POST-NONLINEAR SPEECH MIXTURE IDENTIFICATION USING SINGLE-SOURCE TEMPORAL ZONES & CURVE CLUSTERING

Matthieu Puigt[1], *Anthony Griffin*[1], *and Athanasios Mouchtaris*[1,2]

[1]Foundation for Research and Technology – Hellas , Institute of Computer Science, Heraklion, Crete, Greece, GR-70013
[2] University of Crete, Department of Computer Science, Heraklion, Crete, Greece, GR-71409
email: {mpuigt,agriffin,mouchtar}@ics.forth.gr

ABSTRACT

In this paper, we propose a method for estimating the nonlinearities which hold in post-nonlinear source separation. In particular and contrary to the state-of-art methods, our proposed approach uses a weak joint-sparsity sources assumption: we look for tiny temporal zones where only one source is active. This method is well suited to non-stationary signals such as speech. The main novelty of our work consists of using nonlinear single-source confidence measures and curve clustering. Such an approach may be seen as an extension of linear instantaneous sparse component analysis to post-nonlinear mixtures. The performance of the approach is illustrated with some tests showing that the nonlinear functions are estimated accurately, with mean square errors around 4e-5 when the sources are "strongly" mixed.

1. INTRODUCTION

Blind Source Separation (BSS) consists of estimating a set of N unknown source signals s_j from a set of P observations x_i resulting from mixtures of these sources through unknown propagation channels [1]. Among all the proposed approaches, the ones based on sources joint-sparsity, known under the name of Sparse Component Analysis (SCA) methods, have met with great interest from the community in the last decade (see e.g. Chap. 10 of [1]). Indeed, they are naturally adapted to stationary, non-stationary and/or dependent signals and are thus an alternative to classical Independent Component Analysis (ICA) approaches which assume source mutual independence. Moreover, they allow processing of the underdetermined case where $N > P$.

Most of the SCA approaches have been proposed for linear mixtures, i.e. linear instantaneous (LI), anechoic or convolutive mixtures. While many methods assume the sources to be (approximately) W-disjoint orthogonal (WDO) in an analysis domain[1] [2], several other methods highly relax this assumption, by looking for "single-source zones" (i.e. zones where one source is dominant over the others) [3–6]. Interestingly, while many SCA methods have been proposed for linear mixtures, only a few sparsity-based methods process nonlinear configurations [7–10]. In [7, 8], the authors consider post-nonlinear (PNL) mixtures (i.e. a special configuration where linear mixes of sources are distorted by a function which models data acquisition/sensor nonlinearities, such as saturation), and assume the sources to be approximately WDO[2]. Unfortunately, these approaches are not tested with real source signals, mainly because of the strong sparsity assumption. In [9, 10], the authors extend the measures for finding single-source zones to other classes of nonlinear mixtures but restrict their approach to overdetermined or determined mixtures.

This work is funded by the Marie Curie IAPP "AVID MODE" grant within the 7th European Community Framework Program.

[1]The WDO assumption means that in each atom of an analysis domain (time, time-frequency, time-scale domain), at most one source is non-zero.

[2]Actually, in [7], the authors assume the sources to be $(P-1)$-sparse, which is equivalent to WDO if $P = 2$. In [8], the approximate WDO is not explicitly assumed but is needed by authors and satisfied in their tests.

In this paper, we propose an approach for identifying PNL mixtures based on single-source zones, as in [9,10], and which possibly processes the underdetermined case, as in [7, 8]. We thus avoid the strong source joint-sparsity assumption of [7, 8] while treating the same class of mixtures and applying our approach to mixtures of real speech signals. We only focus on the identification of mixtures and not on the whole separation. Indeed, it is proved that a good estimation of the nonlinearities is crucial for a good separation (see e.g. Chapter 14 of [1]), hence the importance of an accurate approach. Once the identification of the mixtures is done, the separation is straightforward, e.g. using methods proposed in [7, 8]. The main novelty of the paper consists of combining single-source zones and curve clustering and may be seen as a way for extending linear SCA [3–6] to PNL-SCA.

The remainder of the paper is structured as follows: in Section 2, we describe the considered BSS problem. We then introduce our proposed method in Section 3. Section 4 provides an experimental validation of the approach while we conclude and discuss future directions of the incoming work in Section 5.

2. PROBLEM STATEMENT

In this paper, we assume that N real source signals $\underline{s}(n) = [s_1(n), \ldots, s_N(n)]^T$ are mixed by a LI unknown $P \times N$ mixing matrix A, thus providing a set of linearly mixed signals

$$\underline{z}(n) = A\underline{s}(n), \tag{1}$$

to which a nonlinear componentwise vector mapping $\underline{f} = [f_1, \ldots, f_P]^T$, assumed to be invertible, is applied. It can e.g. model data acquisition/sensor nonlinearities such as saturation. Such a situation e.g. arises in audio processing with small and cheap microphones in a mobile device. Observed signals $\underline{x}(n)$ thus read

$$\underline{x}(n) = \underline{f}(\underline{z}(n)) = \underline{f}(A\underline{s}(n)). \tag{2}$$

We aim to estimate the source signals $\underline{s}(n)$, up to a scale coefficient/permutation indeterminacy. This means that we want to suppress or reduce as much as possible the distortions introduced by the nonlinear mappings f_i. For that purpose, we use a separating structure which is the *mirror* of the mixing one (see e.g. Chap.14 of [1]): we first have to estimate g_i, the inverse of the nonlinear mappings f_i, and to apply them to the observations. We then obtain a linear problem comparable to (1) and we process a linear SCA approach to estimate the sources. The global mixing and separating structure is drawn in Fig. 1. The proposed separating structure may be summarised as follows:

1. We first look for temporal zones where one source is dominant over the others (See Section 3.3).
2. We then estimate the nonlinear mappings f_i (See Sections 3.2 and 3.4)
3. We then invert the nonlinearities and get a LI-BSS problem, that we solve using a LI-SCA approach (see Section 3.5).

Let us emphasize that we here focus on the first two above items and that we will not perform the inversion of the nonlinearities.

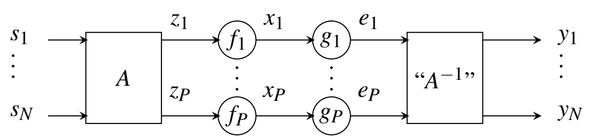

Figure 1: PNL Mixing-separating structure.

3. PROPOSED APPROACH

3.1 Definitions and assumptions

We first present the only assumptions that we make in this approach, and the associated definitions.

Definition 1 ([7]) *Let $A = [a_{ij}]_{i,j}$ be an $P \times N$ matrix. Then A is said to be "mixing" if A has at least 2 nonzero entries in each row. And A is said to be "absolutely degenerate" if there are two normalised columns a_{ik} and a_{il} with $k \neq l$ such that $a_{ik}^2 = a_{il}^2$, i.e. a_{ik} and a_{il} differ only by the sign of the entries.*

Assumption 1 (Mixing assumptions) *1. A has nonzero entries on the first row and at least one nonzero entry in each other row. We cannot find two collinear vectors a_{ik}^2 and a_{il}^2, with $k \neq l$.*
2. In the underdetermined case when $N > P$, every $P \times P$-submatrix of A is invertible.
3. We also assume that, for each index i, $f_i(0) = 0$.

Assumption 1.1 is needed for the following reasons [7]: if A is not "mixing" (according to Definition 1), then this means that there is an index i such that the i-th row of A contains only one non-zero element and consequently f_i cannot be identified. As an extreme case, let us imagine that A is diagonal (up to a permutation in its columns order), then each observed signal reads

$$x_i(n) = f_i(a_{ik}s_k(n)), \tag{3}$$

i.e. we already get separated signals for which we cannot do nothing more without extra information. If A is absolutely degenerate, it can be estimated but the nonlinear mappings cannot [7]. Assumption 1.2 is a classical assumption in underdetermined BSS. This means that locally, if only P sources are active, we get a determined BSS problem which needs to be separable. Lastly, Assumption 1.3 is not limiting for practical applications and is shared by many PNL-BSS approaches.

Definition 2 *A "temporal analysis zone" is a set T of M adjacent samples $T = \{n_1, \ldots, n_M\}$ in the time domain.*

Definition 3 *A source is said to be "isolated" in a temporal analysis zone T if only this source (among all considered mixed sources) has a nonzero variance in this zone. We then say that this zone is "single-source".*

This definition corresponds to the theoretical point of view. From a practical point of view, this means that the variances of all other sources are negligible with respect to the variance of the source that is isolated.

Definition 4 *A source is said to be "accessible" in the time domain if there exist at least one temporal analysis zone where it is isolated.*

Assumption 2 (Source assumptions) *1. Source signals are mutually independent.*
2. At least P sources are accessible in the time domain.
3. By considering several *single-source zones associated with the same source, the amplitude of the observations spans a "wide" range allowing the estimation of the nonlinear functions f_i.*

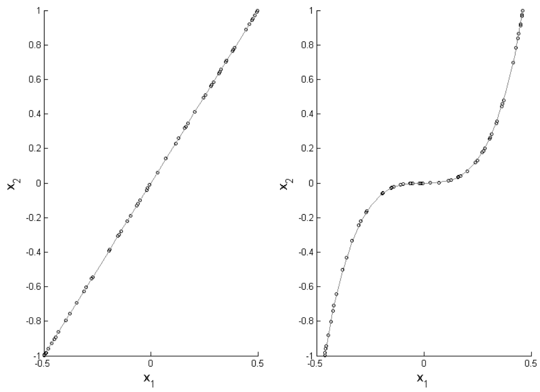

Figure 2: Scatter plots of observations on a single-source zone and theoretical curves. *Left:* LI mixtures. *Right:* PNL mixtures.

Note that, contrary to linear BSS methods [5] which needed source linear independence, we here need source mutual independence. This is due to the more complex mixing model, as we will see in Section 3.3. We need P isolated sources in order to be able to invert the P nonlinear functions f_i [7]. Assumption 2.3 is needed because we want to estimate the nonlinear mappings f_i on their whole domains. In the case when we should be able to estimate the functions f_i on a subset of their domain, the whole estimation might be coarse, thus yielding a poor separation quality.

3.2 Geometrical behavior

Before introducing the proposed method, let us focus on the intuitive idea behind it. In LI-SCA methods, the common main principle can be seen from a geometrical point of view. Let us first recall that in that case, the mapping $\underline{f}$ in Eq. (2) is set to the identity function, i.e. the observations $\underline{x}(n)$ are equal to the signals $\underline{z}(n)$ defined in Eq. (1). If one source, say s_k, is isolated in several adjacent samples $T = \{n_1, \ldots, n_M\}$, the local scatter plot of observations is concentrated in a *line*, whose slope is given by the mixing column of A associated with this source (see Fig. 2 for an example with $P = 2$ observations). Indeed, in that case:

$$x_i(n) = a_{ik}s_k(n) \quad \forall i \in \{1, \ldots, P\}, \forall n \in T \tag{4}$$

and we get the following relationship between observations x_1 and x_i:

$$a_{ik}x_1(n) - a_{1k}x_i(n) = 0 \quad \forall i \in \{1, \ldots, P\}, \forall n \in T. \tag{5}$$

In [3–6], authors propose to find such single-source zones by means of a "single-source confidence measure" (resp. based on ratios of observations [3], correlation [4, 5], and local PCA [6]). In these zones, they estimate the mixing parameters. However, several zones may lead to the same mixing column and they thus have to discard the multiple estimations, e.g. using a threshold distance between mixing columns [3, 4] or by clustering estimated scale coefficients [5, 6].

For PNL mixtures, in the same case when only one source, say s_k, is isolated in an analysis zone T, observations (2) read:

$$x_i(n) = f_i\left(a_{ik}s_k(n)\right) \quad \forall i \in \{1, \ldots, P\}, \forall n \in T \tag{6}$$

and we obtain, assuming that $a_{ik} \neq 0$,

$$s_k(n) = \frac{f_i^{-1}\left(x_i(n)\right)}{a_{ik}} \quad \forall i \in \{1, \ldots, P\}, \forall n \in T. \tag{7}$$

We thus have the following relationship between observations x_1 and x_i:

$$\begin{aligned} x_i(n) &= f_i\left(\frac{a_{ik}}{a_{1k}} f_1^{-1}\left(x_1(n)\right)\right) && (8) \\ &= \phi_{ik}(x_1(n)) \quad \forall i \in \{1, \ldots, P\}, \forall n \in T, && (9) \end{aligned}$$

where the functions ϕ_{ik} are defined as:

$$\phi_{ik}(t) = f_i\left(\frac{a_{ik}}{a_{1k}} f_1^{-1}(t)\right) \tag{10}$$

The main difference with respect to LI mixtures is that we now do not have to estimate a slope which describes a line, but a *function* which describes the nonlinear dependencies between observations. This can e.g. be done by interpolation or approximation. As different analysis zones might be associated with the same source, we propose to cluster these estimated functions. Indeed, clustering usually provides more robustness to noise. Moreover, and contrary to LI mixtures, the accuracy of the curve fitting also depends on the range of $x_1(n)$ in the local scatter plot. In particular, we need our observed points $(x_1(n), x_i(n))$ to span the whole curve. This may not be guaranteed in a unique single-source analysis zone associated with one source, but may happen if we interpolate the function on *several* zones associated with the same source. Merging single-source zones associated with the same source may be done by *curve clustering* and we will present such an approach in Section 3.4.

3.3 Nonlinear single-source confidence measures

We now detail how to find single-source zones in PNL mixtures. Let us first go back to the simpler LI problem. If a source, say s_k is isolated, then Eq. (4) holds and we can see that the observations are proportional. A way to find such zones consists of estimating the correlation coefficient of a pair of observations [4, 5]. Indeed, this coefficient is equal to 1 in absolute value when one source is isolated and is much lower otherwise.

In the considered PNL mixture, we need to measure the *nonlinear* correlation between observations. Mutual Information $\mathscr{I}(\underline{x})$, defined as:

$$\mathscr{I}(\underline{x}) = -\mathbb{E}\left\{\log \frac{\prod_{i=1}^{P} \mathbb{P}_{x_i}(x_i)}{\mathbb{P}_{\underline{x}}(\underline{x})}\right\}, \tag{11}$$

where $\mathbb{E}\{.\}$ stands for expectation, $\mathbb{P}_{\underline{x}}$ and $\mathbb{P}_{x_i}$ ($i \in \{1, \ldots, P\}$) are resp. the joint and marginal probability density functions of the observations, provides such a measure [11]: it takes null values when variables are independent and much higher values otherwise. However, if we want this measure to have the same behavior as linear correlation, we need to normalise it, which is classically done as follows:

$$\mathscr{I}_{\text{norm}}(\underline{x}) = \sqrt{1 - e^{-2\mathscr{I}(\underline{x})}}. \tag{12}$$

This measure has also been used in [10], for another class of nonlinear mixtures and we use it in the same way as [10]: a source is isolated in an analysis zone iff $\mathscr{I}_{\text{norm}}(\underline{x}) = 1$. We thus only consider the analysis zones which maximise Eq. (12).

However, a problem may appear if in a zone T, sources $s_j(n)$ are constant, whose value is denoted $\overline{s_j} \neq 0$ [3]: in that case, we still have $\mathscr{I}_{\text{norm}}(\underline{x}) = 1$ but Eq. (6) then becomes

$$x_i(n) = f_i\left(a_{ik} s_k(n) + \alpha_i(T)\right) \qquad \forall i \in \{1, \ldots, P\}, \tag{13}$$

where $\alpha_i(T) = \sum_{j \neq k} a_{ij} \overline{s_j}$. Eq. (7) and (8) then resp. read

$$s_k(n) = \frac{f_i^{-1}\left(x_i(n) - \alpha_i(T)\right)}{a_{ik}} \qquad \forall i \in \{1, \ldots, P\} \tag{14}$$

and

$$x_i(n) = f_i\left(\frac{a_{ik}}{a_{1k}}\left(f_1^{-1}\left(x_1(n)\right)\right) - a_{ik}\alpha_1(T) + \alpha_i(T)\right). \tag{15}$$

Let us recall that we are looking for zones where all the constant coefficients $\alpha_i(T)$ are zero. As we are applying the proposed approach to speech signals, the situation when one can find an index i such that $\alpha_i(T) \neq 0$ should not occur. Additionally, due to Assumption 1.3, we know the value of each nonlinear function f_i is zero at zero and we can estimate $\widehat{\phi_{ik}}$, the nonlinear relationship between observations defined in Eq. (10) (see Section 3.4) and discard the zones where $\widehat{\phi_{ik}}(0) \neq 0$.

Finally, we look for analysis zones which satisfy:

$$\begin{cases} \mathscr{I}_{\text{norm}}(\underline{x}) & > & 1 - \varepsilon_1 \\ \left|\widehat{\phi_{ik}}(0)\right| & < & \varepsilon_2 \qquad \forall i \in \{2, \ldots, P\} \end{cases}, \tag{16}$$

where ε_1 and ε_2 are user-defined thresholds.

3.4 Curve clustering

If we consider all the single-source analysis zones, then we get a subset of the original observations where the approximate WDO assumption holds. We thus can use the clustering techniques proposed in [7, 8]. However, in [7], the authors propose a geometrical preprocessing which is not robust to noise in general and in particular to *non-ideal* single-source zones[4]. On the other hand, [8] proposes the use of a spectral clustering technique in order to separate the curves (and so the sources). Spectral clustering techniques are well suited to nonlinearities in the data and are more robust to noise than the approach proposed in [7]. However, such techniques are sensitive to the distance between the curves and do not allow the clusters to overlap. This last criterion is obviously not satisfied in the BSS framework and the authors of [8] proposed a solution: they remove the points of $\underline{x}$ which are close to zero, and try to find $2N$ clusters. By assuming that the nonlinear mappings are almost linear for the lowest values of $\underline{x}$, they link the half-clusters.

In this paper, we propose to take advantage of the single-source analysis zones to cluster our data. Indeed, in each single-source zone, as we saw above (see Fig. 2), all the data points belong to the same curve and give us extra information which is not provided in [7, 8]. We thus can cluster the data according to these zones. The underlying idea can be seen as an extension of scale-coefficients clustering in [5, 6] to nonlinear mixtures: while the linear relationships between observations were limited to scale coefficients to be clustered, we here have to cluster the whole curves, i.e. to estimate some parameters adequately describing the functions ϕ_{ik} defined in Eq. (10) to realise a cluster. Such techniques are named curve clustering [12] which we summarise as follows in the framework of our problem.

Given an interval $[x_1(n_b), x_1(n_e)]$, we define a subdivision $\xi_0 = x_1(n_b) \leq \xi_1 \leq \xi_2 \leq \ldots \leq \xi_K \leq \xi_{K+1} = x_1(n_e)$. The points ξ_l are named *knots*. Note that a same knot may be repeated several times, say p times. We then say that it is a multiple knot of *multiplicity order p*. We aim to fit the curve $\{(x_1(n_j), x_i(n_j)\}_{j=1,\ldots,M}$ on such interval by using splines. A spline is a polynomial of degree d (or order $d+1$) on any interval $[\xi_{l-1}, \xi_l)$ which has $d+1$ continuous derivatives on the open interval $[x_1(n_b), x_1(n_e))$. For a fixed sequence of knots, the set of such splines is a linear space of functions with $K+d+1$ free parameters. A useful basis $(B_{1,d}, \ldots, B_{K+d+1,d})$ for this linear space is given by B-splines, recursively defined as

$$\begin{cases} B_{l,0}(t) & = & \begin{cases} 1 & \text{if } \xi_l \leq t < \xi_{l+1} \\ 0 & \text{otherwise} \end{cases}, \\ B_{l,u}(t) & = & \frac{t - \xi_l}{\xi_{l+u} - \xi_l} B_{l,u-1}(t) + \frac{\xi_{l+u+1} - t}{\xi_{l+u+1} - \xi_{l+1}} B_{l+1,u-1}(t). \end{cases} . \tag{17}$$

A spline, denoted $\zeta(x_1, \underline{\beta})$ hereafter, can now be written with respect to the above basis:

$$\zeta(x_1, \underline{\beta}) = \sum_{l=1}^{K+d+1} \beta_l B_{l,d}(x_1), \tag{18}$$

[3] Such a scenario is not a problem in LI-SCA: observations can be locally centred in each analysis zone, thus zeroing the constant signals [4].

[4] An ideal single-source zone is an analysis zone where all the sources except one are *exactly* zero.

where $\underline{\beta} = [\beta_1, \ldots, \beta_{K+d+1}]^T$ are the spline coefficients. For a set of fixed knots, coefficients $\underline{\beta}$ may be estimated as a linear problem. Let $\{(x_1(n_j), x_i(n_j))\}_{j=1,\ldots,M}$ be a regression type data set of M measurements of the curve ϕ_{ik} defined in Eq. (10), ranging over $[x_1(n_e), x_1(n_b)] \times \mathbb{R}$. The spline coefficients are estimated as:

$$\widehat{\beta}_l = \arg\min_{\beta_l} \frac{1}{M} \sum_{j=1}^{M} \left(x_i(n_j) - \zeta(x_1(n_j), \underline{\beta}) \right)^2 \tag{19}$$

The main interest of B-splines is due to the fact that the B-spline basis is only knot-dependent. Once the knots are fixed, the estimated coefficients $\widehat{\beta}_l$ describe the curve shape. If we use the *same* knots for *all* the single-source analysis zones selected in Section 3.3, then all the $K+d+1$ B-spline coefficients have the same meaning and may be compared. If two curves have close estimated B-splines coefficients, then they should be associated with the same source. Otherwise, they should be associated with different sources. Clustering techniques can be applied to such coefficients [12]: while the authors of [12] used K-means to this end, we propose to use the median-based version of K-means, named K-medians which has been used in [5]. Other approaches, such as DEMIX [6], may also be employed.

Once the classification is performed, we group all the curves associated with the same source s_k and we estimate each function ϕ_{ik} as a spline with optimised knot locations and multiplicity order.

3.5 Nonlinear inversion and linear identification

Once the nonlinear functions are estimated, we have to invert them and apply them to the observations $\underline{x}(n)$, in order to get the linearly mixed signals $\underline{e}(n) = [e_1(n), \ldots, e_P(n)]^T$ (see Fig. 1). This is straightforward by e.g. applying one of the neural-network-based methods, proposed in [7, 8], which are based on the same property but are differently implemented. The underlying common property was first defined for PNL-ICA methods and is adapted to PNL-SCA as follows: we estimate a nonlinear mapping $\underline{g} = [g_1, \ldots, g_P]^T$ such that for all indices i, k, $g_i \circ \phi_{ik}$ is linear. To this end, [7] proposes finding a linear relationship between the same components of different clusters while [8] suggests finding a linear relationship between different components in the same cluster.

Once the nonlinearities are inverted, we thus obtain a classical LI-SCA problem. The estimation of the linear mixing parameters is then straightforward if we estimated the N nonlinear curves: once we have linearised the clusters obtained in Section 3.4, each cluster fits a line whose parameters, defined in Eq. (5), may be estimated using a criterion proposed in [3–6].

If we estimated more than P curves but less than N, we are still able to invert the nonlinearities. However, we now do not have all the linearised curves and we will thus have to estimate the linear mixing parameters thanks to a *whole* linear SCA approach, and probably by first applying a linear sparsifying transform to $\underline{e}(n)$, in order to find a zone associated with each source.

Let us remind the reader that the main novelty of our proposed approach consists of the first stages of the whole approach, and that we only test these first stages in Section 4.

4. TESTS

In this section, we test the performance of our proposed approach on PNL mixtures of speech. Indeed, and contrary to [7, 8], we test our approach on simulations made using real speech signals which can be locally sparse in the time domain, due to silence of speakers [4].

We first employ a configuration involving $N = 3$ sources and $P = 2$ sensors, i.e. an underdetermined mixture. The source signals were three speech signals, sampled at 20 kHz, which last 5 s and contain silent parts. These sources are presented in Fig. 3.

We then mixed them with the following mixing matrix:

$$A = \begin{bmatrix} 1 & 1 & 0.9 \\ -0.9 & 0.5 & 1 \end{bmatrix} \tag{20}$$

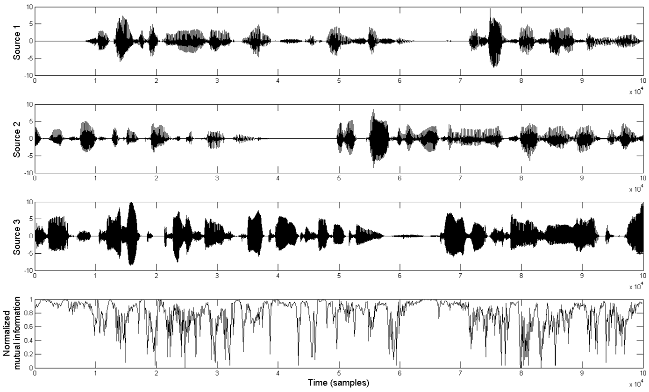

Figure 3: Normalised mutual information measures between two PNL mixtures (below) of three speech sources (plotted above).

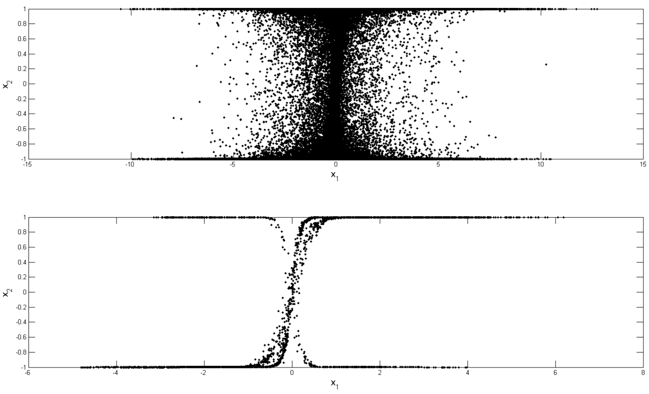

Figure 4: Scatter plots between observations. *Top:* on the full signals. *Bottom:* on the single-source analysis zones.

and then applied the following nonlinear mappings to the resulting signals $\underline{z}(t)$:

$$\left\{ \begin{array}{lcl} f_1(t) & = & \tanh(t) + t \\ f_2(t) & = & \tanh(10t) \end{array} \right. . \tag{21}$$

Please note that the mixing matrix A is close to being an absolutely degenerate matrix, and thus the configuration under consideration is challenging. Moreover, the nonlinear functions have been chosen so that they can model audio effects like *soft-clipping*.

We set the size of our temporal analysis zones to 100 samples, which is long enough to guarantee the independence of the source signals [13]. Mutual information was estimated using the approach in [14]. Fig. 3 shows the plot of speech sources and the obtained normalised mutual information measures. We can see that $\mathscr{I}_n(\underline{x})$ is close to 1 when one source is isolated. We then considered all the zones where ε_1 and ε_2, defined in Eq. (16), are resp. set to 0.01 and 0.1. Fig. 4 shows two scatter plots: on the top, we provided the scatter plot of the original observations. It is clear that the sparsity assumptions needed in [7, 8] are not satisfied at all. In the bottom, we only plotted the scatter plots obtained from zones satisfying Eq. (16). Now, we can clearly see three curves associated with nonlinearities. This thus shows the relevance of the single-source confidence measures and an easy way to improve [7, 8]. We then estimated the different curves on the local scatter plots using B-spline approximations. Because the choice of the knots is data-dependent, we decided to perform a "coarse" fitting, i.e. an approximation whose knot locations and B-spline order are not necessarily optimised but that allow us to separate the curves of the functions ϕ_{ik} defined in Eq. (10). In the example provided here, we used the

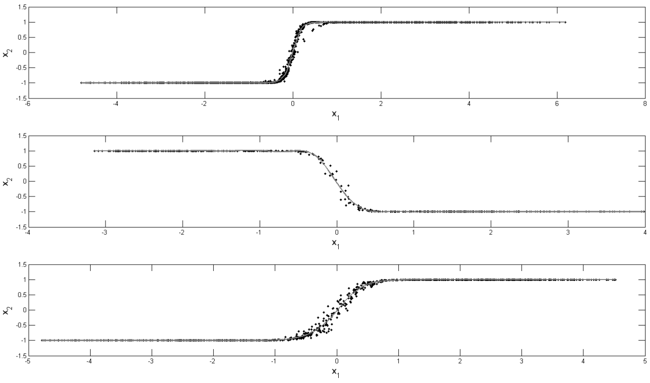

Figure 5: Scatter plots of the separated curves after the clustering procedure (in black) and estimated nonlinear functions (in gray).

λ	0.1	0.5	0.9
Mean MSE	6.7e-2	4.5e-5	4.6e-5
Std.	2.3e-1	5.9e-5	1.0e-4

Table 1: MSE and standard deviation, vs. the value of λ

following knots $\xi_i = -1.5 + 0.3i$ for $i \in \{0, \ldots, 10\}$, without multiplicity order. We set the degree of the B-spline to 4, in order to obtain smooth estimates of the curves. We then obtain the B-spline coefficients $\underline{\widehat{\beta}}$ that we then cluster using K-medians. Fig. 5 shows the separated curves obtained after classification, i.e. the superposition of the local scatter plots on the zones belonging to the same cluster. Such separated curves then allow us to estimate the inverse nonlinear mappings.

We do this using B-splines functions of order 4, with 20 knots. The obtained curves, drawn in Fig. 5, fit very well the scatter plots and are really close to the theoretical ones, as the mean square errors (MSEs) on the sampled curves are resp. set to 2.5e-4, 5.3e-5, and 2.1e-5.

Lastly, we generated a series of determined mixture simulations with $P = N = 2$ sources and sensors: the mixing matrix was set to $A = \begin{bmatrix} 1 & \lambda \\ -\lambda & 1 \end{bmatrix}$, with $\lambda = 0.9$, 0.5, and 0.1, and the nonlinear functions were set as in Eq. (21). We chose 49 pairs of sources among 8 real male and female speech sources, sampled at 20 kHz, lasting 5 s and containing silent parts. We fixed the values of all the parameters as in the previous underdetermined exampled. We aimed to measure the influence of the mixing parameters on the global quality of estimation of the nonlinear mappings. Our proposed approach succeeded in estimating the nonlinear functions in all the tests. The average MSEs and their associated standard deviations are given in Tab. 1. The obtained performance shows a really good accuracy of estimation of the nonlinear mappings, except when the sources are "weakly" mixed, which is consistent with the performance of other PNL-ICA methods (see Chap. 14 of [1]) and the one obtained in [4] for an LI-SCA method.

5. CONCLUSION AND FUTURE WORK

In this paper, we introduced a PNL mixture identification method which uses weak sparsity assumptions in order to estimate the mixing parameters. The main novelty of the proposed method is that we combine single-source confidence measures with curve clustering using B-splines and classical clustering techniques. The proposed approach thus improves on the previously proposed PNL-SCA methods which assume strong joint-sparsity assumptions and cannot be applied to real signals. We presented some tests showing the performance and the interest of our approach. In future work, we will also test the influence of the knots in the clustering procedure. We will propose an approach for inverting the nonlinearities. Indeed, we focused on the estimation of the nonlinear functions, whose accuracy has a direct consequence on the final separation. We could use the inversion approaches presented in [7, 8] but both use neural networks, whose drawbacks are the speed of convergence and the possibility to reach a local optimum. Another future direction will consist of investigating sparsifying transforms well-suited to nonlinear mixtures. Indeed, while the proposed approach may be applied to speech signals, the required sparsity assumption is not met with music signals. Lastly, we will extend our approach to other nonlinear mixtures like PNL convolutive mixtures.

REFERENCES

[1] P. Comon and C. Jutten, *Handbook of blind source separation. Independent component analysis and applications*. Academic press, 2010

[2] O. Yilmaz and S. Rickard, "Blind separation of speech mixtures via time-frequency masking", *IEEE Transactions on Signal Processing*, Vol. 52, No. 7, pp. 1830–1847, 2004.

[3] F. Abrard and Y. Deville, "A time-frequency blind signal separation method applicable to underdetermined mixtures of dependent sources", *Signal Processing*, Vol. 85, No. 7, pp. 1389–1403, 2005.

[4] Y. Deville and M. Puigt, "Temporal and time-frequency correlation-based blind source separation methods. Part I: determined and underdetermined linear instantaneous mixtures", *Signal Processing*, Vol. 87, No. 3, pp. 374–407, Mar. 2007.

[5] M. Puigt and Y. Deville, "Iterative-shift cluster-based time-frequency BSS for fractional-time-delay mixtures, " in *Proc. of ICA 2009*, LNCS, vol. 5441, pp. 306–313, Paraty, Brazil, 2009.

[6] S. Arberet, R. Gribonval, and F. Bimbot, "A robust method to count and locate audio sources in a multichannel underdetermined mixture", *IEEE Transactions on Signal Processing*, Vol. 58, No. 1, Jan. 2010.

[7] F. Theis and S. Amari, "Postnonlinear overcomplete blind source separation using sparse sources," in *Proc. of ICA 2004*, LNCS, vol. 3195, pp. 718–725, Granada, Spain, 2004.

[8] S. Van Vaerenbergh and I. Santamaría, "A spectral clustering approach to underdetermined postnonlinear blind source separation of sparse sources", *IEEE Transactions on Neural Networks*, vol. 17, No. 3, pp. 811–814, 2006.

[9] Y. Deville and S. Hosseini, "Blind identification and separation methods for linear-quadratic mixtures and/or linearly independent non-stationary signals," in *Proc. of ISSPA 2007*, Sharjah, United Arab Emirates, Feb, 12–15, 2007.

[10] L. T. Duarte and C. Jutten, "A nonlinear source separation approach for the Nicolsky-Eisenman model," in *Proc. of EUSIPCO 2008*, Lausanne, Switzerland, 2008.

[11] A. Dionisio, R. Menezes, and D. A. Mendes, "Mutual information: a measure of dependency for nonlinear time series", *Physica A: Statistical Mechanics and its Applications*, Vol. 344, No. 1–2, pp. 326–329, 2004.

[12] C. Abraham, P. A. Cornillon, E. Matzner-Løber, and N. Molinari, "Unsupervised curve clustering using B-splines", *Scandinavian Journal of Statistics*, Vol. 30, No. 3, pp. 581–595, Sep. 2003.

[13] M. Puigt, E. Vincent, and Y. Deville "Validity of the independence assumption for the separation of instantaneous and convolutive mixtures of speech and music sources," in *Proc. of ICA 2009*, LNCS, vol. 5441, pp. 613–620, Paraty, Brazil, 2009.

[14] G.A. Darbellay and I. Vajda, "Estimation of the information by an adaptive partitioning of the observation space", *IEEE Transactions on Information Theory*, Vol. 45, No. 4, pp. 1315-1321, May 1999.