A STUDY ON BLIND SOURCE SEPARATION FOR PREPROCESSING OF AN ACOUSTIC ECHO CANCELLER

Yoshihiro Sakai[1 2] *Wataru Mitsuhashi*[1]

[1]Department of Information and Communication Engineering,
The University of Electro-Communications, Tokyo, Japan,
[2]E-mail: sakai@tsuyama-ct.ac.jp

ABSTRACT

The authors have previously reported a novel method for effectively subtracting the echo replica in double-talk condition by preprocessing blind source separation in advance of echo cancellation. In this method a sequential updating algorithm is adopted for blind source separation (BSS). We discuss in this paper an alternative algorithm in which an auxiliary term is added for improving the convergence rate. The improvement effect of convergence speed was the double, when the operation in the case in which this algorithm was meanwhile applied to the updating of the coefficient of separation circuit.

Index Terms— echo canellar, double talk, blind source separation

1. INTRODUCTION

Echo cancellation is a promising technique in teleconference systems for reducing the energy of the acoustic echo. In most algorithms for echo cancellation a replica waveform corresponding to the echo path is adaptively estimated and is subtracted from the transmitting channel. In double-talk condition, however, the adaptation for estimating the echo replica could fail and some amount of echo energy might be sent back to the receiving channel. In the double-talk condition, updating the adaptive estimation for the echo path change should be inhibited, thereby reducing the convergence rate. This will provide a prevention against deterioration of adaptation. However, the echo path change during double-talk is disregarded in this situation.

The authors have previously reported a novel method for effectively subtracting the echo replica in double-talk condition by preprocessing blind source separation in advance of echo cancellation[2]. In this method a sequential updating algorithm is adopted for blind source separation (BSS). The updating coefficient μ_S controls both the rate of convergence and the crosstalk attenuation, thus affecting the characteristics of echo subtraction. In terms of the value of μ_S there is a trade-off between the rate of convergence and the crosstalk attenuation. This property of the previous proposed method results in a drawback unacceptable in adaptive echo cancellation; high cross-talk attenuation is inconsistent with high rate of conversion.

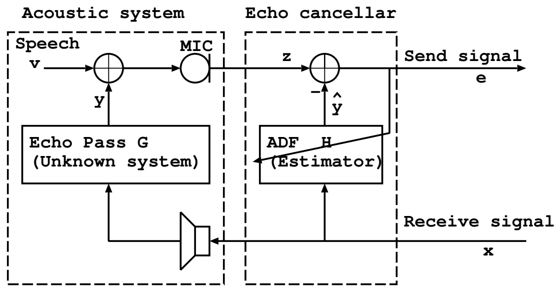

Fig. 1. Block diagram of echo canceller as an Application of ADF.

We discuss in this paper an alternative algorithm in which an auxiliary term is added for improving the convergence rate. Updating the coefficients with these auxiliary terms enables us to improve the convergence rate of adaptation in the estimation of echo path change. This process is quite similar to the independent components analysis with supervisors described in time domain. By using supervisors, convergence of the separation circuit improved to twice as fast as the conventional approach.

2. PROBLEM IN THE ECHO CANCELLER

The structure of teleconference systems that applies echo canceller is shown in Fig. 1. By subtracting echo replica $\hat{y}$ from the sound signal z of the microphone which include echo component y, only the speaker voice v would be included in transmission signal e. The echo replica $\hat{y}$ is generated from received signal x by Adaptive Digital Filter (ADF). In the many practical systems, the structure of ADF is FIR type, and the coefficient of it is adaptively decided using the normalized LMS method expressed in the following formula:

$$h_{n+1}(k) = h_n(k) + \mu \frac{e(t)x(t-k)}{\langle x_2(t) \rangle} \quad . \tag{1}$$

In the double-talk condition, however, the speaker signal v works as a disturbance for the algorithm, so that the estimated error occurs, and the residual echo is thus observed in the output of the echo canceller. The echo return loss enhancement (ERLE) which is one measure of the performance of the echo canceller will decreases to the following formula:

$$ERLE = S/N + 10\log_{10}(\frac{2}{\mu} - 1) \ , \qquad (2)$$

where, S denotes the level of echo signal, and N denotes the level of speaker signal. The speaker signal v is not permitted to be removed, because it should be transmitted to the other side of the teleconference line. Therefore, in the double-talk condition, the echo canceller is required that maintains the estimation accuracy of the unknown system and prevents the increase in the echo[2]. One of the answer for the requirement is level comparison method. In this method, the adaptation is stopped for the period in which the level of the speaker signal is larger than the criteria calculated from the level of received signal. The logic of this adaptive control is shown in the following equation:

$$\begin{cases} \mu = \mu_A & for\ P_z < P_x + \delta \\ \mu = 0 & for\ P_z > P_x + \delta \ , \end{cases} \qquad (3)$$

where P_z is the level of the microphone signal, and P_x is the level of the received signal. By the level comparison method, the increase in residual echo is controlled to some level. However, under the condition of the double-talk with simultaneous change of the impulse response $g(t)$ of the echo path, the coefficient can't follow the change so that residual echo will increase. Therefore, a different method dealing with the double-talk is required to keep the speech quality.

3. NOVEL METHOD USING A BLIND SOURCE SEPARATION CIRCUIT

The authors have previously proposed a novel method for effectively subtracting the echo replica in double-talk condition by preprocessing blind source separation. In the proposed method, even in the double-talk condition, the adaptation is carried out to deal with the change of echo pass, so that the estimation accuracy should be kept. The structure of the proposed method is shown in Fig. 2.

In the proposed method, the system consists of the blind source separation circuit and the adaptive filter which uses the output of it. Microphone signal $y_2(t)$ is the sum of echo signal $x_2(t)$ and speaker signal $x_1(t)$.

$$y_2(t) = x_1(t) + x_2(t) \qquad (4)$$

In the source separation circuit, the microphone signal $y_2(t)$ is subtracted by estimated signal $x_1'(t)$, and it forms the output $s_2(t)$.

$$s_2(t) = y_2(t) - x_1'(t) = x_2(t) + (x_1(t) - x_1'(t)) \qquad (5)$$

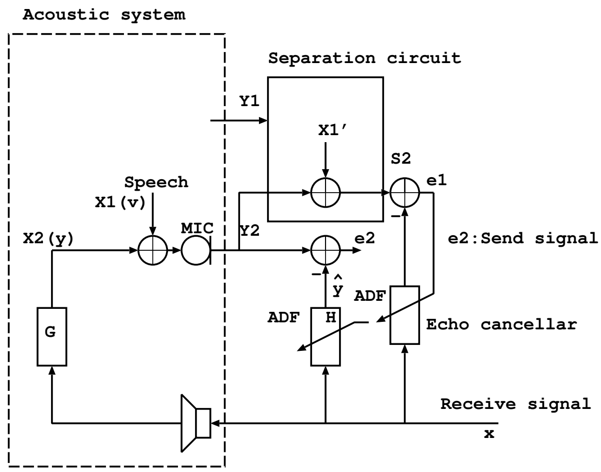

Fig. 2. Echo canceller using separation circuit.

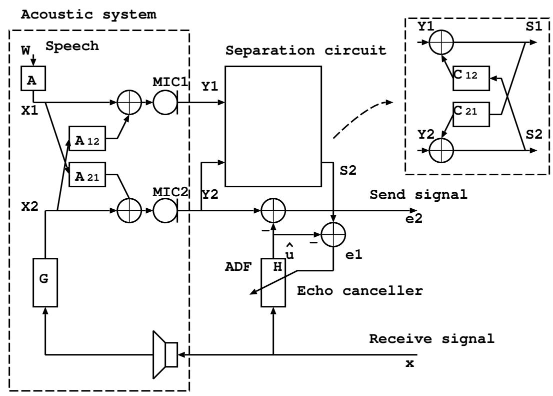

Fig. 3. Echo canceller using blind separation circuit.

If the subtraction signal $x_1'(t)$ is good estimation of speaker signal $x_1(t)$, the output signal $s_2(t)$ becomes a separated echo signal $x_2(t)$,

$$s_2(t) \rightarrow x_2(t)$$

By using this output $s_2(t)$ as the input of ADF, even if the speaker signal is overlapped in the microphone input, the signal is not disturbance for the adaptation, and estimation accuracy of the echo path does not lower. It is possible to carry out echo canceling well without producing an increase in the residual echo, when the filter coefficient gotten in this way is copied to the ADF of the echo canceller.

4. STRUCTURE OF BLIND SIGNAL SEPARATION CIRCUIT

Recently, various approach for realization of blind source separation circuit are studied. In the proposed method, the role is

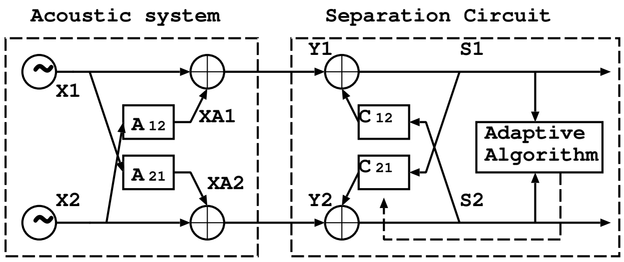

Fig. 4. Structure of blind source separation circuit.

the preprocessing of the echo canceller, it is desirable that the circuit is linear processing and does not affect the adaptation algorithm of echo canceller. Then, we employ the time domain filter structure proposed by Jutten et al [3]. The system is simply explained in the following.

In Fig. 4, the part from signal x to y shows the equivalent circuit from the acoustic signal source to the microphone, and the part following it shows the blind source separation circuit. The input signal $X_1(z)$ and $X_2(z)$ are considered that they are unknown and independent. There are two kind of path from the signal source to the microphone. One is the path reaching each microphone directly, and another is the path reaching the microphone of the opposite side with the transfer function of $A_{12}(z)$ and $A_{21}(z)$. These are called mixing systems. If the signal of each microphone is made to be $Y_1(z)$ and $Y_2(z)$, the mixing system is modeled as follows:

$$\begin{cases} Y_1(z) = X_1(z) + A_{12}(z)X_2(z) \\ Y_2(z) = A_{21}(z)X_1(z) + X_2(z) \, . \end{cases} \tag{6}$$

On the contrary, the input-output function of the separation circuit is expressed in the following equations:

$$\begin{cases} S_1(z) = Y_1(z) - C_{12}(z)S_2(z) \\ S_2(z) = -C_{21}(z)S_1(z) + Y_2(z) \, . \end{cases} \tag{7}$$

In order to separate the signal, the coefficient of the separation circuit should be selected as follows:

$$C_{12}(z) = A_{12}(z), \;\; C_{21}(z) = A_{21}(z) \, . \tag{8}$$

4.1. The Separation Algorithm

It is impossible in general to obtaine the filter coefficients by the equation (8), because the transfer functions $A_{12}(z)$ and $A_{21}(z)$ of the mixing system are unknown. Then, the filter coefficient $c_{ij}(k)$ must be adjusted by the iterative estimation, as shown in the following equation:

$$c_{ij}(n+1,k) = c_{ij}(n,k) - \mu_S \phi_{ij}(c_{ij}(n,k)) \, , \tag{9}$$

where $c_{ij}(n,k)$ is the coefficient obtained by n times iterative calculation, $\phi_{ij}(c_{ij}(n,k))$ is the function which evaluates the separation circuit. and μ_s denotes the coefficient for

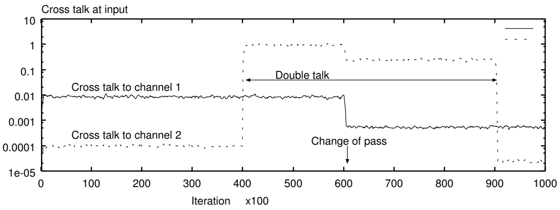

Fig. 5. Signal level of input.

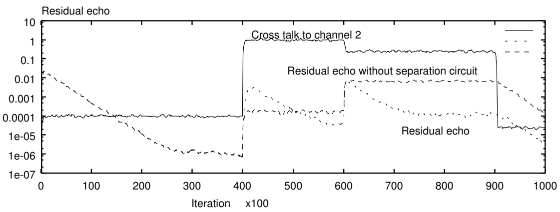

Fig. 6. Residual echo in the output of echo canceller.

determining the filter updating speed. The coefficient μ_s controls the stability of the algorithm. The evaluation function $\phi_{ij}(c_{ij}(n,k))$ can be substituted by cross cumulants,

$$\phi_{ij}(c_{ij}(n,k)) = \frac{\mathrm{Cum}_{31}(s_i(n), s_j(n-k))}{\mathrm{Cum}_{40}(s_i(n), s_j(n-k))}, \tag{10}$$

where $\mathrm{Cum}_{31}(\cdot)$ is a $(3,1)$ order cross cumulant, and $\mathrm{Cum}_{40}(\cdot)$ is a $(4,0)$ order cross cumulant. $\mathrm{Cum}_{40}(\cdot)$ is a factor maintaning the stability of the sequential adaptation against the change of the signal level. The algorithm is expressed as follows:

$$\begin{aligned} c_{ij}(n+1,k) \quad &= c_{ij}(n,k) \\ &\quad - \mu_S \frac{\mathrm{Cum}_{31}(s_i(n), s_j(n-k))}{\mathrm{Cum}_{40}(s_i(n), s_j(n-k))} \, . \end{aligned} \tag{11}$$

The effect of this algorithm is illustrated in Fig. 6. In Fig. 6, the dotted line 'Residual' in the "residual echo" indicates an output of this system. The horizontal axis shows an iteration, and after 40,000 iterations, the speaker signal level increase (double-talk condition). In addition, at 60,000 iterations, echo path changes.

The residual echo once increases in the effect of the double-talk at 40,000 iterationsp, but it decreases (improves) afterwards so that echo canceller may continue the adaptation. In 60,000 iterations, the residual echo increases again caused by the effect of the echo path change. It decreases(improves) again after the time since echo canceller continues the adaptation. This shows that the system can deal with the echo path change, even in the double-talk condition.

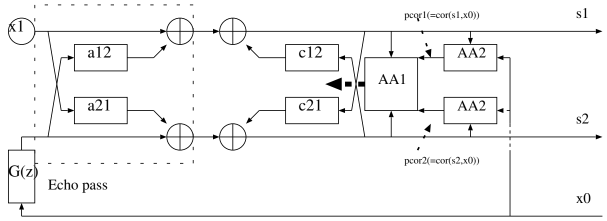

Fig. 7. Structure of the proposed system.

5. COEFFICIENT UPDATING USING THE ADDITION TERM

Though the convergence property of blind source separation circuit can be optimized by adjusting the updating coefficient, the convergence speed and the residual errors are inconsistent with each other. Accordingly, the convergence property using the additional term is examined on the basis of the algorithm shown in equation (11). In the proposed system, a strong correlation exists between the receive signal x and one of the output s_2 shown in Fig.3. The separation property can be improved if the coefficient c_{ij} is updated along the direction in which x and s_2 are highly correlated. The same idea can be applied to the relationship between the receive signal x and the other output s_1 shown in Fig.3. In this case, the separation property is improved if the coefficient c_{ij} is updated along the direction in which x and s_1 are mutually uncorrelated.

There is a striking similarity in the function between the proposed echo canceller mentioned above and an algorithm of independent component analysis with supervisors in the time domain. Fig.7 illustrates the structure of the proposed system. Differentiating the correlation ρ_i between x_0 and s_i, i=1 or 2 with respect to $c_i j$ will yield a clue to the direction in which the correlation has a maximum or a minimum.

The correlation will be written here as the following.

$$\rho_i = r(x_0, s_i) \tag{12}$$

It is differentiated with respect to the coefficient c_{ij} in order to obtain the direction which maximizes this correlation.

$$g_i = \frac{\partial \rho_i}{\partial c_{ij}} = k \cdot r(x_0, s_i) \tag{13}$$

In the relation between signal x_0 and separation circuit output s_1, the correlation should be minimized. The algorithm is following,

$$\begin{aligned} c_{ij}(n+1,k) \quad &= c_{ij}(n,k) \\ &-\mu_S \frac{\mathrm{Cum}_{31}(s_i(n), s_j(n-k))}{\mathrm{Cum}_{40}(s_i(n), s_j(n-k))} \\ &-\mu_1 r(x_o, s_1(k)) \end{aligned} \tag{14}$$

The operation applied this algorithm to the updating of (c_{21}) of the coefficient of the separation circuit, and updating of coefficient c_{12} is carried out by the algorithm shown in equation (11). The result is indicated in Fig. 8.

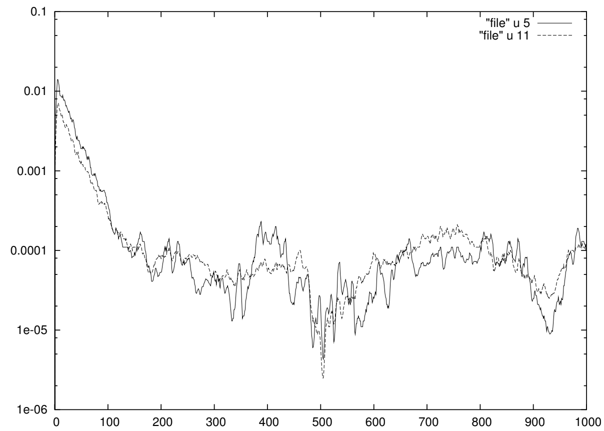

Fig. 8. Residual cross-talk in the output of separation circuit.

By applying to both of coefficient c_{12} and c_{21} of separation circuit, the convergence speed is improved with the double.

6. CONCLUSION

We discussed in this paper an alternative algorithm in which an auxiliary term was added for improving the convergence rate. Updating the coefficients with these auxiliary terms enables us to improve the convergence rate of adaptation in the estimation of echo path change. This process is quite similar to the independent components analysis with supervisors described in time domain. By using supervisors, convergence of the separation circuit improved twice as fast as the conventional approach .

7. REFERENCES

[1] Y.Sakai, K.Takahashi, "Control Method for Blind source separation as Pre-processing for Echo Canceller"(in Jpanese), Journal of Signal Processing Vol.11, No.2, pp.171–177, 2007.

[2] CCITT, "Recommendation G.165", Yellow book, 1980.

[3] H.N.Thi, C.Jutten, "Blind source separation for convolution mixtures", Signal Processing 45,pp.209–229, 1995.