❏   274

Blind Signal Separation Algorithm for Acoustic Echo Cancellation

Haengwoo Lee
Namseoul University/91, Daehak-Ro, Cheonan-Si,
Chung-Nam, 82-10-2693-5563, Korea Republic of

Article Info	ABSTRACT
	This paper is to the blind signal separation algorithm applied to acoustic echo cancellation. This algorithm doesn't degrade the performance of echo cancellation even in the double-talk. In the closed echo environment, the mixing model of acoustic signals has multi-channel, so the convolutive blind signal separation method is applied. And the mixing coefficients are computed by using the feedback model without directly calculating the separation coefficients. The coefficient updating is performed by iterative computations based on the second-order statistical properties, thus estimating the near-end speech. Many simulations have been performed to verify the performance of the proposed blind signal separation. Simulation results show that the proposed acoustic echo canceller operates safely regardless of double-talk, and the PESQ is improved by 0.6 point compared with the general adaptive FIR filter structure.

Corresponding Author:

Haengwoo Lee,
Namseoul University/91, Daehak-Ro, Cheonan-Si,
Chung-Nam, 82-10-2693-5563, Korea Republic of
Email: haengwoolee@hanmail.net

1.   INTRODUCTION

In a speaker-phone system such as a hands-free system, an acoustic echo signal is received by a speaker from a speaker and then reflected on a wall surface or the like to be coupled with a transmission signal of the speaker from the microphone. In order to remove the echo signal, an acoustic echo canceller using an adaptive filter is used. This circuit operates normally only in a one-way call in which only a far-end speaker signal is present, and has the ability to remove an echo signal in a simultaneous call interval in which a self- Sudden degradation occurs. Therefore, this simultaneous call problem must be solved.

Generally, there is a method of preventing the divergence of coefficients by detecting the simultaneous call state and stopping the coefficient update of the echo canceller. However, this method has a relatively long detection time, so that the coefficients of the echo canceller may diverge before detection. In this paper, we use sound source separation technology that can solve these problems fundamentally.

The sound source separation technique can separate each sound source signal from the observed mixed sound. In particular, there is a Blind Signal Separation (BSS) method [1] that separates signals without knowing prior information such as statistical characteristics of input signals. This is a method of separating several sound sources on the assumption that the statistical characteristics between the sounds are independent. The BSS algorithm proposed by Jutten and Herault has announced that it can simply separate linear mixed signals and several subsequent studies have been done since then. Jutten and Herault proposed a BSS algorithm [2] that separates the noise and the signal in the time domain. However, this algorithm uses thousands of taps, so it requires a lot of memory and computational complexity. Bell and Sejnowski proposed a method for estimating delay using Independent Component Analysis (ICA) [3]. However, the speech signal may not converge to the optimal solution depending on the initial value in estimating the delay due to the

periodic characteristic. To solve this problem, Torkkoa estimated the delay using the information maximization algorithm [4]. However, if the amount of computation is large and background noise is present, the delay can not be accurately estimated. Recently, there have been many researches on blind source separation [5] [6] in the frequency domain. In addition, the problem of separating signals in a multi-channel environment where echo signals are present in a closed space requires a lot of computational complexity of multidimensional [7]-[12].

The proposed sound source separation technique uses the difference of signal size and arrival time between microphones. In other words, the algorithm implements a number of nonlinear mathematical equations. A plurality of acoustic signals are mixed and interfered with each other due to overlapping in three dimensions. Three dimensions refer to time, frequency, and arrival direction. Until now, signals superimposed in time domain have been distinguished by using frequency spectrum and arrival direction. However, this technique can distinguish signals that are simultaneously superimposed on time, spectrum, and arrival direction. In other words, in an environment where there are mixed sounds generated from two sound sources, each sound source is separated using a dual microphone. This is because the size and arrival time of the incoming signal to each microphone are different, so the sound source is obtained through iterative calculation by using these characteristics.

This paper is organized as follows. In Section 2, we discuss the concept of the blind signal separation algorithm. In Section 3, we applied the blind signal separation algorithm in the acoustic echo environment with multiple channels. Simulation results are presented in Section 4, and conclusions are drawn in Section 5.

2.   RESEARCH METHOD

In the environment where two sound source signals are mixed, the sound signals generated from the respective sound sources are micro-influenced in size and time. Therefore, if the sound source is $s(k)$, and the microphone input signal is $x(k)$, then the mixture model of each micro incoming signal is expressed as Equation 1 and 2.

$$x_1(k) = a_{11}(k)s_1(k) + a_{12}(k)s_2(k - d_2) \tag{1}$$

$$x_2(k) = a_{21}(k)s_1(k - d_1) + a_{22}(k)s_2(k) \tag{2}$$

Where $a(k)$ is the mixing coefficient in inverse proportion to the distance from each sound source to the microphone. If the sampling frequency is 8 kHz, the delay time d_1, d_2 can be ignored if the distance between the two microphones is within about 4 cm. Since the size of the sound separated from the BSS output is not the same as the size of the sound source, the equation is simplified by normalizing each sound source to $s'_1(k) = s_1(k)/a_{11}(k)$, $s'_2(k) = s_2(k)/a_{22}(k)$. Therefore;

$$x_1(k) = s'_1(k) + a_{12}(k)s'_2(k) \tag{3}$$

$$x_2(k) = a_{21}(k)s'_1(k) + s'_2(k) \tag{4}$$

In other words, the blind signal separation has the structure as shown in Figure 1.
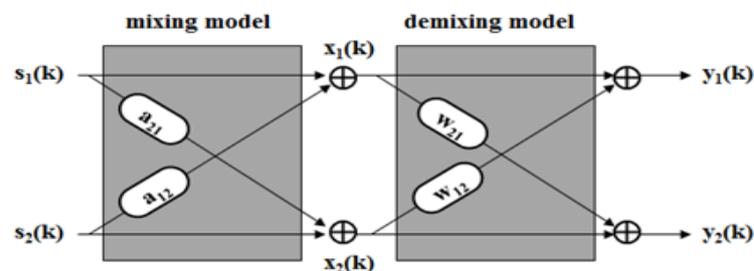

Figure 1. Feedforward model of blind signal separation

where $w(k)$ is the separation coefficient for the mixed signal. By summing up the formulas for estimating the two input signals

$$\tilde{s}_1(k) = x_1(k) - \tilde{a}_{12}(k)\tilde{s}_2(k) \tag{5}$$

$$\tilde{s}_2(k) = x_2(k) - \tilde{a}_{21}(k)\tilde{s}_1(k) \tag{6}$$

where the estimate of the mixing coefficient $\tilde{a}_{12}(k)$ and $\tilde{a}_{21}(k)$ can be solved by iterative calculations. That is, when the acoustic signals $s_1(k)$, $s_2(k)$ are independent of each other, the second-order statistical characteristic of each signal is used to obtain the steepest descent method. On the other hand, since the output signal is an estimate of the sound source $y_1(k) = \tilde{s}_1(k)$, $y_2(k) = \tilde{s}_2(k)$, the separation model of the output signal from Equation 5 and 6 are;

$$\tilde{s}_1(k) = x_1(k) - \tilde{a}_{12}(k)\tilde{s}_2(k) \tag{7}$$

$$\tilde{s}_2(k) = x_2(k) - \tilde{a}_{21}(k)\tilde{s}_1(k) \tag{8}$$

Instead of obtaining the separation coefficient, the mixing coefficient can be directly obtained from Equation 7 and 8. To reuse the mixed model to use the error signal as a cost function

$$x_1(k) - y_1(k) - \tilde{a}_{12}(k)y_2(k) = \varepsilon_1(k) \tag{9}$$

$$x_2(k) - \tilde{a}_{21}(k)y_1(k) - y_2(k) = \varepsilon_2(k) \tag{10}$$

Therefore, the update value using the LMS algorithm is

$$\nabla\rho_1(k) = \frac{\partial\{\varepsilon_1(k)\}^2}{\partial a_{12}(k)} = -2[x_1(k) - y_1(k) - \tilde{a}_{12}(k)y_2(k)]y_2(k) \tag{11}$$

$$\nabla\rho_2(k) = \frac{\partial\{\varepsilon_2(k)\}^2}{\partial a_{21}(k)} = -2[x_2(k) - \tilde{a}_{21}(k)y_1(k) - y_2(k)]y_1(k) \tag{12}$$

And the new coefficient update formula

$$\tilde{a}_{12}(k+1) = \tilde{a}_{12}(k) - \mu_1\nabla\rho_1(k) = \tilde{a}_{12}(k) + 2\mu_1[x_1(k) - y_1(k) - \tilde{a}_{12}(k)y_2(k)]y_2(k) \tag{13}$$

$$\tilde{a}_{21}(k+1) = \tilde{a}_{21}(k) - \mu_2\nabla\rho_2(k) = \tilde{a}_{21}(k) + 2\mu_2[x_2(k) - \tilde{a}_{21}(k)y_1(k) - y_2(k)]y_1(k) \tag{14}$$

where μ denotes the step size that determines the convergence speed. Figure 2 shows feedback model of blind signal separation. Figure 3 shows the calculation process of this coefficient schematically. Expressing Equation 7 and 8 again using the BSS output;

$$y_1(k+1) = x_1(k) - \tilde{a}_{12}(k)y_2(k) \tag{15}$$

$$y_2(k+1) = x_2(k) - \tilde{a}_{21}(k)y_1(k) \tag{16}$$

Therefore, the sound sources to be separated can be obtained by repeatedly calculating Equation 13, 14 and 15, 16 for each sample.
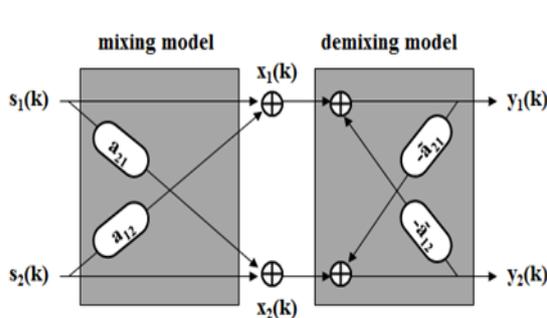

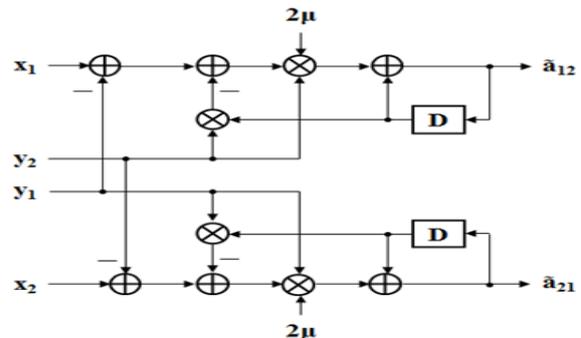Figure 2. Feedback model of blind signal separation   Figure 3. Coefficient updating procedure

3.   ACOUSTIC ECHO CANCELLATION BASED ON BLIND SIGNAL SEPARATION METHOD

In a speaker-phone system, a hands-free call causes acoustic echo. So we use a circuit to eliminate unwanted echo signals. However, in the case of simultaneous calls in which the far-end speaker and the near-end speaker speak simultaneously, the micro-inflowing near-end speaker signal should be calculated. To solve this double-talk problem, we propose a BSS based acoustic echo cancellation system. This is shown in Figure 4. Figure 5 shows feedforward structure of blind signal separation with echo path
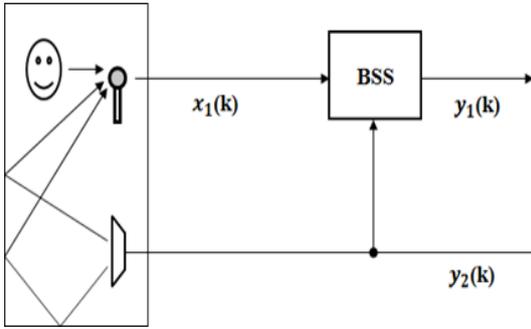

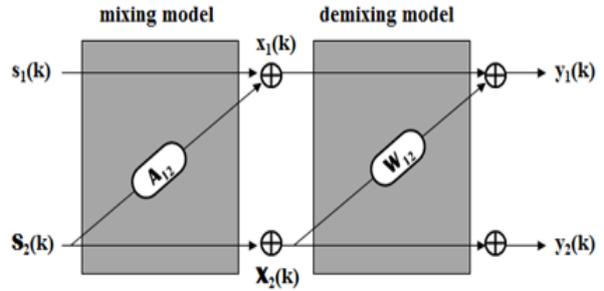Figure 4. Acoustic echo canceller with BSS

Figure 5. Feedforward structure of blind signal separation with echo path

Explaining The acoustic echo signal is the sum of the multi-channel signals. In other words

$$e(k) = \sum_{n=0}^{N-1} g_n(k) y_2(k - n) \tag{17}$$

where $g_n(k)$ denote the response of the echo path and N the number of samples of the response. The acoustic echo signal generated by the loudspeaker is micro-influenced through the echo path of multiple channels. The mixture model is expressed by Equation 18 and 19, where $s_1(k)$ is a near-end speaker signal, $\mathbf{S}_2(k)$ is a matrix of speaker signals, and $\mathbf{A}_{12}(k)$ a matrix of mixing coefficients for echo signals.

$$x_1(k) = s_1(k) + \mathbf{A}_{12}^T(k)\mathbf{S}_2(k) \tag{18}$$

$$x_2(k) = s_2(k) \tag{19}$$

The mixing coefficient of the echo signal is the same as the response of the echo path. So if you rewrite the mixed model

$$x_1(k) = s_1(k) + \mathbf{G}^T(k)\mathbf{S}_2(k) \tag{20}$$

$$x_2(k) = s_2(k) \tag{21}$$

If the input signal is obtained by Equation 20

$$\tilde{s}_1(k) = x_1(k) - \mathbf{G}^T(k)\mathbf{S}_2(k) \tag{22}$$

$$\tilde{s}_2(k) = x_2(k) \tag{23}$$

Since the output signals are $y_1(k) = \tilde{s}_1(k)$, $y_2(k) = \tilde{s}_2(k)$. Equation 24 and 25 show a feedback separation model.

$$y_1(k) = x_1(k) - \mathbf{G}^T(k)\mathbf{Y}_2(k) \tag{24}$$
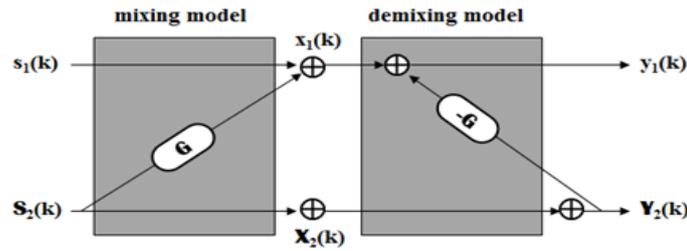
$$y_2(k) = x_2(k) \tag{25}$$

Figure 6. Feedback structure of blind signal separation with echo path

Using the error signal from Equation 24 as a cost function to obtain the echo path response,

$$x_1(k) - \mathbf{G}^T(k)\mathbf{Y}_2(k) - y_1(k) = \varepsilon(k) \qquad (26)$$

Therefore, the update value of the LMS algorithm is

$$\nabla\rho(k) = \frac{\partial\{\varepsilon(k)\}^2}{\partial g(k)} = -2[x_1(k) - y_1(k) - \mathbf{G}^T(k)\mathbf{Y}_2(k)]y_2(k) \qquad (27)$$

Therefore, the update is made as follows.

$$g_n(k+1) = g_n(k) - \mu\nabla\rho(k) = g_n(k) + 2\mu[x_1(k) - y_1(k) - \mathbf{G}^T(k)\mathbf{Y}_2(k)]y_{2n}(k) \qquad (28)$$

for n = 0, ⋯, N − 1

4. SIMULATION RESULTS

In order to verify the separation performance of the convolutional blind signal separation method proposed in this paper, we developed a simulation program that expresses the operation of the system using Matlab on PC.

The input uses the impulse response to the echo path and the far-end speaker signal that generates the echo signal and the own speech call that causes the double-talk. The input signal is sampled at 8 kHz and the coefficients of the BSS used in the experiment were N=128, and the size of the adaptation step is set to be μ=0.0005 when the coefficient $g_n(k)$ is adaptively calculated. In this experiment, the following mathematical model was used for echo path response.

$$h(i) = 0.95h(i-1) + 0.19h(i-2) + 0.09h(i-3) - 0.05h(i-4) + 0.02h(i-5) \qquad (29)$$

In this study, we tried to separate signal when two acoustic signals are mixed and induced into microphone. The waveform of the used speech signal is shown in Figure 7. Figure 8 shows the waveform of the microphone input signal, which is a mixture of near-end speaker signal and echo signal. And the waveform of the separated BSS output signal is shown in Figure 9.
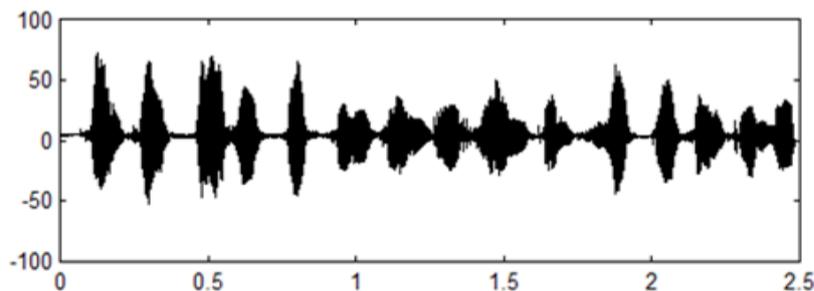

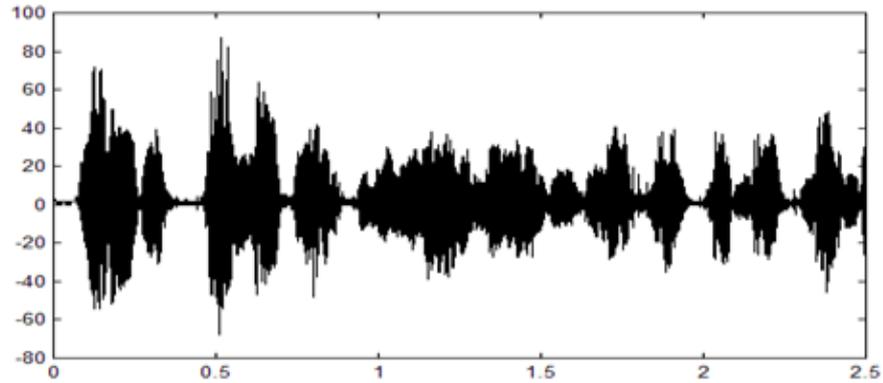Figure 7. Waveform of speech source signal
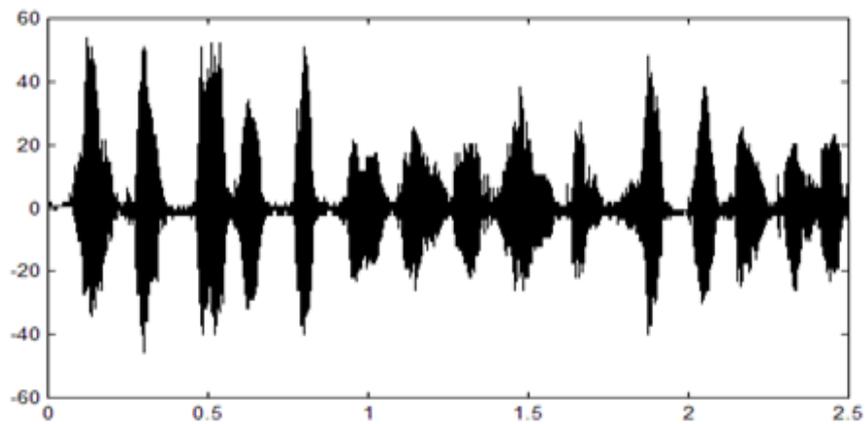
Figure 8. Waveform of mixing signal

Figure 9. Waveform of BSS output signal separated

The PESQ standard is described in ITU-T P.862 [13] as a perceptual evaluation tool for voice quality. The heart of the PESQ work is to use a perceptual model similar to the original signal and the psychological representation of the degraded signal in the human auditory system. The output of the PESQ is a measure of the subjective quality of the degraded signal and is a value between 0~5 point (5=Excellent; 4=Good; 3=Moderate; 2=Bad; 1=Not available). In this paper, the similarity between the original speech signal and the restored speech signal is measured for the convolutive blind signal separation method and the adaptive FIR filter, and the separation performance is evaluated by the PESQ score. From the results of Table 1, it can be seen that the proposed convolutional blind signal separation method achieves better performance than the adaptive FIR filter.

Table 1. PESQ of Separated Signal $y_1(k)$

Structure	Adaptive FIR filter	Convolutive BSS
PESQ Score	3.1	3.7

5. CONCLUSION

This paper deals with a convolutive blind signal separation method for separating the echo signal and the near-end speaker signal from the mixed signal in the acoustic echo environment. This method can remove the echo signal normally even during a double-talk. In a closed echo environment, the blind model of blind signal separation is multi-channel, so the source signal is estimated using the feedback model of the mixed signal based on the second order statistics without directly calculating the separation coefficients. Simulation results show that the proposed method works well regardless of double-talk. This method shows that the PESQ is improved by 0.6 points compared to the general adaptive FIR filter.

REFERENCES
[1]   Comon P, Independent component analysis, A new concept, *Signal Processing*, 1994, vol. 36, no. 3, pp. 287-314.
[2]   Jutten C, Herault J, Blind separation of sources, Part I: An adaptive algorithm based on neuromimetic architecture, *Signal Processing*, 1991, vol. 24, no. 1, pp. 1-10.
[3]   Bell AJ, Sejnowski TJ, An information maximization approach to blind separation and blind deconvolution, *Neural Computation*, 1995, vol. 7, pp. 1129-1159.
[4]   Kari T, *Blind separation of convolved sources based on information maximization*, Neural Networks for Signal Processing, IEEE Signal Processing Society Workshop, 1996.
[5]   Schobben DWE, Sommen PCW, A frequency domain blind signal separation method based on decorrelation, *IEEE Trans. Signal Process*, 2002, vol. 50, no. 8, pp. 1855-1865.
[6]   Schobben DWE, Sommen PCW, *A new algorithm for joint blind signal separation and acoustic echo canceling*, in Proc. Int. Conf. Signal and Image Process, Oct. 1998, pp. 564-569.
[7]   Kawamoto M, Matsuoka K, Ohnishi N, A method of blind separation for convolved non-stationary signals, *Neuro-computing*, 1998, vol. 22, no. 1-3, pp. 157-171.
[8]   Parra L, Spence C, Convolutive blind separation of non-stationary sources, *IEEE Trans. Speech Audio Process*, 2000, vol. 8, no. 3, pp. 320-327.
[9]   Weinstein E, Feder M, Oppenheim AV, Multi-channel signal separation by decorrelation, *IEEE Trans. Speech Audio Process.*, 1993, vol. 1, no. 4, pp. 405-413.
[10] Yellin D, Weinstein E, Multichannel signal separation: methods and analysis, *IEEE Trans. Signal Process*, 1996, vol. 44, no. 1, pp. 106-118.
[11] Lambert RH, Bell AJ, *Blind separation of multiple speakers in a multipath environment*, IEEE international conference on acoustic, speech and signal processing, ICASSP-97, Apr. 1997, vol. 1, pp. 423-426.
[12] Vinaya AA, Arifianto D, Multi-way array decomposition on acoustic source separation for fault diagnosis of a motor-pump system, *International journal of electrical and computer engineering (IJECE)*, 2017, vol. 7, no. 6, pp. 3052-3059.
[13] ITU-T Recommend P.862, *Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone network and speech codecs*, May 2000.