ENHANCED FREQUENCY-DOMAIN ADAPTIVE ALGORITHM FOR STEREO ECHO CANCELLATION

Satoru Emura[1], Yoichi Haneda[1], and Shoji Makino[2]

[1]NTT Cyber Space Laboratories
NTT Corporation
3-9-11, Midoricho, Musashino-shi,
Tokyo 180-8585, Japan

[2]NTT Communication Science Laboratories
NTT Corporation
2-4, Hikaridai, Seika-cho,
Kyoto 619-0237, Japan

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

ABSTRACT

Highly cross-correlated input signals create the problem of slow convergence of misalignment in stereo echo cancellation even after undergoing non-linear preprocessing. We propose a new frequency-domain adaptive algorithm that improves the convergence rate by increasing the contribution of non-linearity in the adjustment vector. Computer simulation showed that it is effective when the non-linearity gain is small.

1. INTRODUCTION

Stereo echo cancellation is indispensable for advanced teleconferencing that implements sound transmission through two channels and full-duplex communication. It is well known that the non-uniqueness problem arises when stereo received signals come from a speaker via two acoustic paths and are highly cross correlated. In this situation, the normal equation to be solved by the adaptive filter is singular [1].

Non-linear preprocessing of received signals was proposed to overcome this problem [1][2]. However, this solution is effective only when combined with a powerful adaptive algorithm, such as the FRLS algorithm [1] and the higher-order affine projection algorithm [2]. These time-domain adaptive algorithms are computationally heavy compared with the conventional NLMS algorithm.

It was shown in [3] that the enhanced time-domain adaptive algorithm for stereo echo cancellation can drastically improve the convergence rate of misalignment, though its amount of computation is almost the same as that of the NLMS algorithm.

In this paper, we extend the idea behind this algorithm to the frequency domain to reduce the amount of computation, and propose the enhanced frequency-domain adaptive algorithm for stereo echo cancellation. This algorithm reduces the effects of inter-channel cross correlation by increasing the contribution of the non-linearity in the frequency-domain adjustment vector. We also discuss the difference between the proposed algorithm and the unconstrained frequency-domain adaptive algorithm for stereo echo cancellation derived by using a frequency-domain recursive least squares criterion [4].

2. REVIEW OF ENHANCED TIME-DOMAIN ADAPTIVE ALGORITHM

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

$$\begin{aligned} x_1(k) &= u_1(k) + g_1[u_1(k)], \\ x_2(k) &= u_2(k) + g_2[u_2(k)], \end{aligned} \tag{1}$$

where k is the time index, and $g_1[\bullet]$ and $g_2[\bullet]$ are non-linear functions used to avoid the non-uniqueness problem [1][2]. The distortion due to this non-linearity has to be hardly perceptible. We hereafter refer to $g_1[u_1(k)]$ and $g_2[u_2(k)]$ as additive signals.

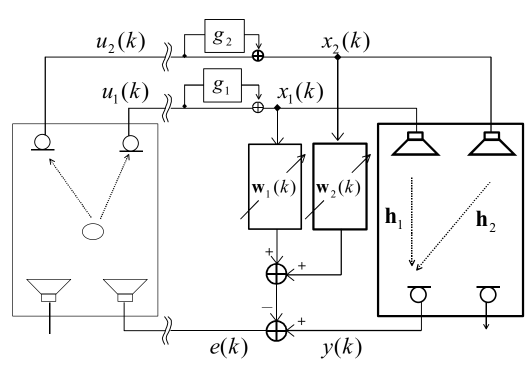

Fig.1 Diagram of system for stereo echo cancellation.

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

$$\begin{gathered} \mathbf{x}(k) = \begin{bmatrix} \mathbf{x}_1(k) \\ \mathbf{x}_2(k) \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1(k) \\ \mathbf{u}_2(k) \end{bmatrix} + \begin{bmatrix} \mathbf{u}_{g1}(k) \\ \mathbf{u}_{g2}(k) \end{bmatrix}, \\ \mathbf{u}_j(k) = \left[u_j(k) \ \ \cdots \ \ u_j(k-L+1) \right]^T \ (j = 1, 2), \\ \mathbf{u}_{gj}(k) = \left[g_j[u_j(k)] \ \ \cdots \ \ g_j[u_j(k-L+1)] \right]^T \ (j = 1, 2), \end{gathered} \tag{3}$$

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

$$e(k) = y(k) - \hat{\mathbf{h}}^T(k)\mathbf{x}(k), \tag{4}$$

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 $\hat{\mathbf{h}}_2(k)$.

The additive signals $g_1[u_1(k)]$ and $g_2[u_2(k)]$ in Fig. 1 are designed to have a much less inter-channel cross correlation than $x_1(k)$ and $x_2(k)$ [5]. Increasing the contribution of these decorrelated additive signals in the adjustment vector should improve the convergence rate. This is the basic idea of the enhanced adaptive algorithm [3]. When this idea is applied to the NLMS algorithm, the adaptive filters are updated according to

$$\begin{gathered} \hat{\mathbf{h}}(k+1) = \hat{\mathbf{h}}(k) + \mu \ \Delta\hat{\mathbf{h}}(k), \\ \Delta\hat{\mathbf{h}}(k) = \frac{e(k)}{\mathbf{z}^T(k)\mathbf{x}(k) + \delta} \mathbf{z}(k), \\ e(k) = y(k) - \mathbf{x}^T(k)\hat{\mathbf{h}}(k), \end{gathered} \tag{5}$$

where

$$\mathbf{z}(k) = \beta \begin{bmatrix} \mathbf{u}_1(k) \\ \mathbf{u}_2(k) \end{bmatrix} + \begin{bmatrix} \mathbf{u}_{g1}(k) \\ \mathbf{u}_{g2}(k) \end{bmatrix}, \tag{6}$$

$\beta \, (0 < \beta < 1)$ is a pre-determined attenuation factor, μ is the step-size, and δ is a small positive constant. Note that $\Delta\hat{\mathbf{h}}(k)$ is obtained by forcing $\hat{\mathbf{h}}(k) + \Delta\hat{\mathbf{h}}(k)$ to satisfy the last input-output relationships as in the conventional NLMS algorithm.

The amount of computation needed for this enhanced adaptive algorithm is almost the same for the conventional NLMS algorithm because the vector $\mathbf{z}(k)$ can be generated from signal samples

$$z_j(k) = \beta \ u_j(k) + g_j[u_j(k)] \, (j = 1,2).$$

Simple convergence analysis was also given in [3], where it was shown that the attenuation factor β can directly control the eigenvalue spread that decides the convergence rate of misalignment.

3. BLOCK ENHANCED ADAPTIVE ALGORITHM

In this section, we derive a block version of the enhanced time-domain adaptive algorithm, which will be the foundation of our frequency-domain adaptive algorithm.

We define the block error signal (of length L) as

$$\mathbf{e}(m) = \mathbf{y}(m) - \hat{\mathbf{y}}_1(m) - \hat{\mathbf{y}}_2(m), \tag{7}$$

where m is the block time index, and

$$\begin{gathered} \mathbf{e}(m) = \left[e(mL) \ \ \cdots \ \ e(mL+L-1) \right]^T, \\ \mathbf{y}(m) = \left[y(mL) \ \ \cdots \ \ y(mL+L-1) \right]^T. \end{gathered}$$

We can easily show that

$$\hat{\mathbf{y}}_j(m) = T_{xj}(m) \, \hat{\mathbf{h}}_j(m), \tag{8}$$

where

$$T_{xj}(m) = \begin{bmatrix} x_j(mL) & \cdots & x_j(mL-L+1) \\ \vdots & \ddots & \\ x_j(mL+L-1) & & x_j(mL) \end{bmatrix},$$

is an $L \times L$ Toeplitz matrix and $\hat{\mathbf{h}}_j(m) \, (j = 1, 2)$ are the adaptive-filter-coefficient vectors of length L.

The adaptive filters are updated according to

$$\hat{\mathbf{h}}_j(m+1) = \hat{\mathbf{h}}_j(m) + \mu \, T_{zj}^T(m)\mathbf{e}(m), \tag{9}$$

where

$$T_{zj}(m) = \begin{bmatrix} z_j(mL) & \cdots & z_j(mL-L+1) \\ \vdots & \ddots & \\ z_j(mL+L-1) & & z_j(mL) \end{bmatrix}.$$

4. PROPOSED ALGORITHM

It is well known that by doubling its size, a Toeplitz matrix $T_{xj}(m)$ can be transformed to a circulant matrix

$$C_{xj}(m) = \begin{bmatrix} T'_{xj}(m) & T_{xj}(m) \\ T_{xj}(m) & T'_{xj}(m) \end{bmatrix}, \tag{10}$$

where $T'_{xj}(m)$ is also a Toeplitz matrix.

$$T'_{xj}(m) = \begin{bmatrix} x_j(mL-L) & \cdots & x_j(mL+1) \\ \vdots & \ddots & \\ x_j(mL-1) & & x_j(mL-L) \end{bmatrix}$$

It is also well known that a circulant matrix is easily decomposed as

$$C_{xj}(m) = \mathbf{F}^{-1}\mathbf{X}_j(m)\mathbf{F}, \tag{11}$$

where matrix $\mathbf{X}_j(m)$ is a diagonal matrix whose elements are the Fourier transform of the first column of the Toeplitz matrix $T_{xj}(m)$, and matrix $\mathbf{F}$ is the Fourier matrix whose elements are given by

$$F_{kl} = \exp(-i\frac{2\pi \, k \, l}{2L}), \ 0 \le k, l \le 2L-1. \tag{12}$$

The block error in the frequency domain is given by

$$\mathbf{E}(m)=\mathbf{Y}-\mathbf{F}\begin{bmatrix}\mathbf{0}_{L\times L} & \mathbf{0}_{L\times L}\\ \mathbf{0}_{L\times L} & \mathbf{I}_{L\times L}\end{bmatrix}\mathbf{F}^{-1}\ [\ \hat{\mathbf{Y}}_1(m)+\hat{\mathbf{Y}}_2(m)\ ],\tag{13}$$

$$\hat{\mathbf{Y}}_j(m)=\mathbf{X}_j(m)\hat{\mathbf{H}}_j(m),$$

where

$$\mathbf{E}(m)=\mathbf{F}\begin{bmatrix}\mathbf{0}_{L\times L}\\ \mathbf{e}(m)\end{bmatrix},\ \mathbf{Y}(m)=\mathbf{F}\begin{bmatrix}\mathbf{0}_{L\times L}\\ \mathbf{y}(m)\end{bmatrix},\ \hat{\mathbf{H}}_j(m)=\mathbf{F}\begin{bmatrix}\hat{\mathbf{h}}_j(m)\\ \mathbf{0}_{L\times L}\end{bmatrix}.$$

The adaptive filter coefficients in the frequency domain are updated according to

$$\hat{\mathbf{H}}_j(m+1)=\hat{\mathbf{H}}_j(m)+\mu\ \mathbf{F}\begin{bmatrix}\mathbf{I}_{L\times L} & \mathbf{0}_{L\times L}\\ \mathbf{0}_{L\times L} & \mathbf{0}_{L\times L}\end{bmatrix}\mathbf{F}^{-1}\,\mathbf{Z}_j(m)\mathbf{E}(m),\tag{14}$$

where $\mathbf{Z}_j(m)$ is a diagonal matrix whose elements are the Fourier transform of the first column of the Toeplitz matrix $T_{z\,j}(m)$.

We also propose to combine "self-orthogonalization"[7] with the proposed algorithm to achieve faster convergence for the important case of correlated input signals. In this case, the adaptive filters are updated according to

$$\hat{\mathbf{H}}_j(m+1)=\hat{\mathbf{H}}_j(m)+\mu\,\mathbf{Q}(m)\mathbf{Z}_j(m)\mathbf{E}(m),\tag{15}$$

where

$$\mathbf{Q}(m)=\left[\frac{1}{q(m,1)+\delta}\ \cdots\ \frac{1}{q(m,2L)+\delta}\right],$$

$$q(m,l)=\lambda\ q(m-1,l)+(1-\lambda\ )\sum_{j=1}^{2}\left|\mathbf{Z}_j^*(m,l)\mathbf{X}_j(m,l)\right|$$

$$(1\le l\le 2L),$$

and λ is a smoothing factor. $\mathbf{Z}_j(m,l)$ and $\mathbf{X}_j(m,l)$ are the l-th element of the vectors $\mathbf{Z}_j(m)$ and $\mathbf{X}_j(m)$, respectively. The step-size μ is normalized by using an estimate of the sum of the cross-power spectrum $\Sigma_j\mathbf{Z}_j^*\mathbf{X}_j$ instead of the sum of the power spectrum $\Sigma_j\mathbf{X}_j^*\mathbf{X}_j$. We can also derive a constrained version by putting $\mathbf{Q}(m)$ between $\mathbf{F}^{-1}$ and $\mathbf{Z}_j(m)$ in (14).

Note that the proposed algorithm reduces the effect of inter-channel cross correlation by obtaining the adjustment vector from $\mathbf{Z}_j(m)$. The proposed algorithm is expected to be less sensitive to the estimation error of cross-power spectra than the multi-channel frequency-domain adaptive algorithm [4] that requires, for reducing the effect of inter-channel cross correlation, the inverse of the matrix whose elements are the estimates of the power spectra and cross-power spectra of all input signals.

The proposed algorithm can be easily generalized to the generalized multidelay filter (GMDFα) [8][9], where α is the overlap factor. This structure is very attractive, because the filter coefficients are updated more frequently (every L/α samples instead of every L samples). As a result, a faster convergence rate and better tracking are expected. For simplicity, we have derived the two-channel adaptive algorithm assuming no overlap (α=1). It is also straightforward to extend this two-channel adaptive algorithm to the multi-channel adaptive algorithm.

5. SIMULATION

We confirmed the validity of the proposed algorithm through computer simulation. The signal source s in the transmission room was a 20-s speech. The two microphone signals were obtained by convolving s with two impulse responses of length 700, which were measured in an actual room. The microphone output signal y in the receiving room was obtained by summing the two convolutions $(h_1 * x_1)$ and $(h_2 * x_2)$, where h_1 and h_2 were also measured in an actual room and were truncated to 700 taps. The sampling frequency was 8 kHz. A white-noise signal with 40-dB SNR was added to $y(k)$ as ambient noise.

We used the following parameters: $L=512$, $\alpha=4$. With these values of L and α, the proposed algorithm is 3.5 times less complex than the two-channel NLMS algorithm, and is as efficient as the two-channel unconstrained frequency-domain adaptive algorithm (two-channel UFLMS) [4]. We added a half-wave rectifier non-linearity [10] to the received signals such as

$$g_1[u]=\gamma\frac{(u+|u|)}{2},\ g_2[u]=\gamma\frac{(u-|u|)}{2},\tag{16}$$

where γ is the non-linearity gain.

We evaluated the performance of (A) the NLMS algorithm, (B) the two-channel UFLMS algorithm [4], and (C) the proposed algorithm (unconstrained, with self-orthogonalization) in terms of the misalignment defined by

$$\frac{\|\mathbf{h}_1-\hat{\mathbf{h}}_1(m)\|^2+\|\mathbf{h}_2-\hat{\mathbf{h}}_2(m)\|^2}{\|\mathbf{h}_1\|^2+\|\mathbf{h}_2\|^2}.$$

As for the two-channel UFLMS algorithm, the exponential forgetting factor λ_{f} was determined according to the relation $\mu=2(1-\lambda_{\mathrm{f}})$ derived in [4]. As for the proposed algorithm, the smoothing factor λ was determined in the same manner.

Figure 2 shows the relationship between step-size and the mean misalignment between t =15 and 20 s. This graph shows that the best step-size for the two-channel UFLMS was μ=0.1 ($\gamma=0.5$) and μ=0.2 ($\gamma=0.25$). The best step-size for the proposed algorithm (β=0.06) was μ=0.4 ($\gamma=0.5$) and μ=0.5 ($\gamma=0.25$).

Figure 3 shows the behavior of misalignment for (A) the conventional NLMS algorithm (μ=0.5), (B) the two-channel UFLMS algorithm (μ=0.1), and (C) the proposed algorithm (μ=0.4, β=0.06) when the non-linearity gain γ

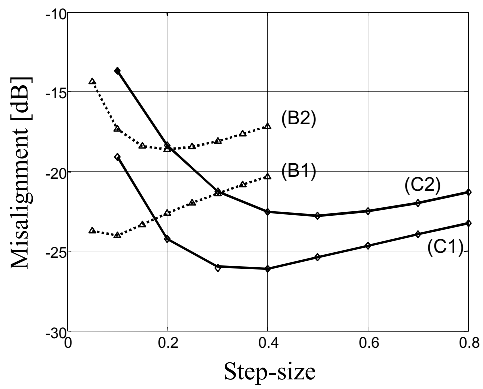

Fig. 2 Misalignment after 20-s speech for various step-sizes: (B1) the UFLMS algorithm $(\gamma = 0.5)$, (B2) the UFLMS algorithm $(\gamma = 0.25)$, (C1) the proposed algorithm $(\gamma = 0.5)$, and (C2) the proposed algorithm $(\gamma = 0.25)$.

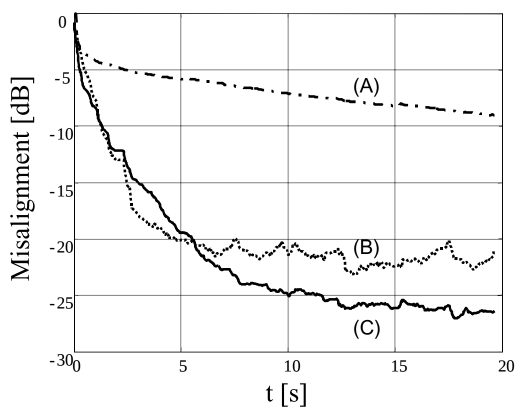

Fig. 3 Behavior of misalignment for (A) the NLMS algorithm, (B) the two-channel UFLMS algorithm, and (C) the proposed algorithm ($\gamma = 0.5$).

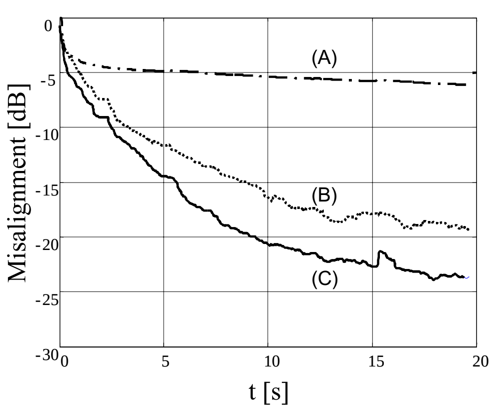

Fig.4 Behavior of misalignment for (A) the NLMS algorithm, (B) the two-channel UFLMS algorithm, and (C) the proposed algorithm ($\gamma = 0.25$).

is 0.5. The convergence rates of misalignment for the frequency-domain adaptive algorithms (B) and (C) were much better than for (A) the NLMS algorithm. The proposed algorithm (C) was as fast as (B) in terms of the time to achieve 20dB of misalignment reduction, and was better in terms of the misalignment after t =10 s.

Figure 4 shows the behavior of misalignment for (A) the conventional NLMS algorithm (μ=0.5), (B) the two-channel UFLMS algorithm (μ=0.2), and (C) the proposed algorithm (μ=0.5, β=0.06) when the non-linearity gain γ is 0.25. The proposed algorithm (C) was better in terms of both the convergence rate and the misalignment.

6. SUMMARY

We propose a new frequency-domain adaptive algorithm for stereo echo cancellation, in which the contribution of additive signals to the adjustment vector is increased. Computer simulation demonstrated that this algorithm works well when the non-linearity gain is small.

REFERENCES

[1] 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).

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

[3] S. Emura, Y. Haneda, and S. Makino, "Adaptive Algorithm Enhancing Decorrelated Additive Signals for Stereo Echo Cancellation," *Proc. IWAENC2001*, pp. 67-70 (2001).

[4] J. Benesty and D. R. Morgan, "Frequency-domain Adaptive Filtering Revisited, Generalization to the Multi-channel Case, and Application to Acoustic Echo Cancellation," *Proc. ICASSP2000*, pp. 789-792 (2000).

[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] E. R. Ferrara, Jr., "Fast Implementation of LMS Adaptive Filter," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-28, pp. 474-475 (1980).

[7] D. Mansour and A. H. Gray, "Unconstrained Frequency Domain Adaptive Filter," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-30, pp. 726-734 (1982).

[8] J. S. Soo and K. K. Pang, "Multidelay Block Frequency Domain Adaptive Filter," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. ASSP-38, pp. 373-376 (1990).

[9] E. Moulines, O. Ait Amrane, and Y. Grenier, "The Generalized Multidelay Adaptive Filter: Structure and Convergence Analysis," *IEEE Trans. Signal Processing*, vol. 43, pp. 14-28 (1995).

[10] P. Eneroth, T. Gaensler, S. Gay, and J. Benesty, "Studies of a Wideband Stereophonic Acoustic Echo Canceller," *Proc. IWAENC1999*, pp. 207-210 (1999).