2003 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS 2003)
Awaji Island, Japan, December 7-10, 2003 C3-4

DOUBLE-TALK RESISTANT ACOUSTIC ECHO CANCELLER WITH DOUBLE FILTERS

Akihiro Hirano, Kenji Nakayama and Shinichiro Ushimaru

Kanazawa University
40-20, Kodatsuno 2-Chome, Kanazawa, 920-8667, Japan

ABSTRACT

This paper proposes double-talk resistant echo canceller with double filters, one for echo-path estimation and one for echo cancellation. An adaptive step-size algorithm based on the cross-correlation between the echo replica and the near-end speech is used for the echo-path identification. The filter coefficients are copied from the estimation filter to the cancellation filter only when the cross-correlation is small enough. Computer simulation results show that the proposed algorithm successfully reduces echoes in a double-talk period. The tracking capability of the proposed algorithm for echo path changes is almost comparable to that of a previously proposed double-filter algorithm with fast transversal filter.

1. INTRODUCTION

Acoustic echo cancellers (AEC's) are used to reduce echoes which disturb comfortable conversation [1]. In AEC, adaptation in a double-talk period is an important problem[2], [3], [4]. As a fast convergence and stable AEC, combined fast adaptive filter (CFAF) algorithm, which uses fast transversal filter (FTS) and normalized least mean squares (NLMS)[5], has been proposed [3], [4]. Though CFAF successfully reduce echoes in a double-talk period, tracking performance to an echo-path change within a double-talk period is not enough.

This paper proposes double-talk resistant AEC with double filters, one for echo-path estimation and one for echo cancellation. An adaptive step-size algorithm based on the cross-correlation between the echo replica and the near-end speech is used for the echo-path identification. Section 2 describes the influence of the cross-correlation on the adaptation, followed by a cross-correlation estimation algorithm. An AEC with double filters is proposed, and its performance is shown by computer simulations using real speech signals.

2. DOUBLE-TALK IN ECHO CANCELLATION

A teleconferencing using an AEC shown in Fig. 1, the echo $r_1(n)$ is generated by propagation of the far-end speech signal $s_1(n)$ from the loudspeaker to the microphone in the near-end room. The AEC suppress echoes by subtracting echo replica $\hat{r}_1(n)$ from the mixture of the echo $r_1(n)$ and the near-end talker speech $s_2(n)$.

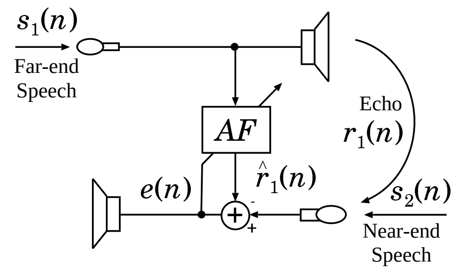


Fig. 1. Teleconferencing using AEC

Assuming an N-tap FIR adaptive filter, the echo replica $\hat{r}_1(n)$ is calculated by

$$\hat{r}_1(n) = \mathbf{w}^T(n)\mathbf{x}(n). \tag{1}$$

$\mathbf{w}(n)$ is an N-th order coefficient vector, $\mathbf{x}(n)$ is an input signal vector consists of $s_1(n) \cdots s_1(n-N+1)$, $\mathbf{w}^T(n)$ denotes the transpose of the vector $\mathbf{w}(n)$. The error signal $e(n)$, which is the AEC output, is generated by

$$e(n) = r_1(n) + s_2(n) - \hat{r}_1(n). \tag{2}$$

Assuming a normalized least mean square (NLMS)[5] algorithm, the filter coefficient vector $\mathbf{w}(n)$ is updated by

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \frac{\mu e(n)\mathbf{x}(n)}{\mathbf{x}^T(n)\mathbf{x}(n)}. \tag{3}$$

μ is a constant known as a step-size, which controls the adaptation speed and the accuracy.

In NLMS, the filter coefficients are so updated as to minimize the mean squared error $E[e^2(n)]$. In double-talk situations, $E[e^2(n)]$ becomes

$$\begin{aligned} E[e^2(n)] &= E[(r_1(n) - \hat{r}_1(n))^2] \\ &\quad + 2\underline{E[(r_1(n) - \hat{r}_1(n))s_2(n)]} \\ &\quad + E[s_2^2(n)]. \end{aligned} \tag{4}$$

In (4), the 3rd term of the right-hand side is a constant because it is independent of the filter characteristics. The 2nd term will be zero if $r_1(n) - \hat{r}_1(n)$ and $s_2(n)$ have no cross-correlation. In this case, minimizing $E[e^2(n)]$ is equivalent to minimizing the 1st term. Thus the adaptive filter can estimate the echo path.

On the other hand, the 2nd term will not be zero if $r_1(n) - \hat{r}_1(n)$ and $s_2(n)$ are cross-correlated. Usually, the far-end speech $r_1(n)$ and the near-end speech $s_2(n)$ are statistically independent. However, a large step-size might result in a strong cross-correlation because of a short-term averaging of such terms.

Minimizing $E[e^2(n)]$ in such a situation might cause a poor estimate of the echo path. Therefore, adaptation control based on the cross correlation between the near-end speech and the far-end speech is necessary.

3. ESTIMATION OF CROSS-CORRELATION

3.1. Approximation

In actual application, it is impossible to calculate the cross-correlation between the near-end speech and the far-end speech. Therefore, it should be estimated using available signals. The cross-correlation can be approximated by using $\alpha(n)$ and $\beta(n)$ defined by

$$\begin{aligned} \alpha(n) &= E[\hat{r}_1(n)(r_1(n) + s_2(n))] \\ &= E[\hat{r}_1(n)r_1(n)] + E[\hat{r}_1(n)s_2(n)] \end{aligned} \qquad (5)$$

$$\beta(n) = E[\hat{r}_1^2(n)]. \qquad (6)$$

After the convergence of the adaptive filter, we can assume

$$E[\hat{r}_1^2(n)] \simeq E[\hat{r}_1(n)r_1(n)]. \qquad (7)$$

In this case, the difference between $\alpha(n)$ and $\beta(n)$, say $\gamma(n)$, leads us to

$$\begin{aligned} \gamma(n) &= \alpha(n) - \beta(n) \\ &= E[\hat{r}_1(n)r_1(n)] + E[\hat{r}_1(n)s_2(n)] \\ &\quad -E[\hat{r}_1^2(n)] \\ &\simeq E[\hat{r}_1(n)s_2(n)]. \end{aligned} \qquad (8)$$

Thus we can estimate the cross-correlation between the echo replica $\hat{r}_1(n)$ and the near-end speech $s_2(n)$. The cross-correlation normalized by the power of the far-end speech $s_1(n)$,

$$\Gamma(n) = \frac{\gamma(n)}{E[s_1^2(n)]}, \qquad (9)$$

will be used to adaptation control.

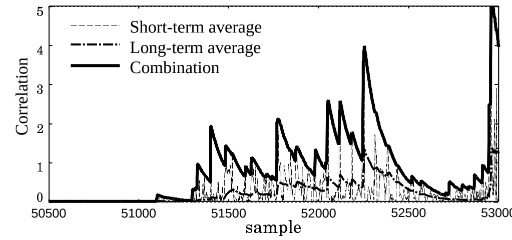


Fig. 2. Example of Correlation Estimation

3.2. Accurate Estimation and Fast Tracking by Short-Term and Long-Term Averages

In the estimation of the cross-correlation, time-averaging is used instead of the ensemble average. Thus we have trade-off between the accuracy and the tracking speed. To overcome such trade-off, a short-term and a long-term averaging are combined.

By a 1st-order leaky integrator, the estimation by the short-term average, $\Gamma_1(n)$, is calculated by the following equations.

$$\begin{aligned} \alpha(n) &= \delta_1\alpha(n-1) \\ &+ (1-\delta_1)\hat{r}_1(n)(r_1(n) + s_2(n)) \end{aligned} \qquad (10)$$

$$\beta(n) = \delta_1\beta(n-1) + (1-\delta_1)\hat{r}_1^2(n) \qquad (11)$$

$$\sigma_{s_1}(n) = \delta_1\sigma_{s_1}(n-1) + (1-\delta_1)s_1^2(n) \qquad (12)$$

$$\Gamma_1(n) = \frac{\alpha(n) - \beta(n)}{\sigma_{s_1}(n)} \qquad (13)$$

δ_1 is a constant which satisfies $1 > \delta_1 > 0$. The long-term average is calculated by

$$\Gamma_2(n) = \delta_2\Gamma_2(n-1) + (1-\delta_1)\Gamma_1(n), \qquad (14)$$

where another constant δ_2 which satisfies $1 > \delta_2 > \delta_1 > 0$ is used.

In the beginning of the double-talk periods, the cross-correlation rapidly grows. To track such change, $\Gamma_2(n)$ is replaced by $\Gamma_1(n)$ when $|\Gamma_1(n)| > |\Gamma_2(n)|$. After the end of the double-talk periods, $\Gamma_2(n)$ might keep a larger value because of the long time-constant δ_2 of the integrator. Larger estimate causes slower convergence. Therefore, $|\Gamma_2(n)|$ is replaced by the maximum value of $|\Gamma_1(n)|$ in the past T_2 samples if $|\Gamma_1(n)| < |\Gamma_2(n)|$ is satisfied in T_1 samples.

Figure 2 demonstrates the estimation by the proposed method. Fast tracking and stable estimate can be achieved by the combination of the short-term and long-term averages.

4. AEC USING DOUBLE FILTERS

4.1. Structure

Figure 3 depicts the structure of the proposed AEC. Two filters, AF1 and AF2, are used. The filter coefficients for AF2

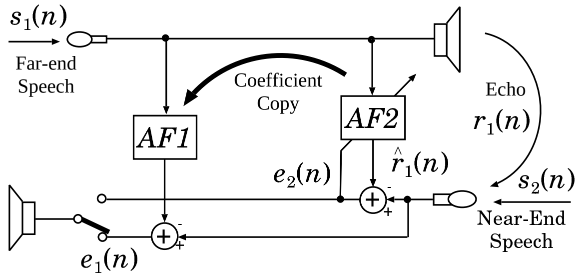


Fig. 3. AEC using double filters

are updated by an adaptive step-size NLMS algorithm. The step-size of AF2 is controlled based on the cross-correlation $\Gamma_2(n)$. The filter coefficients for AF1 is transferred from AF2. Either the AF1 output $e_1(n)$ or the AF2 output $e_2(n)$ is used as the AEC output.

4.2. Adaptation Control Based on Cross Correlation

The step-size $\mu(n)$ for AF2 is selected from μ_H and μ_L, where

$$1 \geq \mu_H > \mu_L \geq 0. \tag{15}$$

The larger step-size μ_H is selected when the cross-correlation $\Gamma_2(n)$ is smaller than the threshold θ_1. The coefficient transfer is carried out when both

$$\Gamma_2(n) < \theta_2 \tag{16}$$

$$\frac{1}{T_3}\sum_{i=0}^{T_3-1} s_1^2(n-i) > \theta_3 \tag{17}$$

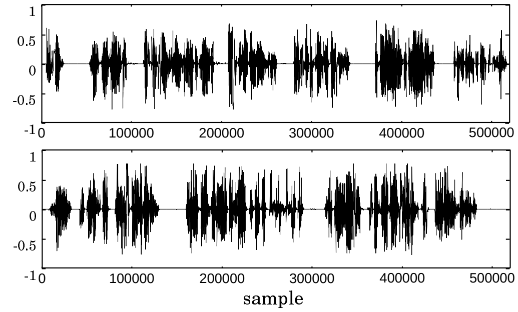
are satisfied for T_4 sample periods. θ_2 is greater than θ_1, thus the coefficient transfer requires more strict condition than the adaptation.

4.3. Echo-Path Change Detection

In a double-filter structure, the performance would be degraded between an echo-path change and the next coefficient transfer. To overcome this, an echo-path change detection is introduced. Both the step-size and the AEC output are controlled.

Between the echo-path change and the coefficient transfer, AF2 error $e_2(n)$ is expected to be smaller than AF1 error $e_1(n)$. In the double-talk periods, however, $e_2(n)$ could become smaller because of the near-end speech cancellation caused by the cross-correlation. Since the duration of such near-end speech cancellation is not so long, the averaged error power $P_i(n)$ $(i = 1, 2)$ defined by

$$P_i(n) = \frac{1}{T_5}\sum_{i=0}^{T_5} e_i^2(n-i) \tag{18}$$



$s_1(n)$ (Upper) and $s_2(n)$ (Lower)

Fig. 4. Near-End and Far-End Speech Signals

can be used to echo-path change detection. The echo-path change is detected when

$$P_1(n) > P_2(n) \tag{19}$$

is true for T_6 sample periods.

If the echo-path change occurs in single-talk periods, it is desirable to use a large step-size and to user $e_2(n)$ as the AEC output for the fast tracking. On the other hand, the stability is more important within double-talk periods. Therefore, if an echo-path change is detected in a single-talk period, i.e.,

$$\Gamma_2(n) < \theta_4 \tag{20}$$

$$\theta_4 < \theta_2 \tag{21}$$

then $e_2(n)$ is used as the AEC output. The maximum step-size

$$\mu(n) = \mu_{Max} = 1 \tag{22}$$

is used. The AEC returns to the normal operation if

$$\Gamma_2(n) \geq \theta_4. \tag{23}$$

If an echo-path change is detected in a double-talk period, the AEC works as usual.

4.4. Computational Costs

For N-tap FIR filters, the proposed structure requires $O(3N)$ operations: N for AF1 and $2N$ for AF2. Even for multiprocessor implementation, no coefficient copy operations are necessary. By preparing two coefficient buffers and by switching one from two, copy operations can be replaced by switching from a buffer to another. In multi-processor case, AF1 and AF2 are divided into cascaded subsections. Each processor performs computation for for both AF1 and AF2 which shares the same coefficients.

5. COMPUTER SIMULATIONS

Simulations have been carried out to show the performance of the proposed AEC. The proposed algorithm is compared

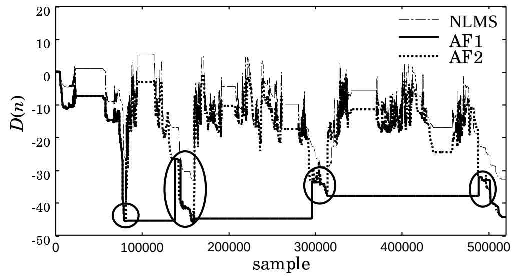
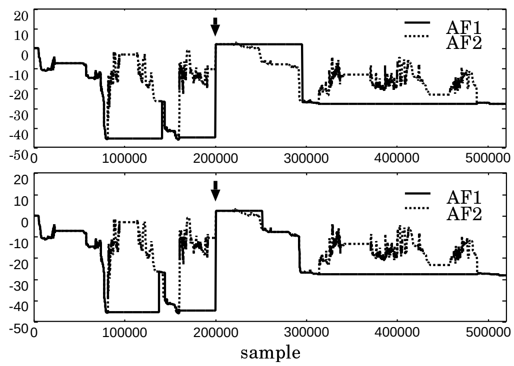


Fig. 5. Coefficient Error for Double-Talk



Without Detection (Upper) and With Detection (Lower)

Fig. 6. Tracking Performance for Echo-Path Change

with the NLMS algorithm. As a structure with multiple filters, The CFAF algorithm which uses FTF and NLMS[3], [4] is also compared.

Figure 4 depicts the far-end speech $s_1(n)$ and the near-end speech $s_2(n)$. Recorded real speech signals have been used. The echo path is a Butteworth low-pass filter whose cut-off frequency is 0.4. As AF1 and AF2, 64-tap FIR filters are used.

Parameters have been so selected as to achieve the smallest coefficient error norm. Normalized coefficient error norm defined by

$$D(n) = 10 \log \frac{|\mathbf{h} - \mathbf{w}(n)|^2}{|\mathbf{h}|^2} \tag{24}$$

is used, where $\mathbf{h}$ is the echo-path impulse response. For the proposed algorithm, $\delta_1 = 0.9$, $\delta_2 = 0.998$, $T_1 = 150$, $T_2 = 64$, $T_3 = 64$, $T_4 = 100$, $T_4 = 4000$, $T_5 = 200$, $\theta_1 = 0.1$, $\theta_2 = 0.002$, $\theta_3 = 0.05$, $\theta_4 = 0.05$ $\mu_H = 0.3$ and $\mu_L = 0.01$ have been used. For the NLMS, $\mu = 0.03$ has been chosen. Both in the proposed and the NLMS, adaptation is not carried out if $|\mathbf{x}(n)|^2 < 1000$ to avoid the degradation when the far-end speech is too small[2]. Figure 5 demonstrates the performance in double-talk periods. The coefficient errors for the NLMS, AF1 and AF2 are compared. The NLMS is unstable. Though the coefficient error for AF2 are degraded within some critical periods shown by the circles, AF1 which generates the AEC output works fine.

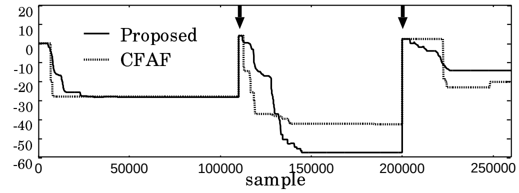


Fig. 7. Comparison with CFAV

The tracking performance for echo-path changes with and without the detection are compared by Fig. 6. The arrow shows the echo-path change. The tracking speed with the detection is twice as fast as that without the detection.

Figure 7 compares the performance of the proposed and CFAF. The proposed algorithm achieves 20dB higher performance between 150000 and 200000 samples. The time-delay between second change and the tracking start for the proposed algorithm is almost 1/4 compared with that for CFAF. Note that the computational costs for CFAF, $O(9N)$, is three times larger than that for the proposed algorithm.

6. CONCLUSION

A double-talk resistant echo canceller with double filters has been proposed. The adaptation is controlled by the cross-correlation between the echo replica and the near-end speech. Computer simulation results show that the proposed algorithm successfully reduces echoes in a double-talk period. The tracking capability of the proposed algorithm for echo path changes is almost comparable to that of a previously proposed double-filter algorithm with fast transversal filter.

7. REFERENCES

[1] Simon Haykin, "Adaptive Filter Theory Third Edition," Prentice Hall, 1996.

[2] A. Hirano and A. Sugiyama, "A Modified Normalized LMS Algorithm based on a Long-Term Average of the Reference Signal Power," IEICE Trans. on Fundamentals, Vol. E78-A, No. 8, pp. 915-920, 1995.

[3] Y. Wang, K. Nakayama, "A Combined Fast Adaptive Filter Algorithm with an Automatic Switching Method," IEICE Trans. Vol. E77-A, No.1, pp.247–256, 1994.

[4] Y. Wang, Y. Terada, M. Matsui, K. Iida and K. Nakayama, "Depelopment of high quality acoustic subband echo canceller using dual-filter structure and fast recursive least squares algorithm," Proc. of ICASSP'2000, pp.VI-3674–3677, 2000.

[5] J. Nagumo and A. Noda, "A Learning Method for System Identification," IEEE Trans. on Automatic Control, Vol. AC-12, No. 3, pp. 282–287, 1967.