A NEW SOURCE SEPARATION CONCEPT AND ITS VALIDATION ON A PRELIMINARY SPEECH ENHANCEMENT CONFIGURATION

Yannick DEVILLE, Frédéric ABRARD, Mohammed BENALI

Université Paul Sabatier (Toulouse III)
Laboratoire d'Acoustique, de Métrologie, d'Instrumentation (LAMI)
38 Rue des 36 Ponts, 31400 Toulouse, France. E-mail: ydeville@cict.fr

ABSTRACT

Classical source separation methods only apply to the case when the number n_o of observed signals is (at least) equal to the number n_s of source signals. In this paper, we first introduce a concept which starts from such methods and which allows to derive corresponding differential (partial) source separation methods applicable to the case when $n_o < n_s$, thus allowing to reduce the number of sensors. We then illustrate this concept on a new separation criterion based on differential normalized kurtosis and we experimentally validate it on a preliminary speech enhancement configuration.

1. INTRODUCTION

Blind source separation is a signal processing problem which consists in estimating a set of n_s independent source signals $X_i(t)$ from a set of n_o observed signals $Y_i(t)$, which are mixtures of these source signals [1]-[3]. The mixed signals $Y_i(t)$ are typically provided by sensors, and the mixing phenomenon then results from the propagation of all signals from their emission locations to all sensors. In the so-called linear instantaneous mixture model, each source-to-observation propagation is represented by a scalar coefficient a_{ij} applied to the considered source signal. This coefficient is related to attenuation during propagation. The overall relationship between the column vectors $X(t)$ and $Y(t)$ of sources and observations then reads:

$$Y(t) = AX(t), \tag{1}$$

where the mixing matrix A consists of the coefficients a_{ij}. Most investigations have been performed in the case when: i) $n_o = n_s$, so that the matrix A is square, and ii) this matrix is invertible. The source separation problem then basically consists in determining an estimate of the inverse of A: consider a vector $S(t)$ of outputs signals of a source separation system, obtained by multiplying the available mixed signals by a separating matrix C, i.e:

$$S(t) = CY(t). \tag{2}$$

Combining (1) and (2) shows that, when C is made equal to A^{-1}, all the output signals $S_i(t)$ resp. become exactly equal to all the source signals $X_i(t)$ to be restored. Various methods have been proposed to estimate A^{-1}, based on the assumed statistical independence of the source signals (see surveys in [1]-[2]). Most of them consist of maximizing or cancelling parameters of the outputs signals such as their moments, cumulants or normalized kurtosis, which are classical quantities in the higher-order statistics field [4]-[5].

Other methods have been developed for so-called convolutive mixtures, where each source-to-observation overall propagation path is represented by a filter. For a survey of these methods, which esp. apply to acoustic signals, see e.g. [3]. Most of these investigations were also performed under the assumption $n_o = n_s$. In many practical situations however, only a limited number of sensors is acceptable, due e.g. to cost constraints or physical configuration, whereas these sensors receive a larger number of sources. This paper concerns this case when $n_o < n_s$: the inadequacy of classical methods is shown, and a new approach is proposed and validated.

2. SOURCE SEPARATION LIMITATION AND PROPOSED CONCEPT

For the sake of simplicity, we focus on the case of linear instantaneous mixtures in this paper. We showed above that if $n_o = n_s$ an exact separation of all sources is theoretically possible (and is achieved when C is made exactly equal to A^{-1}). It is clearly also possible when $n_o > n_s$, but not when $n_o < n_s$ as will now be shown. In the latter case, A is not square. Each output signal $S_i(t)$ defined by (2) is a linear instantaneous combination of the n_o mixed signals. The highest number of source contributions that may be cancelled in such a (non-zero) combination is $n_o - 1$. This optimum case, where each output is still a mixture of the remaining $(n_s - n_o + 1)$ sources, is called partial source separation hereafter.

This discussion only shows that partial source separation is theoretically possible, in the sense that it is achieved for adequate combination coefficients c_{ij} in the matrix C. To actually achieve it in practice, algorithms which are able to estimate these adequate values of the coefficients c_{ij} are then needed. It may be shown that the classical methods, that have been developed for the case when $n_o = n_s$, do not meet this requirement: whereas their principles (such as the above-mentioned cumulant maximization) coincide with the separation of the source signals when $n_o = n_s$, they yield outputs signals which are still mixtures of all sources when applied to arbitrary observed signals such that $n_o < n_s$.

This paper therefore aims at introducing a concept which achieves the above-defined partial source separation when $n_o < n_s$. By "concept", we mean that we do not propose a single criterion (and/or algorithm) but a general way to derive new partial source separation methods from various existing approaches developed for the case when $n_o = n_s$. We therefore start from one of the latter methods, based on a given parameter, such as those defined in Section 1. We set the following additional constraints on the source signals: i) two occurences of the considered parameter should be available and ii) the considered sources should consist of two types with respect to corresponding source parameters, i.e:

1. first type of sources: for (at most) n_o sources, the corresponding source parameter should take different values in the two occurences,
2. second type of sources: for the other $n_s - n_o$ sources (at least), this parameter should take the same value in the two occurences.

The above-mentioned two "occurences" may be obtained in various ways, allowing one to derive various methods from the concept that we propose. For example, they may correspond to the two values of a parameter, such as a cumulant, resp. obtained for two time domains $\mathcal{D}_1$ and $\mathcal{D}_2$. These domains may be defined as follows. When considering these theoretical cumulants themselves, which consist of mathematical expectations of signals [4]-[5], each of these time domains $\mathcal{D}_i$ is restricted to the time position t_i when the corresponding mathematical expectations are considered. On the contrary, the corresponding cumulant estimates are resp. obtained by time averaging over two domains $\mathcal{D}_1$ and $\mathcal{D}_2$ which then consist of non-overlapping bounded time intervals, assuming that each source exhibits stationarity over each such domain (this is referred to as "short-term stationarity" below). When the two parameter occurences are defined in such a way, the second type of sources consists of sources whose cumulant values do not vary from one of the considered time domains to the other, i.e. sources which in addition exhibit "long-term stationarity". On the contrary, first-type sources then consist of long-term non-stationary sources.

Whatever the origin of these occurences, the proposed concept consists in deriving new signal parameters from the initial ones in such a way that the effect of the second type of sources desappears in them, using a method that is detailed further in this section. The initial problem is thus transformed in a problem where (at most) n_o sources are visible (i.e. the first-type sources) from the point of the view of the new parameter, and n_o mixtures of these sources are available. In other words, thanks to this approach we get back in the classical configuration involving as many observed signals as source signals. Adapting the initial source separation method to the optimization of the new parameter then yields separating coefficients c_{ij} such that each output signal $S_i(t)$ contains a contribution from only one of the sources seen by the approach, i.e. of the first-type sources. Of course, each such signal $S_i(t)$ also contains contributions from all second-type sources, corresponding to the combination of the mixed signals $Y_i(t)$ by means of the above-mentioned coefficients c_{ij}, as shown in (2). This approach therefore achieves the above-defined partial source separation, and more specifically the sources that are thus completely separated are the first-type one. These sources are the signals of interest in this partial source separation problem, as opposed to the "noise signals", i.e. the second-type sources, which are still present in all outputs.

We now explain how to derive new parameters in which the second-type sources are invisible. For the sake of simplicity, we assume that the initial parameter is a linear function of the corresponding source parameters (for example, an output cumulant is indeed a linear function of the cumulants of the independent sources). The new parameter that we then define exploits the difference of behavior between the two types of sources concerning the variations of their initial parameter from one occurence to the other. More precisely, we here focus on the case when the new parameter that we define is the difference between the two values of the initial parameter resp. associated with the two available occurences (e.g. the difference between the two output cumulants corresponding to the two time domains). As each second-type source yields the same contribution in the two initial parameters, its effect is cancelled in the corresponding parameter difference (see the example in the next section), i.e. the above-defined goal is thus reached.

To summarize, the proposed concept is a differential source separation approach, which uses differential parameters as optimization criteria, so as to make some sources invisible in these criteria and to perform an exact separation of the other sources only. In this section, we expressed this concept in the most general terms to show that it may be applied to various existing initial (i.e. non-differential) parameters and source separation methods in order to derive their differential counterparts. In the next section, we explicitly apply these ideas to a specific parameter and an associated method.

3. A CRITERION BASED ON DIFFERENTIAL NORMALIZED KURTOSIS

We now apply the above concept to the source separation criterion that we defined in [6]-[7] for the case $n_o = 2$ mixtures of $n_s = 2$ sources. Briefly, this criterion consists of adapting each coefficient of the source separation system so that the normalized kurtosis of the corresponding system output reaches a specific type of extremum. So, we now have to define the new parameter used in the differential source separation method derived below, i.e. the differential version of the normalized kurtosis of a signal. Given an arbitrary signal u and two time domains $\mathcal{D}_1$ and $\mathcal{D}_2$ defined in Section 2, we define the corresponding (zero-lag) m^{th}-order differential cumulant as:

$$\begin{aligned} DCUM_m(u, \mathcal{D}_1, \mathcal{D}_2) &= CUM_m(u, \mathcal{D}_2) \\ &\quad - CUM_m(u, \mathcal{D}_1), \end{aligned} \tag{3}$$

where $CUM_m(u, \mathcal{D}_i)$ are the corresponding classical (i.e. non-differential) cumulants, which may be found e.g. in [4]-[5]. Based on the expression of the classical (i.e. non-differential) normalized kurtosis [4]-[5], we then define the corresponding differential normalized kurtosis as:

$$DK(u, \mathcal{D}_1, \mathcal{D}_2) = \frac{DCUM_4(u, \mathcal{D}_1, \mathcal{D}_2)}{[DCUM_2(u, \mathcal{D}_1, \mathcal{D}_2)]^2}. \tag{4}$$

The source separation configuration that we then consider corresponds to $n_o = 2$ observed signals $Y_i(t)$ which are linear instantaneous mixtures of an arbitrary number n_s of sources signals. Two of these source signals, denoted $X_1(t)$ and $X_2(t)$, have long-term non-stationarity: these are the signals of interest to be separated. The other source signals, denoted $V_3(t)$ to $V_{n_s}(t)$, have long-term stationarity: these are the noise signals. In the remainder of this section, we consider the centered versions of all signals, denoted with lower-case letters. The (centered) mixed signals then read:

$$y_i(t) = a_{i1}x_1(t) + a_{i2}x_2(t) + \sum_{j=3}^{n_s} a_{ij}v_j(t), \quad i \in 1, 2. \tag{5}$$

Now consider a linear instantaneous combination of the source signals:

$$z(t) = \alpha_1 x_1(t) + \alpha_2 x_2(t) + \sum_{j=3}^{n_s} \alpha_j v_j(t). \tag{6}$$

This signal may e.g. be an output of the considered source separation system, created as a linear instantaneous combination of the available two mixed signals and here expressed as:

$$z(t) = y_1(t) - cy_2(t). \tag{7}$$

The α coefficients in (6) then combine the mixing and separating coefficients a_{ij} and c. The corresponding m^{th}-order cumulant at each time position t_i is then derived by using the properties of cumulants [4]-[5] and the assumed statistical independence of the source signals. This yields:

$$\begin{aligned} CUM_m(z, t_i) &= \alpha_1^m CUM_m(x_1, t_i) \\ &\quad + \alpha_2^m CUM_m(x_2, t_i) \\ &\quad + \sum_{j=3}^{n_s} \alpha_j^m CUM_m(v_j, t_i). \end{aligned} \tag{8}$$

The corresponding differential cumulant $DCUM_m(z, t_1, t_2)$ then follows from the difference defined in (3). The contribution of each noise signal desappears in this difference, as its cumulants $CUM_m(v_j, t_i)$ corresponding to the two time positions t_1 and t_2 have the same value. All noise signals therefore also desappear in the differential normalized kurtosis, which reads:

$$\begin{aligned} &DK(z, t_1, t_2) \\ &= \frac{\alpha_1^4 DCUM_4(x_1, t_1, t_2) + \alpha_2^4 DCUM_4(x_2, t_1, t_2)}{[\alpha_1^2 DCUM_2(x_1, t_1, t_2) + \alpha_2^2 DCUM_2(x_2, t_1, t_2)]^2} \end{aligned} \tag{9}$$

or equivalently:

$$DK(z, t_1, t_2) = \frac{DK(x_1, t_1, t_2) + q^2 DK(x_2, t_1, t_2)}{[1 + q]^2} \tag{10}$$

where we introduce the variable q defined as:

$$q = \frac{\alpha_2^2 DCUM_2(x_2, t_1, t_2)}{\alpha_1^2 DCUM_2(x_1, t_1, t_2)}. \tag{11}$$

Comparing (9)-(11) with the expressions what we derived in [6]-[7] for our initial (i.e. non-differential) source separation criterion shows that they are identical, except that: i) the initial (i.e. non-differential) cumulants and normalized kurtosis are replaced by their differential counterparts, and therefore ii) the variable p that we introduced in our initial criterion is replaced by the variable q. Note that, whereas p always covers the range $[0, +\infty]$, q covers the same range or $[-\infty, 0]$ depending on the signs of the differential 2^{nd}-order cumulants of the considered sources. Due to space limitations, we here restrict ourselves to the case when it covers the range $[0, +\infty]$. The solution that we then propose for the partial source separation problem considered here is therefore the differential counterpart of the initial one, i.e: the separation of the signals of interest is guaranteed to be achieved by adjusting each coefficient c of the separation system so as to reach the extrema of $DK(z, t_1, t_2)$, where the types of extrema to be considered depend on the signs of the here-differential cumulants of the sources, as explained in [6]-[7].

4. EXPERIMENTAL RESULTS

We now validate the above approach on a preliminary speech enhancement configuration. The mixed signals to be processed contain three sources, i.e. two real speech signals $X_1(t)$ and $X_2(t)$ and an artificial stationary uniform noise signal $V_3(t)$. All signal values ranges from -1 to + 1. The speech signals are sampled at 8 kHz and considered over two 50-ms time intervals, over which they exhibit short-term stationarity, whereas they have a long-term non-stationarity from one interval to the other. We then aim at creating output signals in which one of these speech signals is suppressed (and stationary noise is still present but may then be reduced by traditional techniques which are not considered here). The tests reported here are preliminary in the sense that, for the specific method of Section 3 to be applicable, we created artificial linear instantaneous mixtures of these sources (as opposed to actual convolutive speech mixtures), i.e specifically:

$$Y_1(t) = X_1(t) + 0.9X_2(t) + 0.2V_3(t) \quad (12)$$
$$Y_2(t) = 0.8X_1(t) + X_2(t) + 0.3V_3(t). \quad (13)$$

Adapting the theoretical results of [6]-[7] to the differential version of the approach considered here shows that, for the considered positive-differential-cumulant speech signals, the differential normalized kurtosis of the centered source separation system output, i.e. $DK(z, \mathcal{D}_1, \mathcal{D}_2)$, should have two maxima, and that these maxima should correspond to the two values of c for which one of the speech signals is suppressed in the system output, i.e. $c = a_{12}/a_{22} = 0.9$ and $c = a_{11}/a_{21} = 1.25$. Fig. 1 shows that the considered signals actually lead to this behavior and are therefore indeed separated when using the criterion which consists in maximizing $DK(z, \mathcal{D}_1, \mathcal{D}_2)$.

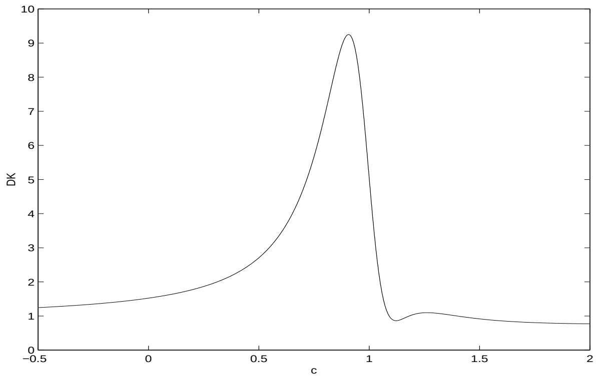

Figure 1: Variations of differential normalized kurtosis of output of source separation system vs. its tunable coefficient c.

5. CONCLUSIONS AND FUTURE WORK

In this paper, we have introduced a concept which allows to perform partial source separation with a reduced number of sensors. We developed this differential approach up to experimental validation, but two types of extensions are still to be made for practical application, i.e: i) optimization algorithms, e.g. gradient-based methods, will be derived from the proposed separation criterion and ii) for speech enhancement applications, the proposed concept will be applied to convolutive source separation methods.

6. REFERENCES

[1] J.-F. Cardoso, "Blind signal separation: statistical principles", Proceedings of the IEEE, vol. 86, no. 10, pp. 2009-2025, Oct. 1998.

[2] K. J. Pope, R.E. Bogner, "Blind signal separation I. Linear, instantaneous combinations", Digital Signal Processing, vol. 6, pp. 5-16, 1996.

[3] K. J. Pope, R.E. Bogner, "Blind signal separation II. Linear, convolutive combinations", Digital Signal Processing, vol. 6 pp. 17-28, 1996.

[4] J.M. Mendel, "Tutorial on higher-order statistics (spectra) in signal processing and system theory: theoretical results and some applications", Proceedings of the IEEE, vol. 79, no. 3, pp. 278-305, March 1991.

[5] C.L. Nikias, J.M. Mendel, "Signal processing with higher-order spectra", IEEE Signal Processing Magazine, pp. 10-37, July 1993.

[6] Y. Deville, "A source separation criterion based on signed normalized kurtosis", Proceedings of the 4th International Worshop on Electronics, Control, Measurement and Signals (ECMS'99), pp. 143-146, Liberec, Czech Republic, May 31 - June 1, 1999.

[7] Y. Deville, M. Benali, "Un critère de séparation de sources fondé sur les kurtosis normalisés signés", Proceedings of GRETSI'99 symposium, vol. 4, pp. 1145-1148, Vannes, France, Sept. 13-17, 1999.