NORSIG The 4th International Workshop on Acoustic Echo and Noise Control
Røros, Norway, 21 - 23 June 1995
THE INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, INC NORWAY SECTION

Implementation and Evaluation of an Acoustic Echo Canceller using the Duo-Filter Control System

Yoichi Haneda, Shoji Makino, Junji Kojima and Suehiro Shimauchi

NTT Human Interface Laboratories
3-9-11, Midori-cho, Mushashino-shi, Tokyo, 180, Japan
Tel.: +81 422 59 3108; Fax.: +81 422 60 7811
E-mail: haneda@ac-lab.ntt.jp

ABSTRACT

The developed acoustic echo canceller uses an exponentially weighted step-size projection algorithm and a duo-filter control system to achieve fast convergence speed and high speech quality. The duo-filter control system has an adaptive filter and a fixed filter (foreground/background structure), and uses variable loss insertions. Evaluation of this system with multi-channel A/D and D/A converters showed that (1) the convergence speed is under 1.5 seconds for speech input when the adaptive filter length is 125 ms, (2) the residual echo level is nearly as low as the ambient noise level (average: under -20 dB; maximum: under -35 dB), and (3) near-end speech is sent with no disturbance during double talk.

1. INTRODUCTION

Acoustic echo cancellers (AECs) with adaptive filters are widely used for full-duplex hands-free communication. However, there are several problems with implementing these AECs in an actual environment [1].

One problem is how to improve the convergence speed of the adaptive filter for speech input while minimizing computational complexity. The conventional NLMS algorithm is widely used in AECs due to its ease of computation, but its convergence speed is too slow for speech input. This is because the conventional NLMS algorithm does not decorrelate the speech signal. Many algorithms have been proposed to improve the convergence speed for speech input that are based on the characteristics of the speech signal. The exponentially weighted step-size projection (ES projection) algorithm [2] is a very promising algorithm for speech input.

A more difficult problem is when near-end and far-end speakers talk simultaneously (double talk). If the adaptive filter updates the filter coefficients during double talk, it adds an unnecessary echo signal. Therefore, during double talk, the adaptive filter should stop updating the filter coefficients. However, detecting double talk is difficult because distinguishing it from an echo path change is not easy. Furthermore, an AEC should detect double talk before the adaptive filter becomes misadjusted. Several methods have been proposed for detecting double talk, including the use of the acoustic coupling value [3] and the use of an adaptive filter and fixed filter [4].

While adaptive filters can achieve full-duplex communication, they are limited by their convergence speed and the echo cancelling level. Therefore, other echo-suppression methods, such as variable loss insertions and center clipping, are used to augment the echo cancelling.

Considering these three basic problems, we have developed an acoustic echo canceller that uses the ES projection algorithm [2]. This AEC has a foreground/background (FG/BG) structure [4] and variable loss insertions. We call this acoustic echo cancelling system the "duo-filter control system". In this paper we will describe our system and evaluate its performance.

2. SYSTEM DESCRIPTION

2.1. ES projection algorithm

To achieve fast convergence for speech input, our AEC uses the second-order ES projection algorithm. This algorithm uses both the exponentially weighted step size based on the expected variation in a room impulse response and the effect of speech whitening. While its computational power is the same as that of the conventional NLMS, it converges four times faster [2].

The algorithm is expressed as

$$\mathbf{z}(k+1) = \mathbf{z}(k) + \mu\mathbf{A}\,[\beta_1(k-1) + \beta_2(k)]\mathbf{x}(k-1),$$

$$\hat{y}(k) = \mathbf{z}(k)^T\mathbf{x}(k) + \mu\beta_1(k-1)r_{10},$$

and

$$e(k) = y(k) - \hat{y}(k) + n(k),$$

where $\mathbf{z}(k)$ is the intermediate variable,

$$\beta_1(k) = \frac{e(k)r_{11} - (1-\mu)e(k-1)r_{10}}{r_{00}r_{11} - r_{10}r_{10} + \delta}$$

and

$$\beta_2(k) = \frac{(1-\mu)e(k-1)r_{00} - e(k)r_{10}}{r_{00}r_{11} - r_{10}r_{10} + \delta}.$$

Here,

$$r_{00} = \mathbf{x}(k)^T\mathbf{A}\,\mathbf{x}(k),$$

$$r_{10} = \mathbf{x}(k-1)^T\mathbf{A}\,\mathbf{x}(k),$$

$$r_{11} = \mathbf{x}(k-1)^T\mathbf{A}\,\mathbf{x}(k-1),$$

and

$$\mathbf{A} = \begin{pmatrix} \alpha_1 & & & 0 \\ & \alpha_2 & & \\ & & \ddots & \\ 0 & & & \alpha_L \end{pmatrix},$$

where $\alpha_i = \alpha_0\gamma^{i-1}$ (i=0, $\cdots$, L) [γ : exponential attenuation ratio].

Intermediate variable $\mathbf{z}(k)$ is related to impulse response replica $\hat{\mathbf{h}}(k)$:

$$\mathbf{z}(k) = \hat{\mathbf{h}}(k) - \mu\mathbf{A}\beta_1(k-1)\mathbf{x}(k-1).$$

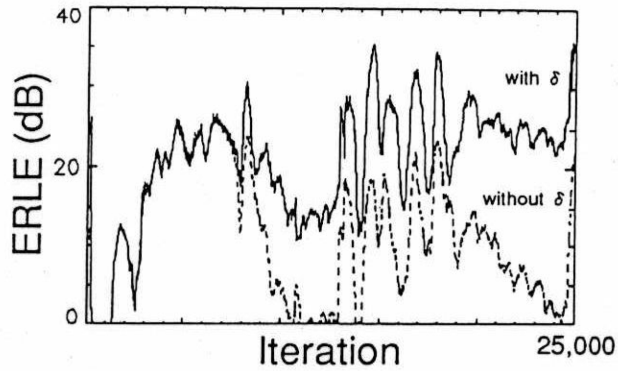

Figure 1. Effect of δ.

The δ in β_1(k) and β_2(k) is a small positive value to avoid dividing by a small denominator. Figure 1 shows the effect of δ. The algorithm with δ receives a stable echo-return loss enhancement (ERLE), the one without δ does not. Adding δ is thus effective for speech with a rapidly changing dynamic range.

2.2. Duo-Filter Structure

An AEC has to distinguish between double talk and an echo path change to determine whether to freeze or update the coefficients.

To handle this requirement, we use a foreground/ background (FG/BG) structure, which has a fixed filter and an adaptive filter. The fixed filter reduces the echo, and the adaptive filter estimates the room impulse response. The coefficients for the fixed filter are transferred from the adaptive filter when the adaptive filter converges. During double talk, the coefficients of the adaptive filter may become misadjusted, increasing the mean-squared error in the adaptive filter. If this occurs, the filter coefficients of the adaptive filter are not transferred to the fixed filter. The filter coefficients of the fixed filter are therefore not updated during double talk, and the echo cancelling level before double talk is maintained.

We use the ES projection algorithm for the adaptive filter in the FG/BG structure. This algorithm uses the intermediate variable $\mathbf{z}$(k) and the fixed filter uses the impulse response replica $\hat{\mathbf{h}}(k)$, so we call this system the "Duo-filter" system. A block diagram of this system is shown in Fig. 2. Intermediate variable $\mathbf{z}$(k) is converted to impulse

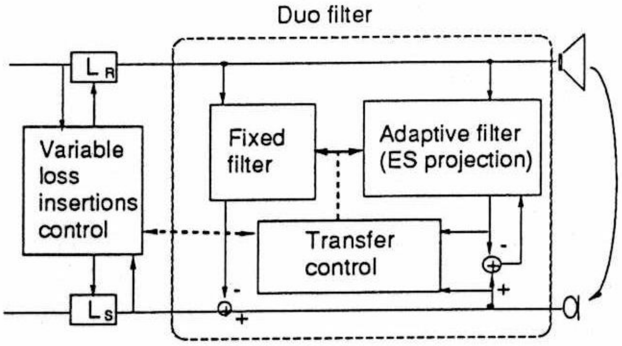

Figure 2. Duo filter and variable loss insertions.

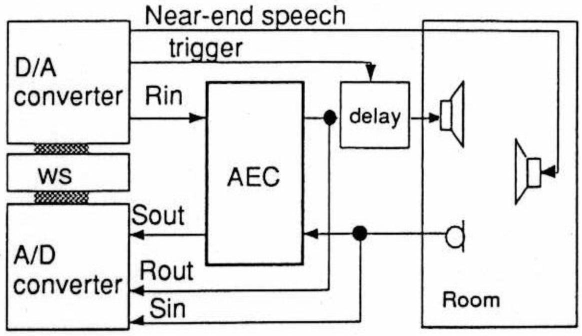

Figure 3. AEC evaluation system.

response replica $\hat{h}(k)$ before the filter coefficients are transferred.

2.3. Variable Loss Insertions

Our AEC system is designed to be used without any prior training sequence and with a maximum acoustic coupling level between the loudspeaker and microphone of + 9 dB. When the acoustic coupling level is above 0 dB, the AEC may begin howling immediately after it is turned on if there is no prior training. It may also begin howling when the echo path changes rapidly.

To prevent howling, variable losses are inserted into the system, as shown in Fig. 2. If the loss levels are too large, they may produce subjective dissatisfaction, so they are kept as small as the howling margin level. The howling margin level is determined from the acoustic coupling level, which should be measured only when far-end speech exists. Our AEC system determines that only far-end speech is active if the adaptive filter has converged. Once the duo-filter convergence level is brought below the howling margin level, the variable loss levels are equal to 0 dB.

3. EVALUATION

We implemented our AEC system with a frequency range from 100 Hz to 7 kHz on DSPs. The filter length was 125 ms. The speech transmission delay was less than 3 msec in each direction. We measured the convergence speed and performance during double talk [5, 6] with the multiple-channel A/D and D/A converter configuration shown in Fig. 3. The D/A converter output the received signal and the near-end speech signal; the A/D converter measured the Rout, Sin, and Sout. The room was 70 m^3 with a reverberation time of 250 msec. The loudspeaker and microphone were placed according to ITU Recommendation G.165 [5].

3.1. Convergence Speed

The convergence speed after an echo path change was measured by using several input signals. To simulate an echo path change, the delay time was changed by a trigger output from another channel of the D/A converter.

The measurement sequence was as follows:

(1) converge the AEC
(2) input received signal into Rin
(3) measure Sin and Sout
(4) after 2 sec, change delay time to simulate echo path change.

Figure 4 shows the convergence speed for three Rin input signals (far-end speech signals) and the average of 50 Japanese and English speech trials. The convergence speed for a decay in the mean-squared error level of -20 dB was under 1.5 s in all cases. The ES projection algorithm is thus suitable for processing speech input signals.

3.2. Performance during double talk

To simulate double talk, the D/A converter output the received signal into Rin and the near-end speech signal into the mouth simulator, while the A/D converter measured Rout, Sin, and Sout. Sout(dt) is the Sout signal during double talk, and Sout(st) is the Sout signal with only near-end speech. As shown in Fig. 5, the difference between Sout(dt) and Sout(st) was very small, and subjective quality was good.

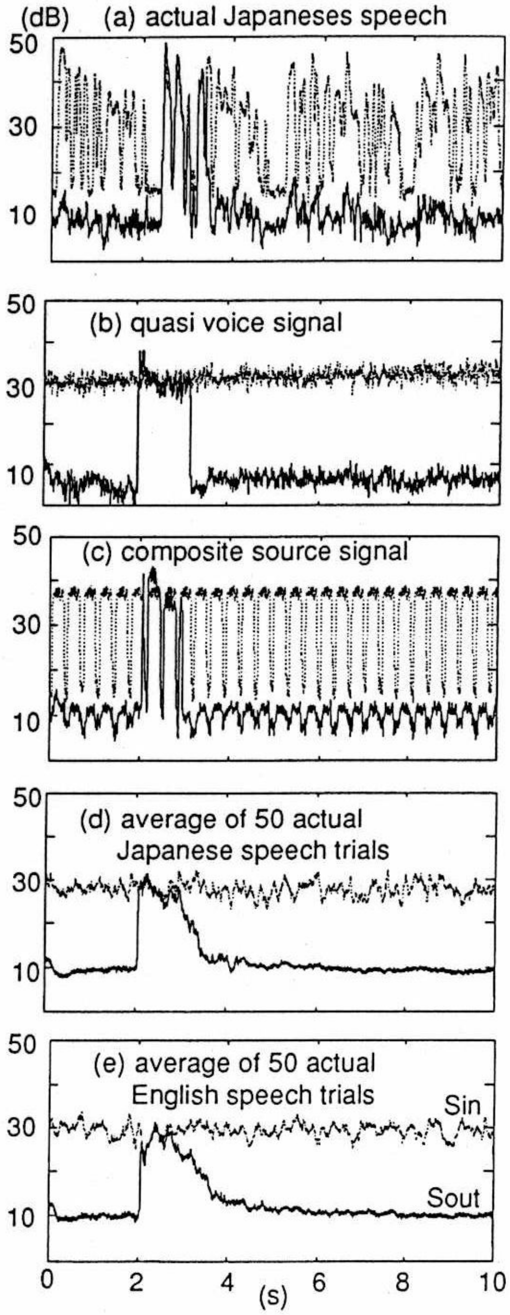

Figure 4. Convergence speeds for four input signals.

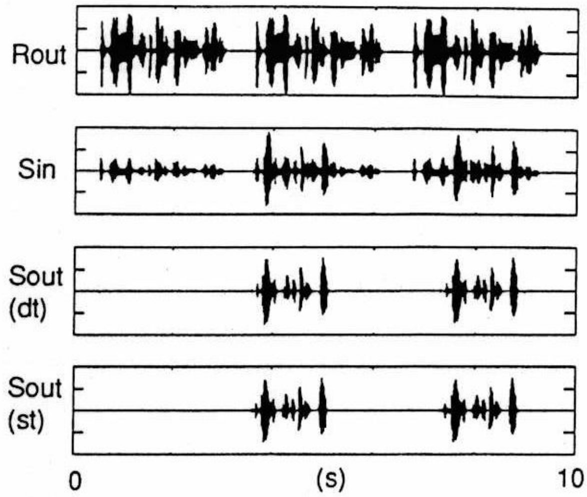

Figure 5. Performance during double talk.

4. CONCLUSION

We have developed an acoustic echo canceller that uses an exponentially weighted step-size projection algorithm and a duo-filter control system. Evaluation of this AEC with multi-channel A/D and D/A converters showed that it has fast convergence speed and good performance during double talk.

ACKNOWLEGEMENTS

We are very grateful to Dr. Nobuhiko Kitawaki, Dr. Yutaka Kaneda, Mr. Masashi Tanaka and Mr. Hisafumi Nishi for their support and constructive suggestions.

References

[1] E. Haensler, "The hands-free telephone problem: an annotated bibliography update, " Ann. Telecommun., 49, no. 7-8, pp. 360-367, 1994.
[2] S. Makino and Y. Kaneda, "Exponentially weighted step-size projection algorithm for acoustic echo cancellers," Trans. IEICE Japan, vol. E75-A, no. 11, pp. 1500-1508, 1992.
[3] P. L. Chu, "Weaver SSB subband acoustic echo canceller," Proceedings 3rd IWAEC in France, pp. 173-176, 1993.
[4] K. Ochiai, T. Araseki and T. Ogihara, "Echo Canceller with Two Echo Path Models," IEEE Trans. Commun. COM-32, pp. 589-595, 1977.
[5] Contribution of ITU, G.165 and G.167.
[6] A. Gillore, "Performance evaluation of acoustic echo controls required values and measurement procedures, " Ann. Telecommun., vol. 49, no. 7-8, pp. 368-372, 1994.