Evaluation of blind source separation methods in acoustics

Roger Boustany, Jérôme Antoni

Laboratoire Roberval, CNRS, Université de Technologie de Compiègne , 60205 Compiègne, France, Email: boustany@utc.fr

Introduction

Blind source separation (BSS) is a signal processing technique which consists in recovering a set of unobserved signals (called *sources*) given only a set of measured signals (called *observations*) arising when the sources are mixed by passage through some unknown medium. Although the term blind indicates the fact that neither the sources nor the mixing structure are known *a priori*, several assumptions must be made regarding both. Most of the existing methods rest on the mutual statistical independence of the sources and on the linearity of the mixing channel. Moreover, depending on the field of application, the mixture model may be either *instantaneous* or *convolutive*.

BSS Problem Models

The BSS problem consists in retrieving m unknown sources, denoted by an $(m \times 1)$ source vector $\mathbf{s}(t) = (s_1(t), s_2(t), \cdots, s_m(t))^t$, from n recorded mixture signals denoted by an $(n \times 1)$ observation vector $\mathbf{x}(t) = (x_1(t), x_2(t), \cdots, x_n(t))^t$. In this study, we will suppose that $n \geq m$.

Assuming a linear mixing channel, the noiseless mixture can be modeled as

$$x_j(t) = \sum_{i=1}^{m} h_{ji}(t) * s_i(t), \qquad 1 \leq j \leq n \tag{1}$$

or in a matrix form

$$\mathbf{x}(t) = \mathbf{h}(t) * \mathbf{s}(t) \tag{2}$$

where $*$ is the convolution product and $\mathbf{h}$ is a matrix whose elements $h_{ji}(t)$ are finite impulse response (FIR) filters representing the effect of the passage of the source i through the medium to sensor j. The mixture is then said to be *convolutive* (or *dynamic*). It is the case of most of the problems encountered in acoustics and in vibration analysis.

A particular case is when recordings are done in an anechoic environment. The model reduces to the delayed mixture model

$$x_j(t) = \sum_{i=1}^{m} a_{ji} s_i(t - \tau_{ji}), \qquad 1 \leq j \leq n \tag{3}$$

where τ_{ji} is the time delay from source j to sensor i.

In other applications, the mixing channel has no memory effect. The filter impulse response reduces then to a scalar. The mixture is said to be *instantaneous* (or *static*) and can be modeled by

$$x_j(t) = \sum_{i=1}^{m} a_{ji} s_i(t), \qquad 1 \leq j \leq n \tag{4}$$

BSS Methods and Indeterminacies

Separation of instantaneous mixtures

In order to separate instantaneous mixtures, most of the approaches exploit the property of statistical independence of the sources as in [1]. These approaches suffer from two main indeterminacies. The sources are recovered neither in order nor in their actual magnitudes. Moreover, the instantaneous model (4) does not reflect a real acoustic environment.

Separation of convolutive mixtures

As a matter of fact, signal mixtures in acoustics are either convolutive or delayed. Some of the methods for separating convolutive use the time representation of the mixture and have the disadvantage of being computationally expensive. Another approach is to transform the time domain convolutive BSS problem into multiple independent frequency domain (FD) BSS problems through Fourier transform

$$X(f) = A(f)S(f), \qquad 1 \leq f \leq N \tag{5}$$

where f is the frequency bin and N is the length of the DFT. This approach presents the advantage of a simple computational and mathematical analysis. However, most of the FD methods suffer from the permutation problem when reconstructing the spectrum of each source. Moreover, the convolutive BSS main indeterminacy is that sources can only be found up to a filter as explained in [2].

Evaluation of the Methods

In order to evaluate the performance of the algorithms, we will use measures of distortion and separation quality as proposed in [4].

Measure of distortion

The distortion of the j^{th} separated output is given by

$$D_j = 10 \log_{10} \left(\frac{E\{(x_{j,s_j} - \alpha_j y_j)^2\}}{E\{(x_{j,s_j})^2\}} \right) \tag{6}$$

where x_{j,s_j} is the contribution of the j^{th} source to the j^{th} sensor, y_j corresponds to the j^{th} recovered source, $E\{.\}$ is the expectation operator and $\alpha_j = E\{(x_{j,s_j})^2\}/E\{(y_j)^2\}$ denotes a scaling factor used to take into consideration the scaling of the retrieved sources. All though this distortion estimate is valid only for static mixtures, it will also be applied to convolutive mixtures in the experiments section.

Quality of separation

The quality of separation of the j^{th} output can be measured by

$$Q_j = 10\log_{10}\left(\frac{E\{(y_{j,s_j})^2\}}{E\{(\sum_{i\neq j} y_{j,s_i})^2\}}\right) \tag{7}$$

with y_{j,s_i} is the j^{th} output of the cascaded mixing/unmixing channel when only s_i is active.

Experiments and Results

In this section, Echo Cancelling and Blind Source Separation (ECoBliSS) [3], Blind Source Separation in the Time-Frequency Domain (BSSTFD) [5], Lambert's FIR Matrix Algebra BSS (FIRMABSS) [6] and Frequency Domain Blind Identification (FDBI) [7] algorithms are tested (Matlab codes are available in [8]). The task is to recover two speech signals of an english-speaking man and a french-speaking woman from two microphone recordings with a sampling frequency of 16 kHz. First, synthetic noiseless mixtures are generated using respectively a scalar and a 4-tap-filter mixing channels in order to test the separation capability of the algorithms. Tables 1 and 2 show a good performance of convolutive BSS algorithms. The results in the static case can be compared to those of JADE algorithm [1].

Method	D_j (dB)		Q_j (dB)	
	y_1	y_2	y_1	y_2
JADE	-1.63	1.99	55.28	53.05
ECoBliSS	0.17	8.40	32.77	21.52
BSSTFD	-9.31	-12.31	13.97	19.41
FIRMABSS	14.07	13.99	23.09	20.33
FDBI	-2.57	-3.21	41.29	39.42

Table 1: Evaluation of the separation of synthetic instantaneous mixtures.

Method	D_j (dB)		Q_j (dB)	
	y_1	y_2	y_1	y_2
ECoBliSS	-0.10	-1.88	16.12	11.48
BSSTFD	-5.81	-9.82	15.15	12.52
FIRMABSS	5.90	9.80	4.85	5.64
FDBI	3.95	6.81	7.34	5.25

Table 2: Evaluation of the separation of synthetic convolutive mixtures.

Experiments are then carried out on real-room audio signals. The two previously cited persons simultaneously read sentences in an anechoic and a reverberant rooms with background noise. Table 3 shows a low performance of all the methods as considerable crosstalk remains in the recovered sources. The task of separating mixtures in a reverberant room turns to be much more difficult as the quality indicator tends to 0 dB for all of the tested methods (table 4).

Method	D_j (dB)		Q_j (dB)	
	y_1	y_2	y_1	y_2
ECoBliSS	-0.00	-0.03	2.66	2.34
BSSTFD	2.23	3.43	2.63	6.76
FIRMABSS	8.59	7.00	3.80	3.88
FDBI	6.51	5.01	7.38	3.88

Table 3: Evaluation of the separation of audio mixed in an anechoic room.

Method	D_j (dB)		Q_j (dB)	
	y_1	y_2	y_1	y_2
ECoBliSS	0.20	0.01	0.89	0.53
BSSTFD	-1.17	-1.34	0.44	0.86
FIRMABSS	8.33	7.07	0.54	0.85
FDBI	3.66	3.32	0.08	0.07

Table 4: Evaluation of the separation of audio mixed in a reverberant room.

Conclusions

In this paper, audio blind source separation was discussed. Evaluation criteria proved that well-performing algorithms failed to give satisfactory separation of speech signals recorded in difficult environments. Future works will include improving the results and applying BSS to vibroacoustical signals like rotating machines'.

References

[1] J-F. Cardoso and A. Souloumiac. Blind beamforming for non gaussian signals. IEE Proceedings-F, **140(6)** (1993), 362-370

[2] E. Weinstein, M. Feder and A.V. Oppenheim. Multichannel signal separation by decorrelation. IEEE Trans. on Speech and Audio Processing **1(4)** (1993), 405-413

[3] D.W.E. Schobben and P.C.W. Sommen. A frequency domain blind signal separation method based on decorrelation. IEEE Trans. on Signal Processing **50(8)** (2002), 1855-1865

[4] D.W.E. Schobben, K. Torkkola and P. Smaragdis. Evaluation of blind signal separation methods. Proc. ICA'99, Aussois, France (1999), 889-892

[5] N. Murata, S. Ikeda and A. Ziehe. An approach to blind source separation based on temporal structure of speech signals. Neurocomputing **41** (2001), 1-24

[6] R.H. Lambert. Multichannel blind deconvolution: FIR matrix algebra and separation of multipath mixtures. PhD thesis, Univ. of Southern California (1996)

[7] K. Diamantaras, A. Petropulu and B. Chen. Blind two-input-two-output FIR channel identification based on frequency domain second-order statistics. IEEE Trans. on Signal Processing **48(2)** (2000), 534-542

[8] Independent Component Analysis forum. URL: `http://www.tsi.enst.fr/icacentral`