Authors' final version of a paper published in "Applied Signal Processing"

Paper reference:
Y. Deville, S. Roblin, "A feature extraction method for convolutively mixed signals, with applications to power estimation", Applied Signal Processing, vol. 6, no. 1, pp. 2-12, 1999.

A feature extraction method for convolutively mixed signals, with applications to power estimation

Y. Deville [†] (1), (2) [††], S. Roblin (2)

(1) Université Paul Sabatier
Laboratoire d'Acoustique, de Métrologie et d'Instrumentation (LAMI)
38, Rue des 36 Ponts
31400 Toulouse
France.

(2) Laboratoires d'Electronique Philips S.A.S. (LEP)
22, Avenue Descartes
BP 15
94453 Limeil-Brévannes Cedex
France.

[†] corresponding author:
tel. : + 33 5 61 55 65 33
fax. : + 33 5 61 25 94 78
e-mail: ydeville@cict.fr

[††] This work was performed when the first author was with (2). He is now with (1).

Abstract : This paper presents a method for processing convolutive mixtures of source signals. The presented method aims at estimating various features of these sources, with simpler means than the techniques traditionnally required for completely separating these signals. The mixed signals are first processed in order to obtain "associated mixed signals", which are only linear instantaneous mixtures of "associated source signals" (at least to a first order). These associated mixtures are then processed by a linear instantaneous source separation device. This yields estimates of the associated sources, which are eventually used to derive features of the original sources. The performance of this approach has been checked experimentally in the case when: i) the considered signals are real mixtures of acoustic sources (speech and/or music), and ii) the extracted features are the average powers of the source signals. The resulting power estimation errors are lower than 4%, which is quite acceptable in the considered applications.

Keywords : acoustic signals, autocorrelation function, blind source separation, convolutive mixtures, higher-order statistics, power estimation.

1 Introduction

Blind source separation is a generic signal processing problem found in many applications such as antenna or microphone array processing [1], [2]. More precisely, several classes of such problems may be distinguished, depending on the nature of the mixtures which occur between the considered source signals. The simplest class is the separation of unknown source signals $X_j(t)$ only available through sensor signals $E_i(t)$ which are linear instantaneous unknown mixtures of these source signals. For this Linear Instantaneous Source Separation (LISS) problem, the basic configuration corresponds to two sources and two sensors (Fig. 1) and is represented by:

$$E_1(t) = a_{11}X_1(t) + a_{12}X_2(t) \quad (1)$$
$$E_2(t) = a_{21}X_1(t) + a_{22}X_2(t) \quad (2)$$

where the terms a_{ij} are unknown constant mixture coefficients, which may represent the attenuation of the sources occuring during their propagation to the sensors. The goal is then to estimate the source signals $X_j(t)$ (up to a permutation and a scaling factor) from the measured signals $E_i(t)$.

A more general class of Attenuation Delay Source Separation (ADSS) corresponds to mixtures which include the attenuation coefficients of Eqs. (1) and (2) and also elements of delays to yield the basic configuration:

$$E_1(t) = a_{11}X_1(t-\theta_{11}) + a_{12}X_2(t-\theta_{12}) \quad (3)$$
$$E_2(t) = a_{21}X_1(t-\theta_{21}) + a_{22}X_2(t-\theta_{22}) \quad (4)$$

where the delays θ_{ij} correspond to the unknown propagation times between the emission locations of the source signals and the sensors.

Finally, the most general class considered in this paper is Convolutive Source Separation (CSS) which corresponds to wideband convolutive mixtures, i.e. in the basic configuration:

$$E_1(t) = h_{11}(t) * X_1(t) + h_{12}(t) * X_2(t) \quad (5)$$
$$E_2(t) = h_{21}(t) * X_1(t) + h_{22}(t) * X_2(t) \quad (6)$$

where "$*$" represents convolution and $h_{ij}(t)$ are the impulse responses of unknown constant mixture filters which, for example, model the case when wideband sources propagate with a frequency-dependent attenuation.

A link between the above-defined classes of source separation problems is created as follows in this paper. Convolutively mixed signals are considered. For such signals, a classical CSS approach would aim at restoring precisely each source, but could reach this goal only at the expense of considerable processing means (i.e. typically a set of adaptive filters: see e.g. [3]-[10]). On the contrary, the method considered in this paper is based on a performance/cost trade-off. The function to be performed by the system under investigation is restricted to the estimation of source features, such as their average powers in successive time windows. But this is achieved with simpler means than a CSS device: apart from a pre-processing stage, these means only consist of a LISS device (where each complete filter of a CSS device is replaced by a single coefficient).

This method applies to acoustic signals, since the propagation of a number of acoustic sources to a set of microphones provides signals which are superpositions of filtered versions of these sources, i.e. convolutive mixtures of these sources. Especially, when reflections

may be neglected (e.g. in the case of free-field propagation), this source filtering effect resulting from propagation becomes restricted to an attenuation and a delay, as in the above-defined ADSS problem. Acoustic signals are the type of sources that motivated our investigation and that are used in this paper to illustrate the performance of the proposed approach.

The remainder of this paper is organized as follows. The detailed principles of the method outlined above are presented in Section 2. Its performance for artificial mixtures of acoustic signals is provided in Section 3, whereas the results obtained with real recordings are reported in Section 4. Section 5 first describes some potential applications of the proposed method, as they may help understand the motivations of this investigation. Section 5 then presents the conclusions drawn from this investigation and its potential extensions.

2 Principles of the proposed approach

2.1 Overall system structure

In this subsection, we consider two types of acoustic source signal mixtures, i.e. respectively the ones corresponding to ADSS and CSS problems. For each case, the proposed processing method is depicted.

In the case when the available mixed signals are the same as in ADSS problems (see (3)-(4)), the method proposed in this paper consists of the following three steps (Fig. 2):

1. Estimated mean values $< E_i(t) >$ of the mixed signals $E_i(t)$ are computed over the considered time window. Estimated zero-mean versions $e_i(t)$ of the supposedly stationary mixed signals are then derived as: $e_i(t) = E_i(t) - < E_i(t) >$. These signals $e_i(t)$ are related to the zero-mean versions $x_j(t)$ of the source signals $X_j(t)$ according to the same equations as the ones that exist between the non-zero-mean signals, i.e:

$$e_1(t) = a_{11}x_1(t-\theta_{11}) + a_{12}x_2(t-\theta_{12}) \quad (7)$$
$$e_2(t) = a_{21}x_1(t-\theta_{21}) + a_{22}x_2(t-\theta_{22}). \quad (8)$$

2. The zero-mean mixed signals $e_i(t)$ are then processed so as to obtain signals called the "associated mixed signals" and denoted $\tilde{e}_i(v)$ hereafter, which depend on a variable v defined below. This processing step aims at providing signals $\tilde{e}_i(v)$ which are linear instantaneous mixtures of some signals, called the "associated source signals" and denoted $\tilde{x}_j(v)$ below. This may be achieved by various processing approaches, so that the overall generic method described here may be used to define various types of systems. Especially, a possible approach consists in computing the autocorrelation function $\Gamma_{e_i}(\tau)$ of each zero-mean mixed signal $e_i(t)$. The validity of this specific approach results from the fact that, due to (7)-(8), these autocorrelation functions $\Gamma_{e_i}(\tau)$ may be expressed as follows with respect to the autocorrelation functions $\Gamma_{x_j}(\tau)$ of the supposedly uncorrelated zero-mean source signals $x_j(t)$:

$$\Gamma_{e_1}(\tau) = a_{11}^2\Gamma_{x_1}(\tau) + a_{12}^2\Gamma_{x_2}(\tau) \quad (9)$$
$$\Gamma_{e_2}(\tau) = a_{21}^2\Gamma_{x_1}(\tau) + a_{22}^2\Gamma_{x_2}(\tau). \quad (10)$$

Comparing these equations to (1)-(2) shows that the autocorrelation functions $\Gamma_{e_i}(\tau)$ are indeed linear instantaneous mixtures of the autocorrelation functions $\Gamma_{x_j}(\tau)$. This paper mainly deals with this case when the associated mixed and source signals $\tilde{e}_i(v)$ and $\tilde{x}_j(v)$ are respectively the autocorrelation functions $\Gamma_{e_i}(\tau)$ and $\Gamma_{x_j}(\tau)$ (the variable v on which $\tilde{e}_i$ and $\tilde{x}_j$ depend is then the argument τ of the corresponding autocorrelation functions). [11] depicts other cases, especially based on the power spectral densities of the signals.

3. The above-defined second processing step transformed the initial problem into a LISS problem (concerning the associated source signals). The third step of the overall proposed method therefore consists in solving the latter problem. To this end, any classical or original LISS structure may be used. In our investigation, we considered various such structures. Their principles and application to our approach are described in the following subsections.

It should be noted that in classical situations the considered source separation unit directly receives the mixed signals $E_i(t)$ (or their centered versions $e_i(t)$) and aims at providing output signals $S_i(t)$ (or their centered versions $s_i(t)$) which are estimates of the original sources. On the contrary, in the overall approach proposed in this paper, the source separation unit receives the associated mixed signals $\tilde{e}_i(v)$. This method therefore yields a restriction, i.e. its outputs $\tilde{s}_i(v)$ are only estimates of the associated source signals, but not estimates $S_i(t)$ of the initial source signals. More precisely, if the second step of the proposed approach consists in computing autocorrelation functions, the system provides estimates of the autocorrelation functions $\Gamma_{x_j}(\tau)$ of the centered source signals[1]. In particular, the values of these outputs for $\tau = 0$ are estimates of the (average) powers of the centered source signals[2]. These powers are the major parameters to be determined with the proposed approach.

In the case when the available mixed signals are the same as in CSS problems (see (5)-(6)), two alternative approaches may be used. The first one consists in only using the above-defined method, which then only provides an approximate solution. The second approach consists of: i) splitting the problem in sub-bands, so as to reduce it in each sub-band to the basic ADSS-like problem that we described above, and ii) "gathering" the results obtained in each sub-band (e.g. by adding the source signal powers estimated in each sub-band) in order to solve the overall problem. As shown below, "good performance" (with respect to the accuracy and product cost required in the considered applications) may be achieved without splitting the signals in sub-bands, even in real situations where the mixed signals are expected to be of the same general type as in CSS problems.

2.2 Hérault, Jutten and/or Nguyen LISS devices

As stated above, this subsection summarizes the major principles of the LISS devices used in this paper. Many other LISS devices exist however. A survey of such devices may be found e.g. in [2]. The considered devices are based on the recursive structure shown in Figure 1. They are supposed to receive linear instantaneous mixtures of statistically independent source signals (see Fig. 1)[3]. In the first version, proposed by Hérault and Jutten

[1] Up to a permutation and scaling factor, as explained below.

[2] Or more precisely, of the normalized centered source signals (see Subsection 2.3).

[3] In fact, these devices also apply to sources which only meet some requirements which are less stringent than statistical independence. For example, let us consider the first version of these devices described

[1], [12]-[14], the adaptive weights c_{12} and c_{21} of this structure are updated according to the following adaptation rule:

$$\Delta c_{ij}(t) = \mu f[s_i(t)]g[s_j(t)], \tag{11}$$

where μ is a positive adaptation gain, $s_i(t)$ and $s_j(t)$ are the (estimated) zero-mean signals corresponding to the outputs $S_i(t)$ and $S_j(t)$ of this structure, and f and g are functions which should meet some requirements [1],[10]. The most commonly used functions are $f = (.)^3$ and $g = (.)$. The rule (11) then becomes:

$$\Delta c_{ij}(t) = \mu[s_i(t)]^3 s_j(t). \tag{12}$$

Each weight c_{ij} is thus updated according to a rule which performs a stochastic cancellation of the expression $E\{[s_i(t)]^3 s_j(t)\}$, where $E\{\}$ stands for mathematical expectation. This expression is the 3.1 cross-moment of the (ordered) couple of zero-mean output signals corresponding to the considered weight, i.e. outputs i and j respectively (see [15] for general information about cross-moments of signals; "3.1" here refers to the powers 3 and 1 with which outputs i and j respectively appear in the considered moment). The structure based on these functions however has a limited field of application, since it can only separate globally sub-Gaussian sources [16]-[18] i.e. sources such that

$$E\{x_1^4\}E\{x_2^4\} < 9(E\{x_1^2\})^2(E\{x_2^2\})^2. \tag{13}$$

It may be shown that globally super-Gaussian sources, i.e. sources such that

$$E\{x_1^4\}E\{x_2^4\} > 9(E\{x_1^2\})^2(E\{x_2^2\})^2, \tag{14}$$

are separated by using $f = (.)$ and $g = (.)^3$ (see [19]-[20]; this may be shown by adapting the approach of [16] to the functions considered here). With such functions, each weight c_{ij} is updated according to the rule:

$$\Delta c_{ij}(t) = \mu s_i(t)[s_j(t)]^3, \tag{15}$$

which performs a stochastic cancellation of $E\{s_i(t)[s_j(t)]^3\}$, which is the 1.3 cross-moment of the couple of zero-mean output signals corresponding to this weight.

The last version of the LISS device used in this paper is the one which has been proposed by Jutten and Nguyen [4],[6]. It adapts each weight c_{ij} of the above-defined recursive structure so as to cancel the 3.1 cross-cumulant [15] of the couple of zero-mean output signals corresponding to this weight. This cross-cumulant cancellation is based on a stochastic gradient descent, using the square of this cumulant as the cost function, with a constant or adaptive gain (see [4],[6] for more details). This approach was claimed

to apply to any type of (non-Gaussian) sources, thus avoiding the above-mentioned restrictions of the cross-moment-based rules. However, it results in a higher computational complexity.

---

below, which corresponds to (11), and let us restrict ourselves to the case when its functions are set to $f = (.)^3$ and $g = (.)$. Then, to be separated exactly, the sources only have to be such that the 3.1 cross-moments (defined below) of the zero-mean versions of these sources are zero (in addition to the condition (13)). Using $f = (.)$ and $g = (.)^3$ leads to an equivalent condition. Moreover, non-zero source cross-moments may be acceptable, provided they only entail minor deviations for the convergence points of these algorithms, as explained in Sub-section 3.3.

For all these versions of the LISS device, when the weight values are such that exact source separation without a permutation is achieved, the proportionality coefficients between the sources and the device outputs have specific values, i.e:

$$S_1(t) = a_{11}X_1(t) \quad (16)$$
$$S_2(t) = a_{22}X_2(t). \quad (17)$$

Comparing these equations to (1)-(2) shows that each device output $S_i(t)$ is then equal exactly (i.e. not only up to an arbitrary factor) to the contribution of source $X_i(t)$ in the mixed signal $E_i(t)$. This property is of importance in the remainder of this paper, where the signals $S_i(t) = a_{ii}X_i(t)$ are called the "normalized source signals" (this "normalization" refers to the signal magnitudes).

2.3 Application to the proposed method

The application of any of the above-defined LISS devices to the overall method proposed in this paper deserves the following comments. This device here receives the associated mixed signals $\tilde{e}_i(v)$, not the initial mixed signals $E_i(t)$ (see Fig. 2). Especially, in the major version considered hereafter, it receives the autocorrelation functions $\Gamma_{e_i}(\tau)$. Each input sample provided to this device thus corresponds to a specific value of the parameter τ of the functions $\Gamma_{e_i}(\tau)$. From these mixed signals, defined by (9)-(10) in the case of ADSS, the device derives (estimates of) the associated normalized source signals, i.e:

$$\tilde{s}_1(v) = a_{11}^2\Gamma_{x_1}(\tau) \quad (18)$$
$$\tilde{s}_2(v) = a_{22}^2\Gamma_{x_2}(\tau) \quad (19)$$

and, as already stated above, the variable v on which these signals depend is then the argument τ of autocorrelation functions. Especially, for $\tau = 0$, each device output i is equal to $a_{ii}^2\Gamma_{x_i}(0)$, which is the (average) power of the normalized source i, i.e. the power of the component corresponding to source i contained by the zero-mean mixed signal i. It should be noted that this power of the normalized source is obtained exactly, i.e. not up to an arbitrary factor, which would of course be a useless result.

3 Performance with artificial mixtures

3.1 Goal and principles of the tests

The first series of tests aimed at validating the basic principles of the proposed approach and was performed in the following conditions. The considered sources are three real speech signals, recorded separately and resp. corresponding to the French words "bonjour" and "parle" and to the short sentence "le camp d'été s'est passé" [6]. The mixed signals, provided to the processing system proposed in this paper, are artificially created. More precisely, the tests described in this section aim at investigating the performance of the proposed approach for the basic type of mixtures for which it was developed, i.e. for mixed signals which meet exactly (3)-(4). These tests therefore primarily consist in selecting two of the above-mentioned real source signals, and numerically combining them according to (3)-(4), so as to create two mixed signals. Associated mixed signals $\Gamma_{e_i}(\tau)$ are then derived from the above mixed signals, using the first two steps of the approach described in Sub-section 2.1. In the third step of this approach, the associated mixed signals are eventually separated with a LISS device.

In fact, the considered practical implementation is slightly different from the above-defined basic principles: as the associated mixed signals $\Gamma_{e_i}(\tau)$ meet exactly (9)-(10) in the considered conditions, the approach which is actually used consists of the following steps. The associated source signals $\Gamma_{x_j}(\tau)$ are first computed (the signal means and autocorrelation functions were computed with the complete available source signals in the tests reported below[4]). The associated mixed signals $\Gamma_{e_i}(\tau)$ are then numerically derived according to (9)-(10). The latter signals are eventually separated with a LISS device. As compared to the basic approach described in the previous paragraph, this modified version avoids the need to introduce the parameters θ_{ij}, which are not relevant here since their influence subsequently disappears when processing the signals with the proposed method[5]. As for the structure of the LISS device, all the versions described in Section 2 are successively used here. For each of them, the tests carried out aim at checking that the adaptive weights c_{12} and c_{21} of the considered LISS device actually converge to the values corresponding to exact source separation without a permutation, i.e. that these weights converge to the following theoretical values (which are derived by adapting the results in [1] to the mixture equations (9)-(10) considered here):

$$c_{ij}^0 = \left(\frac{a_{ij}}{a_{jj}} \right)^2 . \tag{20}$$

The mixture coefficient values a_{ij} used in these tests are selected according to two criteria. On the one hand, ratios of such coefficients define: i) the magnitude of the "mixture rate" which occurs between the source signals, and ii) the weight values required for the LISS device to separate these mixed signals (see (20)). From that point of view, the tests are performed in the case of moderately high mixture rates, corresponding to theoretical weight values $c_{12}^0 = 1/2$ and $c_{21}^0 = 1/2$. On the other hand, applying a common scaling factor to all these coefficients a_{ij} allows us to set the magnitudes of the associated mixed signals. This factor is selected so as to scale these signal magnitudes to about 1.

3.2 Inadequacy of the 3.1 cross-moment-based algorithm

Tests were first performed with the version of the LISS device based on the adaptation rule (12). The weights of this device then do not converge to their theoretical values. As shown below, this results from the nature of the sources processed by this device, i.e. in the current case, it results from the nature of the associated source signals $\Gamma_{x_j}(\tau)$.

Generally speaking, the normalized kurtosis (or normalized 4th-order zero-lag cumulant) of a zero-mean stationary signal s is defined as [21]:

$$\gamma = \frac{cum4(s)}{(E\{s^2\})^2} = \frac{E\{s^4\}}{(E\{s^2\})^2} - 3. \tag{21}$$

A super-Gaussian signal [17] corresponds to: $\gamma > 0$. Moreover, if two signals are super-Gaussian, the couple that they form is globally super-Gaussian[6]. As stated in Sub-section

[4] The considered signals are thus processed as if they were stationary. They do not meet this requirement exactly from a theoretical point of view, but at least their non-stationarity is limited by the fact that they consist of relatively short periods of speech without silences. The proposed approach therefore contains an approximation. This approximation is accepted here because a simple feature extraction method is sought, at the expense of limited accuracy, as already mentioned at the end of Subsection 2.1 (see additional comments in Section 5).

[5] This version also allows us to compute autocorrelation functions only once, for each zero-mean source.

[6] If both signals are such that $\gamma > 0$, one easily derives that their couple meets (14).

2.2, such signals cannot be separated by the LISS algorithm considered here.

These general principles may be applied as follows to the three associated source signals $\Gamma_{x_j}(\tau)$ used here. Their normalized kurtosis[7] range from 6 to 13. All these signals are therefore strongly super-Gaussian, which explains why they are not separated in the tests considered here. This interpretation is confirmed by the results presented in the next sub-section.

3.3 Performance of the 1.3 cross-moment-based algorithm

The above discussion also shows that the LISS device based on 1.3 cross-moment cancellation is very well suited to the type of source signals considered here, unlike the previous version of this device. This is confirmed by the tests that we performed with this version, as the weights then converge close to their theoretical values.

More precisely, the relative difference between the observed and theoretical weight convergence values typically ranges from 4 to 25%. This non-negligible deviation results from the fact that the LISS device is designed so as to converge to a point corresponding to the cancellation of the 1.3 cross-moments of the zero-mean output signals, which is a bit different from the point corresponding to exact source separation in the specific case considered in this paper, since the 1.3 cross-moments of the source signals processed here are not exactly zero: the normalized cross-moments[8] corresponding to the three associated source signals $\Gamma_{x_j}(\tau)$ used here range from 0.1 to 0.2. These non-negligible moment values result from the specific nature of the source signals considered in this investigation, i.e. autocorrelation functions, which especially all contain a high peak at the origin. Anyway, the resulting weight errors and the associated source signal power errors are small enough to be acceptable in our target applications (such as those described in Subsection 5.1 and in [11]).

3.4 Performance of the 3.1 cumulant-based algorithm

Tests were then performed with the LISS device based on the 3.1 cumulants of the output signals. A constant gain was first used. The weights thus again converged close to their theoretical values. The difference between the observed and theoretical values typically ranges from 2 to 10%, which is lower than in the previous case. The version of this algorithm based on an adaptive gain results in a weight error ranging from 2 to 20%. These cumulant-based algorithms are studied in more detail in the next section, where their performance for real signal mixtures is analyzed.

3.5 Preliminary conclusions

At this stage of the investigation, the following preliminary conclusions may be drawn from the above-described tests:

[7] These kurtosis are computed for zero-mean versions of these signals, over the entire signals, i.e. again as if they were stationary.

[8] By "normalized" 1.3 cross-moments, we here mean a quantity which does not depend on the magnitudes of the considered u and v signals, i.e. the correlation coefficient of u and v^3 (the numerator of which is the non-normalized 1.3 cross-moment):

$$\frac{E\{uv^3\}}{\sqrt{E\{u^2\}E\{v^6\}}}. \tag{22}$$

It should be noted that the source signals u and v eventually considered in this paper are autocorrelation functions, i.e. second-order moments, so that the corresponding moments are "moments of moments".

- From a general point of view, the overall set of tests allowed us to validate the basic principles of the proposed approach on an example, as the weights of (some versions of) the LISS device actually converge close to their theoretical values.

- More precisely, the most classical LISS device, based on 3.1 cross-moments, cannot be used here, due to the specific nature of the signals to be processed. Other versions of this device are therefore required. As stated above, these versions actually succeed in separating the sources in the considered conditions, which is in agreement with their above-defined theoretical field of application.

- Among these successful versions, the 3.1 cumulant-based ones yield better performance than the 1.3 moment-based one and presumably avoid restrictions on the nature of the sources. Therefore, only these cumulant-based versions are considered below.

4 Performance with real mixtures

4.1 Goal and principles of the tests

The second series of tests aims at validating the proposed approach in completely realistic experimental conditions, i.e. with two microphones which are situated in a standard room and which receive signals emitted by two loudspeakers (Fig. 3)[9]. Two sets of source signals are considered: 1) two speech signals, then 2) a speech signal and a music signal[10].

Unlike in the previous section, the analysis is not based on comparing the experimental weight values to theoretical ones here, as the latter depend on the values of the mixture parameters, which are unknown here. Instead, the performance criterion used here is the one which matters in practical applications of the proposed approach, i.e. the relative error Δ_{ir} made in the estimation of the power of each zero-mean normalized source i, i.e:

$$\Delta_{1r} = \frac{\hat{\Gamma}_{e_{1A}}(0) - \Gamma_{e_{1A}}(0)}{\Gamma_{e_{1A}}(0)} \quad (23)$$

$$\Delta_{2r} = \frac{\hat{\Gamma}_{e_{2B}}(0) - \Gamma_{e_{2B}}(0)}{\Gamma_{e_{2B}}(0)}. \quad (24)$$

The quantities which appear in the above expressions are defined as follows:

- $\Gamma_{e_{1A}}(0)$ and $\Gamma_{e_{2B}}(0)$ are the powers of the zero-mean versions of the normalized source signals. These normalized source signals are denoted $E_{1A}(t)$ and $E_{2B}(t)$ hereafter and are obtained as follows. In a first measurement step, denoted A, only the source corresponding to loudspeaker 1 is turned on, and two resulting signals $E_{1A}(t)$ and $E_{2A}(t)$ are resp. measured by microphones 1 and 2. Similarly, in a subsequent measurement step, denoted B, only the source corresponding to loudspeaker 2 is turned on, and two signals $E_{1B}(t)$ and $E_{2B}(t)$ are resp. measured by microphones 1 and 2. Each overall mixed signal $E_i(t)$ corresponding to microphone i is then derived by superposing the two contributions which were previously obtained for

[9] Here again, the proposed approach is applied without splitting the signals in sub-bands.

[10] Means and autocorrelations are again computed over the entire signals, which last 10 seconds and contain no silences.

this microphone by turning on a single source[11]:

$$E_i(t) = E_{iA}(t) + E_{iB}(t). \tag{25}$$

- $\hat{\Gamma}_{e_{1A}}(0)$ and $\hat{\Gamma}_{e_{2B}}(0)$ are the estimates of the powers $\Gamma_{e_{1A}}(0)$ and $\Gamma_{e_{2B}}(0)$ provided by the LISS device during the "resolution phase", i.e. they are the output values of this device obtained when providing it with the inputs corresponding to $\tau = 0$, after adapting it and freezing its weights.

In addition to the above-defined final performance criterion Δ_{ir}, the magnitude of the improvement provided by the proposed approach is also analyzed hereafter. To this end, the errors Δ_{ir} are compared to the errors Δ_{it} which would be obtained by only using each (mixed, i.e. "noisy") microphone signal $E_i(t)$ as a rough estimate of the corresponding source:

$$\Delta_{1t} = \frac{\Gamma_{e_1}(0) - \Gamma_{e_{1A}}(0)}{\Gamma_{e_{1A}}(0)} \tag{26}$$

$$\Delta_{2t} = \frac{\Gamma_{e_2}(0) - \Gamma_{e_{2B}}(0)}{\Gamma_{e_{2B}}(0)}. \tag{27}$$

4.2 Performance of the constant-gain 3.1 cumulant-based algorithm

The results obtained with the constant-gain 3.1 cumulant-based algorithm are provided in Figures 4 and 5. These figures represent the variations of the errors Δ_{ir} vs the adaptation gain μ. They show that the source power estimation errors Δ_{ir} are lower than 4% in a range of adaptation gains which applies to all considered signals and which is relatively large (from 0.1 to 4.5).

This should be contrasted with the errors Δ_{it}, which would be obtained if not inserting the proposed system after the microphones: these errors range from 60% to 150%, which shows that the proposed approach yields a major performance improvement.

4.3 Performance of the adaptive-gain 3.1 cumulant-based algorithm

Figures 6 and 7 represent the results obtained with the adaptive-gain 3.1 cumulant-based algorithm. With this algorithm too, the normalized source powers are estimated with an accuracy better than 4% in a large range of parameter values. Moreover, this parameter range again applies to all considered signals.

These errors Δ_{ir} should be compared to the errors Δ_{it}, which are here the same as in Sub-section 4.2. This version of the approach therefore also yields a major performance improvement.

5 Applications and conclusion

5.1 Typical applications of the proposed approach

This subsection briefly describes a few typical potential applications of the proposed approach[12]. It aims at showing clearly the motivations of the investigation presented in this

[11] This method should be distinguished from the artificial mixtures used in the previous section: it only assumes that each microphone response is an additive function of the received sources, but it actually takes into account the convolutive transfer functions which exist between each source and each microphone in the considered experimental setup.

[12] For more details about these and other applications, the reader should refer to [11].

paper and at allowing us to then draw conclusions about the applicability of this approach in such applications. Generally speaking, this method is motivated by the fact that the following conditions are met in various products:

- Some sensors provide convolutively mixed signals.
- Some features related to each of the signals contained by these mixtures should be extracted. However, a CSS device cannot be used to this end, because its complexity and therefore its cost are too high for the considered product.

The solution to this problem therefore consists in using the system proposed in this paper, which is attractive because it is simpler than a CSS device. The proposed approach can only extract some source features, but these features make it possible to provide the considered product with interesting functions, such as those described below.

As stated above, various types of source features may thus be extracted. Hereafter, we first consider the case when these features are the "mean levels" of the (normalized) sources during various time periods, i.e. their average powers in successive time windows. The resulting applications especially include control functions in audio products. A first application of this type is an improved version of automatic volume control in car radios. A first generation of radios including a basic control function was recently launched. These radios contain a microphone, which measures the overall ambient sound level inside the car. The considered control unit increases the volume of the car radio when the ambient sound level increases. This control rule is motivated by the fact that when the car speed increases, the overall ambient sound level increases (due to the increase of the ambient noise level) and that the car radio volume should then be increased, so that the passengers may keep on listening to the car radio with the same comfort. However, the actual effect of this control rule is the opposite of the desired one when the passengers start talking: their speech is part of the overall ambient sound level, so that the car radio considers it like noise and tries to drown it, whereas it should reduce its level as long as the passengers keep on talking. In order to solve this problem, a specific control law should be used for each type of acoustic source present in the microphone signals (i.e. passengers, engine noise, wind noise ... and the car radio output itself, which should have no influence on the car radio volume). The approach proposed in this paper makes it possible to design such an improved control unit, as it provides estimates of each source level separately (or at least of the major sources to be taken into account, the microphones being placed close to these sources).

Similarly, the estimation of the levels of acoustic sources may be used to detect when a person is speaking in a microphone, while being insensitive to the ambient "noise" (this "noise" may include other people's speech). This allows to transmit the microphone signal of the considered person only when this person is actually speaking. This may be used to reduce the data rate of the transmitted signal and/or to improve the perceived comfort of the people who are listening to this signal, by not transmitting the noise periods. This approach may be used in hand-free car phones, during conversation but also to detect when the user is uttering the phone number to be called, in order to then trigger a voice recognition unit. It also allows to automatically switch between participants in audio control products for (single or multiple) conference rooms.

The proposed approach may also be used to extract the power spectral densities (PSD) of all the sources contained in microphone signals [11]. This especially makes it possible to extract the PSD of the signal received by a microphone from a car radio, or more

generally speaking from an audio product situated in a room or a car. Comparing the audio PSD in the microphone signal to the initial output PSD of the audio product would then allow to characterize the acoustic channel and the response of the loudspeaker, and to automatically control the tone of the audio product, so as to achieve an adequate acoustic frequency response in this room or car.

5.2 Conclusion

In this paper, we proposed a signal processing method for convolutively mixed sources, which aims at extracting specific features of the considered source signals with simple processing means. The principles and performance of the proposed approach were especially detailed in the case when the parameters to be extracted are estimated powers of the (normalized) source signals. The approach has been experimentally validated in this case. The power estimation errors thus achieved for real acoustic signals are lower than 4%, which is quite satisfactory for the considered applications of this approach. This investigation also highlighted some properties of several well-known LISS algorithms, especially related to the types of sources that they can separate (these results were summarized in Sub-section 3.5 and are therefore not repeated here).

Beyond the validation of the proposed approach that was thus achieved, various extensions of this investigation might be considered. They may especially consist in using a larger database, so as to check more extensively the estimation accuracy of the proposed approach and to extend the signal nature analysis which was presented above. These investigations should preferably be performed in connection with one of the target applications of the approach presented above and in [11]. Taking this application into account would have the following consequences:

- It would make it possible to define precisely in which conditions the mixed signals should be measured.
- It would result in splitting the signals into successive time windows and in using the proposed approach so as to estimate the mean source signal powers in these windows, as these successive powers are the parameters used to control real products.

Acknowledgements

The authors would like to thank the members of the TIRF laboratory at INP-Grenoble, and especially Mrs L. Nguyen and MM. J. Hérault and C. Jutten, for providing the speech signals used in the tests presented in Section 3 of this paper. They would also like to thank the following LEP members: J. Damour (for measuring the signals used in Section 4), N. Charkani, J.C. Boissy and L. Andry (for fruitful discussions about source separation).

References

[1] Jutten C. Hérault J. Blind separation of sources, Part I: An adaptive algorithm based on neuromimetic architecture. Signal Processing 1991; 24:1-10

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

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

[4] Jutten C, Nguyen Thi H.L. New algorithms for separation of sources. In: Congrès Satellite du congrès Européen de Mathématiques, Aspects Théoriques des Réseaux de Neurones. Paris. 1991

[5] Nguyen Thi H.L, Jutten C, Caelen J. Séparation aveugle de parole et de bruit dans un mélange convolutif. In: Treizième colloque GRETSI. Juan-Les-Pins. 1991. pp. 737-740

[6] Nguyen Thi H.L. Séparation aveugle de sources à large bande dans un mélange convolutif. Thesis, Institut National Polytechnique de Grenoble, Grenoble (France), 22 Jan. 1993

[7] Nguyen Thi H.L. Jutten C. Blind source separation for convolutive mixtures. Signal Processing 1995; 45:209-229

[8] Charkani N, Deville Y, Hérault J. Stability analysis and optimization of time-domain convolutive source separation algorithms. In: Viberg M, Cardoso J.-F. (eds). First IEEE Signal Processing Workshop on Signal Processing Advances in Wireless Communications (SPAWC '97). IEEE Press. Paris. 1997. pp. 73-77

[9] Charkani N, Deville Y. Optimization of the asymptotic performance of time-domain convolutive source separation algorithms. In: Verleysen M (ed). Fifth European Symposium on Artificial Neural Networks (ESANN'97). D Facto Publications. Bruges. 1997. pp. 273-278

[10] Deville Y, Charkani N. Analysis of the stability of time-domain source separation algorithms for convolutively mixed signals. In: 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 97). IEEE Press. Munich. 1997. pp. 1835-1838

[11] Deville Y. Système de caractérisation de sources de signaux. French Patent, no. 94 03 078, filed on March 16, 1994

[12] Hérault J, Jutten C. Space or time adaptive signal processing by neural network models. In: Neural Networks for Computing. American Institute of Physics. Snowbird. 1986. pp. 206-211

[13] Jutten C. Calcul neuromimétique et traitement du signal; analyse en composantes indépendantes. Thesis, Grenoble 1987

[14] Jutten C. Hérault J. Une solution neuromimétique au problème de séparation de sources. Traitement du Signal 1988; 5:389-403

[15] A. Papoulis, Probability, random variables, and stochastic processes, McGraw-Hill, Singapore, 1984.

[16] Sorouchyari E. Blind separation of sources, Part III: Stability analysis. Signal Processing 1991; 24:21-29

[17] Fort J.C. Stabilité de l'algorithme de séparation de sources de Jutten et Hérault. Traitement du Signal 1991; 8:35-42

[18] Deville Y. A unified stability analysis of the Hérault-Jutten source separation neural network. Signal Processing 1996; 51:229-233

[19] Deville Y. Andry L. Application of blind source separation techniques to multi-tag contactless identification systems. IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences 1996; E79-A:1694-1699

[20] Deville Y, Andry L. Application of blind source separation techniques to multi-tag contactless identification systems. In: 1995 International Symposium on Nonlinear Theory and Its Applications (NOLTA '95). vol 1. Research Society of Nonlinear Theory and its Applications, IEICE. Las Vegas. 1995. pp. 73-78

[21] Nikias C. L. Mendel J. M. Signal processing with higher-order spectra. IEEE Signal Processing Magazine 1993; July:10-37

List of Figures

1 Basic configuration for the LISS problem and solution proposed by Hérault and Jutten. 18
2 Proposed processing method. 18
3 Experimental configuration for the tests performed with real signals. $D_1 =$ 1 m, $D_2 = 10$ cm, $D_3 = 1$ m. 18
4 Considered conditions : 3.1 cumulant-based source separation algorithm with a constant adaptation gain μ; real mixtures of two speech sources. Results shown: variations of the errors Δ_{ir} vs adaptation gain μ. 19
5 Same principle as Figure 4 for real speech/music mixtures. 19
6 Considered conditions: adaptive-gain 3.1 cumulant-based source separation algorithm; the parameter d of the rule used to adapt the gain is constant ($d = 0.1$), the parameter ϵ of this rule is varied (these parameters are defined in [4]); real mixtures of two speech sources. Results shown: variations of the errors Δ_{ir} vs parameter ϵ. 20
7 Same principle as Figure 6 for real speech/music mixtures. 20

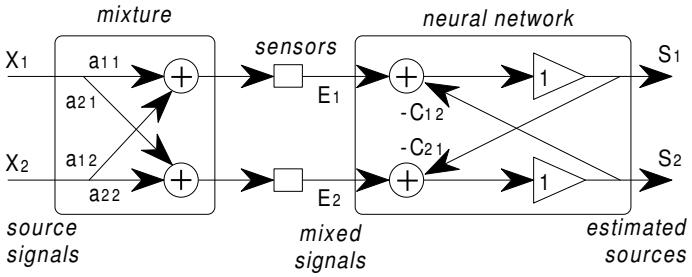

Figure 1: Basic configuration for the LISS problem and solution proposed by Hérault and Jutten.

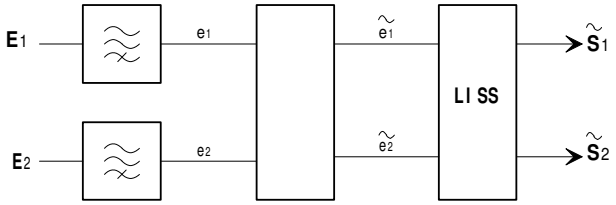

Figure 2: Proposed processing method.

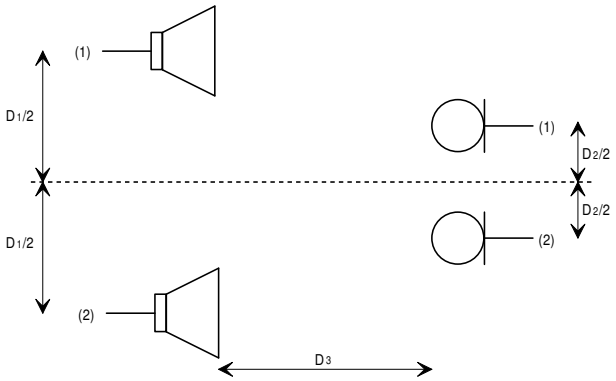

Figure 3: Experimental configuration for the tests performed with real signals. $D_1 = 1$ m, $D_2 = 10$ cm, $D_3 = 1$ m.

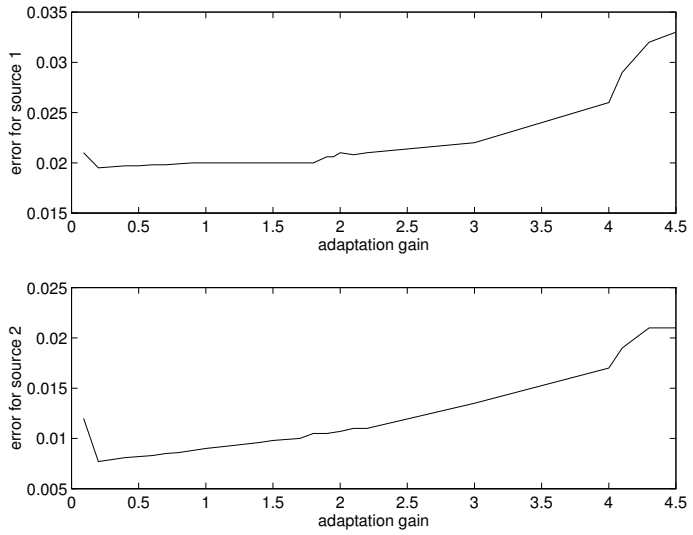

Figure 4: Considered conditions : 3.1 cumulant-based source separation algorithm with a constant adaptation gain μ; real mixtures of two speech sources. Results shown: variations of the errors Δ_{ir} vs adaptation gain μ.

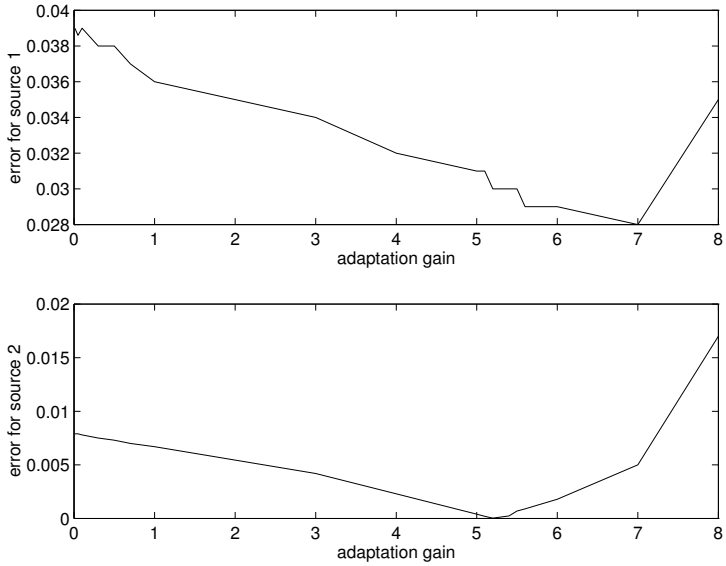

Figure 5: Same principle as Figure 4 for real speech/music mixtures.

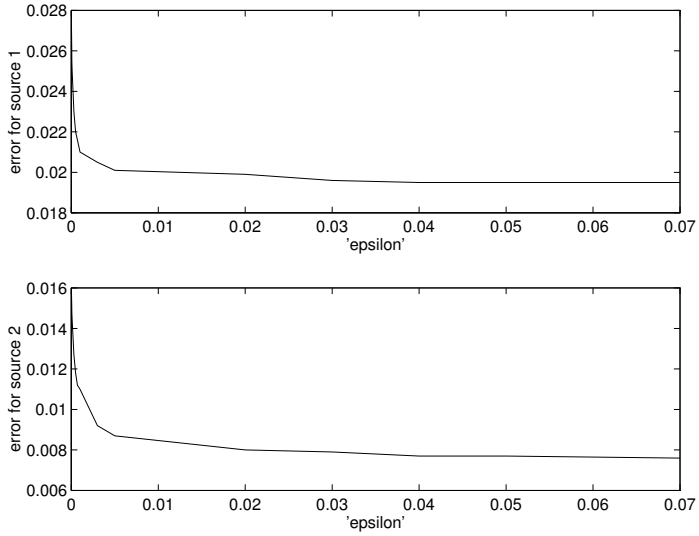

Figure 6: Considered conditions: adaptive-gain 3.1 cumulant-based source separation algorithm; the parameter d of the rule used to adapt the gain is constant ($d = 0.1$), the parameter ϵ of this rule is varied (these parameters are defined in [4]); real mixtures of two speech sources. Results shown: variations of the errors Δ_{ir} vs parameter ϵ.

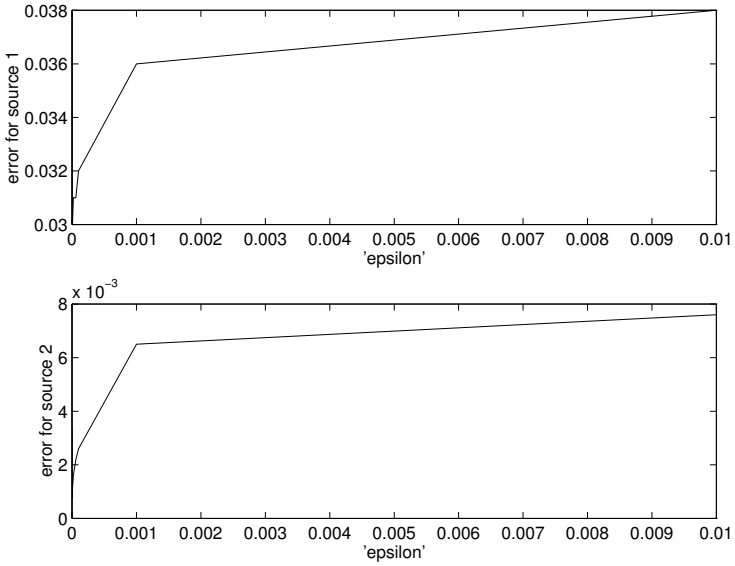

Figure 7: Same principle as Figure 6 for real speech/music mixtures.