SEMI–BLIND SOURCE SEPARATION FOR CONVOLUTIVE MIXTURES BASED ON FREQUENCY INVARIANT TRANSFORMATION

Wei Liu and Danilo P. Mandic

Communications & Signal Processing Research Group
Department of Electrical & Electronic Engineering
Imperial College London, U.K.
Email:{wei.liu, d.mandic}@imperial.ac.uk

ABSTRACT

A novel method for separation of a class of convolutive mixtures is proposed, in which the received sensor signals are first transformed into instantaneous mixtures and then standard blind source separation (BSS) algorithms for instantaneous mixtures are applied. Since partial information about the mixing mechanism is required in the design of the transformation, the proposed method is strictly speaking semi-blind. From the beamforming viewpoint, the proposed approach represents a blind broadband beamforming method. As the separation is performed in fullband and only one separation is needed, the permutation problem associated with the frequency-domain BSS is avoided and the separation can be easily implemented online. Simulation results verify the usefulness of the proposed method.

1. INTRODUCTION

In the last decade or so, blind source separation (BSS) for instantaneously mixed sources has been studied extensively and a number of algorithms have been proposed for this purpose [1, 2]. However, for convolutive mixtures, the direct time-domain extension of BSS algorithms from instantaneous mixtures to the convolutive case is difficult and computationally very expensive [3, 4]. To circumvent this problem, it is convenient to transform the received sensor signals into the frequency domain, where many separating algorithms for instantaneous mixtures can be applied directly. This is possible since convolutive mixing in the time domain corresponds to the instantaneous one in the frequency domain [5]. However, frequency-domain BSS introduces the well-known permutation problem, and an online implementation of these algorithms is very difficult if not prohibitive for cases with more than three sources [6].

Following the idea of transforming the convolutive mixing into an instantaneous one, we here propose a novel transformation method for separation of a special class of convolutive mixtures. In traditional blind source separation, we always ignore the information about the mixing mechanism, even if it is available, such as in the cases of sonar, radar, and microphone arrays, where we know the sensor positions and often assume the impinging signals are plane waves. We therefore set out to exploit this information and proceed to design a frequency invariant beamforming (FIB) network [7, 8]. Broadband signals arrive at the adjacent sensors with specific delays, which conforms to the principle of convolutive mixing. When these signals progress through that beamforming network, convolutive mixing will be transformed into an instantaneous one, due to the frequency invariant property of the network. After that, according to statistical properties of the original sources, an appropriate blind source separation algorithm for instantaneous mixtures is applied to obtain the original sources. As we require partial information about the mixing mechanism, the method is strictly speaking not blind, but semi-blind. Due to the lack of knowledge about the directions of arrival (DOAs), we still do not have full information about the mixing. From the beamforming viewpoint, we can refer to our method as blind broadband beamforming, which is different from the previously studied blind beamforming scenarios [9], where only narrowband signals are considered which therefore represents just a instantaneous mixing problem. Since the separation is performed in fullband and only one separation is needed, the permutation problem associated with the frequency-domain BSS is avoided, which makes it possible for the algorithm to be implemented online, for an arbitrary number of sources.

This paper is organized as follows. The class of blind source separation problems considered here is given in Section 2 and the FIB technique is briefly reviewed in Section 3. We then propose a novel semi-blind separation method in Section 4. Simulation results and conclusions are given in Sections 5 and 6, respectively.

2. CONVOLUTIVE MIXING FOR BROADBAND ARRAYS

An array of sensors with plane-wave input is shown in Fig 1. Suppose $s_l(t)$, $l = 0, \ldots, L-1$ are the L impinging plane-wave signals that would be received at the origin of the coordinate system, then the signal received at the m-th sensor will be $x_m(t) = \sum_{l=0}^{L-1} s_l(t - \tau_{m,l})$, where $\tau_{m,l}$ is the delay from the m-th sensor to the origin of the coordinate system for the signal $s_l(t)$. For narrowband signals, this delay can be expressed as a complex number and $x_m(t)$ will be a weighted sum of the signals $s_l(t)$, $l = 0, \ldots, L-1$, which represents an instantaneous mixing problem. For broadband signals, this delay can be expressed as a convolution of $\delta(t - \tau_{m,l})$ and $s_l(t)$, which turns the problem into a convolutive mixing one. In a vectorial form, we have

$$\mathbf{x}(t) = \mathbf{A} * \mathbf{s}(t), \tag{1}$$

with

$$\begin{aligned} \mathbf{s}(t) &= [s_0(t)\ s_1(t)\ \cdots\ s_{L-1}(t)]^{\mathrm{T}} \\ \mathbf{x}(t) &= [x_0(t)\ x_1(t)\ \cdots\ x_{M-1}(t)]^{\mathrm{T}} \\ [\mathbf{A}]_{m,l} &= a_{m,l} = \delta(t - \tau_{m,l})\ . \end{aligned} \tag{2}$$

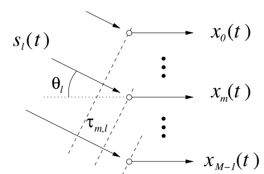

Fig. 1: A general array with a plane wave $s_l(t)$ impinging from an angle θ_l.

The discrete-time form of expression (1) is given by

$$\begin{aligned} \mathbf{s}[n] &= [s_0[n]\ s_1[n]\ \cdots\ s_{L-1}[n]]^{\mathrm{T}} \\ \mathbf{x}[n] &= [x_0[n]\ x_1[n]\ \cdots\ x_{M-1}[n]]^{\mathrm{T}} \ , \end{aligned} \tag{3}$$

and consequently the entries $a_{m,l}$ of $\mathbf{A}$ become a series of filters in the form of sinc functions. The filter $h_{m,l}[n]$ at the position $a_{m,l}$ is given by

$$h_{m,l}[n] = \frac{\sin((n - \frac{\tau_{m,l}}{T})\pi)}{(n - \frac{\tau_{m,l}}{T})\pi} \ , \tag{4}$$

where T is the sampling period.

To recover the original signals $s_l[n]$, $l = 0, \ldots, L-1$, we can apply blind source separation algorithms for convolutive mixtures directly [3, 4] or alternatively use frequency-domain methods [5], which are either computationally very expensive or suffer from other serious problems such as permutation ambiguity. As there are many cases where we know the positions of the sensors and the propagation properties of the incoming signals, we exploit this information and design a frequency invariant network to transform the convolutive mixing of (1) into instantaneous mixing directly, thereafter a BSS algorithm for instantaneous mixing can be applied. In the next section, we will give a brief review of the related FIB technique proposed in [7, 8].

3. FREQUENCY INVARIANT BEAMFORMING (FIB)

Frequency invariant beamforming is a technique for broadband array design to form a response only as a function of the direction of arrival of the impinging signals, independent of the signal frequency. We will use the beamforming method proposed in [8], which exploits the Fourier transform relationship between the array's spatio-temporal distribution and its beam pattern, and can be easily applied to 1-D, 2-D and 3-D broadband arrays with either discrete or continuous aperture. A previously proposed frequency invariant linear array [7] can be regarded as a special case of this new class of broadband arrays. For simplicity and without loss of generality, we here focus only on the uniformly spaced linear arrays, but our semi-blind separation method can be extend to any arrays with frequency invariant beam patterns described in [8] and elsewhere such in [10, 11].

For a uniformly spaced linear array with element spacing of d_x and signal sampling period T, as shown in Fig. 2, its output $y[n]$ is given by

$$y[n] = \sum_{m=0}^{M-1} \sum_{k=0}^{J-1} w_{m,k} \cdot x_m[n-k] \ , \tag{5}$$

and its beam pattern $P(\omega, \theta)$ is given by

$$P(\omega, \theta) = \sum_{m=0}^{M-1} \sum_{k=0}^{J-1} w_{m,k} \cdot e^{-jm\omega\Delta\tau} \cdot e^{-jk\omega T} \ , \tag{6}$$

where c is the wave propagation speed and $\Delta\tau = \frac{d_x \sin\theta}{c}$. With the normalized angular frequency $\Omega = \omega T$, (6) can be rewritten as

$$P(\Omega, \theta) = \sum_{m=0}^{M-1} \sum_{k=0}^{J-1} w_{m,k} \cdot e^{-jm\mu\Omega\sin\theta} \cdot e^{-jk\Omega} \text{ with } \mu = \frac{d_x}{cT} \ . \tag{7}$$

Substituting $\Omega_1 = \mu\Omega\sin\theta$ and $\Omega_2 = \Omega$ into (7) yields

$$P(\Omega_1, \Omega_2) = \sum_{m=0}^{M-1} \sum_{k=0}^{J-1} w_{m,k} \cdot e^{-jk\Omega_1} \cdot e^{-jm\Omega_2} \ . \tag{8}$$

As the spatio-temporal spectrum of the impinging signal lies on the line defined by $\Omega_1 = \mu\Omega_2 \sin\theta$, we can replace $\sin\theta$ by $\frac{\Omega_2}{\mu\Omega_1}$ in the desired frequency invariant beam pattern $P(\sin\theta)$, and apply a 2-D inverse Fourier transform to the resultant $P(\Omega_1, \Omega_2)$ to obtain the desired coefficients $w_{m,k}$, $m = 0, \ldots, M-1$ and $k = 0, \ldots, J-1$. For more details, please refer to [8].

As, in general, the sampling period is half that of the signal component with highest frequency and array spacing is half the wavelength of the highest signal frequency, we have $d_x = \frac{1}{2} \cdot c \cdot (2T) = cT$ and $\mu = 1$. Therefore, without loss of generality, we will only consider the case with $\mu = 1$ in the design and simulations. Fig. 3 shows a design result with $M = 18$ sensors and a tapped-delay line length of $J = 24$. The approximate frequency invariance property is clearly visible over the band $\Omega \in [0.30\pi; \pi]$.

4. SEMI–BLIND SOURCE SEPARATION WITH FREQUENCY INVARIANT TRANSFORMATION

Suppose we have obtained N sets of array coefficients $\mathbf{W}_i$, $i = 0, \ldots, N-1$, with

$$\mathbf{W}_i = \begin{bmatrix} w_{i,0,0} & w_{i,0,1} & \cdots & w_{i,0,J-1} \\ w_{i,1,0} & w_{i,1,1} & \cdots & w_{i,1,J-1} \\ \vdots & \vdots & \ddots & \vdots \\ w_{i,M-1,0} & w_{i,M-1,1} & \cdots & w_{i,M-1,J-1} \end{bmatrix} \ , \tag{9}$$

as shown in Fig. 4. Each set of coefficients $\mathbf{W}_i$ plays the same role as those in Fig. 2 and forms a frequency invariant

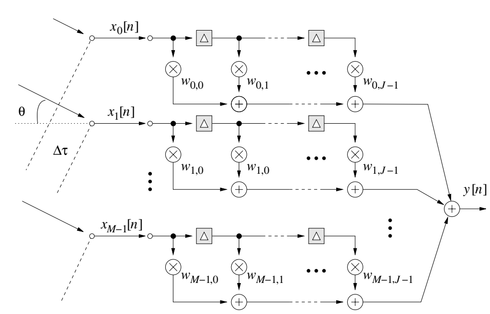

Fig. 2: A signal impinges from an angle θ onto a uniformly spaced broadband linear array with M sensors, each followed by a J-tap filter.

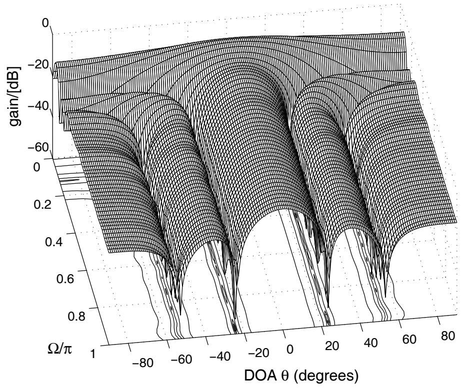

Fig. 3: A design example with 18 × 24 coefficients for an equispaced linear array with a broadside mainbeam.

response $P_i(\theta)$ with an output $y_i[n]$, which is given by

$$y_i[n] = \sum_{m=0}^{M-1} \sum_{k=0}^{J-1} w_{i,m,k} \cdot x_m[n-k] \; . \qquad (10)$$

The temporal counterpart of $P_i(\theta)$ is $P_i(\theta)\delta(n)$, where n is the time index. Note that there is no delay in these responses which means that the whole beamforming network is noncausal. To obtain a causal system, we can simply shift the system by some delay Δn and all the response will change to $P_i(\theta)\delta(n - \Delta n)$. As we can set the same delay for the whole set of N beamforming sub-systems, for simplicity, in the subsequent analysis we will ignore this delay part. Suppose DOA angles of the L sources are respectively $\theta_0, \theta_1, \ldots, \theta_{L-1}$. Then the outputs of the N frequency invariant beamformers can be expressed as

$$\mathbf{y}[n] = \mathbf{B} \cdot \mathbf{s}[n], \qquad (11)$$

with

$$\begin{aligned} \mathbf{y}[n] &= [y_0[n] \; y_1[n] \; \cdots \; y_{N-1}[n]]^{\mathrm{T}} \\ [\mathbf{B}]_{i,l} &= P_i(\theta_l) \; . \end{aligned} \qquad (12)$$

We see that, in this way, the convolutive mixing is transformed into an instantaneous mixing problem and $y_i[n]$, $i = 0, 1, \ldots, N-1$ are the new instantaneous mixtures. To solve this problem, depending on the nature of the sources, we can employ the corresponding instantaneous BSS algorithms such as those employing second-order and higher-order statistics [12, 2]. In Section 5, we will employ a density matching BSS algorithm using natural gradient adaptation [1] with its update given by

$$\mathbf{D}[n+1] = \mathbf{D}[n] + \mu \left[\mathbf{I} - \mathbf{f}(\hat{\mathbf{s}}[n])\hat{\mathbf{s}}^T[n] \right] \mathbf{D}[n] \; , \qquad (13)$$

with

$$\hat{\mathbf{s}}[n] = [\hat{s}_0[n], \ldots, \hat{s}_{N-1}[n]]^T = \mathbf{D}[n]\mathbf{y}[n] \; , \qquad (14)$$

and

$$\mathbf{f}(\hat{\mathbf{s}}[n]) = \left[\hat{s}_0^3[n], \ldots, \hat{s}_{N-1}^3[n] \right]^T \; , \qquad (15)$$

where $\mathbf{D}$ is the separation matrix and $\hat{\mathbf{s}}[n]$ will be the separated sources.

A network with N sets of frequency invariant beamformers is shown in Fig. 4. As well known, in theory, we can only successfully separate at most M sources with M

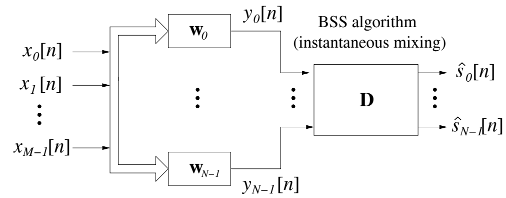

Fig. 4: A frequency invariant beamforming network for instantaneous BSS algorithm.

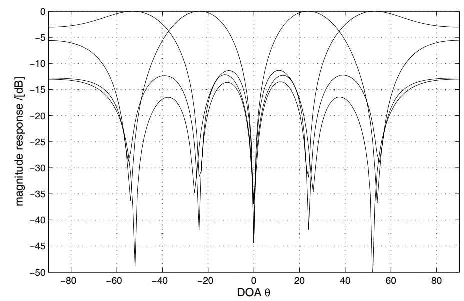

Fig. 5: The beam shapes of the four frequency invariant beamformers.

sensors, thus we can only design $N <= M$ independent transformations, i.e. frequency invariant beamformers. In our structure, the N frequency invariant beamformers will have their main beam directions equally distributed over the DOA range $[-90^\circ; 90^\circ)$.

The obvious advantage of this BSS approach is that it transforms the convolutive mixing problem into an instantaneous mixing problem and greatly simplifies the separation. Since there is no decomposition of the source signals into any domain, we will not have the permutation problem of the frequency-domain method and the estimated source signals can be recovered directly by the separation matrix $\mathbf{D}$. However, there are also some disadvantages. The key issue in this new structure is to design a FIB network, which will not be possible if we only have a few sensors. To have a good frequency invariance property, we may need a number of sensors (for example, in Fig. 3, we have 18 sensors). Therefore, if there are only several sources, we will need more sensors to separate them than when using the ordinary time-domain convolutive BSS algorithms or the frequency-domain algorithms. When there are many sources to separate, such as 16, our method will exhibit great superiority, while the other methods may become very complicated and even fail.

5. SIMULATIONS

In our simulations, we have four sources coming respectively from the DOA angles of $-60^\circ, -20^\circ, 20^\circ$ and 60°. Note that the DOA information is not available to the FIB network and the separation algorithm. We have $M = 18$ sensors to receive those sources. As we have only four sources, there is no need to design 18 frequency invariant beamformers to transform them into instantaneous mixtures. For simplicity, we design only four frequency invariant beamformers with their beam shapes shown in Fig. 5. The four sets of coefficients $\mathbf{W}_i$ have a dimension of 18×24, i.e. $J = 24$. The outputs $y_0[n], y_1[n], y_2[n], y_3[n]$

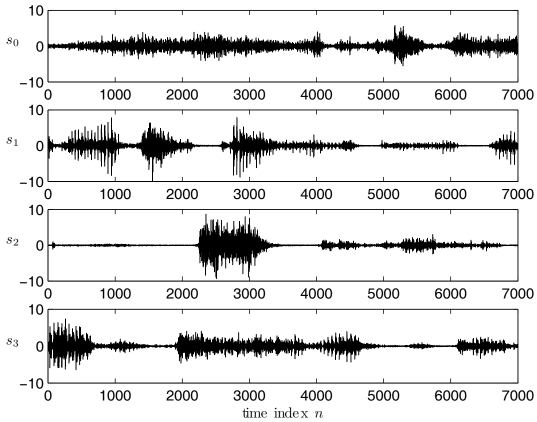

Fig. 6: The four source signals.

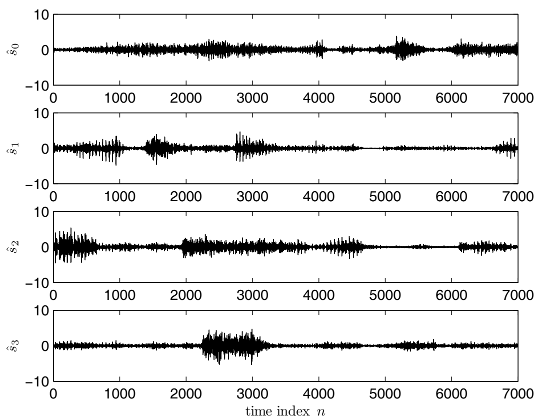

Fig. 7: The four separated source signals.

of this network are the new instantaneously mixed signals. We perform separation by the algorithm shown in (13). The chosen stepsize was $\mu = 0.0001$. After the steady state of the adaptation has been reached, we use the resultant separation matrix $\mathbf{D}$ to calculate the separated signals.

The original soure signals are plotted in Fig. 6. As the frequency invariant property of the four beamformers is not good enough in both the lower frequency band and the band around π, we have filtered the source signals, which limits their bandwidth to $[0.3\pi; 0.95\pi]$. The four separated signals are shown in Fig. 7. We see that although the order of the separated signals is different from the order of the sources, there is a clear match between the original and separated signals.

6. CONCLUSIONS

We have proposed a new method for separation of convolutive mixtures, where first an FIB network, which transforms the convolutive mixing into instantaneous mixing, is designed and then standard BSS algorithms for instantaneous mixing are applied. As we exploit the knowledge of the sensor positions and the plane wave assumption of the sources in the design of the FIB network, the proposed approach is not totally blind as the traditional BSS algorithm. Compared with the frequency-domain method, as there is no decomposition of the signals, we have avoided the difficult permutation problem of the classical frequency-domain method. Simulation results show that this new method can successfully separate the convolutively mixed signals.

7. REFERENCES

[1] S. C. Douglas, "Blind signal separation and blind deconvolution," in *Handbook of Neural Network Signal Processing*, Y. H. Hu and J. N. Hwang, Eds., chapter 7. CRC Press, Boca Raton, Florida, 2002.

[2] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing*, John Wiley & Sons, Inc., New York, 2003.

[3] K. Matsuoka and S. Nakashima, "Minimal distortion principle for blind source separation," in *Proc. of The Fourth International Symposium on Independent Component Analysis and Blind Signal Separation*, December 2001, pp. 722–727.

[4] S. C. Douglas and X. Sun, "Convolutive blind separation of speech mixtures using the natural gradient," *Speech Communication*, vol. 39, no. 1, pp. 65–78, January 2003.

[5] P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," *Nerocomputing*, vol. 22, pp. 21–34, 1998.

[6] H. Sawada, R. Mukai, S. Araki, and S. Makino, "Convolutive blind source separation for more than two sources in the frequency domain," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Montreal, CA, May 2004, vol. III, pp. 885–888.

[7] T. Sekiguchi and Y. Karasawa, "Wideband Beamspace Adaptive Array Utilizing FIR Fan Filters for Multibeam Forming," *IEEE Transactions on Signal Processing*, vol. 48, no. 1, pp. 277–284, January 2000.

[8] W. Liu and S. Weiss, "A new class of broadband arrays with frequency invariant beam patterns," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Montreal, Canada, May 2004, vol. 2, pp. 185–188.

[9] J. Sheinvald, "On blind beamforming for multiple nongaussian signals and the constant-modulus algorithm," *IEEE Transactions on Signal Processing*, vol. 46, pp. 1878–1885, July 1998.

[10] T. Chou, "Frequency-independent beamformer with low response error," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Detroit, USA, May 1995, vol. 5, pp. 2995–2998.

[11] D. B. Ward, R. A. Kennedy, and R. C. Williamson, "Theory and design of broadband sensor arrays with frequency invariant far-field beam patterns," *Journal of the Acoustic Society of America*, vol. 97, no. 2, pp. 1023–1034, February 1995.

[12] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley & Sons, Inc., New York, 2001.