

Analysis of Signal Separation and Distortion Analysis in Feedforward Blind Source Separation for Convolutive Mixture

Kenji Nakayama Akihiro Hirano Yasuhiro Dejima
Division of Electrical Engineering and Computer Science
Graduate School of Natural Science and Technology, Kanazawa Univ.
2-40-20, Kodatsuno, Kanazawa, 920-8667, JAPAN
e-mail:nakayama@t.kanazawa-u.ac.jp

Abstract— This paper analyzes signal separation and distortion in a feedforward blind source separation (BSS) applied to convolutive mixture environment. In the BSS learning, a separation block is adjusted so as to make the output signals statistically independent. One direction for this adjustment is to extract only one signal source at a single output. The other direction is to make the output signal spectra white. The second direction causes signal distortion and conflicts with the first direction. This confliction prevent signal separation. These relations are also dependent on a length of the impulse responses in the mixture and the filter length in the separation block. These properties are theoretically analyzed and are investigated through simulations using the white signal sources and the speech signal sources.

I. INTRODUCTION

Signal processing including noise cancelation, echo cancelation, equalization of transmission lines, estimation and restoration of signals have been becoming very important technology. In some cases, we do not have enough information about signals and interference. Furthermore, their mixing and transmission processes are not well known in advance. Under these situations, blind source separation (BSS) technology using statistical property of the signal sources have become very important [1].

Since, in many applications, mixing processes are convolutive mixtures, FIR or IIR filters are required in unmixing processes. Several methods in the time domain and the frequency domain have been proposed. Two kinds of network structures have been proposed for the BSS, that is a feedforward type and a feedback type. Separation performance is highly dependent on the transfer functions in the mixture [2]-[8]. The BSS learning algorithms make the output signals to be statistically independent. Therefore, signal distortion, caused by linear transfer functions, cannot be controlled [9],[10].

This paper analyzes both signal separation and signal distortion in the feedforward BSS (FF-BSS) applied to the convolutive mixture [11],[12]. The FF-BSS has a degree of freedom to modify the output signal spectra. Two kinds of directions of adjusting the separation block are analyzed. Furthermore, conditions under which two directions conflict with each other, and prevent signal source separation, are investigated. These properties are theoretically analyzed and are confirmed through simulations by using the white signal sources and the speech signal sources.

II. FEEDFORWARD BSS FOR CONVOLUTIVE MIXTURE

A. Network Structure and Equations

For simplicity, 2 signal sources and 2 sensors are used. The frequency domain BSS [11] is taken into account. A block diagram is shown in Fig.1.

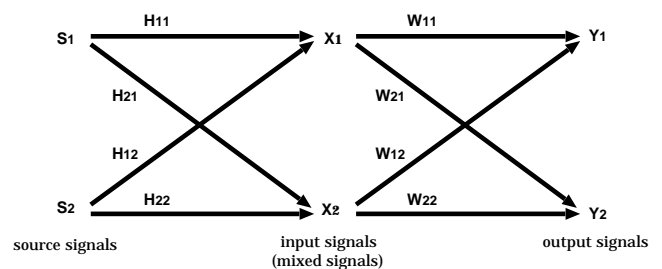


Fig. 1. Feedforward BSS with 2 signal sources and 2 sensors.

$$x_i(n) = \sum_{i=1}^2 \sum_{k=0}^{K-1} h_{ji}(k) s_i(n-k), i = 1, 2 \quad (1)$$

The observed signals are first transformed into the frequency domain, and are processed in the frequency domain. The equations are expressed as follows:

$$\begin{bmatrix} X_1(e^{j\omega}) \\ X_2(e^{j\omega}) \end{bmatrix} = \begin{bmatrix} H_{11}(e^{j\omega}) & H_{12}(e^{j\omega}) \\ H_{21}(e^{j\omega}) & H_{22}(e^{j\omega}) \end{bmatrix} \begin{bmatrix} S_1(e^{j\omega}) \\ S_2(e^{j\omega}) \end{bmatrix} \quad (2)$$

$$\begin{aligned} \begin{bmatrix} Y_1(e^{j\omega}) \\ Y_2(e^{j\omega}) \end{bmatrix} &= \begin{bmatrix} W_{11}(e^{j\omega}) & W_{12}(e^{j\omega}) \\ W_{21}(e^{j\omega}) & W_{22}(e^{j\omega}) \end{bmatrix} \begin{bmatrix} X_1(e^{j\omega}) \\ X_2(e^{j\omega}) \end{bmatrix} \\ &= \begin{bmatrix} W_{11}(e^{j\omega}) & W_{12}(e^{j\omega}) \\ W_{21}(e^{j\omega}) & W_{22}(e^{j\omega}) \end{bmatrix} \\ &\times \begin{bmatrix} H_{11}(e^{j\omega}) & H_{12}(e^{j\omega}) \\ H_{21}(e^{j\omega}) & H_{22}(e^{j\omega}) \end{bmatrix} \begin{bmatrix} S_1(e^{j\omega}) \\ S_2(e^{j\omega}) \end{bmatrix} \quad (3) \\ &= \begin{bmatrix} A_{11}(e^{j\omega}) & A_{12}(e^{j\omega}) \\ A_{21}(e^{j\omega}) & A_{22}(e^{j\omega}) \end{bmatrix} \begin{bmatrix} S_1(e^{j\omega}) \\ S_2(e^{j\omega}) \end{bmatrix} \quad (4) \end{aligned}$$

Separation is complete if the following conditions are satisfied.

$$\begin{aligned} \begin{bmatrix} A_{11}(e^{j\omega}) & A_{12}(e^{j\omega}) \\ A_{21}(e^{j\omega}) & A_{22}(e^{j\omega}) \end{bmatrix} &= \begin{bmatrix} C_{11}(e^{j\omega}) & 0 \\ 0 & C_{22}(e^{j\omega}) \end{bmatrix} \\ \text{or} &\begin{bmatrix} 0 & C_{12}(e^{j\omega}) \\ C_{21}(e^{j\omega}) & 0 \end{bmatrix} \quad (5) \end{aligned}$$

B. Learning Algorithm

Filter coefficients in the separation block are trained following [10],[11],

$$W(t+1) = W(t) + \eta [I - \langle \Phi(Y) Y^H \rangle] W(t) \quad (6)$$

$$\Phi(Y) = \frac{1}{1 + \exp(-Y^{(R)})} + j \frac{1}{1 + \exp(-Y^{(I)})} \quad (7)$$

III. SIGNAL SEPARATION AND DISTORTION

A. BSS Learning Process

The BSS learning adjusts the separation block so as to make the output signals statistically independent. There are two kinds of directions for this adjustment. The first direction is to extract only one signal source at a single output. The second direction is to make the output signal spectra to be flat, just as the white noise, in order to reduce correlation among the output signals. When there exists a degree of freedom to adjust the separation block, these two directions can be simultaneously satisfied. The second direction causes signal distortion. On the other hand, if a degree of freedom does not exist, two directions conflict with each other, and prevent signal separation.

In the network shown in Fig.1, if the separation conditions Eq.(5) are satisfied, the signal sources are completely separated, that is the first direction is satisfied. The second direction is to adjust $|Y_i(e^{j\omega})| = |C_{ij}(e^{j\omega})S_i(e^{j\omega})|$ to be uncorrected to each other. As a result, $Y_i(e^{j\omega})$ are deviated from $H_{ij}(e^{j\omega})S_i(e^{j\omega})$, which should be extracted at the output. Effects of $H_{ij}(e^{j\omega})$ cannot be compensated for.

B. Conditions for Signal Separation

Assume S_1 and S_2 are separated into Y_1 and Y_2 , respectively. This assumption does not lose any generality. The conditions for complete separation are given by

$$W_{11}(z)H_{12}(z) + W_{12}(z)H_{22}(z) = 0 \quad (8)$$

$$W_{21}(z)H_{11}(z) + W_{22}(z)H_{21}(z) = 0 \quad (9)$$

Let $W_{ij}(z)$ and $H_{ij}(z)$ be expressed by FIR filters, and their order be $N - 1$ and $K - 1$, respectively. The above equations include $N + K - 1$ terms, $z^0, z^{-1}, \dots, z^{-(N+K-2)}$, whose coefficients should be zero, and $2N$ coefficients of $W_{ij}(z)$, which are variables of the equations. The number of equations is $N + K - 1$ and that of the variables is $2N$. Therefore, if $N + K - 1 < 2N$, the solution of the equations become uncertainty solutions. The uncertain solutions mean there exit a degree of freedom to adjust $W_{ij}(z)$ toward the second direction. On the other hand, if $N + K - 1 > 2N$, then separation is not enough. In this case, the separation block is adjusted toward the second direction to some extent. In other words, Two directions conflict with each other, and prevent the signal separation as a result.

As described above, $W_{ij}(z)$ are determined so as to satisfy the separation conditions Eqs.(8) and (9), and to make $|Y(e^{j\omega})|$ flat. If there is a degree of freedom, these two kinds of directions are simultaneously satisfied. This causes sever signal distortions. Especially, when the signal sources are speech signals, their spectrum are similar, then signal distortion becomes a sever problem.

C. Simple Example

A simple example is demonstrated here using the circuit shown in Fig.2. In this case, $A_{ij}(e^{j\omega})$ are given by

$$A_{11}(e^{j\omega}) = 1 - 0.81e^{-j2\omega} \quad (10)$$

$$A_{12}(e^{j\omega}) = 0 \quad (11)$$

$$A_{21}(e^{j\omega}) = 0 \quad (12)$$

$$A_{22}(e^{j\omega}) = 1 - 0.81e^{-j2\omega} \quad (13)$$

Eventhough $X_i(e^{j\omega})$ include $S_i(e^{j\omega})$ without any distortion. However, the BSS outputs $Y_i(e^{j\omega})$ are distorted by $1 - 0.81e^{-j2\omega}$ from $X_i(e^{j\omega})$. This is the signal distortion discussed in this paper.

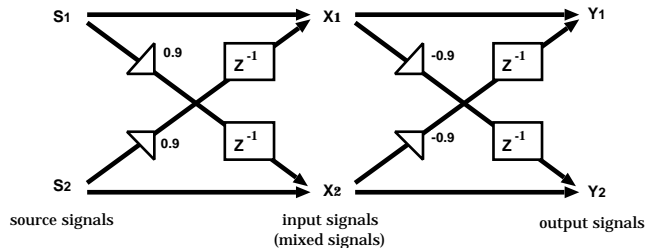


Fig. 2. A simple example of BSS with 2 signal sources and 2 sensors.

D. Limitation of Signal Separation

Assume the signal sources are the white noises. Complete separation requires the conditions Eqs.(10)-(13). However, these conditions cause signal distortion, that is $|Y_i(e^{j\omega})|$ deviates from the white spectrum. This deviation causes another correlation between Y_1 and Y_2 . In another words, Y_1 and Y_2 cannot be independent. This is the conflict of two directions, mentioned above, and prevent the signal separation.

IV. SIMULATION AND DISCUSSIONS

A. Simulation Conditions

Impulse responses of the mixture are shown in Fig.3.

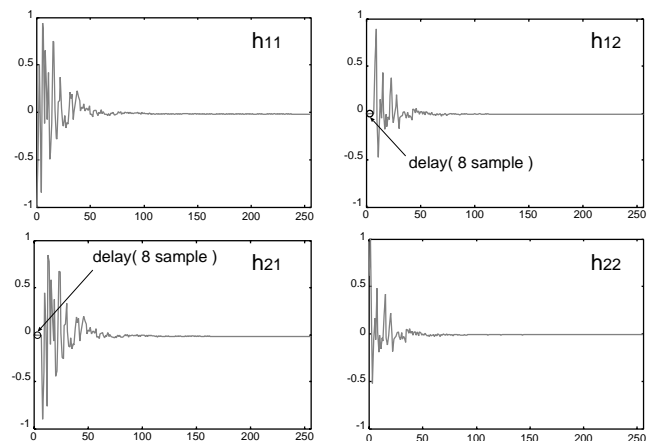


Fig. 3. Impulse responses of mixture. Cross paths $h_{12}(n)$ and $h_{21}(n)$ include 8 sample delay.

$h_{21} = 0.9h_{11}$ and $h_{12} = 0.9h_{22}$. $h_{ij}(n)$ have time delay by 1, 2, 4, 8, 16 samples. FFT size is 128 points, the stepsize $\mu = 0.0002$, and the initial guess of the separation block are $W_{11}(z) = W_{22}(z) = 1$ and $W_{ij}(z) = 0, i \neq j$.

Separation performance is evaluated by

$$SNR = 10 \log \frac{\sum_{\omega} |A_{11}(e^{j\omega})|^2 + \sum_{\omega} |A_{22}(e^{j\omega})|^2}{\sum_{\omega} |A_{12}(e^{j\omega})|^2 + \sum_{\omega} |A_{21}(e^{j\omega})|^2} \quad (14)$$

B. Separation Performance

Learning curves for the white signal sources and the speech signal sources are shown in Figs.4 and 5, respectively. The horizontal axis is iteration number, and the vertical axis is SNR , respectively. When the cross paths have no time delay, the separation performance is good. However, if they include some time delay, the separation is not good. The directions of canceling S_1 and S_2 in Y_2 and Y_1 , respectively, and making the output signal spectra flat conflict with each other.

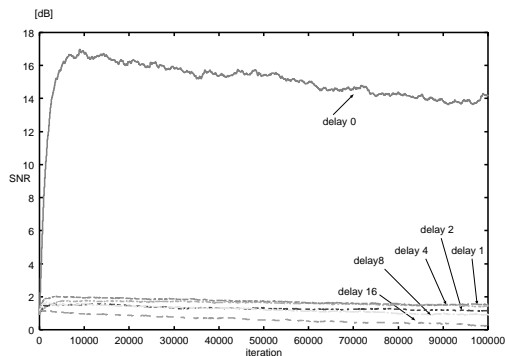


Fig. 4. Separation performances for white signal sources.

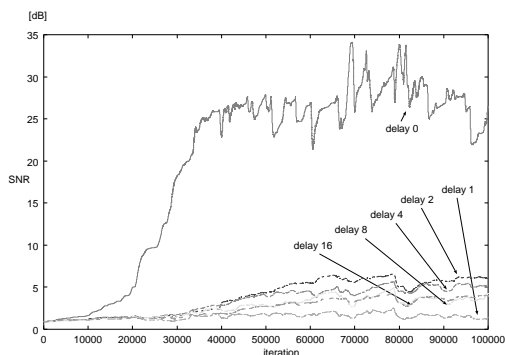


Fig. 5. Separation performances for speech signal sources.

C. Transfer Functions $A_{ij}(e^{j\omega})$ and Output Spectra

In the case of the white signal sources and without time delay, $|A_{ij}(e^{j\omega})|$ are shown in Fig.6. The horizontal axis indicates the frequency axis $[0, f_s]$, f_s is the sampling frequency. In this figure, $|A_{11}(e^{j\omega})| = |A_{22}(e^{j\omega})| \simeq 0.15$ and $|A_{12}(e^{j\omega})| = |A_{21}(e^{j\omega})| \simeq 0.02$, thus, good separation is obtained.

On the other hand, when the cross paths have one sample delay, $|A_{ij}(e^{j\omega})|, i \neq j$ are not well reduced due to the confliction as shown in Fig.7.

$|A_{ij}(e^{j\omega})|$ for the speech signal sources without and with one sample delay are shown in Figs.8 and 9, respectively. The same properties are demonstrated. However, in this case, the spectra in high frequency band are amplified for no time delay. This means canceling S_1 and S_2 in Y_2 and Y_1 are well achieved, at the same time, whitening the output spectra are accomplished by using a degree of freedom. As a result, the speech waveforms at the outputs are extremely distorted.

Figure 10 shows $|S_1(e^{j\omega})|$, $|A_{11}(e^{j\omega})|$ and $|Y_1(e^{j\omega})|$. The

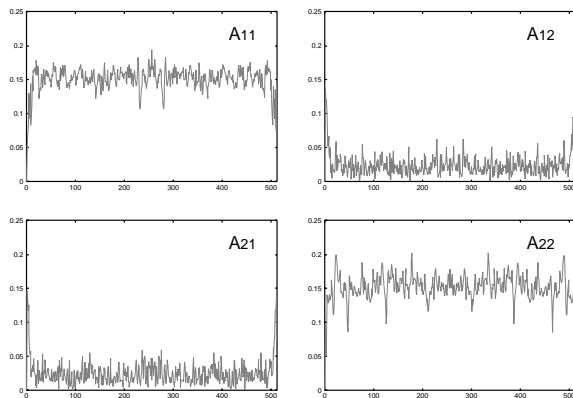


Fig. 6. $|A_{ij}(e^{j\omega})|$ for white signal sources without time delay.

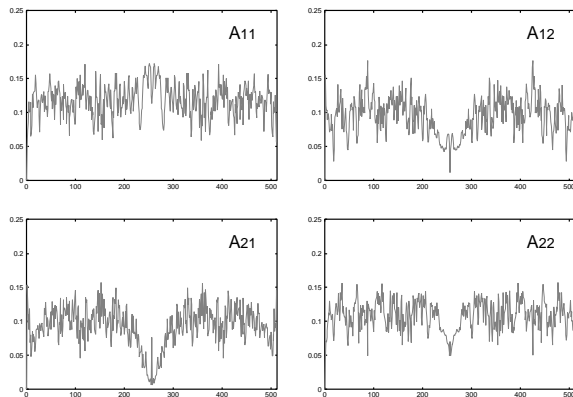


Fig. 7. $|A_{ij}(e^{j\omega})|$ for white signal sources with one sample delay.

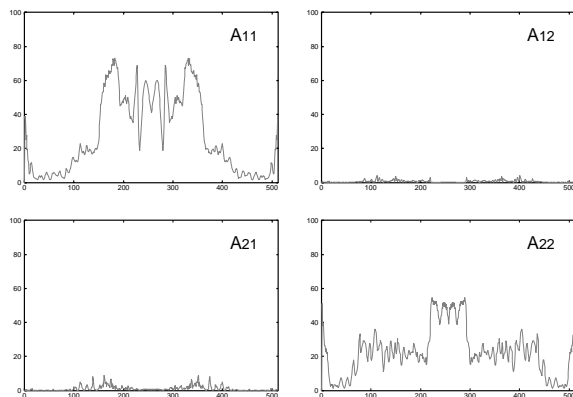


Fig. 8. $|A_{ij}(e^{j\omega})|$ for speech signal sources without time delay.

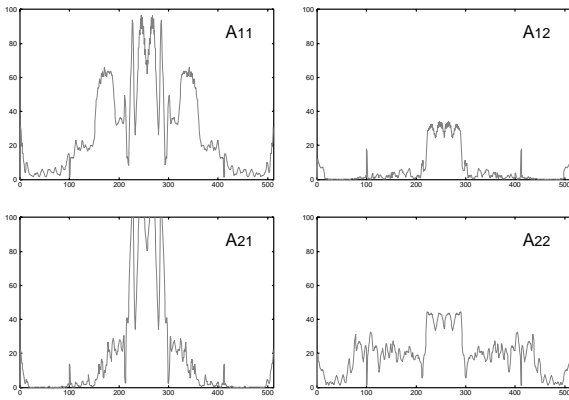


Fig. 9. $|A_{ij}(e^{j\omega})|$ for speech signal sources with one sample delay.

signal source spectrum is flattened at the output. Figure 11 shows the waveform of $s_1(n)$ and $y_1(n)$. $y_1(n)$ is severely distorted from $s_1(n)$. This distortion cannot be compensated for in the BSS learning algorithm given by Eqs.(6) and (7). We must develop another regularization for reducing the signal distortion.

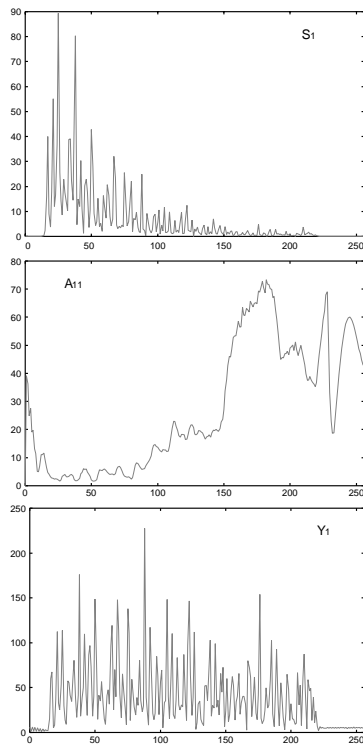


Fig. 10. Whitening output spectra for speech signal sources without time delay.

V. CONCLUSIONS

In this paper, signal separation performance and signal distortion in the feedforward BSS have been analyzed. Through the BSS learning process, the separation block is adjusted toward two directions, that is, canceling the interference signal sources and making the output signal spectra to be white. These directions easily conflict with each other, and limitation of separation and sever signal

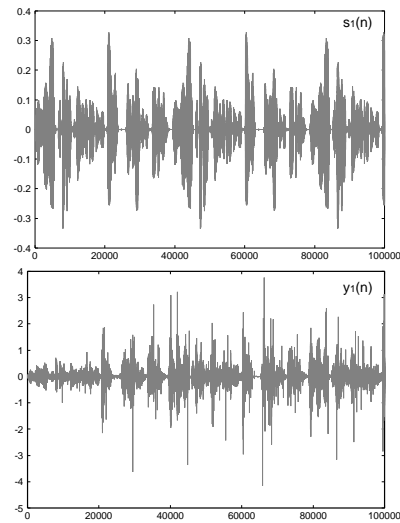


Fig. 11. Waveform of $s_1(n)$ and $y_1(n)$, under no time delay.

distortion occur. Another regularization for reducing signal distortion should be developed.

REFERENCES

- [1] C.Jutten and Jeanny Herault, "Blind separation of sources, Part I: An adaptive algorithm based on neuromimetic architecture", *Signal Processing*, 24, pp.1-10, 1991.
- [2] H.L.Nguyen Thi and C.Jutten, "Blind source separation for convolutive mixtures", *Signal Processing*, vol.45, no.2, pp.209-229, March 1995.
- [3] C.Simon, G.d' Urso, C.Vignat, Ph.Loubaton and C.Jutten, "On the convolutive mixture source separation by the decorrelation approach", *IEEE Proc. ICASSP'98*, Seattle, pp.IV-2109-2112, May 1998.
- [4] S.Cruces and L.Castedo, "A Gauss-Newton methods for blind source separation of convolutive mixtures", *IEEE Proc. ICASSP'98*, Seattle, pp.IV2093-2096, May 1998. 2002.
- [5] K.Nakayama, A.Hirano and A.Horita, "A learning algorithm for convolutive blind source separation with transmission delay constraint", *IEEEINNS, Proc. IJCNN'2002*, Honolulu, Hawaii, pp.1287-1292, May 2002.
- [6] K.Nakayama, A.Hirano and A.Horita, "A learning algorithm with adaptive exponential stepsize for blind source separation of sonvolutive mixtures with reverberations", *Proc. IJCNN'2003* July 2003.
- [7] K. Nakayama, A.Hirano and T.Sakai, "An adaptive nonlinear function controlled by estimated output pdf for blind source separation", *4th International Symposium on Independent Component Analysis and Blind Source Separation (ICA2003)*, Nara, Japan, pp.427-432, April 2003.
- [8] K.Nakayama, A.Hirano and T.Nishiwaki", "A cascade form blind source separation connecting source separation and linearization for nonlinear mixtures", *IEEE&INNS Proc. IJCNN'03*, Portland, Oregon, July 2003.
- [9] A.Cichocki, S.Amari, M.Adachi, W.Kasprzak, "Self-adaptive neural networks for blind separation of sources", *Proc. IS-CAS'96*, Atlanta, pp.157-161, 1996.
- [10] S.Amari, T.Chen and A.Cichocki, "Stability analysis of learning algorithms for blind source separation", *Neural Networks*, vol.10, no.8, pp.1345-1351, 1997.
- [11] S.Araki, S.Makino, T.Nishikawa and H.Saruwatari, "Fundamental limitation of frequency domain blind source separation for convolutive mixture of speech", *IEEE Proc. ICASSP'01*, Salt Lake City, MULT-P2.3, May 2001.
- [12] I.Kopriva, Z.Devcic and H.Szu, "An adaptive short-time frequency domain algorithm for blind separation of nonstationary convolved mixtures", *IEEE INNS Proc. IJCNN'01*, pp.424-429, July 2001.