ALTERNATIVE STRUCTURES AND POWER SPECTRUM CRITERIA FOR BLIND SEGMENTATION AND SEPARATION OF CONVOLUTIVE SPEECH MIXTURES

Benoit Albouy, Yannick Deville

Laboratoire d'Acoustique, de Métrologie et d'Instrumentation,
Université Toulouse III
Bât. 3R1B2, 118 Route de Narbonne, 31062 Toulouse Cedex, France
albouy@cict.fr, ydeville@cict.fr

ABSTRACT

This paper deals with the blind separation of convolutively mixed speech sources. The proposed methods take advantage of the a priori knowledge that speech signals contain silences. They consist in first detecting silence phases in these source signals and then identifying each filter of the considered separating systems in such a phase. The criteria used in both stages of these approaches are based on the power (cross)-spectra of the observations: their time-segmented coherence function is first used to detect silence phases and the filters to be identified are then expressed as the ratios of observation power (cross)-spectra. This general approach is applied to various separating systems, depending i) whether the considered structures are symmetrical, asymmetrical, or asymmetrical with a complementary part, and ii) whether they include or not a post-processing stage for filtering the extracted sources. The performance of all these approaches and of two methods from the literature is investigated by means of experimental tests performed with speech sources mixed by means of real acoustical in-car transfer functions. This shows that the proposed approaches yield an interesting performance/complexity trade-off as compared to previously reported methods.

1. INTRODUCTION

All situations related to the processing of data resulting from the reception of several source signals by an array of sensors lead to a difficult analysis [2, 9]. Blind Source Separation (BSS) methods allow to treat the observed sensor signals resulting from the mixing of source signals, so as to estimate the sources from these mixed signals. In recent years, BSS became one of the exciting new topics in advanced statistics and signal processing, and it applies to major fields such as audio, seismology and even within the medical framework.

Two kinds of mixtures exist : i) linear instantaneous mixtures, which apply to narrow-band telecommunication problems for example, and ii) convolutive mixtures, which are the most realistic model e.g. for acoustical signals, due to multipath propagation with non-negligible time delays. But the latter mixtures remain more difficult to treat.

Many convolutive BSS methods assume the statistical independence of the sources and take advantage of various properties. Surveys of such methods may be found e.g. in [15, 16]. Various approaches are related to information theory (see e.g. [1, 10, 11, 12, 14]). Some approaches were initially introduced by considering non-linear functions. This includes the original convolutive version [13] of the Herault-Jutten approach [7] and its optimized extension by Charkani and Deville [3, 4, 5] which is considered in the benchmark reported in this paper. Tests performed with such approaches [4, 5] showed that their performance is quite good for simple artificial mixtures but gets much lower when real mixtures of acoustical signals are considered, despite the sophisticated principles and important computational load of these methods.

This paper therefore also addresses the convolutive BSS problem, with main emphasis on real mixtures of speech signals but with a specific goal: we aim at achieving a better performance/complexity trade-off than the previous methods for the considered class of signals. To this end, we take advantage of the a priori knowledge that speech sources contain silences.

The remainder of this paper is organized as follows. A first BSS approach with a symmetrical structure is introduced in Section 2, where we especially detail signal segmentation using coherence functions and filter identification. In Section 3, we present an alternative BSS approach based on an asymmetrical structure, which yields various options. We then present experimental results in Section 4 and draw conclusions from this work in Section 5.

2. FIRST PROPOSED BSS APPROACH

2.1. Classical symmetrical BSS structure

We consider the classical BSS configuration shown in Fig. 1, where two signals, $x_1(n)$ and $x_2(n)$, are provided by two microphones. They are generated by the propagation of two speech source signals $s_1(n)$ and $s_2(n)$ which are assumed to be centered and uncorrelated hereafter. These microphone signals $x_i(n)$ are then convolutive mixtures of the two source signals and the signal propagation may be modelled as:

$$\begin{cases} X_1(z) &= A_{11}(z).S_1(z) + A_{12}(z).S_2(z) \\ X_2(z) &= A_{21}(z).S_1(z) + A_{22}(z).S_2(z) \end{cases} \quad (1)$$

where the transfer functions of the $Moving\ Average$ (MA) mixing filters are denoted $A_{ij}(z)$.

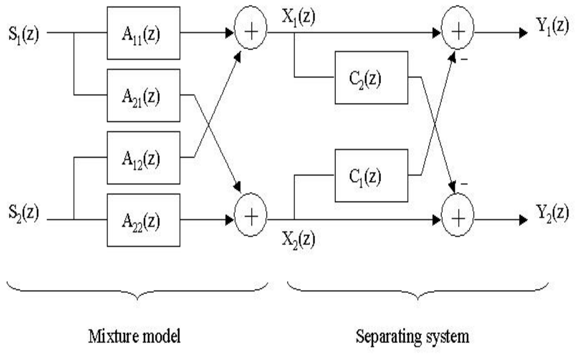

Fig. 1. Classical BSS configuration.

The outputs of the considered separating system read:

$$\begin{cases} Y_1(z) &= X_1(z) - C_1(z).X_2(z) \\ Y_2(z) &= X_2(z) - C_2(z).X_1(z) \end{cases} \tag{2}$$

If we insert the expressions of the observations provided in Eq. (1) into Eq. (2), we obtain the following expressions for the outputs:

$$\begin{cases} Y_1(z) &= S_1(z).[A_{11}(z) - C_1(z).A_{21}(z)] \\ & \quad +S_2(z).[A_{12}(z) - C_1(z).A_{22}(z)] \\ Y_2(z) &= S_1(z).[A_{21}(z) - C_2(z).A_{11}(z)] \\ & \quad + S_2(z).[A_{22}(z) - C_2(z).A_{12}(z)] \end{cases} \tag{3}$$

BSS methods based on this classical structure aim at selecting its filters $C_1(z)$ and $C_2(z)$ so that the following condition (C1) is fulfilled: the system outputs $Y_1(z)$ and $Y_2(z)$ depend on sources $S_1(z)$ and $S_2(z)$ respectively, or on $S_2(z)$ and $S_1(z)$ respectively.

Various methods have been proposed in the literature for adapting the separating filters $C_1(z)$ and $C_2(z)$ so as to reach condition (C1). We will now introduce an alternative method to identify these filters.

2.2. Filter identification procedure

2.2.1. Preliminary version of the identification procedure

As shown above, BSS requires to find the separating filters $C_1(z)$ and $C_2(z)$ which reach condition (C1). Eq. (3) shows that this is achieved for the two couples of filters defined by:

$$\begin{cases} C_1(z) &= \frac{A_{1i}(z)}{A_{2i}(z)} \\ C_2(z) &= \frac{A_{2l}(z)}{A_{1l}(z)} \end{cases} \tag{4}$$

where $(i,l) = (1,2)$ or $(i,l) = (2,1)$.

The general idea of the methods that we introduce in this paper is to benefit from a priori knowledge about speech signals, i.e. more precisely to take advantage of their silence phases. For the sake of simplicity, we first present a preliminary version of the approach that we propose for the above symmetrical BSS structure. This version uses the Fourier transforms of the observations over silence phases as follows:

- We consider the first detected silence phase (we explain in Subsection 2.3 below how we detect it) and use it to assign C_1 as follows:

$$C_1(e^{j\omega}) = \frac{X_1(e^{j\omega})}{X_2(e^{j\omega})} \tag{5}$$

By denoting i the index of the source which is active over this phase, (1) yields:

$$\begin{cases} X_1(e^{j\omega}) &= A_{1i}(e^{j\omega}).S_i(e^{j\omega}) \\ X_2(e^{j\omega}) &= A_{2i}(e^{j\omega}).S_i(e^{j\omega}) \end{cases} \tag{6}$$

so that (5) results in

$$C_1(e^{j\omega}) = \frac{A_{1i}(e^{j\omega})}{A_{2i}(e^{j\omega})} \tag{7}$$

- Similarly, we then consider a silence phase associated to the other source, and use it to assign C_2 as follows:

$$C_2(e^{j\omega}) = \frac{X_2(e^{j\omega})}{X_1(e^{j\omega})} \tag{8}$$

By denoting l the index of the source which is active over this phase, with $l \neq i$, we get in the same way as above:

$$C_2(e^{j\omega}) = \frac{A_{2l}(e^{j\omega})}{A_{1l}(e^{j\omega})} \tag{9}$$

By setting the frequency responses of the filters C_1 and C_2 to the values defined by (7) and (9), with $l \neq i$, these filters are therefore actually made equal to one of the couples of target filters defined in (4).

2.2.2. Final filter identification procedure

Instead of the Fourier transforms of the observations (over time windows) considered above, the eventual approach that we propose to identify the above filters is based on the power spectral densities $S_{x_ix_i}(\omega,k)$ and power cross-spectral densities $S_{x_ix_j}(\omega,k)$ of the observations over time windows indexed by k. By selecting such a window in a silence phase where only S_i is active, Eq. (1) yields:

$$\begin{cases} S_{x_1x_2}(\omega,k) &= A_{1i}(e^{j\omega}).A^*_{2i}(e^{j\omega}).S_{s_is_i}(\omega) \\ S_{x_2x_2}(\omega,k) &= A_{2i}(e^{j\omega}).A^*_{2i}(e^{j\omega}).S_{s_is_i}(\omega) \end{cases} \tag{10}$$

so that:

$$\frac{S_{x_1x_2}(\omega,k)}{S_{x_2x_2}(\omega,k)} = \frac{A_{1i}(e^{j\omega})}{A_{2i}(e^{j\omega})} \tag{11}$$

Deriving the left-hand term of Eq. (11) from the observed signals therefore yields an estimate of the same frequency response $C_1(e^{j\omega})$ as in (7). Similarly, the filter $C_2(e^{j\omega})$ is estimated during a silence phase where only S_l is active, using:

$$\frac{S_{x_2x_1}(\omega,k)}{S_{x_1x_1}(\omega,k)} = \frac{A_{2l}(e^{j\omega})}{A_{1l}(e^{j\omega})} \tag{12}$$

which yields the same filter as (9).

2.3. Silence detection

We now explain how to detect the phases when one source disappears. Our method is based on the real coherence function of the observations $\Gamma(\omega,k)$, measured over half-overlap ping time windows, which are indexed by k. This function is defined as:

$$\Gamma(\omega,k) = \frac{\left| S_{x_ix_j}(\omega,k) \right|^2}{S_{x_ix_i}(\omega,k).S_{x_jx_j}(\omega,k)} \tag{13}$$

where the same notations as above are used.

In the phases when one source disappears, the two observations are equal up to a filter (see Eq. (6)), so that $\Gamma(\omega, k) = 1$. On the contrary, $\Gamma(\omega, k)$ gets significantly lower than 1 for the phases and frequencies when the two sources are active and have a frequency overlap. The magnitude of $\Gamma(\omega, k)$ therefore allows one to detect silence phases. The criterion that we use in practice is the mean of $\Gamma(\omega, k)$ over the frequency range [0-800Hz]. This range was selected because it includes a major part of the energy of speech signals (i.e. pitch and first formant) and $\Gamma(\omega, k)$ is therefore accurately estimated on this range.

The above filter identification and silence detection principles are combined as follows in the resulting overall BSS method. We first compute the above mean coherence functions for all time windows of the mixed signals. We then form an ordered list of the time windows corresponding to silence phases, so that they correspond to decreasing values of the corresponding means of $\Gamma(\omega, k)$. We consider the first window in this ordered list and identify the corresponding value of the separating filter $C_1(z)$. We then use the next time windows in the above ordered list as follows. We identify the value of the filter $C_2(z)$ in another time window and we check if the distance between the frequency responses of $C_1(z)$ and $C_2(z)$ is above a user-defined threshold, thus indicating that $C_2(z)$ does not correspond to the same source as $C_1(z)$. If this condition is met, both target filters of a couple of filters defined in Eq. (4) were identified and the filter identification procedure ends. Otherwise, the above steps are repeated for the next time windows in the ordered list, until the target filter $C_2(z)$ corresponding to the other source is reached.

2.4. Post-processing

The method that we introduced at this stage makes it possible to identify a couple of target filters defined in Eq. (4) and to derive the corresponding separating system outputs according to Eq. (2).

This approach yields two cases, depending on which sources among $S_1(e^{j\omega})$ and $S_2(e^{j\omega})$ are active respectively in the first and second silence phases, i.e. depending whether $(i, l) = (1, 2)$ or $(i, l) = (2, 1)$. More precisely, the resulting signals read:

i) For $(i, l) = (1, 2)$:

$$\left\{ \begin{array}{lcl} Y_1(z) & = & S_2(z).[A_{12}(z) - \frac{A_{11}(z)}{A_{21}(z)}.A_{22}(z)] \\ Y_2(z) & = & S_1(z).[A_{21}(z) - \frac{A_{22}(z)}{A_{12}(z)}.A_{11}(z)] \end{array} \right. \tag{14}$$

ii) For $(i, l) = (2, 1)$:

$$\left\{ \begin{array}{lcl} Y_1(z) & = & S_1(z).[A_{11}(z) - \frac{A_{12}(z)}{A_{22}(z)}.A_{21}(z)] \\ Y_2(z) & = & S_2(z).[A_{22}(z) - \frac{A_{21}(z)}{A_{11}(z)}.A_{12}(z)] \end{array} \right. \tag{15}$$

These signals are therefore not equal to the sources signals $S_l(z)$ themselves, nor to their contributions $A_{il}(z).S_l(z)$ in the measured signals, but to specific filtered versions of the sources. Such a frequency distorsion is a drawback for various applications, especially in the speech processing field. To avoid it, we add a post-filtering stage realized by two filters $D_m(z)$ respectively applied to the output signals $Y_m(z)$, and expressed as:

$$D_m(z) \quad = \quad \frac{1}{\frac{1}{C_n(z)} - C_m(z)} \qquad m \neq n \in \{1, 2\} \tag{16}$$

This stage can be realized if and only if we have identified the two separating filters $C_1(z)$ and $C_2(z)$. With the post-processing filters defined in Eq. (16), the final, i.e. post-processed, outputs read:

i) For $(i, l) = (1, 2)$:

$$\left\{ \begin{array}{lclcl} P_1(z) & = & Y_1(z).D_1(z) & = & A_{22}.S_2(z) \\ P_2(z) & = & Y_2(z).D_2(z) & = & A_{11}.S_1(z) \end{array} \right. \tag{17}$$

ii) For $(i, l) = (2, 1)$:

$$\left\{ \begin{array}{lclcl} P_1(z) & = & Y_1(z).D_1(z) & = & A_{21}.S_1(z) \\ P_2(z) & = & Y_2(z).D_2(z) & = & A_{12}.S_2(z) \end{array} \right. \tag{18}$$

They are therefore equal to the source contributions in the sensor signals in both cases.

3. SECOND PROPOSED BSS APPROACH

The BSS approach that we here propose again uses silence phases in each source to identify separating filters. But these filters are here included in a modified BSS structure.

3.1. Proposed asymmetrical BSS structure

We here consider the same mixture model as in Section 2, but we now introduce a new BSS structure. This structure again yields two signals $Y_1(z)$ and $Y_2(z)$.

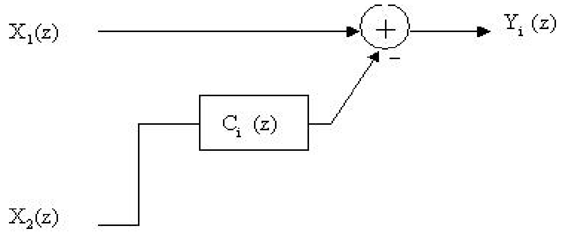

Fig. 2. Proposed structure for deriving each of the separating system outputs.

Each of these signals is here obtained as shown in Fig. 2 and therefore reads:

$$Y_i(z) \quad = \quad X_1(z) - C_i(z).X_2(z) \qquad i \in [1, 2] \tag{19}$$

or:

$$\begin{array}{lcl} Y_i(z) & = & S_1(z).[A_{11}(z) - C_i(z).A_{21}(z)] \\ & & +S_2(z).[A_{12}(z) - C_i(z).A_{22}(z)] \\ & & i \in [1, 2] \end{array} \tag{20}$$

The two values of the separating filter $C_i(z)$ that make it possible to respectively extract $S_1(z)$ and $S_2(z)$ are therefore:

$$\left\{ \begin{array}{lcl} C_i(z) & = & \frac{A_{12}(z)}{A_{22}(z)} \\ C_i(z) & = & \frac{A_{11}(z)}{A_{21}(z)} \end{array} \right. \tag{21}$$

3.2. Filter identification and silence detection

The initial BSS problem is then reformulated as a filter identification problem, i.e: if we are able to identify the two values of $C_i(z)$ defined in Eq. (21), then by computing the corresponding signals $Y_i(z)$ according to Eq. (19) we extract the two source signals. The method that we propose to this end is a direct adaptation of the method introduced in Section 2 for identifying the filter C_1 of the symmetrical structure. More precisely, we here again detect silence phases by means of the procedure described in Subsection 2.3. In such a phase, we set one of the two filters C_i to:

$$C_i(e^{j\omega}) \quad = \quad \frac{S_{x_1x_2}(\omega,k)}{S_{x_2x_2}(\omega,k)} \qquad (22)$$

If only source S_l is active in this phase, the same calculations as in Subsection 2.2.2 here yield:

$$C_i(e^{j\omega}) = \frac{A_{1l}(e^{j\omega})}{A_{2l}(e^{j\omega})} \qquad (23)$$

The two values of C_i thus obtained, respectively associated to $l = 1$ and 2 (again distinguished as explained in Subsection 2.3), are therefore actually equal to the two target values defined in (21).

3.3. Post-processing

At this stage, the outputs signals resulting from this method read explicitly:

$$\left\{ \begin{array}{lcl} Y_i(z) & = & S_1(z).[A_{11}(z) - \frac{A_{12}(z)}{A_{22}(z)}.A_{21}(z)] \\ Y_i(z) & = & S_2(z).[A_{12}(z) - \frac{A_{11}(z)}{A_{21}(z)}.A_{22}(z)] \end{array} \right. \qquad (24)$$

Here again, these signals are not equal to the sources signals $S_i(z)$ themselves, nor to their contributions $A_{il}(z).S_l(z)$ in the measured signals, but to specific filtered versions of the sources. As in our first method, we again need to add a post-filtering stage. That is realized by filters $D_m(z)$ applied to each output signal $Y_m(z)$, and here defined as:

$$D_m(z) \quad = \quad \frac{1}{C_n(z) - C_m(z)} \qquad (25)$$
$$m \neq n \in \{1, 2\}$$

Hereafter, the post-processing filters can again be realized if and only if we have identified the two separating filters $C_i(z)$. With the post-processing filters defined in Eq. (25), the final outputs read:

$$\left\{ \begin{array}{lcl} P_i(z) & = & A_{21}(z).S_1(z) \\ P_i(z) & = & A_{22}(z).S_2(z) \end{array} \right. \qquad (26)$$

This enables us to obtain the contributions of each source on sensor 2, if separation is carried out as in Eq. (19).

3.4. Extensions

3.4.1. Source contributions on the first sensor

A complementary version of the above method, still based on the identification of the ratio of mixing filters during a silence of each source, makes it possible to extract the components from each source on the other sensor. The corresponding separating system provides the signals:

$$Y_i(z) \quad = \quad X_2(z) - C_i(z).X_1(z) \qquad (27)$$
$$i \in \{1, 2\}$$

If we write these new output expressions as in Eq. (20), we can then derive the expressions of the new separating filters, defined by:

$$\left\{ \begin{array}{lcl} C_i(z) & = & \frac{A_{22}(z)}{A_{12}(z)} \\ C_i(z) & = & \frac{A_{21}(z)}{A_{11}(z)} \end{array} \right. \qquad (28)$$

We must note that these filters in Eq. (28) are the inverse of those obtained in (21) for the initial asymmetrical structure. They are therefore directly available from the previous identifications. The post-processing filters $D_m(z)$ used here have the same expression as in Eq .(25), but they now include the filters defined in Eq. (28).

This enables us to obtain as final outputs the contributions of each source on sensor 1, defined by:

$$\left\{ \begin{array}{lcl} P_i(z) & = & A_{11}(z).S_1(z) \\ P_i(z) & = & A_{12}(z).S_2(z) \end{array} \right. \qquad (29)$$

3.4.2. Selection of extracted sources

The four output signals defined in Eq. (26) and in Eq. (29) yield two extracted versions for each source. We should then keep the couple of outputs corresponding to the best separated signals in practice. To this end, we propose to introduce another stage in this method, which calculates the cross-correlation coefficients, before the post-processing stage, between the above four intermediate outputs taken two by two. We then keep the couple of signals which yields the lowest cross-correlation coefficient.

3.4.3. Post-processing Options

As explained in Subsection 3.3, some applications need a post-filtered version of the extracted signals and the corresponding post-processing stage requires both target values of $C_i(z)$ to be identified. We showed in Subsection 3.4.1, that the complementary version of the proposed approach has the same target filters $C_i(z)$ as the initial version (up to an inversion). This means that, by implementing both the initial and complementary versions, we get two estimates of each target filter. We should then select which estimate we keep for each filter. Based on the related discussion that we provided in Subsection 3.4.2., the selection method that we here propose consists in keeping the filter estimates and the couple of outputs, which yield the least correlated signals before the post-processing stage. Then we can apply the post-filtering with the filters thus kept.

4. EXPERIMENTS AND RESULTS

The experiments described in this section concern the cocktail party problem. The two considered speech signals from the English Multext data base are centered, rescaled so as to have similar powers and limited to a length of 100000 samples.

The performance of the proposed approaches has been tested with various mixing filters lengths, with main emphasis on real measured acoustical transfer functions. More precisely, we have first used 256-tap MA filters derived from measurements performed in a car with two microphones, which had an inter-microphone distance equal to 25 cm. The above source signals have also been applied to 8, 16, 32, 64, 128-tap MA filters derived from subsampled in-car measurements. This aims at testing the proposed approach for various mixing conditions.

Eq. (11,12,13,22) require estimates of the power spectral densities $S_{x_i x_i}(\omega, k)$ and power cross-spectral densities $S_{x_i x_j}(\omega, k)$ of the observations over a time window indexed by k. The method used to this end in this paper is a modified version of averaged periodogram. This method is called the Welch-WOSA method [6, 8]. Each time window k of signals is segmented by means of overlapping weighting sub-windows. The power spectral densities and power cross-spectral densities of the segmented signals are calculated over each sub-window, and by averaging them we extract the required estimates. More precisely we use a Hamming window and an overlap of 75%.

Two other methods are tested in order to compare them with our methods. The first of them is noted N1 by Charkani and Deville [3, 4, 5]. It's a decorrelation method with a simple normalization. The other, more sophisticated, method is called the Optimized method and uses sub-optimum separating functions [3, 4, 5]. These two methods are based on a modified version of the stochastic algorithm, corresponding to the Herault-Jutten [7] rule, previously extended by Nguyen Thi et al. [13] to the convolutive domain. The performance reported below for these adaptive algorithms has been measured after a period allocated to their convergence. To this end, the mixed signals were presented twice to these BSS systems and performance was only measured during their second occurence.

The signal/noise ratio (SNR) available before BSS processing, i.e. in each observed signal i, is denoted $SNR_{in}(i)$ hereafter. Performance is then assessed in terms of : i) the SNR measured in the outputs of the BSS system, and denoted $SNR_{out}(i)$ below, and ii) the SNR improvement ($SNRI$) achieved by the system, i.e: $SNRI(i)$=$SNR_{out}(i)$ / $SNR_{in}(i)$. This SNRI is averaged over the two channels of the BSS system, so that the eventual performance criterion is: $SNRI$=$\sqrt{SNRI(1).SNRI(2)}$.

	First Method		Second Method	
filter order	MD	AD	MD	AD
8	15.8	10.5	14.9	9.4
16	22.9	15.6	17	15.2
32	22.2	5.6	19.7	9.0
64	5.6	2.5	5.5	6.3
128	6.0	4.5	4.2	4.7
256	7.6	6.3	6.1	6.0

Table 1. $SNRI$ (dB) obtained with the two proposed methods for different mixing filters, without post-processing, and with automatic (AD) or manual (MD) silence detection.

Table 1 presents the results of experiments performed with the two proposed methods[1] without post-processing, i.e. when considering the signals $Y_i(z)$ as outputs. In each case, we test: i) the complete approaches i.e. including their automatic silence detection (AD) procedure described in Subsection 2.3 and ii) a restricted version where we manually detect the silence phases (MD). Table 1 shows that the two approaches proposed in this paper have similar performance and yield a significant $SNRI$ even in hard mixing conditions. Moreover, they always yield much better performance than the decorrelation approach presented in Table 3, while their computational load is quite limited[2]. The performance of our approaches is much closer to that of the complex optimized approach also presented in Table 3 and even better for various mixing filters, including the actual 256-tap filters. It should also be noted that the modified version of our approaches with manual detection of silence phases yields much better performance than the automated version. These approaches could therefore be improved by developing a better silence detection procedure.

	First Method		Second Method	
filter order	MD	AD	MD	AD
8	12.1	6.0	15.3	9.3
16	16.4	15.3	0.6	0.4
32	22.1	6.5	-1.8	-2.6
64	5.0	4.5	-4.3	-4.9
128	5.8	5.3	0.3	-0.74
256	7.9	6.9	2.5	1.5

Table 2. $SNRI$(dB) obtained with the two proposed methods for different mixing filters, with post-processing, and with automatic (AD) or manual (MD) silence detection.

Table 2 contains the results of experiments performed with the two proposed methods with post-processing, i.e. when considering the output signals $P_i(z)$, in the same conditions as above. This shows that even in hard mixing conditions the first proposed approach yields a $SNRI$ most often equivalent to that of the sophisticated method and always much better than the simple decorrelation method. Here again this approach therefore yields a much better complexity/performance trade-off than the previously reported methods. For some filters, the post-processing stage of the second proposed approach yields significantly degraded performance as compared to the results reported in Table 1. This phenomenon, which is to be further investigated, might be related to the noncausality of some target separating filters.

	Filter order					
	8	16	32	64	128	256
Decorrelation method (N1)	-2.4	2.3	0.1	1.3	0.8	-0.9
Optimized method	10.1	6.7	5.6	4.3	8.1	5.7

Table 3. $SNRI$ (dB) obtained with two classical methods for different mixing filters.

As an example, we now detail the results obtained in an experiment with the two speech source signals and real 256^{th}-order MA filters. Automatic silence detection is made with half-overlapping time windows containing 4096 samples. The 100000-sample source signals are thus segmented in 47 windows indexed by $k = 1, ..., 47$. The coherence functions of the observations over some of these windows are represented in Fig. 3. A silence phase (in source 1) is

[1] All the results reported in this section for the second method concern its overall version, i.e. including its complementary part defined in Subsection 3.4.1. At this stage of our tests, the selection of the version of the extracted sources and associated filters, defined in Subsections 3.4.2 and 3.4.3, was performed as follows: for each source, we manually selected the version which yields the best $SNRI(i)$ before post-processing.

[2] It should be remembered however in this comparison that the methods without post-processing proposed in this paper and the two methods from Charkani and Deville do not yield the same filtered version of the sources.

detected by our method around $k = 17$. This is coherent with the fact that the actual silence phase corresponds to $k \in [16, 22]$. The (filtered) speech signals $P_i(z)$ thus extracted by the first presented BSS method are shown in the two bottom plots of Fig. 4, with the mixed signals represented in the middle plots and the source signals in the top plots.

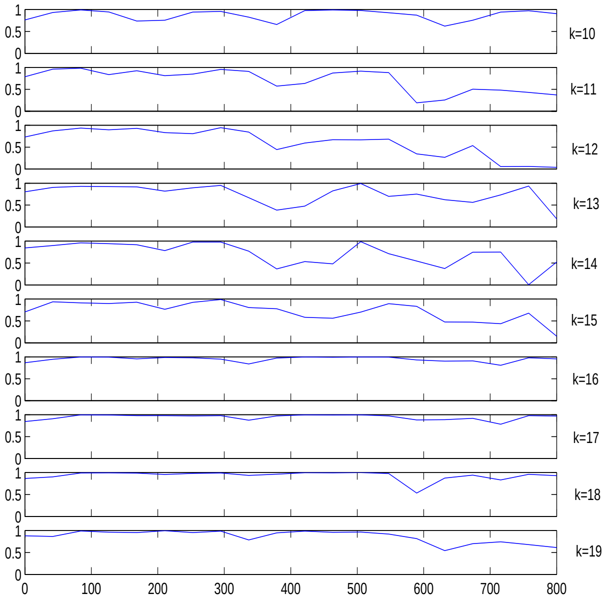

Fig. 3. Coherence function in the range [0 Hz,800 Hz] over time windows indexed by k for 256^{th}-order MA filters.

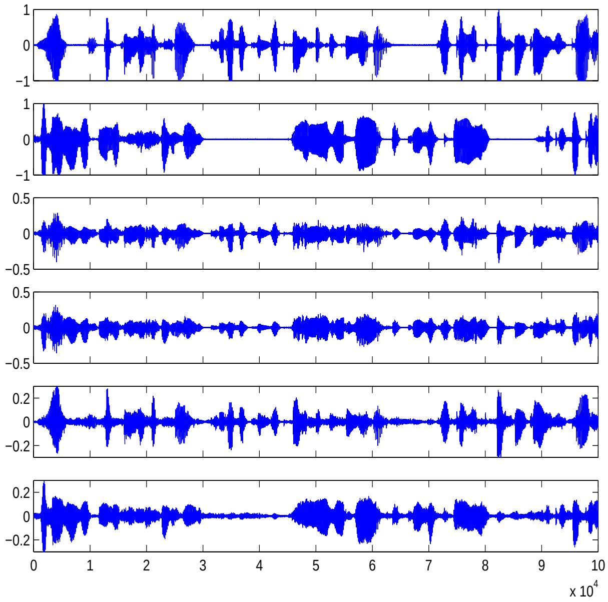

Fig. 4. Temporal signals : Source signals (top), mixed signals (middle), output signals (bottom) for 256^{th}-order MA filters.

5. CONCLUSIONS

In this paper, we presented two main convolutive BSS methods intended for speech separation and based on silence detection and subsequent frequency-domain filter identification. These methods give rise to several versions, especially depending whether or not the extracted sources are post-filtered to obtain the source contributions in sensor signals. Experimental tests with real high-order acoustical mixing filters showed that this post-processing stage may degrade the performance of one the proposed methods. This phenomenon might be related to the fact the two considered structures do not identify the same filters and therefore do not yield the same causality contraints. These topics will be further investigated. Anyway, the two methods without post-processing introduced at this stage yield an interesting performance/complexity trade-off as compared to the two approaches from the literature to which they were experimentally compared.

6. REFERENCES

[1] A.J. Bell, T.J. Sejnowski: *Fast blind separation based on information theory*, Proc. Internat. Symp. on Nonlinear Theory and its Applications, pp. 43-47, Las Vegas, USA, 1995.

[2] J.F. Cardoso: *Blind Signal Separation : Statistical Principles*, Proceedings of the IEEE, vol.86, no.10, 1998.

[3] N. Charkani, Y. Deville: *Self-adaptive separation of convolutively mixed signals with a recursive structure, Part I*, Signal Processing, vol. 73, no. 3, pp. 225-254, 1999.

[4] N. Charkani, Y. Deville: *Self-adaptive separation of convolutively mixed signals with a recursive structure, Part II*, Signal Processing, vol. 75, no. 2, pp. 117-140, 1999.

[5] N. Charkani: *Séparation auto-adaptative de sources pour des mélanges convolutifs : application à la téléphonie mains-libres dans les voitures*,Ph. D. Thesis, Institut National Polytechnique de Grenoble, France, Nov.1996.

[6] M. Durnerin, N. Martin: *Démarche d'analyse spectrale en vue d'une interprétation automatique, application à un signal d'engrenages*, Seizième colloque GRETSI, vol. 1, p. 539-542, Grenoble, France, septembre 1997.

[7] C. Jutten, J. Herault: *Blind separation of sources, Part I: an adaptive algorithm based on neuromimetic architecture*, Signal Processing, vol. 24, pp. 1-10, 1991.

[8] P.J. Johnson, D.G. Long: *The Probability Density of Spectral Estimates Based on Modified Periodogram Averages*, Signal Processing, vol. 47, no. 5, pp. 1255-1261, 1999.

[9] A. Hyvarinen, J. Karhunen, E. Oja: *Independent Component Analysis* , John Wiley, 2001.

[10] K. Matsuoka, M. Kawamoto: *Blind signal separation based on mutual information criterion*, Proceedings of NOLTA, pp. 85-90, Las Vegas, USA, 10 -14 December 1995.

[11] E. Moulines, J.F. Cardoso, E. Gassiat: *Maximum likelihood for blind separation and deconvolution of noisy signals using mixture models*, Proceedings of ICASSP, pp. 3617-3620, Munich, Germany, 21-24 April 1997.

[12] D.T. Pham, P. Garat, C. Jutten: *Separation of a mixture of independent sources through a maximum likelihood approach*, Signal Processing VI, pp. 771-774, 1992.

[13] H-L.N. Thi, C. Jutten: *Blind source separation for convolutive mixtures*, Signal Processing 45, pp. 209-229, 1995.

[14] K. Torkkola: *Blind separation of convolved sources based on information maximisation*, Proceedings of the IEEE Workshop on Neural Network for Signal Processing , Kyoto, Japan, 4-6 September 1996.

[15] K. Torkkola: *Blind separation for audio signals - are we there yet?*, Proceedings of ICA, Aussois, France, 11-15 January 1999.

[16] D. Yellin, E. Weinstein: *Criteria for multichannel signal separation*, IEEE Trans. on Signal Processing, vol. 42, no. 8, pp. 2158-2167, 1994.