

A learning algorithm for convolutive blind source separation with transmission delay constraint

著者	Nakayama Kenji, Hirano Akihiro, Horita Akihide
journal or publication title	IEEE&INNS, Proc. IJCNN'2002, Honolulu, Hawai
page range	1287-1292
year	2002-05-01
URL	http://hdl.handle.net/2297/6844

A Learning Algorithm for Convolutional Blind Source Separation with Transmission Delay Constraint

Kenji NAKAYAMA

Akihiro HIRANO

Akihide HORITA

Dept of Information and Systems Eng., Faculty of Eng., Kanazawa Univ.
2-40-20 Kodatsuno, Kanazawa, 920-8667, Japan
e-mail: nakayama@t.kanazawa-u.ac.jp

Abstract

A learning algorithm is proposed for fully recurrent convolutional blind source separation. Let $s_i(n)$ and $x_j(n)$ be the signal sources and the observations. $H_{ji}(z)$ expresses a transfer function from $s_i(n)$ to $x_j(n)$. It is assumed that transmission delay time of $H_{ji}(z)$, $j \neq i$ is longer than that of $H_{ii}(z)$. In many practical applications, this assumption is acceptable. Based on this assumption, $s_i(n)$ in the output $y_j(n)$, $j \neq i$ of an unmixing block is cancelled through the feedback $C_{ji}(z)$ from the i th output to the j th observation. However, $s_i(n)$ in the output $y_i(n)$ cannot be cancelled, because a noncausal $C_{ij}(z)$ is required. A cost function $E[q(y_j(n))]$ can be used, where $q()$ is an even function with a single minimum point. Coefficients of $C_{ji}(z)$, that is $c_{ji}(l)$ are updated following a gradient descent method. The correction term is expressed $\mu \dot{q}(y_j(n)) y_i(n-l)$. $\dot{q}()$ is a partial derivative of $q()$. Two channel blind source separation has been simulated using speech signals. 100th- and 70th-order FIR filters are used for $C_{12}(z)$ and $C_{21}(z)$, respectively. A power ratio of the main signals and the cross components is about 15dB.

1 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 process and transmission processes are not well known in advance. Under these situations, blind source separation technology using statistical property of the signal sources have become important [1]–[9].

Jutten et al proposed a blind separation algorithm for a fully recurrent network based on statistical independence and symmetrical distribution of the signal sources [4]–[6]. Two stabilization methods have been

proposed [10]. Unstable behavior caused by corruption of symmetrical distribution and imbalance of the signal source levels can be overcome. Furthermore, a pair learning algorithm has been proposed based on Jutten's algorithm [11].

In many applications, mixing processes usually have some characteristics, that is convolutional mixtures. Therefore, unmixing processes should be realized by using FIR or IIR filters. Several methods in a time domain and frequency domain have been proposed. However, when high-order filters are required for the feedbacks $C_{ji}(z)$, a learning process becomes unstable and separation is not enough [12]–[16].

In this paper, a learning algorithm is proposed for fully recurrent blind source separation. Some practical assumption is imposed on transmission delay time of the mixing process. Updating the coefficients requires only the corresponding output $y_j(n)$ and their input $y_k(n-l)$. It is similar to LMS algorithm for adaptive filters. Simulation of two channel blind speech signal separation will be shown in order to confirm usefulness of the proposed method.

2 Network Structure and Equations

2.1 Network Structure

Figure 1 shows a fully recurrent separation model proposed by Jutten et al [4],[12]. The mixing stage has convolutional structure. In this paper, FIR filters are used for the feedback circuits as shown in Fig.2. The number of the signal sources, the sensors and the outputs are all the same.

2.2 Network Equations in Time Domain

The signal sources $s_i(n)$, $i = 1, 2, \dots, N$ are combined through the unknown convolutional mixture block, which has the impulse response $h_{ji}(m)$, and are sensed at N

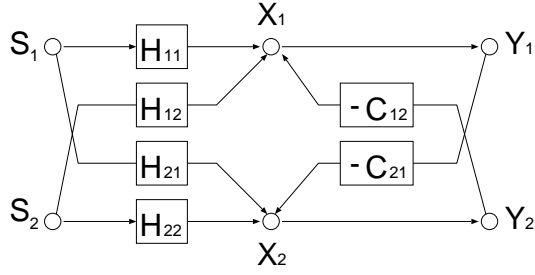


Figure 1: Block diagram of recurrent blind separation.

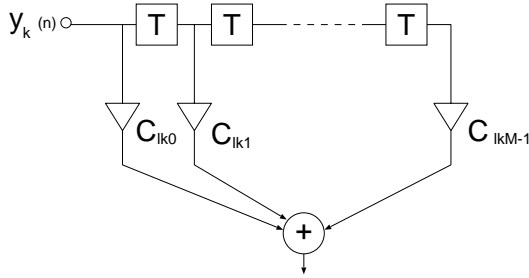


Figure 2: FIR filter used for $C_{21}(z)$ and $C_{12}(z)$ in feedback.

points, resulting in $x_j(n)$.

$$x_j(n) = \sum_{i=1}^N \sum_{m=0}^{M_{ji}-1} h_{ji}(m) s_i(n-m) \quad (1)$$

The output of the separation block $y_j(n)$ is given by

$$y_j(n) = x_j(n) - \sum_{\substack{k=1 \\ k \neq j}}^N \sum_{l=0}^{L_{jk}-1} c_{jk}(l) y_k(n-l) \quad (2)$$

This relation is expressed using vectors and matrices as follows:

$$\mathbf{x}(n) = \mathbf{H}^T \mathbf{s}(n) \quad (3)$$

$$\mathbf{y}(n) = \mathbf{x}(n) - \mathbf{C}^T \tilde{\mathbf{y}}(n) \quad (4)$$

$$\mathbf{s}(n) = [s_1^T(n), s_2^T(n), \dots, s_N^T(n)]^T \quad (5)$$

$$\mathbf{s}_i(n) = [s_i(n), s_i(n-1), \dots, s_i(n-M_i+1)]^T \quad (6)$$

$$\mathbf{x}(n) = [x_1(n), x_2(n), \dots, x_N(n)]^T \quad (7)$$

$$\mathbf{y}(n) = [y_1(n), y_2(n), \dots, y_N(n)]^T \quad (8)$$

$$\tilde{\mathbf{y}}(n) = [\mathbf{y}_1^T(n), \mathbf{y}_2^T(n), \dots, \mathbf{y}_N^T(n)]^T \quad (9)$$

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{21} & \dots & h_{N1} \\ h_{12} & h_{22} & \dots & h_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ h_{1N} & h_{2N} & \dots & h_{NN} \end{bmatrix} \quad (10)$$

$$\mathbf{h}_{ji} = [h_{ji}(0), h_{ji}(1), \dots, h_{ji}(M_{ji}-1)]^T \quad (11)$$

$$\mathbf{C} = \begin{bmatrix} 0 & c_{21} & \dots & c_{N1} \\ c_{12} & 0 & \dots & c_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ c_{1N} & c_{2N} & \dots & 0 \end{bmatrix} \quad (12)$$

$$\mathbf{c}_{jk} = [c_{jk}(0), c_{jk}(1), \dots, c_{jk}(L_{jk}-1)]^T \quad (13)$$

$$(14)$$

2.3 Network Equations in z-Domain

Letting $S_i(z)$, $X_j(z)$ and $Y_k(z)$ be z-transform of $s_i(n)$, $x_j(n)$ and $y_k(n)$, respectively, they are related as follows:

$$\mathbf{X}(z) = \mathbf{H}(z) \mathbf{S}(z) \quad (15)$$

$$\mathbf{Y}(z) = \mathbf{X}(z) - \mathbf{C}(z) \mathbf{Y}(z) \quad (16)$$

$$\mathbf{S}(z) = [S_1(z), S_2(z), \dots, S_N(z)]^T \quad (17)$$

$$\mathbf{X}(z) = [X_1(z), X_2(z), \dots, X_N(z)]^T \quad (18)$$

$$\mathbf{Y}(z) = [Y_1(z), Y_2(z), \dots, Y_N(z)]^T \quad (19)$$

$$\mathbf{H}(z) = \begin{bmatrix} H_{11}(z) & H_{12}(z) & \dots & H_{1N}(z) \\ H_{21}(z) & H_{22}(z) & \dots & H_{2N}(z) \\ \vdots & \vdots & \ddots & \vdots \\ H_{N1}(z) & H_{N2}(z) & \dots & H_{NN}(z) \end{bmatrix} \quad (20)$$

$$\mathbf{C}(z) = \begin{bmatrix} 0 & C_{12}(z) & \dots & C_{1N}(z) \\ C_{21}(z) & 0 & \dots & C_{2N}(z) \\ \vdots & \vdots & \ddots & \vdots \\ C_{N1}(z) & C_{N2}(z) & \dots & 0 \end{bmatrix} \quad (21)$$

From these expressions, a relation between the signal sources and the separation outputs becomes

$$\begin{aligned} \mathbf{Y}(z) &= (\mathbf{I} + \mathbf{C}(z))^{-1} \mathbf{X}(z) \\ &= (\mathbf{I} + \mathbf{C}(z))^{-1} \mathbf{H}(z) \mathbf{S}(z) \end{aligned} \quad (22)$$

The following matrix can be regarded as a separation matrix.

$$\mathbf{W}(z) = (\mathbf{I} + \mathbf{C}(z))^{-1} \quad (23)$$

In order to evaluate separation performance, the following matrix is defined.

$$\mathbf{P}(z) = \mathbf{W}(z) \mathbf{H}(z) \quad (24)$$

If $\mathbf{P}(z)$ takes the next forms, then the signal sources $s_i(n)$ are completely separated at the outputs $y_k(n)$. Two channel case is shown here.

$$\mathbf{P}(z) = \begin{bmatrix} P_{11}(z) & 0 \\ 0 & P_{22}(z) \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} 0 & P_{12}(z) \\ P_{21}(z) & 0 \end{bmatrix} \quad (25)$$

Letting $p_{ki}(n)$ be an impulse response of $P_{ki}(z)$, if

$$p_{ki}(n) = 0, \quad n \neq n_0 \quad (26)$$

then, the separated signal becomes

$$y_k(n) = p_{ki}(n_0)s_i(n - n_0) \quad (27)$$

This output is obtained by amplifying and shifting $s_i(n)$. This means no distortion on the separated signals.

3 Learning Algorithm

3.1 Assumption on Transmission Delay

For simplicity, 2-channel case is taken into account. It is assumed that delay time of $H_{11}(z)$ and $H_{22}(z)$ are shorter than that of $H_{21}(z)$ and $H_{12}(z)$. This means that in Fig.2, the sensor of X_1 is located close to $s_1(n)$, and the sensor of X_2 close to $s_2(n)$.

The separation conditions are given by

$$(1) \quad C_{21}(z) = \frac{H_{21}(z)}{H_{11}(z)} \quad C_{12}(z) = \frac{H_{12}(z)}{H_{22}(z)} \quad (28)$$

$$y_1(n) = \mathbf{h}_{11}^T \mathbf{s}_1(n) \quad y_2(n) = \mathbf{h}_{22}^T \mathbf{s}_2(n) \quad (29)$$

$$(2) \quad C_{21}(z) = \frac{H_{22}(z)}{H_{12}(z)} \quad C_{12}(z) = \frac{H_{11}(z)}{H_{21}(z)} \quad (30)$$

$$y_1(n) = \mathbf{h}_{12}^T \mathbf{s}_2(n) \quad y_2(n) = \mathbf{h}_{21}^T \mathbf{s}_1(n) \quad (31)$$

From the assumption on the transmission delay time in $H_{ji}(z)$, the solutions in (1) become causal systems, which can be physically realized. On the other hand, the solutions in (2) are noncausal, which cannot be realized.

After separation, the output $y_k(n)$ is not exactly the same as the signal sources $s_i(n)$. Effect of $H_{11}(z)$ and $H_{22}(z)$ still remain.

3.2 Cost Function

From Eq.(22), the outputs can be expressed for 2-channel blind separation as follows:

$$\begin{aligned} \begin{bmatrix} Y_1(z) \\ Y_2(z) \end{bmatrix} &= \frac{1}{1 - C_{12}(z)C_{21}(z)} \begin{bmatrix} 1 & -C_{12}(z) \\ -C_{21}(z) & 1 \end{bmatrix} \\ &\times \begin{bmatrix} H_{11}(z) & H_{12}(z) \\ H_{21}(z) & H_{22}(z) \end{bmatrix} \begin{bmatrix} S_1(z) \\ S_2(z) \end{bmatrix} \\ &= \frac{1}{1 - C_{12}(z)C_{21}(z)} \\ &\times \begin{bmatrix} H_{11}(z) - C_{12}(z)H_{21}(z) & H_{12}(z) - C_{12}(z)H_{22}(z) \\ H_{21}(z) - C_{21}(z)H_{11}(z) & H_{22}(z) - C_{21}(z)H_{12}(z) \end{bmatrix} \\ &\times \begin{bmatrix} S_1(z) \\ S_2(z) \end{bmatrix} \end{aligned} \quad (32)$$

As described in Sec.3.1, transmission delay time of $H_{ji}(z)$, $j \neq i$ is longer than that of $H_{ii}(z)$. Also, $C_{12}(z)$ and $C_{21}(z)$ are causal circuits, which have positive transmission delay. Therefore, In Eq.(33), the diagonal elements cannot be zero. On the other hand, the nondiagonal elements can be zero by adjusting the feedback coefficients $C_{12}(z)$ and $C_{21}(z)$.

Therefore, a cost function can be defined as follows:

$$J_j(n) = E[q(y_j(n))] \quad (34)$$

$q()$ is an even function with a single minimum point. By minimizing the above cost function, the nondiagonal elements can be minimized, while the diagonal elements can hold some level. Instead of $E[q(y_j(n))]$, the instantaneous value $q(y_j(n))$ is used like the LMS algorithm for adaptive filters [7].

$$\hat{J}_j(n) = q(y_j(n)) \quad (35)$$

3.3 Update Equation for $C_{jk}(z)$

The gradient of $\hat{J}_j(n)$ becomes

$$\begin{aligned} \frac{\partial \hat{J}_j(n)}{\partial c_{jk}(l)} &= \frac{\partial q(y_j(n))}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial c_{jk}(l)} \\ &= \dot{q}(y_j(n)) y_k(n - l) \end{aligned} \quad (36)$$

$$y_j(n) = x_j(n) - \sum_{l=0}^{L_{jk}-1} c_{jk}(l) y_k(n - l) \quad (37)$$

$\dot{q}()$ is a partial derivative of $q()$, which is an odd function. If $k = 1$, then $j = 2$, and vice versa. Therefore, the update equation of $c_{jk}(l)$ is given by

$$c_{jk}(n + 1, l) = c_{jk}(n, l) + \Delta c_{jk}(n, l) \quad (38)$$

$$\Delta c_{jk}(n, l) = \mu \dot{q}(y_j(n)) y_k(n - l) \quad (39)$$

3.4 Statistical Analysis of Convergence

The probability density function (pdf) of the signal sources are assumed to be even functions. Furthermore, the signal sources are statistically independent to each other. Then, they satisfy

$$\begin{aligned} E[f(s_1(n))g(s_2(n))] &= E[f(s_1(n))]E[g(s_2(n))] \\ &= 0 \end{aligned} \quad (40)$$

$f()$ and $g():$ odd functions

On the other hand, if a very small learning rate μ is used in Eq.(39), the correction term can be regarded as $E[\dot{q}(y_j(n))y_k(n-l)]$. Since, $\dot{q}(y_j(n))$ and $y_k(n-l)$ are also odd functions, then Eq.(40) can be held. This means that as $y_1(n)$ and $y_2(n)$ approach the $\mathbf{h}_{11}^T \mathbf{s}_1(n)$ and $\mathbf{h}_{22}^T \mathbf{s}_2(n)$, respectively, the correction terms can be reduced, finally they can be zero.

3.5 Nonlinear Functions

As shown in Eq.(39), $\dot{q}(y_j(n))$ and $y_j(n-l)$ are used for nonlinear functions. They are odd functions. Furthermore, $y_j(n-l)$ can be replaced by another squashing function in order to achieve stable convergence. Optimum nonlinear functions are also highly dependent on a pdf of the output signal [9],[17]. This subject is not discussed in this paper. Two kinds of combinations of $f()$ and $g()$ shown below are considered in simulation.

$$f(y) = \tanh(\alpha y) \quad g(y) = \tanh(\beta y) \quad (41)$$

$$f(y) = \tanh(\alpha y) \quad g(y) = y \quad (42)$$

3.6 Convergence Property

When the transmission delays satisfy the condition, that is delay of $H_{ji}(z), j \neq i$ is longer than that of $H_{ii}(z)$, $s_j(n)$ can be cancelled in $y_i(n)$, and $s_i(n)$ cannot be cancelled in $y_i(n)$. As a result $s_i(n)$ can be separated in $y_i(n)$. On the other hand, signal separation is highly dependent on the signal levels in the observations [11]. Since, the initial guess for $c_{jk}(n)$ is set to zero. In early stage in a learning process, $x_i(n)$ is mainly extracted in $y_i(n)$. Thus, the signal source $s_j(n)$, whose power is dominant in $x_i(n)$, is also dominant in $y_i(n)$. This $s_j(n)$ will cancel $s_j(n)$ included in the other outputs. However, in the convolutive blind separation, this cancellation is strongly affected by transmission delays. For example, the transmission delays satisfy the above conditions, and $s_i(n)$ is not dominant in $x_i(n)$, convergence is not good. Usually, when the sensor $x_i(n)$ is located close to the signal source $s_i(n)$, the conditions on signal power and transmission delay for convergence can be satisfied.

3.7 Comparison with Other Algorithm

Thi and Jutten proposed a learning algorithm based on the fully recurrent network shown in Fig.1 [12]. The

coefficients $c_{jk}(n, l)$ are updated by cancelling cross-cumulants $\text{Cum}_{22}(y_j(n)y_k(n-l))$. Furthermore, nonlinear functions are generalized as in instantaneous mixtures [4]. On the contrary, in our method, the cost function is given by $E[q(y_j(n))]$, where $q()$ is an even function having a single minimum point. The algorithm is derived following the gradient descent method. Convergence is guaranteed by the assumption on transmission delay, which was not mentioned in [12]. Even though the update equation is the same, the derivation process is different. Furthermore, in our approach, convergence property can be discussed based on signal levels in the observations and transmission delays.

4 Simulation

4.1 Simulation Conditions

Signal Sources and Nonlinear Functions

Two channel blind separation of speech signals was simulated. The speech signals are shown in Fig.3. A learn-

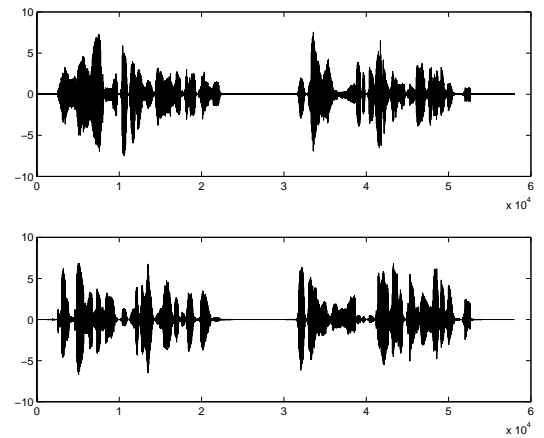


Figure 3: Waveforms of speech signal sources.

ing rate is 0.001. The following nonlinear functions are used.

$$(1) \quad f(y) = \tanh(5y) \quad g(y) = \tanh(y) \quad (43)$$

$$(2) \quad f(y) = \tanh(5y) \quad g(y) = y \quad (44)$$

In the case (2), the learning rate should be reduced for stable convergence. The simulation results are almost the same, then the results using the case (1) are shown in the following.

Measure of Separation

The separation performance is evaluated by the follow-

ing SNR, defined by using $\mathbf{P}(z)$ in Eq.(24)

$$\sigma_s^2 = \sum_{i=1}^2 \frac{1}{2\pi} \int_{-\pi}^{\pi} |P_{ii}(e^{j\omega T})|^2 d\omega T \quad (45)$$

$$\sigma_c^2 = \sum_{j \neq i} \frac{1}{2\pi} \int_{-\pi}^{\pi} |P_{ji}(e^{j\omega T})|^2 d\omega T \quad (46)$$

$$SNR = 10 \log \frac{\sigma_s^2}{\sigma_c^2} \quad [\text{dB}] \quad (47)$$

σ_s^2 expresses power of the selected signals and σ_c^2 is that of the cross components.

Mixing Convolutional Matrix

$H_{ji}(z)$ are 20 tap FIR filters. The ideal impulse response for $C_{12}(z)$ and $C_{21}(z)$ are shown in Figs.4 and 5, respectively. From these figures, 100th- and 70th-order FIR filters are required for $C_{12}(z)$ and $C_{21}(z)$, respectively.

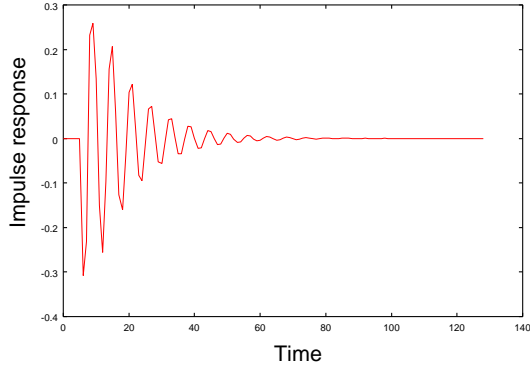


Figure 4: Ideal impulse response of $C_{12}(z) = H_{12}(z)/H_{22}(z)$.

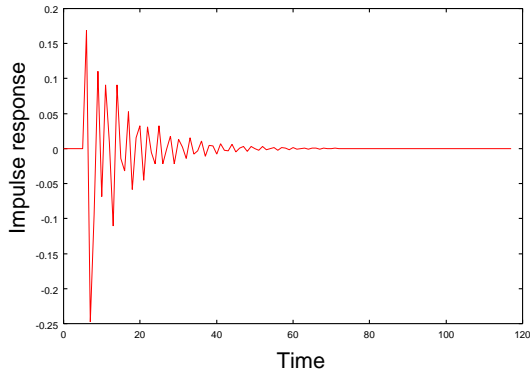


Figure 5: Ideal impulse response of $C_{21}(z) = H_{21}(z)/H_{11}(z)$.

4.2 Separation Performances

SNR defined by Eq.(47) is shown in Fig.6. Approximately, $SNR = 15\text{dB}$ is obtained. Furthermore, the separated waveforms are shown in Fig.7. They are almost the same as the signal sources. Figures 8, 9 show the amplitude response of $P_{ii}(z)$ with a solid line and those of $H_{ii}(z)$ with a dashed line. $P_{ii}(z)$ almost approximates $H_{ii}(z)$. Effect of $H_{ii}(z)$ remains on $y_i(n)$ in this method. In order to equalize $H_{ii}(z)$, another statistical approaches should be employed.

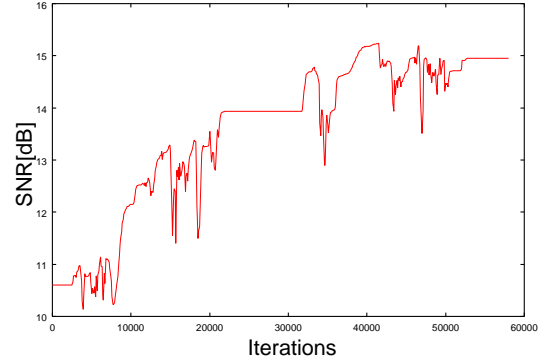


Figure 6: Learning curves of proposed method. SNR is defined by Eq.(47).

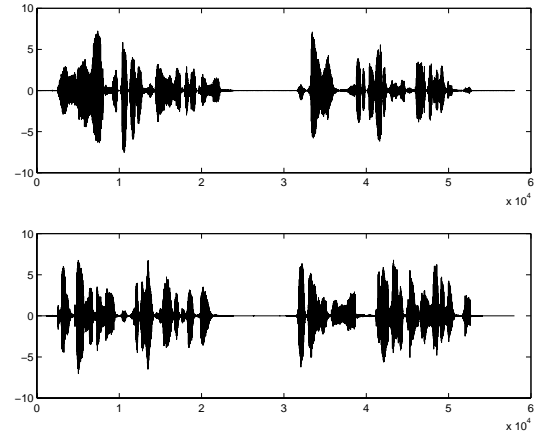


Figure 7: Waveforms of separation block outputs.

5 Conclusions

A simple LMS like learning algorithm has been proposed for a fully recurrent convolutional blind separation. Some assumption is imposed on transmission delay time in the mixing process, which are practically acceptable. A cost

function and the learning algorithm have been derived based on this assumption. Simulation for two channel speech source separation has been carried out. High order FIR filters with 100th- and 70th-order, are used in an unmixing process. A power ratio of the main and the cross components is about 15dB.

References

[1] P.Comon, "Separation of stochastic process whose linear mixtures observed", Proc. ONR- NSF-IEEE Workshop on Higher Spectral Analysis Vail, Colorado, pp.174-179, June 28-30, 1989.

[2] P.Comon, "Separation of sources using higher-order cumulants", SPIE Conference, Vol.1152, Advanced Algorithms and Architectures for Signal Processing IV, San Diego, pp. 170-181, August 6-11, 1989.

[3] J.F.Cardoso, "Eigen structure of the 4th order cumulant tensor with application to the blind source separation problem", ICASSP'90 Proc. pp. 2655-2658, 1990.

[4] C.Jutten and Jeanny Herault, "Blind separation of sources, Part I: An adaptive algorithm based on neuromimetic architecture", Signal Proc., 24, pp.1-10, 1991.

[5] P.Comon, C.Jutten and J.Herault, "Blind separation of sources, Part II: Problems statement", Signal Proc., 24, pp.11-20, 1991.

[6] E.Sorouchyari, "Blind separation of sources, Part III: Stability analysis", Signal Proc., 24, pp.21-29, 1991.

[7] S.Haykin, Adaptive Filter Theory, 3rd ed., Prentice-Hall, Inc. 1996.

[8] A.Cichocki, S.Amari, M.Adachi, W.Kasprzak, "Self-adaptive neural networks for blind separation of sources", Proc. ISCAS'96, Atlanta, pp.157-161, 1996.

[9] 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.

[10] K.Nakayama, A.Hirano and M.Nitta, "A constraint learning algorithm for blind source separation", IEEE INNS, Proc. IJCNN'2000, Como, Italy, pp.24-27, July, 2000.

[11] K.Nakayama, A.Hirano and T.Sakai, "A pair-channel learning algorithm with constraints for multi-channel blind separation", IEEE INNS Proc. IJCNN'01, Washington DC, July 2001.

[12] H.L.Nguyen Thi and C.Jutten, "Blind source separation for convolutive mixtures", Signal Processing, vol.45, no.2, pp.209-229, March 1995.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] H.Mathis and S.C.Douglas, "On optimal and universal nonlinearities for blind signal separation", IEEE Proc. ICASSP'01, MULT-P3.3, May 2001.

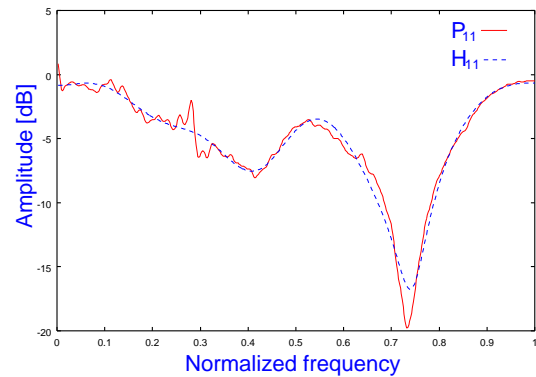


Figure 8: Amplitude response of $P_{11}(z)$ and $H_{11}(z)$.

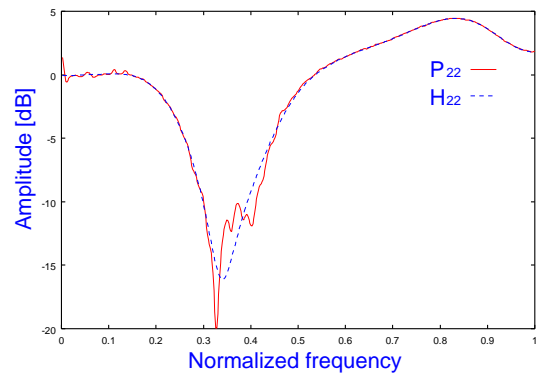


Figure 9: Amplitude response of $P_{22}(z)$ and $H_{22}(z)$.