
Blind Separation of Filtered Sources Using State-Space Approach

Liqing Zhang* and Andrzej Cichocki†
Laboratory for Open Information Systems,
Brain Science Institute, RIKEN
Saitama 351-0198, Wako shi, JAPAN
Email: {zha, cia}@open.brain.riken.go.jp

Abstract

In this paper we present a novel approach to multichannel blind separation/generalized deconvolution, assuming that both mixing and demixing models are described by stable linear state-space systems. We decompose the blind separation problem into two process: separation and state estimation. Based on the minimization of Kullback-Leibler Divergence, we develop a novel learning algorithm to train the matrices in the output equation. To estimate the state of the demixing model, we introduce a new concept, called hidden innovation, to numerically implement the Kalman filter. Computer simulations are given to show the validity and high effectiveness of the state-space approach.

1 Introduction

The field of blind separation and deconvolution has grown dramatically during recent years due to its similarity to the separation feature in human brain, as well as its rapidly growing applications in various fields, such as telecommunication systems, image enhancement and biomedical signal processing. The blind source separation problem is to recover independent sources from sensor outputs without assuming any priori knowledge of the original signals besides certain statistic features. Refer to review papers [1] and [5] for the current state of theory and methods in the field.

Although there exist a number of models and methods, such as the infomax, natural gradient approach and equivariant adaptive algorithms, for separating blindly independent sources, there still are several challenges in generalizing mixture to dy-

*On leave from South China University of Technology, China

†On leave from Warsaw University of Technology, Poland

dynamic and nonlinear systems, as well as in developing more rigorous and effective algorithms with general convergence.[1-9], [11-13]

The state-space description of systems is a new model for blind separation and deconvolution[9,12]. There are several reasons why we use linear state-space systems as blind deconvolution models. Although transfer function models are equivalent to the state-space ones, it is difficult to exploit any common features that may be present in the real dynamic systems. The main advantage of the state space description for blind deconvolution is that it not only gives the internal description of a system, but there are various equivalent types of state-space realizations for a system, such as balanced realization and observable canonical forms. In particular it is known how to parameterize some specific classes of models which are of interest in applications. Also it is much easy to tackle the stability problem of state-space systems using the Kalman Filter. Moreover, the state-space model enables much more general description than standard finite impulse response (FIR) convolutive filtering. All known filtering (dynamic) models, like AR, MA, ARMA, ARMAX and Gamma filterings, could also be considered as special cases of flexible state-space models.

2 Formulation of Problem

Assume that the source signals are a stationary zero-mean i.i.d processes and mutually statistically independent. Let $\mathbf{s}(t) = (s_1(t), \dots, s_n(t))$ be an unknown vector of independent i.i.d. sources. Suppose that the mixing model is described by a stable linear state discrete-time system

$$\bar{\mathbf{x}}(k+1) = \bar{\mathbf{A}}\bar{\mathbf{x}}(k) + \bar{\mathbf{B}}\mathbf{s}(k) + \bar{\mathbf{L}}\xi_P(k), \quad (1)$$

$$\mathbf{u}(k) = \bar{\mathbf{C}}\bar{\mathbf{x}}(k) + \bar{\mathbf{D}}\mathbf{s}(k) + \boldsymbol{\theta}(k), \quad (2)$$

where $\bar{\mathbf{x}} \in \mathbf{R}^r$ is the state vector of system, $\mathbf{s}(k) \in \mathbf{R}^n$ is the vector of source signals and $\mathbf{u}(k) \in \mathbf{R}^m$ is the vector of sensor signals. $\bar{\mathbf{A}}$, $\bar{\mathbf{B}}$, $\bar{\mathbf{C}}$ and $\bar{\mathbf{D}}$ are the mixing matrices of the state space model with consistent dimensions. $\xi_P(k)$ is the process noise and $\boldsymbol{\theta}(k)$ is sensor noise of the mixing system. If we ignore the noise terms in the mixing model, its transfer function matrix is described by a $m \times n$ matrix of the form

$$\mathbf{H}(z) = \bar{\mathbf{C}}(z\mathbf{I} - \bar{\mathbf{A}})^{-1}\bar{\mathbf{B}} + \bar{\mathbf{D}}, \quad (3)$$

where z^{-1} is a delay operator.

We formulate the blind separation problem as a task to recover original signals from observations $\mathbf{u}(t)$ without prior knowledge on the source signals and the state space matrices $[\bar{\mathbf{A}}, \bar{\mathbf{B}}, \bar{\mathbf{C}}, \bar{\mathbf{D}}]$ besides certain statistic features of source signals. We propose that the demixing model here is another linear state-space system, which is described as follows, (see Fig. 1)

$$\mathbf{x}(k+1) = \mathbf{A}\mathbf{x}(k) + \mathbf{B}\mathbf{u}(k) + \mathbf{L}\xi_R(k), \quad (4)$$

$$\mathbf{y}(k) = \mathbf{C}\mathbf{x}(k) + \mathbf{D}\mathbf{u}(k), \quad (5)$$

where the input $\mathbf{u}(k)$ of the demixing model is just the output (sensor signals) of the mixing model and the $\xi_R(k)$ is the reference model noise. \mathbf{A} , \mathbf{B} , \mathbf{C} and \mathbf{D} are the demixing matrices of consistent dimensions. In general, the matrices $\mathbf{W} = [\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}, \mathbf{L}]$ are parameters to be determined in learning process.

For simplicity, we do not consider, at this moment, the noise terms both in the mixing and demixing models. The transfer function of the demixing model is $\mathbf{W}(z) = \mathbf{C}(z\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D}$. The output $\mathbf{y}(k)$ is designed to recover the source signals in the following sense

$$\mathbf{y}(k) = \mathbf{W}(z)\mathbf{H}(z)\mathbf{s}(k) = \mathbf{P}\Lambda(z)\mathbf{s}(k), \quad (6)$$

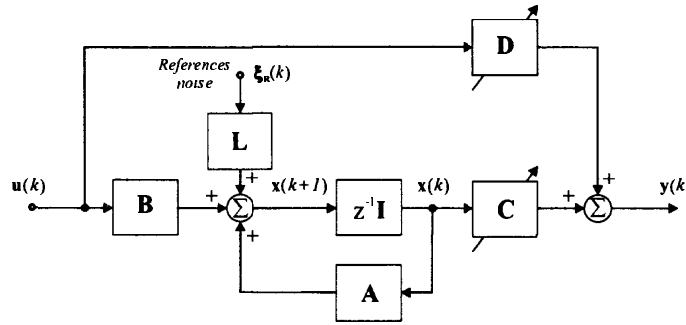


Figure 1: General state-space model for blind deconvolution

where \mathbf{P} is any permutation matrix and $\Lambda(z)$ is a diagonal matrix with $\lambda_i z^{-\tau_i}$ in diagonal entry (i,i), here λ_i is a nonzero constant and τ_i is any nonnegative integer. It is easy to see that the linear state space model mixture is an extension of instantaneous mixture. When both the matrices $\overline{\mathbf{A}}, \overline{\mathbf{B}}, \overline{\mathbf{C}}$ in the mixing model and $\mathbf{A}, \mathbf{B}, \mathbf{C}$ in the demixing model are null matrices, the problem is simplified to standard ICA problem [1-8].

The question here is whether exist matrices $[\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}]$ in the demixing model (4) and (5), such that its transfer function $\mathbf{W}(z)$ satisfies (6). It is proven [12] that if the matrix $\overline{\mathbf{D}}$ in the mixing model is of full rank, $\text{rank}(\overline{\mathbf{D}}) = n$, then there exist matrices $[\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}]$, such that the output signal \mathbf{y} of state-space system (4) and (5) recovers the independent source signal \mathbf{s} in the sense of (6).

3 Learning Algorithm

Assume that $p(\mathbf{y}, \mathbf{W}), p_i(y_i, \mathbf{W})$ are the joint probability density function of \mathbf{y} and marginal pdf of y_i , ($i = 1, \dots, n$) respectively. We employ the mutual information of the output signals, which measures the mutual independence of the output signals $y_i(k)$, as a risk function [1,2]

$$l(\mathbf{W}) = -H(\mathbf{y}, \mathbf{W}) + \sum_{i=1}^n H(y_i, \mathbf{W}), \quad (7)$$

where

$$H(\mathbf{y}, \mathbf{W}) = - \int p(\mathbf{y}, \mathbf{W}) \log p(\mathbf{y}, \mathbf{W}) d\mathbf{y}, \quad H(y_i, \mathbf{W}) = - \int p_i(y_i, \mathbf{W}) \log p_i(y_i, \mathbf{W}) dy_i.$$

In this paper we do not directly develop learning algorithms to update all parameters $\mathbf{W} = [\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}]$ in demixing model. We separate the blind deconvolution problem into two procedures: separation and state-estimation. In the separation procedure we develop a novel learning algorithm, using a new search direction, to update the matrices \mathbf{C} and \mathbf{D} in output equation (5). Then we define a hidden innovation of the output and use Kalman filter to estimate the state vector $\mathbf{x}(k)$.

For simplicity we suppose that the matrix \mathbf{D} in the demixing model (5) is nonsingular $n \times n$ matrix. From the risk function (7), we can obtain a cost function for on line learning

$$l(\mathbf{y}, \mathbf{W}) = -\frac{1}{2} \log \det(\mathbf{D}^T \mathbf{D}) - \sum_{i=1}^n \log p_i(y_i, \mathbf{W}), \quad (8)$$

where $\det(\mathbf{D}^T \mathbf{D})$ is the determinant of symmetric positive definite matrix $\mathbf{D}^T \mathbf{D}$. For the gradient of l with respect to \mathbf{W} , we calculate the total differential dl of $l(\mathbf{y}, \mathbf{W})$ when we takes a differential $d\mathbf{W}$ on \mathbf{W}

$$dl(\mathbf{y}, \mathbf{W}) = l(\mathbf{y}, \mathbf{W} + d\mathbf{W}) - l(\mathbf{y}, \mathbf{W}). \quad (9)$$

Following Amari's derivation for natural gradient methods [1-3], we have

$$dl(\mathbf{y}, \mathbf{W}) = -tr(d\mathbf{D}\mathbf{D}^{-1}) + \boldsymbol{\varphi}^T(\mathbf{y})d\mathbf{y}, \quad (10)$$

where tr is the trace of a matrix and $\boldsymbol{\varphi}(\mathbf{y})$ is a vector of nonlinear activation functions

$$\varphi_i(\mathbf{y}_i) = -\frac{d \log p_i(\mathbf{y}_i)}{d\mathbf{y}_i} = -\frac{p'_i(\mathbf{y}_i)}{p_i(\mathbf{y}_i)}. \quad (11)$$

Taking the derivative on equation (5), we have following approximation

$$d\mathbf{y} = d\mathbf{C}\mathbf{x}(k) + d\mathbf{D}\mathbf{u}(k). \quad (12)$$

On the other hand, from (5), we have

$$\mathbf{u}(k) = \mathbf{D}^{-1}(\mathbf{y}(k) - \mathbf{C}\mathbf{x}(k)) \quad (13)$$

Substituting (13) into (12), we obtain

$$d\mathbf{y} = (d\mathbf{C} - d\mathbf{D}\mathbf{D}^{-1}\mathbf{C})\mathbf{x} + d\mathbf{D}\mathbf{D}^{-1}\mathbf{y}. \quad (14)$$

In order to improve the computing efficiency of learning algorithms, we introduce a new search direction

$$d\mathbf{X}_1 = d\mathbf{C} - d\mathbf{D}\mathbf{D}^{-1}\mathbf{C}, \quad (15)$$

$$d\mathbf{X}_2 = d\mathbf{D}\mathbf{D}^{-1}. \quad (16)$$

Then the total differential dl can be expressed by

$$dl = -tr(d\mathbf{X}_2) + \boldsymbol{\varphi}^T(\mathbf{y})(d\mathbf{X}_1\mathbf{x} + d\mathbf{X}_2\mathbf{y}). \quad (17)$$

It is easy to obtain the derivatives of the cost function l with respect to matrices \mathbf{X}_1 and \mathbf{X}_2 as

$$\frac{\partial l}{\partial \mathbf{X}_1} = \boldsymbol{\varphi}(\mathbf{y}(k))\mathbf{x}^T(k), \quad (18)$$

$$\frac{\partial l}{\partial \mathbf{X}_2} = \boldsymbol{\varphi}(\mathbf{y}(k))\mathbf{y}^T(k) - \mathbf{I}. \quad (19)$$

From (15) and (16), we derive a novel learning algorithm to update matrices \mathbf{C} and \mathbf{D} .

$$\Delta\mathbf{C}(k) = \eta(-\boldsymbol{\varphi}(\mathbf{y}(k))\mathbf{x}^T(k) + (\mathbf{I} - \boldsymbol{\varphi}(\mathbf{y}(k))\mathbf{y}^T(k))\mathbf{C}(k)), \quad (20)$$

$$\Delta\mathbf{D}(k) = \eta(\mathbf{I} - \boldsymbol{\varphi}(\mathbf{y}(k))\mathbf{y}^T(k))\mathbf{D}(k). \quad (21)$$

The equilibrium points of the learning algorithm satisfy the following equations

$$E[\boldsymbol{\varphi}(\mathbf{y}(k))\mathbf{x}^T(k)] = 0, \quad (22)$$

$$E[\mathbf{I} - \boldsymbol{\varphi}(\mathbf{y}(k))\mathbf{y}^T(k)] = 0. \quad (23)$$

This means that separated signals \mathbf{y} could achieve as mutually independent as possible if the nonlinear activation function $\boldsymbol{\varphi}(\mathbf{y})$ are suitably chosen and the state vector $\mathbf{x}(k)$ is well estimated. From (20) and (21), we see that the natural gradient learning algorithm [2] is covered as a special case of the learning algorithm when the mixture is simplified to instantaneous case.

The above derived learning algorithm enable to solve the blind separation problem under assumption that state matrices \mathbf{A} and \mathbf{B} are known or designed appropriately. In the next section instead of adjusting state matrices \mathbf{A} and \mathbf{B} directly, we propose new approaches how to estimate state vector \mathbf{x} .

4 State Estimator

From output equation (5), it is observed that if we can accurately estimate the state vector $\mathbf{x}(k)$ of the system, then we can separate mixed signals using the learning algorithm (20) and (21).

4.1 Kalman Filter

The Kalman filter is a useful technique to estimate the state vector in state-space models. The function of the Kalman Filter is to generate on line the state estimate of the state $\mathbf{x}(k)$. The Kalman filter dynamics are given as follows

$$\mathbf{x}(k+1) = \mathbf{A}\mathbf{x}(k) + \mathbf{B}\mathbf{u}(k) + \mathbf{K}\mathbf{r}(k) + \boldsymbol{\xi}_R(k), \quad (24)$$

where \mathbf{K} is the Kalman filter gain matrix, and $\mathbf{r}(k)$ is the innovation or residual vector which measures the error between the measured (or expected) output $\mathbf{y}(k)$ and the predicted output $\mathbf{C}\mathbf{x}(k) + \mathbf{D}\mathbf{u}(k)$. There are varieties of algorithms to update the Kalman filter gain matrix \mathbf{K} as well as the state $\mathbf{x}(k)$, refer to [10] for more details.

However in the blind deconvolution problem there exists no explicit residual $\mathbf{r}(k)$ to estimate the state vector $\mathbf{x}(k)$ because the expected output $\mathbf{y}(t)$ means the unavailable source signals. In order to solve this problem, we present a new concept called hidden innovation to implement the Kalman filter in blind deconvolution case. Since updating matrices \mathbf{C} and \mathbf{D} will produces an innovation in each learning step, we introduce a hidden innovation as follows

$$\mathbf{r}(k) = \Delta\mathbf{y}(k) = \Delta\mathbf{C}\mathbf{x}(k) + \Delta\mathbf{D}\mathbf{u}(k), \quad (25)$$

where $\Delta\mathbf{C} = \mathbf{C}(k+1) - \mathbf{C}(k)$ and $\Delta\mathbf{D} = \mathbf{D}(k+1) - \mathbf{D}(k)$. The hidden innovation presents the adjusting direction of the output of the demixing system and is used to generate an a posteriori state estimate. Once we define the hidden innovation, we can employ the commonly used Kalman filter to estimate the state vector $\mathbf{x}(k)$, as well as to update Kalman gain matrix \mathbf{K} . The updating rule in this paper is described as follows:

- (1) Compute the Kalman gain matrix

$$\mathbf{K}(k) = \mathbf{P}(k)\mathbf{C}(k)^T(\mathbf{C}(k)\mathbf{P}(k)\mathbf{C}(k)^T + \mathbf{R}(k))^{-1}$$

- (2) Update state vector with hidden innovation

$$\hat{\mathbf{x}}(k) = \mathbf{x}(k) + \mathbf{K}(k)\mathbf{r}(k)$$

- (3) Update the error covariance matrix

$$\hat{\mathbf{P}}(k) = (\mathbf{I} - \mathbf{K}(k)\mathbf{C}(k))\mathbf{P}(k)$$

- (4) evaluate the state vector ahead

$$\mathbf{x}_{k+1} = \mathbf{A}(k)\hat{\mathbf{x}}(k) + \mathbf{B}(k)\mathbf{u}(k)$$

- (5) evaluate the error covariance matrix ahead

$$\mathbf{P}(k) = \mathbf{A}(k)\hat{\mathbf{P}}(k)\mathbf{A}(k)^T + \mathbf{Q}(k)$$

with the initial condition $\mathbf{P}(0) = \mathbf{I}$, where $\mathbf{Q}(k)$, $\mathbf{R}(k)$ are the covariance matrices of the noise vector $\boldsymbol{\xi}_R$ and output measurement noise \mathbf{n}_k .

The theoretic problems, such as convergence and stability, remain to be elaborated. Simulation experiments show that the proposed algorithm, based on the Kalman filter, can separate the convolved signals well.

4.2 Information Back-propagation

Another solution to estimating the state of a system is to propagate backward the mutual information. If we consider the cost function is also a function of the vector \mathbf{x} , then we have the partial derivative of $l(\mathbf{y}, \mathbf{W})$ with respect to \mathbf{x}

$$\frac{\partial l(\mathbf{y}, \mathbf{W})}{\partial \mathbf{x}} = \mathbf{C}^T \boldsymbol{\varphi}(\mathbf{y}). \quad (26)$$

Then we adjust the state vector $\mathbf{x}(k)$ according to the following rule

$$\hat{\mathbf{x}}(k) = \mathbf{x}(k) - \eta \mathbf{C}(k)^T \boldsymbol{\varphi}(\mathbf{y}(k)). \quad (27)$$

Then the estimated state vector is used as a new state of the system.

5 Numerical Implementation

Several numerical simulations have been done to demonstrate the validity and effectiveness of the proposed algorithm. Here we give a typical example

Example 1. Consider the following MIMO mixing model

$$\mathbf{u}(k) + \sum_{i=1}^{10} \mathbf{A}_i \mathbf{u}(k-i) = \mathbf{s}(k) + \sum_{i=1}^{10} \mathbf{B}_i \mathbf{s}(k-i) + \mathbf{v}(k),$$

where $\mathbf{u}, \mathbf{s}, \mathbf{v} \in \mathbf{R}^3$, and

$$\begin{aligned} \mathbf{A}_2 &= \begin{pmatrix} -0.48 & -0.16 & -0.64 \\ -0.16 & -0.48 & -0.24 \\ -0.16 & -0.16 & -0.08 \end{pmatrix}, & \mathbf{A}_8 &= \begin{pmatrix} -0.50 & -0.10 & -0.40 \\ -0.10 & -0.50 & -0.20 \\ -0.10 & -0.10 & -0.10 \end{pmatrix}, \\ \mathbf{A}_{10} &= \begin{pmatrix} 0.32 & 0.19 & 0.38 \\ 0.16 & 0.29 & 0.20 \\ 0.08 & 0.08 & 0.10 \end{pmatrix}, & \mathbf{B}_2 &= \begin{pmatrix} 0.42 & 0.21 & 0.14 \\ 0.10 & 0.56 & 0.14 \\ 0.21 & 0.21 & 0.35 \end{pmatrix}, \\ \mathbf{B}_8 &= \begin{pmatrix} -0.40 & -0.08 & -0.08 \\ -0.08 & -0.40 & -0.16 \\ -0.08 & -0.08 & -0.56 \end{pmatrix}, & \mathbf{B}_{10} &= \begin{pmatrix} -0.19 & -0.15 & -0.10 \\ -0.11 & -0.27 & -0.12 \\ -0.16 & -0.18 & -0.22 \end{pmatrix}, \end{aligned}$$

and other matrices are set to the null matrix. The sources \mathbf{s} are chosen to be i.i.d signals uniformly distributed in the range $(-1,1)$, and \mathbf{v} are the Gaussian noises with zero mean and a covariance matrix $0.1\mathbf{I}$. We employ the state space approach to separate mixing signals. The nonlinear activation function is chosen $\boldsymbol{\varphi}(\mathbf{y}) = \mathbf{y}^3$. The initial value for matrices \mathbf{A} and \mathbf{B} in the state equation are chosen as in canonical controller form. The initial values for matrix \mathbf{C} is set to null matrix or given randomly in the range $(-1,1)$, and $\mathbf{D} = \mathbf{I}_3$. A large number of simulations show that the state space method can easily recover source signals in the sense of $\mathbf{W}(z)\mathbf{H}(z) = \mathbf{P}\mathbf{A}$. Figure 2 illustrates the coefficients of global transfer function $\mathbf{G}(z) = \mathbf{W}(z)\mathbf{H}(z)$ after 3000 iterations, where the (i,j) th sub-figure plots the coefficients of the transfer function $G_{ij}(z) = \sum_{k=0}^{\infty} g_{ijk}z^{-k}$ up to order of 50.

References

- [1] S. Amari and A. Cichocki, "Adaptive blind signal processing—neural network approaches", *Proceedings of the IEEE*, 86(10):2026-2048, 1998.
- [2] S. Amari, A. Cichocki, and H.H. Yang, "A new learning algorithm for blind signal separation", *Advances in Neural Information Processing Systems 1995* (Boston, MA: MIT Press, 1996), pp. 752-763.

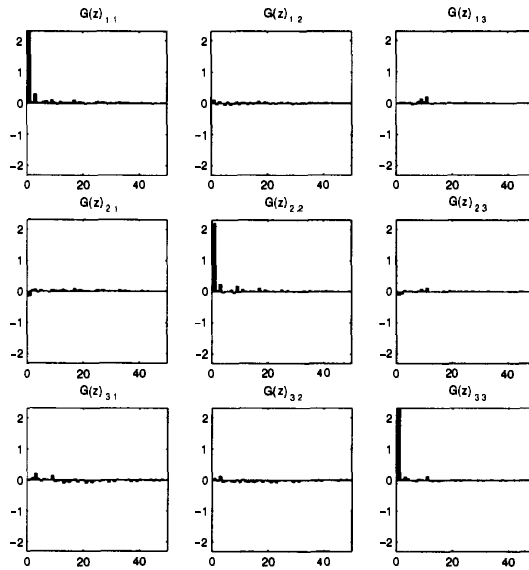


Figure 2: The coefficients of global transfer function after 3000 iterations

- [3] S. Amari “Natural gradient works efficiently in learning”, *Neural Computation*, Vol.10, pp251-276, 1998.
- [4] A. J. Bell and T. J. Sejnowski, “ An information-maximization approach to blind separation and blind deconvolution”, *Neural Computation*, Vol.7, pp 1129-1159, 1995.
- [5] J.-F. Cardoso, “Blind signal separation: statistical principles”, *Proceedings of the IEEE*, 86(10):2009-2025, 1998.
- [6] J.-F. Cardoso and B. Laheld, “Equivariant adaptive source separation,” *IEEE Trans. Signal Processing*, vol. SP-43, pp. 3017–3029, Dec. 1996.
- [7] A.Cichocki and R. Unbehauen, “Robust neural networks with on-line learning for blind identification and blind separation of sources” *IEEE Trans Circuits and Systems I : Fundamentals Theory and Applications*, vol 43, No.11, pp. 894-906, Nov. 1996.
- [8] P. Comon, “Independent component analysis: a new concept?”, *Signal Processing*, vol.36, pp.287–314, 1994.
- [9] A. Gharbi and F. Salam, “ Algorithm for blind signal separation and recovery in static and dynamics environments”, *IEEE Symposium on Circuits and Systems*, Hong Kong, June, 713-716, 1997.
- [10] O. L. R. Jacobs, “Introduction to Control Theory”, Second Edition, Oxford University Press, 1993.
- [11] T. W. Lee, A.J. Bell, and R. Lambert, “Blind separation of delayed and convolved sources”, *NIPS 9*, 1997, MIT Press, Cambridge MA, pp758-764.
- [12] L. -Q. Zhang and A. Cichocki, “Blind deconvolution/equalization using state-space models”, *Proc. '98 IEEE Signal Processing Society Workshop on NNSP*, pp123-131, Cambridge, 1998.
- [13] S. Choi, A. Cichocki and S. Amari, “Blind equalization of simo channels via spatio-temporal anti-Hebbian learning rule”, *Proc. '98 IEEE Signal Processing Society Workshop on NNSP*, pp93-102, Cambridge, 1998.

PART V
IMPLEMENTATION

