A METHOD OF COHERENCE-BASED STEP-SIZE CONTROL FOR ROBUST STEREO ECHO CANCELLATION

Satoru Emura and Yoichi Haneda

NTT Cyber Space Laboratories, NTT Corporation
3-9-11, Midoricho, Musashino-shi , Tokyo 180-8585, Japan
{emura.satoru haneda.yoichi}@lab.ntt.co.jp

ABSTRACT

The fast adaptive algorithm is required for stereo echo cancellation because the strong inter-channel cross-correlation of stereo received signals leads to ill-conditioned normal equation to be solved by the adaptive filter. An adaptive algorithm for this task should also be robust against such disturbances as near-end speech and near-end noise. We propose a coherence-based method of step-size control that provides robustness in stereo echo cancellation. Computer simulation demonstrated that the method was robust against near-end speech and noise, and was capable of quickly tracking change in echo paths.

1. INTRODUCTION

Cancellation of stereo echoes is a prerequisite to providing full-duplex communication with stereophonic sound as part of advanced teleconferencing applications. This is achieved by applying a fast adaptive algorithm to estimate the multiple echo paths between multiple loudspeakers and a microphone. The strong inter-channel cross-correlation of stereo received signals leads to ill-conditioned normal equation to be solved by the adaptive filter [1],[2].

However, a high convergence rate is usually accompanied by strong and rapid divergence in the presence of near-end speech and noise. In general, fast adaptive algorithms are sensitive to such disturbances as near-end speech and near-end noise. Hence, fast adaptive algorithm for stereo echo cancellation must also be robust in this sense.

Step-size control is known to be an effective way of making monaural echo cancellation robust against noisy environments [3]. Although theory provides a derivation of the optimal step-size for the Normalized Least Mean Square (NLMS) algorithm, it is not always possible to obtain the required components from the available signals.

In this paper, we show that coherence analysis provides a numerical method for determining the optimal step-size. We then propose a coherence-based method of step-size control for robustness in stereo echo cancellation. The validity of the proposed method was verified by computer simulation.

2. THE NLMS ALGORITHM IN A NOISY ENVIRONMENT

In this section, we review the NLMS algorithm, its theoretical optimal step-size for a noisy environment, and previously proposed practical forms of step-size control.

2.1. The NLMS algorithm

Figure 1 is a diagram of a typical monaural echo cancellation system. In the NLMS algorithm, the coefficient vector $\hat{\mathbf{h}}(k)$ is updated by using a fixed step-size μ_0 as follows:

$$\hat{\mathbf{h}}(k+1) = \hat{\mathbf{h}}(k) + \mu_0 \frac{e(k)\mathbf{x}(k)}{\mathbf{x}^T(k)\mathbf{x}(k)}, \tag{1}$$

where $\mathbf{x}(k)$ is the received signal vector, defined as

$$\mathbf{x}(k) = \begin{bmatrix} x(k) & \cdots & x(k-L+1) \end{bmatrix},$$

with L as the length of the coefficient vector $\hat{\mathbf{h}}(k)$. The error signal, $e(k)$, is given as the difference between the signal at the microphone, $y(k)$, and the estimated echo signal.

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

2.2. Theoretically optimal step-size

In a noisy environment, the error signal contains a disturbing signal, $z(k)$, as well as the residual echo $r(k) = (\mathbf{h} - \hat{\mathbf{h}}(k))^T \mathbf{x}(k)$.

According to Mader et al. [3], the optimal step-size for the NLMS algorithm in a noisy environment can be theoretically derived as

$$\mu_{opt} = \frac{E[r^2(k)]}{E[r^2(k)] + E[z^2(k)]}. \tag{3}$$

In the noise-free case, this step-size, μ_{opt}, is 1.

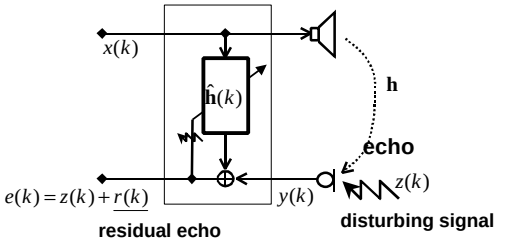

Fig. 1 A typical monaural echo cancellation system

Unfortunately, this optimal step-size is not applicable because it is not always possible to obtain the residual echo signal $r(k)$ from the available signals, $y(k)$ and $e(k)$, in a noisy environment. Despite the impossibility of always calculating optimal step-size, inspection of (2) shows that time-varying step-size should be reduced with increase in the disturbing signal power, with decrease in the input-signal power, and with decrease in the residual echo power in a noisy environment.

2.3. Previously proposed forms of step-size control

Hirano et al. have proposed an alternative equation in which the input signal power and the noise power as estimated by using $e^2(k)$ during $\overline{e^2(k)} > \{\overline{\hat{\mathbf{h}}^T(k)\mathbf{x}(k)\}^2}$ provide the basis for control of the step-size of the NLMS algorithm, thus avoiding the effect of residual echo [4]. Trump has proposed to estimate the power of the disturbing signal during natural pauses in the received speech, and the step-size of the frequency-domain adaptive algorithm is controlled accordingly [5].

3. COHERENCE-BASED STEP-SIZE CONTROL

In this section, we analyze the relation between optimal step-size in a noisy environment and the squared magnitude of coherence. With this analysis as a basis, we propose a coherence-based method of step-size control that provides robustness in stereo echo cancellation. We also show that this method can be extensible to larger numbers of channels.

3.1 Coherence analysis

To obtain the optimal step-size for application in our adaptive filter, the power of the residual echo signal should always be estimated from the available signals. For that purpose, we can use the "squared magnitude of coherence function" for a frequency *f*, which is given as

$$g^2(f) = \frac{|S_{xe}(f)|^2}{S_{xx}(f)S_{ee}(f)}, \quad (4)$$

where $S_{xx}(f)$ and $S_{ee}(f)$ are the power spectral density of the received signal $x(k)$ and the error signal $e(k)$, respectively. $S_{xe}(f)$ is the cross power spectral density of the signals $x(k)$ and $e(k)$ [6],[7].

This squared magnitude of coherence, $g^2(f)$, can be interpreted as the fractional portion of $S_{ee}(f)$ which is linearly due to $x(k)$ at frequency *f*. Since the disturbing signal, $z(k)$, can be considered statistically independent of both the received signal $x(k)$ and the residual echo signal $r(k)$,

$$\begin{aligned} &S_{ee}(f) = S_{(r+z)(r+z)}(f) \cong S_{rr}(f) + S_{zz}(f), \text{ and} \\ &|S_{xe}(f)|^2 = |S_{x(r+z)}(f)|^2 \cong |S_{xr}(f)|^2 \le S_{xx}(f)S_{rr}(f). \end{aligned} \quad (5)$$

Therefore, the relation between the optimal step-size $m_{opt}(f)$ and the coherence function $g^2(f)$ is given as

$$g^2(f) = \frac{|S_{xe}(f)|^2}{S_{xx}(f)S_{ee}(f)} \le \frac{S_{xx}(f)S_{rr}(f)}{S_{xx}(f)\{S_{rr}(f)+S_{zz}(f)\}} = m_{opt}(f). \quad (6)$$

Based on this relation, we propose the squared magnitude of coherence for each frequency bin as the corresponding estimator of optimal step-size. We see that this optimal step-size is directly applicable to the frequency-domain adaptive algorithms.

This form of coherence-based step-size control can be implemented by estimating the power spectral density $S_{xx}(f)$, $S_{ee}(f)$ and $S_{xe}(f)$ as follows:

$$\begin{aligned} S_{xx}(f) &= e[X^*(f)X(f)] \\ S_{ee}(f) &= e[E^*(f)E(f)] \\ S_{xe}(f) &= |e[X^*(f)E(f)]|^2 \end{aligned} \quad (7)$$

where $X(f)$ and $E(f)$ represent the discrete Fourier transforms of the signal frames of $x(k)$ and $e(k)$. $X^*(f)$ denotes the complex conjugate of $X(f)$. $e[\,]$ denotes taking of the short-term expectation.

3.2. Extension to the stereo case

Now consider the typical stereo echo cancellation system depicted in Fig. 2. We propose to extend the above method of coherence-based step-size control to this case.

The squared magnitude of coherence $g^2(f)$ between stereo received signals $x_1(k), x_2(k)$ and the residual echo signal $e(k)$ is defined [6] as

$$g^2(f) = 1 - (1 - g_{1e}^2(f))(1 - g_{2e\bullet1}^2(f)), \quad (8)$$

where $g_{1e}^2(f)$ is the squared magnitude of coherence between $x_1(k)$ and $e(k)$, defined as

$$g_{1e}^2(f) = \frac{|e[X_1^*(f)E(f)]|^2}{e[X_1^*(f)X_1(f)]\,e[E^*(f)E(f)]}. \quad (9)$$

$g_{2e\bullet1}^2(f)$ is the partial coherence function between $x_2(k)$ and $e(k)$, defined as

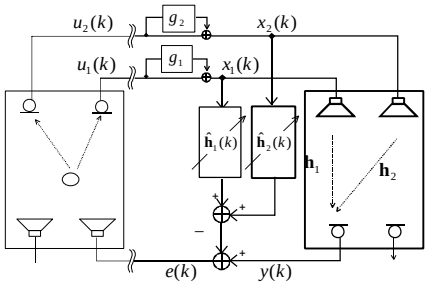

Fig. 2 A typical stereo echo cancellation system

$$g^2_{2e\bullet 1}(f)=\frac{\left|e[X^*_{2\bullet 1}(f)E_{\bullet 1}(f)]\right|^2}{e[X^*_{2\bullet 1}(f)X_{2\bullet 1}(f)]\,e[E^*_{\bullet 1}(f)E_{\bullet 1}(f)]} \quad (10)$$

where $X_{2\bullet 1}(f)$ is obtained by removing the component correlated with $X_1(f)$ from $X_2(f)$ as follows:

$$X_{2\bullet 1}(f)=X_2(f)-X_1(f)\,\frac{e[X^*_1(f)X_2(f)]}{e[X^*_1(f)X_1(f)]}. \quad (11)$$

Similarly, $E_{\bullet 1}(f)$ is obtained by removing the component correlated with $X_1(f)$ from $E(f)$ as follows:

$$E_{\bullet 1}(f)=E(f)-X_1(f)\,\frac{e[X^*_1(f)E(f)]}{e[X^*_1(f)X_1(f)]}. \quad (12)$$

3.3 Extension to the multi-channel case

Analogously with the stereo case, the squared magnitude of coherence for an M-input single-output system is defined [6] as

$$g^2(f)=1-(1-g^2_{1y}(f))\cdots(1-g^2_{My\bullet(M-1)}(f)), \quad (13)$$

where the partial coherence function $g^2_{my\bullet(m-1)}(f)$ between $x_m(k)$ and $e(k)$ is obtained as the coherence between $X_{m\bullet(m-1)}(f)$ and $E_{\bullet(m-1)}(f)$.

$$g^2_{my\bullet(m-1)}(f)=\frac{\left|e[X^*_{m\bullet(m-1)}(f)E_{\bullet(m-1)}(f)]\right|^2}{e[X^*_{m\bullet(m-1)}(f)X_{m\bullet(m-1)}(f)]} \times\frac{1}{e[E^*_{\bullet(m-1)}(f)E_{\bullet(m-1)}(f)]} \quad (14)$$

In (14), $X_{m\bullet(m-1)}(f)$ and $E_{\bullet(m-1)}(f)$ $(m=2,\cdots,M)$ are obtained by removing the components correlated with $X_1(f),\cdots,X_{m-1}(f)$ from $X_m(f)$ and $E(f)$ as follows:

$$X_{m\bullet(m-1)}(f)=X_m(f)-\sum_{i=1}^{m-1}X_{i\bullet(i-1)}(f) \times\frac{e[X^*_{i\bullet(i-1)}(f)X_m(f)]}{e[X^*_{i\bullet(i-1)}(f)X_{i\bullet(i-1)}(f)]}$$

$$E_{\bullet(m-1)}(f)=E(f)-\sum_{i=1}^{m-1}X_{i\bullet(i-1)}(f) \times\frac{e[X^*_{i\bullet(i-1)}(f)E(f)]}{e[X^*_{i\bullet(i-1)}(f)X_{i\bullet(i-1)}(f)]}$$

where $X_{1\bullet 0}(f)$ is equivalent to $X_1(f)$.

4. SIMULATION

We confirmed the validity of the proposed method of coherence-based step-size control through computer simulation. The signal source $\boldsymbol{s}$ in the transmission room was a 20-s speech signal. The two microphone signals were obtained by convolving $\boldsymbol{s}$ with two impulse responses, each 700 taps in length. Both responses 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 half-wave rectifier non-linearity [9] with a gain of 0.3 was used for non-linear functions g_1 and g_2.

We combined the proposed method and the enhanced frequency-domain adaptive algorithm [8] with the following parameters: adaptive-filter-length $L=512$, $m_0=0.4$, overlap factor $a=4$. The filter coefficients were updated every L/a samples. The length of the signal frame for coherence analysis was set to L. The short-term expectation of an expression of the form $V(j,f)$ is

$$e[V(j,f)]=l\,e[V(j-1,f)]+(1-l)V(j,f),$$

where j is the frame index and l is the forgetting factor. We used $l=0.95$. With these values, the computational complexity of the algorithm for stereo echo cancellation is raised by about one-third when we incorporate the proposed form of step-size control. We used the misalignment between responses as defined by

$$\frac{\|\mathbf{h}_1-\hat{\mathbf{h}}_1\|^2+\|\mathbf{h}_2-\hat{\mathbf{h}}_2\|^2}{\|\mathbf{h}_1\|^2+\|\mathbf{h}_2\|^2}$$

and the levels of residual echo power to evaluate the performance.

Firstly, we set the attenuation factor in the enhanced frequency-domain adaptive algorithm to 0.1, and then tested the algorithm's robustness, with and without the proposed form of step-size control against a disturbing signal. We used Hoth-noise varying between –10 dB and –14 dB SNR and near-end speech as the disturbing signals. Figure 3 shows the echo, disturbing, and microphone signals.

Figure 4 shows the behavior of the misalignment and the residual echo power. With no step-size control and the step-size fixed at $m=0.3$, 0.2, or 0.1, the adaptive filter was made to diverge at t=8 s by near-end speech. The graph of residual echo power shows that the near-end speech made the adaptive filter unstable with a fixed step-size ($m=0.1$), and that the proposed method effectively suppressed the echo during near-end speech.

Secondly, we checked the ability of the enhanced frequency-domain adaptive algorithm with the proposed step-size control to track changes in echo paths. The algorithm's attenuation factor was again set to 0.1[8]. In this experiment, we used the near-end noise as the sole disturbing signal.

Figure 5 shows the behavior of misalignment and residual echo power in response to a change in echo paths at t=10 s. The misalignment started to decrease within 128 ms after the echo path change.

5. CONCLUSION

We proposed a coherence-based method of step-size control for robustness in stereo echo cancellation. We also showed that the proposed method is extensible to the multi-channel case. Computer simulation demonstrated that the proposed method provided robustness against near-end speech and near-end noise and quickly tracked changes in echo paths.

REFERENCES

[1] J. Benesty, D. R. Morgan, and M. M. Sondhi, "A Better Understanding and an Improved Solution to the Specific Problems of Stereophonic Acoustic Echo Cancellation," *IEEE Trans. Speech Audio Processin*g, 6, pp. 156–165, 1998.

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

[3] A. Mader, H. Puder, and G. U. Schmidt, "Step-Size Control for Acoustic Echo Cancellation Filters – An Overview," *Signal Processing*, 80, pp. 1697-1719, 2000.

[4] A. Hirano and A. Sugiyama, "A Noise-Robust Stochastic Gradient Algorithm with an Adaptive Step-Size Suitable for Mobile Hands-Free Telephones," *Proc. ICASSP95*, pp. 1392-1395, 1995.

[5] T. Trump, "A Frequency Domain Adaptive Algorithm for Colored Measurement Noise Environment," *Proc. ICASSP98*, pp. 1705-1708, 1998.

[6] J.S. Bendat and A.G. Piersol, *Engineering Applications of Correlation and Spectral Analysis*, John Wiley & Sons, New York, 1980.

[7] G. Enzner, R. Martin, and P. Vary, "Unbiased Residual Echo Power Estimation for Hands-Free Telephony," *Proc. ICASSP2002*, pp. 1893-1896, 2002.

[8] S. Emura, Y. Haneda, and S. Makino, "Enhanced Frequency-Domain Adaptive Algorithm for Stereo Echo Cancellation," *Proc. ICASSP2002*, pp. 1901-1904, 2002.

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

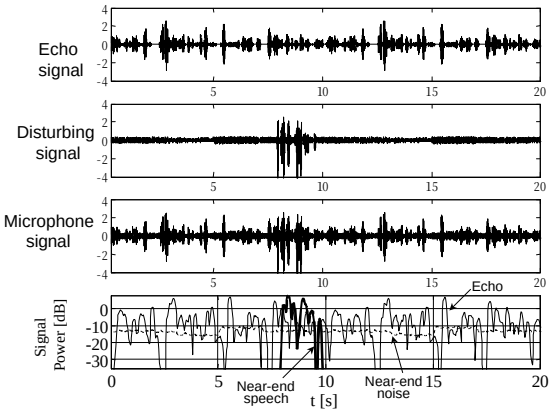

Fig. 3 The echo, near-end speech and noise, and microphone signals used in the simulation

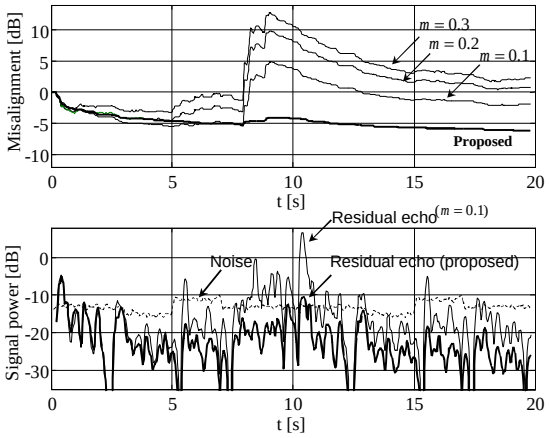

Fig. 4 Misalignment for fixed step-size ($m = 0.3, 0.2, 0.1$) and the proposed method. Residual echo power for fixed step-size $m = 0.1$ and the proposed method.

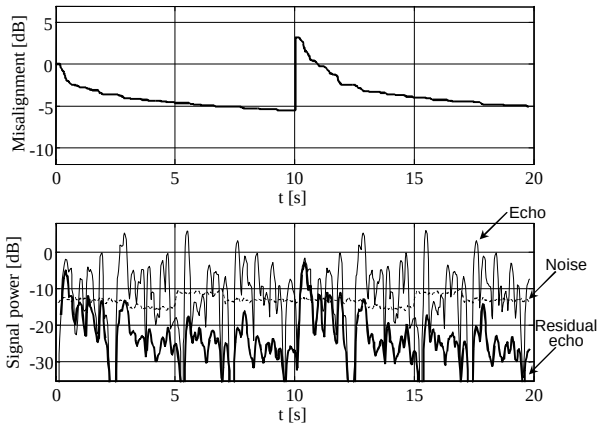

Fig. 5 Behavior of misalignment and residual echo power of the proposed method in response to a change in echo path at t=10 s.