A METHOD OF DIFFRENTIATING BETWEEN DOUBLE-TALK AND ECHO PATH CHANGE

Kensaku Fujii and Juro Ohga
Network systems laboratories, FUJITSU LABORATORIES LTD.
fujiken@flab.fujitsu.co.jp

ABSTRACT

This paper presents a method of differentiating between double-talk and echo path change requisite for holding the acoustic coupling gain stable. In an acoustic echo canceller system, powers of the reference signal and the environmental noise incessantly fluctuate. The conventional system has been designed so as to suspend the estimation process while either power is outside a desirable range. This paper introduces a method to continue the estimation process even in that while, by using the block implementation of the normalized least mean square (NLMS) algorithm and by adjusting that block length.

1. INTRODUCTION

The NLMS algorithm is widely used to estimate the coefficients of an adaptive finite impulse response (FIR) filter synthesizing replicas of acoustic echoes which were returned through acoustic paths from a loud speaker to a microphone. The estimation error is equivalent to the gain of the acoustic paths, which is evaluated as a function of the step gain and the power ratio of reference signal to additive noise. In the acoustic echo canceller system, near-end talker's signal irregularly adds to the additive noise. This addition, which is called 'double-talk', lowers the power ratio, and consequently increases the estimation error. Conventionally, the increase is prevented by suspending the estimation process immediately if the double-talk was detected.

A difficulty is that even a slight reduction of the power ratio must be detected to hold the estimation error at a desirable amount. For holding the acoustic path gain at –40 dB, for example, a threshold for the detection must be situated at 30 dB when the step gain is fixed at 0.1. This means that the near-end talker's signal is almost hidden in the acoustic echoes. The residual echo, left after the acoustic echoes were cancelled, is available for the detection [1]. This detection method, however, requires differentiating between the double-talk and the echo path change because both increase the residual echo. This paper introduces a method for the differentiation [2], [3].

The double-talk detection is naturally accompanied with some delays. The acoustic path gain will probably increases until the estimation process is suspended after an occurrence of double-talk. The simplest way of preventing the increase is to apply the delayed-x NLMS algorithm in which the reference signal and the residual echo are delayed by the time necessary for the detection [4]. However, this way reduces the convergence rate because of requiring the application of a small step gain [5].

This paper also presents a method of continuing the estimation process even while the power ratio is below a predetermined limit to guarantee a desirable acoustic path gain. The method uses the block implementation [6] of the NLMS algorithm and holds the desirable gain by inversely proportioning its length to the power ratio.

2. DIFFERENTIATION METHOD

2.1 Parameter for differentiation

The coefficients of the adaptive FIR filter synthesizing the echo replica, G_j, are fixed immediately after the increase of the residual echo, E_j, was detected. The residual echo can be regarded as independent of the echo replica in the double-talk, if the fixing operation was successfully performed before the coefficients are corrupted by the near-end talker's signal. Then, this product sum,

$$P_{EG}(n) = \sum_{j=nM+1}^{(n+1)M} E_j G_j \, , \tag{1}$$

can be dealt as a random variable whose mean value is approximated to zero under a sufficiently large M, because the near-end talker's signal and the environmental noise can be also supposed to be independent of the echo replica.

On the other hand, the echo path change leaves the echo replica in the residual echo. Therefore the fixing operation provides a nonzero mean value to the product sum. This mean value is equal to the power of the echo replica if the echo and its replica are completely independent of each other. The difference of these two mean values is available for the differentiation between the double-talk and the echo path change. However, the difference fluctuates with the far-end talker's signal power. A small difference makes the differentiation difficult. The difficulty can be removed [2], [3] by the normalization using

$$P_{GG}(n) = \sum_{j=nM+1}^{(n+1)M} G_j^{\,2} \, . \tag{2}$$

The normalized cross-correlation,

$$R_{EG}(n) = P_{EG}(n)/P_{GG}(n), \tag{3}$$

provides the stable difference independently of the far-end talker's signal power fluctuation.

The normalization is also possible by using

$$P_{yy}(n) = \sum_{j=nM+1}^{(n+1)M} y_j^{\,2}. \tag{4}$$

Since the echo canceller input signal, y_j, is approximated to G_j in the echo path change,

$$R_{Ey}(n) = P_{EG}(n)/P_{yy}(n) \tag{5}$$

is similar to $R_{EG}(n)$. This normalization by using $P_{yy}(n)$ has two advantages. One is that the variance of $R_{Ey}(n)$ becomes smaller than that of $R_{EG}(n)$ in the double-talk because $P_{yy}(n)$ including the near-end talker's signal is larger than $P_{GG}(n)$. This smaller variance reduces the probability of causing the differentiation error. Another is that $P_{yy}(n)$ doesn't become zero differently from $P_{GG}(n)$ which may be zero at the initial estimation stage.

Figure 1 shows two frequency distributions of the normalized cross-correlation $R_{Ey}(n)$ correspondent to the double-talk and the echo path change, respectively. These are calculated by using 131,072 (=2[17]) samples each under the conditions that $M = 64$ and $I = 512$. The decrease of the power ratio from 30 to 0 dB is substituted for the double-talk, and the increase of the estimation error from –30 to 0 dB is approximated to the echo path change. These two separate distributions shows that the cross-correlation $R_{Ey}(n)$ is suitable for the differentiation.

In this method, the differentiation error can be reduced by applying a large M. However, the large M delays the differentiation, and consequently makes it difficult to suspend the estimation process before the coefficients are corrupted by the double-talk. Here, it should be noted that the differentiation is first required at $j = L$ when the residual echo increased. Thus, $R_{Ey}(n)$ can be substituted by

$$R_{Ey}(k) = \sum_{j=L}^{L+k-1} E_j G_j \bigg/ \sum_{j=L}^{L+k-1} y_j^{\,2}. \tag{6}$$

In (6), the double-talk reduces $R_{Ey}(k)$ with increase of k, and the echo path change makes it converge on a value related to the degree of the likeness between the echo and its replica. This means that the threshold for the differentiation can be gradually heightened with increase of k.

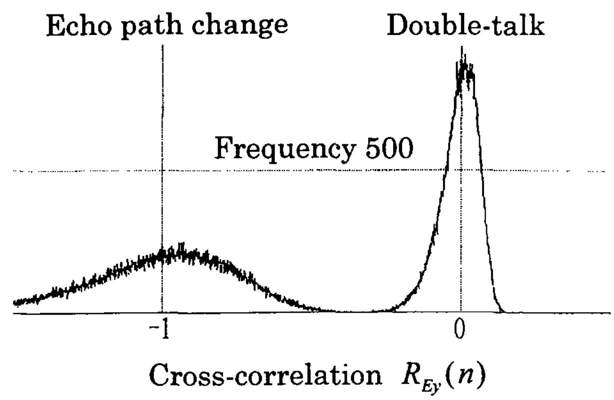

Fig. 1 Frequency distributions of $R_{Ey}(n)$

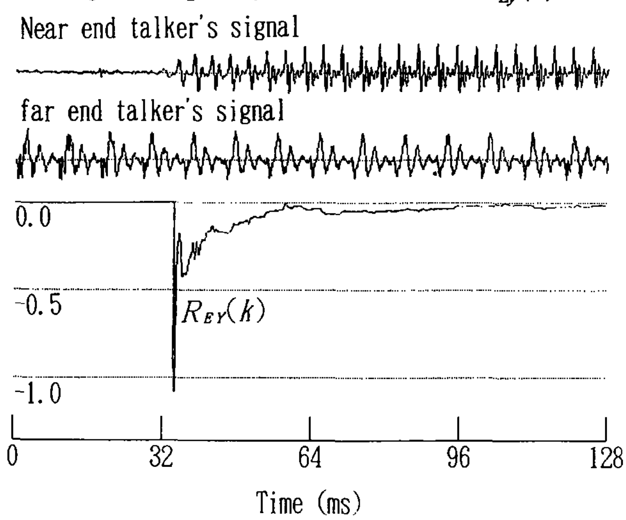

Fig. 2 Transition of $R_{Ey}(k)$ in double-talk.

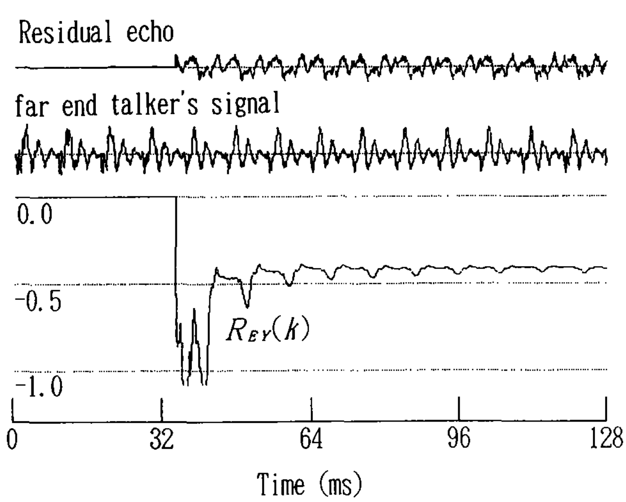

Fig. 3 Transition of $R_{Ey}(k)$ in echo path change.

The echo path change is detected with the delay which is inversely proportional to the degree.

3.2 Simulation using speech signal

Figures 2 and 3 are transitions of $R_{Ey}(k)$ in a double-talk and an echo path change, respectively.

In Fig. 2, the power ratio of the echo and the near-end talker's signal is 0 dB. The estimation process is suspended immediately after the increase of the residual echo was detected. In the double-talk, $R_{Ey}(k)$ grows to zero as shown in Fig.2, and in the echo path change, it converges on a value such as shown in Fig. 3.

This result shows that $R_{Ey}(k)$ is affected by the far-end talker's pitch. Estimating the power of speech signal requires the sample date for a pitch at least. The differentiation should be first tried after an interval equivalent to a pitch. In addition, the different transitions of $R_{Ey}(k)$ in the double-talk and the echo path change, suggest that gradually increasing threshold can be employed for the differentiation. A small echo path change can be detected in time by such increasing threshold [3].

3. HOLDING ACOUSTIC PATH GAIN

3.1 Delayed-x NLMS algorithm

The simplest way of suspending the estimation process before near-end talker's signal corrupts the coefficients, is to apply the residual echo and the reference signal delayed by two sift-registers to the NLMS algorithm as shown in Fig. 4. This delayed-x NLMS algorithm expressed as follows:

$$H_{j+1} = H_j + \mu E_{j-D} X_{j-D} / \|X_{j-D}\|^2 \tag{7}$$

can completely compensates the detection delay [4]. The estimation process can be suspended before the FIR filter coefficient vector, H_j, updated by the residual echo including the near-end talker's signal, S_j, if an occurrence of double-talk can be detected within D sample times. This estimation process is started again after the near-end talker's signal was removed from the delays.

In the delayed-x NLMS algorithm, the step gain should be selected [5] from the range of

$$0 < \mu \le -(D/I) + \sqrt{(D/I)^2 + 1}\,, \tag{8}$$

where I is the order of the FIR filter and

$$\mu_0 = -(D/I) + \sqrt{(D/I)^2 + 1} \tag{9}$$

is the step gain which maximizes the convergence rate. In (9), $D = 0$ provides $\mu_0 = 1$. Such a smaller step gain than that in the conventional NLMS algorithm and the suspension of the estimation process, however, reduce the convergence rate.

3.2 Reference signal power fluctuation

The estimation process is also suspended while the reference signal power is lower than a predetermined limit. This power decrease is equivalent to an occurrence of double-talk. Unfortunately,

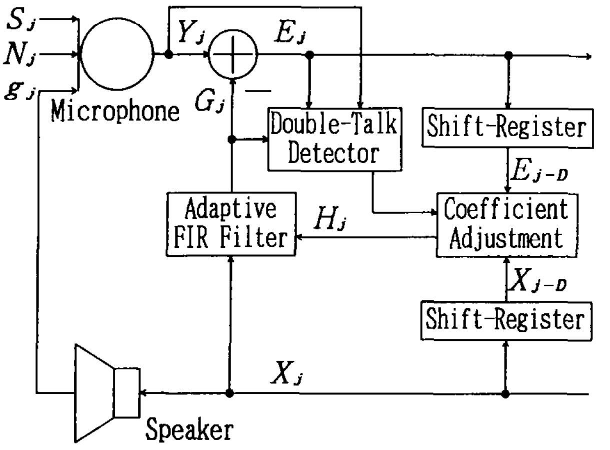

Fig. 4 Echo canceller with two shift-registers compensating double-talk detection delay.

the far-end talker's signal is served as the reference signal in the acoustic echo canceller system. This suspension should be supposed to be frequently applied to the estimation process.

The block implementation of the NLMS algorithm [6],

$$H_{n+1} = H_n + \rho A_n / P_n\,, \tag{10}$$

$$A_n = \sum_{j=nJ+1}^{(n+1)J} E_j X_j \tag{11}$$

$$P_n = \sum_{j=nJ+1}^{(n+1)J} \|X_j\|^2 \tag{12}$$

makes it possible to continue the estimation process even while the reference signal power is below the limit, by adjusting the block length J. Convergence performances of the conventional NLMS algorithm and its block implementation are related with

$$\mu = \rho / J\,. \tag{13}$$

Equation (13) shows that lengthening the block is equivalent to reducing the step gain. The estimation error is held against the reference signal power fluctuation if the block length is properly adjusted [7].

The estimation error is governed by the following equation:

$$\sigma_d^2 = \mu \sigma_N^2 / (2 - \mu) \sigma_X^2\,, \tag{14}$$

where σ_N^2 and σ_X^2 are the environmental noise and the reference signal powers, respectively. In acoustic echo canceller system, this estimation error is equivalent to the acoustic path gain between a microphone and a loudspeaker. This gain must be held below a limit causing howling. Unfortunately, the reference signal is speech signal whose power fluctuates extremely and ceaselessly. The acoustic path gain rises and falls with that power fluctuation if the step gain is fixed.

A method of suspending to update the coefficients while the reference power is less than a predetermined value, is conventionaly applied to the estimation process. This method increases the convergence time due to the suspension. The authors propose a method of continuing the estimation process independent of the reference signal power fluctuation [7]. According to the study [7], the acoustic path gain estimated as the estimation error can be held at a fixed amount, C_0, if the coefficient is updated just when the denominator of (4), P_n, becomes over

$$P_0 = \mu I Q_0 / C_0(2-\mu), \tag{15}$$

where the environmental noise power is supposed to be constant, Q_0. However, this methods also requires suspending the estimation process, when the environmental noise power exceeded Q_0

3.3 Environmental noise power fluctuation

Equation (15) states that the acoustic path gain can be held at C_0 against the environmental noise power of Q_n if the coefficients are updated when P_n exceeded $\mu I Q_n / C_0(2-\mu)$. This environmental noise power, Q_n can be evaluated from the residual echo,

$$E_j = N_j + (g_j - G_j). \tag{16}$$

According to (14), the maximum power of this second term, which is the difference between the echo g_j and its replica G_j, is equal to the environmental noise power, $\sigma_N{}^2$, when the step gain is unity.

The approximation of evaluating the environmental noise power as

$$Q_n \approx \sum_{j=nJ+1}^{(n+1)J} E_j{}^2 \Big/ J \tag{17}$$

reduces the estimation error because

$$Pe_n = \sum_{j=nJ+1}^{(n+1)J} E_j{}^2 \tag{18}$$

is larger than $Q_n J$. The acoustic path gain is held below C_0, if the coefficient is updated when

$$P_n \geq \mu I Pe_n / (2-\mu) C_0 J. \tag{19}$$

3.4 Block length control procedure

Equation (19) shows that the acoustic path gain is evaluated as

$$C_n = \mu I Pe_n / (2-\mu) P_n J \tag{20}$$

when the coefficients were updated with P_n. Therefore, when P_n reached P_0,

(1) if $C_n \leq C_0$, since the environmental noise power is estimated to be equal or less than the minimum Q_0, the coefficients should be updated.

(2) if $C_n > C_0$, since this means an occurrence of double-talk or echo path change, the block should be lengthened and the evaluation of $R_{Ey}(k)$ should be started. After one or two pitch, if $R_{Ey}(k)$ decreased extremely, $C_n > C_0$ means an occurrence of double-talk, inversely if $R_{Ey}(k)$ fixed at a value, $C_n > C_0$ shows an echo path change. In former case, the coefficients can be updated whenever $C_n \leq C_0$. In latter case, the estimation process is performed whenever $P_n = P_0$. This process returns to the usual operation after Pe_n decreases and satisfies

$$Pe_n / J \leq 2Q_0. \tag{21}$$

4. CONCLUSION

We have introduced a method of differentiating the double-talk from the echo path change with the delay of one pitch and a little more, and also have presented a method to continue the estimation process against the double-talk and the power fluctuation of the reference signal.

REFERENCES

[1] S. Minami and T. Kawasaki, "A double talk detection method for echo canceller," *Proc. IEEE Inf. Conf. Communi.*, vol. 3, pp. 1492-1497, 1985.

[2] K. Fujii and J. Ohga, "A method of detecting and discriminating between double-talk and echo path fluctuation," *Proc. 1994 IEICE Spring Conference*, A-204, Mar. 1994.

[3] K. Fujii and J. Ohga, "Discrimination between double-talk and echo path fluctuation," *Technical Report IEICE*, EA94-15, pp. 29-36, May 1994.

[4] K. Fujii and J. Ohga, "Compensation for the double-talk detection delay in echo canceller systems," *IEICE Trans.*, vol. E76-A, no. 7, pp. 1143-1146, July 1993.

[5] K. Fujii and J. Ohga, "Recommended range of the step gain for the 'delayed-E/X' NLMS algorithm," *IEICE Trans.*, vol. E77-A, no.4, pp. 616-625, Apr. 1994.

[6] G. A. Clark, S. K. Mitra, and S. R. Parker, "Block implementation od adaptive digital filters," IEEE Trans. Circuits Syst., vol. CAS-28, pp. 584-592, June 1981.

[7] K. Fujii and J. Ohga, "A method of continuing to adjust adaptive filter coefficients in voiceless noise terms avilable for acoustic echo canceller systems," *IEICE Trans.*, vol. J78-A, no. 11, pp. 1403-1409, Nov. 1995.