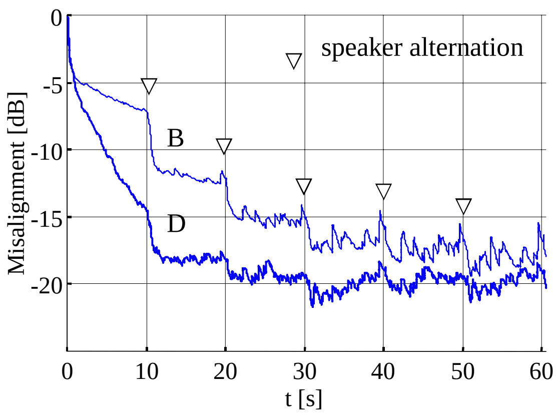

Fig. 4 Behavior of misalignment for (B) the conventional second-order affine projection algorithm with additive signals and (D) the proposed second-order enhanced affine projection algorithm.

The second-order affine projection algorithm (p=2) with additive signals (B) and the proposed enhanced affine projection algorithm (p=2, a=0.06) (D) were used. We see that the convergence of misalignment for the proposed algorithm (D) was achieved more quickly than for the second-order affine projection algorithm (B).

The proposed enhanced affine projection algorithm can improve the convergence speed drastically.

4. CONCLUSION

We propose modifying the affine projection algorithm (including the NLMS algorithm) by enhancing additive signals in the adjustment vector. Computer simulation demonstrated that this method decreased the misalignment of the adaptive filter for stereo echo cancellation much faster than the conventional first- and second-order affine projection algorithms, with almost the same amount of computation as the NLMS algorithm. The mechanism for this improvement was investigated by analyzing the first-order statistical behavior of the proposed algorithm for the projection order p = 1.

REFERENCES

[1] M. M. Sondhi, D. R. Morgan, and J. L. Hall, “Stereophonic Acoustic Echo Cancellation – An Overview of the Fundamental Problem,” *IEEE Signal Processing Letters*, vol. 2, no. 8, pp. 148-151 (1995).

[2] S. Shimauchi and S. Makino, “Stereo Projection Echo Canceller with True Echo Path Estimation,” *Proc. ICASSP95*, pp. 3059-3062 (1995).

[3] J. Benesty, D. R. Morgan, and M. M. Sondhi, “A Better Understanding and an Improved Solution to the Problems of Stereophonic Acoustic Echo Cancellation,” *Proc. ICASSP97*, pp. 303-306 (1997).

[4] J. Benesty, P. Duhamel, and Y. Grenier, “A Multi-Channel Affine Projection Algorithm with Applications to Multi-Channel Acoustic Echo Cancellation,” *IEEE Signal Processing Letters*, vol. 3, no. 2, pp. 35-37 (1996).

[5] S. Shimauchi, Y. Haneda, S. Makino, and Y. Kaneda, “New Configuration for a Stereo Echo Canceller with Nonlinear Pre-Processing,” *Proc. ICASSP98*, pp. 3685-3688 (1998).

[6] M. Tanaka, Y. Kaneda, S. Makino, and J. Kojima, “Fast Projection Algorithm and Its Step Size Control,” *Proc. ICASSP95*, pp. 945-948 (1995).

[7] B. Widrow and S. D. Stearns: *Adaptive Signal Processing*, second edition, pp. 316-331, 1985, Prentice-Hall, New Jersey.

[8] M. Tarrab and A. Feuer, “Convergence and Performance Analysis of the Normalized LMS Algorithm with Uncorrelated Gaussian Data,” *IEEE Trans. IT*, vol. 34, no. 4, pp. 680-691 (1988).

[9] P. Eneroth, T. Gaensler, S. Gay, and J. Benesty, “Studies of a Wideband Stereophonic Acoustic Echo Canceller,” *Proc. IWAENC*, pp. 207-210 (1999).