**7th Annual International IEEE EMBS Conference on Neural Engineering**
**Montpellier, France, 22 - 24 April, 2015**

# A Novel Approach for Detection of Medial Temporal Discharges Using Blind Source Separation Incorporating Dictionary Look Up

Shahrzad Shapoori, Saeid Sanei, and Wenwu Wang

*Abstract*— **In blind source separation (BSS), sparsity is proved to be very advantageous. If data is not sparse in its current domain, it can be modelled as sparse linear combinations of elements of a chosen dictionary. The choice of dictionary that sparsifies the data is very important. In this paper the dictionary is pre-specified based on chirplet modelling of various kinds of real epileptic spikes. Dictionary look up together with source separation is used to extract the closest source to the source of interest from the scalp EEG measurements. The algorithm has been tested on synthetic and real data consisting of epileptic discharges, and the results are compared with those of traditional BSS.**

*Index Terms*— **Blind source separation, sparsity, dictionary look-up, epileptic spikes, chirplet modelling**

## I. INTRODUCTION

Seizure prediction is very important in assessment and treatment of many neurological disorders such as epilepsy. Long before the onset of seizure, there are tiny medial temporal discharges, also known as epileptic spikes, in the Electroencephalogram (EEG). These discharges happen long before the onset of seizure and look much more pronounced in the intracranial EEG than in the scalp EEG. The scalp EEG can be considered as a mixture of intracranial EEGs considering the nonlinearity of head as the medium in which the intracranial signal is going through before reaching the surface of the head [1].

The problem of separating original sources from a set of observations or mixtures with no or very little knowledge is called blind source separation (BSS). This phenomenon can be formulated and shown as:

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{V} \tag{1}$$

where $\mathbf{Y} \in \mathbb{R}^{m\times N}$, is the observation matrix, $\mathbf{X} \in \mathbb{R}^{n\times N}$ is the source matrix, and $\mathbf{A}$ is the $m\times n$ mixing matrix. The additive noise caused by the instrumental noise or imperfection of the model which is denoted by $\mathbf{V}$ is of size $m \times N$. The aim of BSS is to estimate both $\mathbf{X}$ and $\mathbf{A}$ from $\mathbf{Y}$.

There is no exclusive solution for this problem. Solutions can be found by imposing constraints into the process of separation. Statistical independence and non-gaussianity of the sources, which are used by independent component analysis (ICA) [2], are two of these constraints. Also, ICA tries to separate the sources by minimizing the mutual information.

*This work has been supported by the EPSRC, UK, Grant No. EP/K005510/1.

Shahrzad Shapoori, Saeid Sanei, and Wenwu Wang are with The Faculty of Engineering and Physical Sciences, University of Surrey, Guildford, UK (s.shapoori, s.sanei, w.wang)@surrey.ac.uk.

Another constraint to be considered can be sparsity of the sources. The idea is that using a known dictionary, the sources can be sparsely represented . For example, in the wavelet domain the signal can be represented sparsely [3]. In the work by Abolghasemi *et al.* [4], the sparsity of sources has been considered to be very useful in the separation process. However, in their application sources are not sparse in time domain and they have used a dictionary to sparsely represent the images. Dictionary learning is then applied in order to separate the original image sources. Further, they have considered the multichannel case with more observations than sources.

In this paper, the pre-ictal period of intracranial and scalp EEG recordings of epileptic patients is studied. A dictionary of many different epileptic spikes taken from intracranial data and modelled using chirplet transform, is generated. The dictionary is kept fixed and is used in the separation process as a constraint to extract the closest source from the mixtures. The results of the algorithm for synthetic and real data are shown.

## II. METHODOLOGY

### A. Dictionary Look Up

Representation of the signals depends on the choice of dictionary. Dictionary elements also called atoms which are a set of basis signals, are used to decompose the data. Every signal can be uniquely represented as a linear combination of these atoms.

*1) Dictionary Representation:* Dictionaries have different forms such as orthogonal, bi-orthogonal and overcomplete. In the orthogonal case, the representations are inner products of the signal and dictionary atoms. However, in the bi-orthogonal case they are inner products of the signal and the dictionary inverse. Overcomplete dictionaries which have more atoms than the signal dimensionality, are able to represent more features of the signal [5].

Representation of a signal using these dictionaries can be done in two ways: the analysis path, or the synthesis path. In the former case, inner products of the signal with the atoms represent the signal:

$$\gamma_a = \mathbf{D}^T\mathbf{x} \tag{2}$$

In the latter case, a linear combination of the dictionary atoms represents the signals:

$$\mathbf{x} = \mathbf{D}\gamma_s \tag{3}$$

These two definitions overlap, when the dimension of the signal and the number of dictionary atoms are the same. This case is known as the complete case. In this situation the analysis and synthesis dictionaries are both bi-orthogonal. However, the two dictionaries are very different in the general case.

The family of representations, $\gamma_s$ satisfying (3) is actually infinitely large, when the dictionary is ovecomplete. In order to obtain the most informative representation, a cost function is defined as $C(\gamma)$. Therefore, the objective is to minimize $C(\gamma)$.

$$\gamma_s = \underset{\gamma}{\operatorname{argmin}}\, C(\gamma) \;\; \textit{subject to} \quad \mathbf{x} = \mathbf{D}\gamma \tag{4}$$

Sparsity of the representation is encouraged via practical choices of this cost function. Solving Eq. (4) is denoted as sparse coding. Sparsity can be achieved through selecting a robust penalty function that keeps the large coefficient and omits the small near-zero coefficients.

Choosing a proper dictionary for different applications is very challenging. Fourier and wavelet dictionaries are examples of traditional dictionaries that are very easy to use and appropriate for one dimensional signals, but not so suitable for higher dimensional signals and more complex data. Newly developed dictionary methods can be divided into two main groups; dictionaries which are characterised based on a mathematical model of data, or dictionaries which are a group of realizations of the data. The first group of dictionaries are characterized by logical formulation, but the second group has the ability to adapt to and perform best on a specific set of data. Lately, dictionaries which have characteristics of both groups have been desirable [5].

*2) Creating The Dictionary:* The dictionary is pre-specified based on various morphologies of real epileptic discharges in the intracranial EEG scored by clinician experts. These discharges are modelled using a limited number of chirplets. Some of them are shown in Fig. 1.

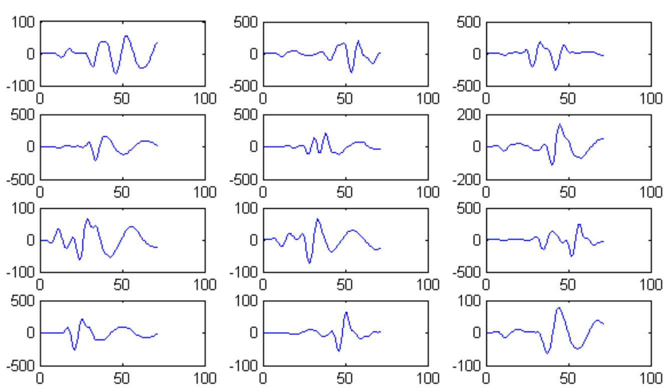

Fig. 1. Some of the dictionary components, which are chirplet modellings of selected pre-ictal discharges in intracranial EEG.

As depicted in Fig. 1, epileptic discharges are usually made up of a sharp peak followed by a slow wave [6].

## B. Sparse Recovery

Sparse recovery is the procedure of calculating the coefficients for representing the signal based on the given signal and the dictionary. Generally, sparse recovery is also referred to as atom decomposition. It requires solving one of the (5) or (6) equations, which is usually done by a pursuit algorithm. [7].

$$(P_0) \;\; \min_{\mathbf{s}} \|\mathbf{s}\|_0 \;\; \textit{subject to} \quad \mathbf{x} = \mathbf{D}\mathbf{s} \tag{5}$$

$$(P_{0,\epsilon}) \;\; \min_{\mathbf{s}} \|\mathbf{s}\|_0 \;\; \textit{subject to} \quad \|\mathbf{x} - \mathbf{D}\mathbf{s}\| \le \epsilon \tag{6}$$

where $\|.\|_0$ shows the $l_0$-norm and counts the number of non-zeros. It is an NP-hard problem to exactly determine the sparsest representation of the signal [8]. Instead of precise solutions, approximate ones are considered. Matching pursuit (MP) [7] and orthogonal matching pursuit (OMP) [9] algorithms select the dictionary atoms one after the other. In these algorithms, inner products between the signal and dictionary atoms are computed, and some least squares solvers are used. Initially, the atom which has the largest inner product with the signal is found, and the contribution due to that atom is subtracted from the signal. This procedure is repeated until the signal is decomposed completely. By changing the algorithm stopping rule, both (5) and (6) are addressed. The main difference between MP and OMP is the process of updating all the extracted coefficients after each step, which is done in OMP. This is done by calculating the orthogonal projection of the signal onto the set of atoms selected after each iteration. OMP can have better results than standard MP, but requires more calculation.

If the required solution is sparse enough, these algorithms and similar techniques are able to recover it accurately [10].

## C. Multichannel Source Separation

In this section the aim is to use the sparse recovery theory and the created dictionary from the previous subsections in order to help extracting the original sources from a set of mixtures. If the BSS model in Eq. (1) is considered, the sources of interest are intracranial epileptic discharges and the mixtures are scalp recordings of epileptic patients. If all the channels of scalp recordings are saved in a matrix $\mathbf{Y}$ then, stacking all the channels in one row makes a single vector $\mathbf{y}$. The extracted desired source $\mathbf{X}$ is also vectorized and shown as $\mathbf{x}$. Therefore, the BSS model in Eq. (1) is changed because the signals here are vectors instead of being matrices. The new equation is:

$$\underline{\mathbf{y}} = (\mathbf{I} \otimes \mathbf{A})\underline{\mathbf{x}} + \underline{\mathbf{v}} \tag{7}$$

In the above equation, $\underline{\mathbf{x}} = vec(\mathbf{X})$ and $\underline{\mathbf{y}} = vec(\mathbf{Y})$ are column vectors of length $nN$ and $mN$ respectively, in which $n$ is the number of scalp channels, $m$ is the number of sources, and $N$ is the number of samples. $(\mathbf{I} \otimes \mathbf{A})$ of size $mN \times nN$ is a block diagonal matrix, and $\otimes$ is the Kronecker product symbol. The noiseless setting is considered. The overall source separation problem is expressed as:

$$\min_{\{s_i\},\underline{x},A} \lambda\|\underline{\mathbf{y}} - (\mathbf{I} \otimes \mathbf{A})\underline{\mathbf{x}}\|_2^2 + \sum_{i=1}^{p}[\mu_i\|\mathbf{s}_i\|_0 + \|\mathbf{D}\mathbf{s}_i - \Re_i\underline{\mathbf{x}}\|_2^2] \tag{8}$$

where $\lambda$ and $\mu$ control the noise power and sparsity degree, respectively. $\mathbf{D}$ is the dictionary of size $r \times k$ which contains normalised columns, also called atoms. $\{\mathbf{s}_i\}$ are sparse coefficients of length $k$. For simplicity the source vector $\mathbf{x}$ is divided into patches of length $r$, and the $i$-th patch from $\mathbf{x}$ is shown by vector $\Re_i\mathbf{x}$. Furthermore, the total number of patches is shown by $p$. $\mathbf{A}$ shows the mixing matrix, which is initialized by a random matrix of suitable size. The source vector $\mathbf{x}$ is then initialized by $\mathbf{x} = \mathbf{A}^T\mathbf{y}$.

The process of minimizing (8) starts with extracting patches of $\mathbf{x}$. Then, the OMP algorithm estimates the sparse coefficients $\{\mathbf{s}_i\}_{i=1}^{p}$ for each patch $\Re_i\mathbf{x}$, having the patch itself and the dictionary $\mathbf{D}$ as inputs. Afterwards, $\{\mathbf{s}_i\}_{i=1}^{p}$ and $\mathbf{A}$ are assumed fixed, and $\mathbf{x}$ has to be estimated. The solution for $\mathbf{x}$ is obtained by taking the gradient of (8) and setting it to zero:

$$0 = \lambda(\mathbf{I}\otimes\mathbf{A})^T((\mathbf{I}\otimes\mathbf{A})\underline{\mathbf{x}} - \underline{\mathbf{y}}) + \sum_{i=1}^{p}\Re_i^T(\Re_i\underline{\mathbf{x}} - \mathbf{D}\mathbf{s}_i) \quad (9)$$

which becomes:

$$\hat{\underline{\mathbf{x}}} = (\lambda(\mathbf{I}\otimes\mathbf{A})^T(\mathbf{I}\otimes\mathbf{A}) + \sum_{i=1}^{p}\Re_i^T\Re_i)^{-1}\cdots$$
$$\cdots(\lambda(\mathbf{I}\otimes\mathbf{A})^T\underline{\mathbf{y}} + \sum_{i=1}^{p}\Re_i^T\mathbf{D}\mathbf{s}_i). \quad (10)$$

The matrix of mixtures $\mathbf{Y}$ and the source matrix $\mathbf{X}$, which is computed in previous step, are used in order to estimate the mixing matrix $\mathbf{A}$. Simplifying (8) by changing the first quadratic term into normal matrix product gives:

$$\min_{\mathbf{A}} \lambda\|\mathbf{Y} - \mathbf{A}\mathbf{X}\|_F^2. \quad (11)$$

This minimization problem is solved very simply by taking the pseudo inverse of $\mathbf{X}$ as:

$$\hat{\mathbf{A}} = \mathbf{Y}\mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1} \quad (12)$$

The steps of estimating $\{\mathbf{s}_i\}$, $\underline{\mathbf{x}}$, and $\mathbf{A}$ have to be repeated for a suitable number of iterations, in order to minimize (8).

## III. Experimental Results

### A. Data

Simulated data consists of linear mixing of a piece of intracranial signal, which contains epileptic discharges, with a random signal, and adding Gaussian noise to the mixtures. This makes the overall mixing system underdetermined.

Real data has been recorded using both scalp and Foramen Ovale (FO) electrodes to measure the scalp mixtures and intracranial sources, respectively. This data has been taken from patients with temporal lobe epilepsy in order to plan for surgical operation. The part of intracranial signals which is recorded before the seizure and contains epileptic discharges, has been selected. These signals are greatly affected by noise and activity of neighbouring neurons. The scalp and intracranial recordings, has been done simultaneously with the sampling rate of 200 samples per second [11].

### B. Results

*1) Simulated Data:* As depicted, the signal in Fig. 2(a), which shows an intracranial discharge is linearly mixed with the signal in Fig. 2(b), and $10dB$ noise is added to them in order to create the mixtures shown in Fig. 2(c) and 2(d). The amplitude of intracranial discharge is 10 times less than the amplitude of the random signal. This has been done in order to make the simulated data more similar to the real scalp signals. In reality, epileptic discharges in the intracranial signals have much smaller amplitudes than the scalp EEG. Conventional BSS, in this case Fast ICA, has been applied to

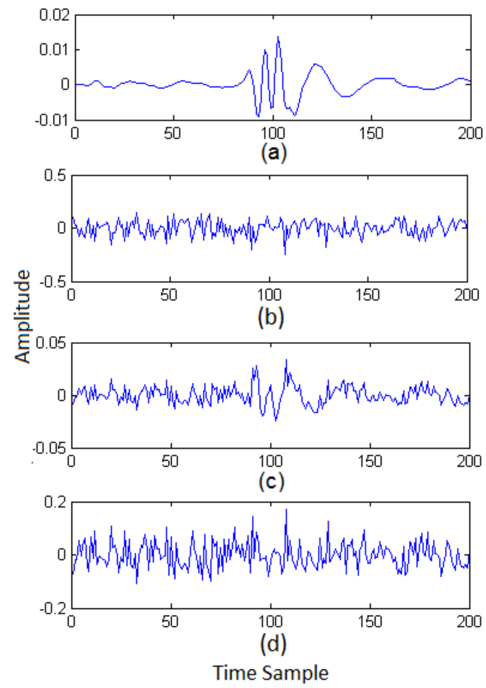


Fig. 2. Simulated original sources and mixtures; (a) Original epileptic discharge (b) Random signal (c) Mixture1 (d) Mixture2.

these mixtures and the result is shown in Fig. 3. As shown, conventional BSS is not able to separate the sources properly, because of two reasons; the epileptic discharge have a much smaller amplitude than the other signal, and the number of mixtures (considering noise) are less than the number of sources. The proposed method has been applied to these

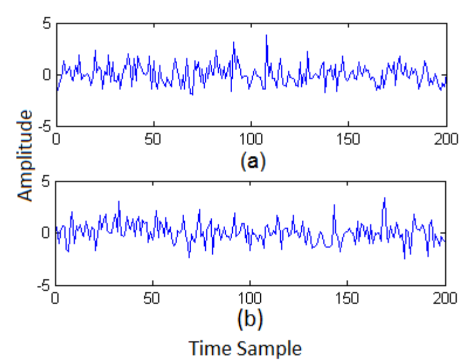


Fig. 3. Separated sources using BSS; (a) Source1 (b) Source2.

mixtures and the output source result is shown in Fig. 4. It is shown that the original source has been extracted from the set of mixtures with just a little noise, and the result is much better than that of the conventional BSS. Also, a comparison

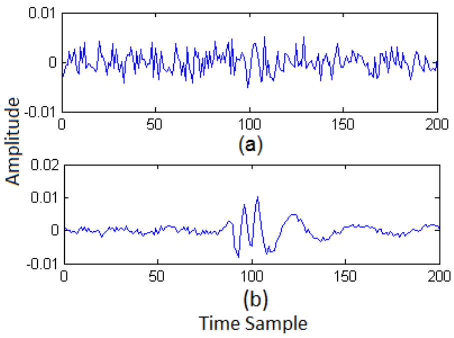


Fig. 4. Proposed method output sources; (a) Source1 (b) Source2.

of the root mean square error (RMSE) between the desired source and the extracted source by the conventional BSS and the proposed method is shown in Table I.

TABLE I
COMPARISON OF THE RMSE BETWEEN THE DESIRED SOURCE AND THE EXTRACTED SOURCE USING THE CONVENTIONAL BSS (FAST ICA) AND THE PROPOSED METHOD FOR SIMULATED SIGNALS.

| Method | Proposed BSS-Dictionary Look Up | Fast ICA |
|---|---|---|
| RMSE | 0.0728 | 0.3536 |

2) *Real Data:* Multiple channels of scalp EEG signals, in this case 16 channels are used as mixtures and the algorithm is tested on them in order to extract the source, which appeared in the same time instant as in the intracranial signal. The original intracranial source, the closest extracted source by traditional BSS (Fast ICA), and the extracted source by the proposed method are shown in Fig. 5(a), 5(b), and 5(c) respectively. Also, a comparison of the root mean square error (RMSE) between the desired source and the closest

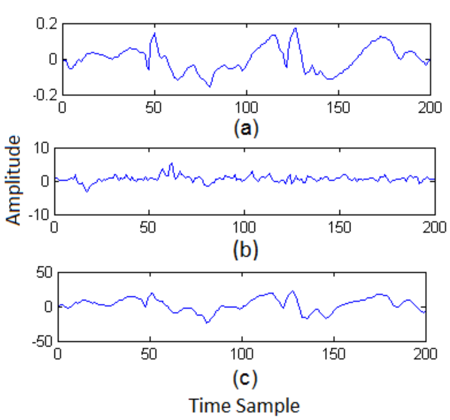


Fig. 5. Original source from real intracranial EEG and output sources; (a) original source from real intracranial EEG (b) traditional BSS (fast ICA) closest output source from real scalp EEG (c) proposed method output source from real scalp EEG.

extracted source by the conventional BSS and the proposed method is shown in Table II.

TABLE II
COMPARISON OF THE RMSE BETWEEN THE DESIRED SOURCE AND THE CLOSEST EXTRACTED SOURCE USING THE CONVENTIONAL BSS (FAST ICA) AND THE PROPOSED METHOD FOR REAL SCALP SIGNALS.

| Method | Proposed BSS-Dictionary Look Up | Fast ICA |
|---|---|---|
| RMSE | 0.1 | 0.2449 |

## IV. CONCLUSION

In this paper, taking advantage of the sparsity of original sources in order to better tackle the BSS problem has been investigated. It has been shown that even if the sources are not sparse in their current domain, it may be possible to model them as sparse linear combinations of elements in a chosen dictionary. The dictionary is pre-specified and contains different models of epileptic discharges. These discharges are produced by chirplet modelling of the actual spikes. Iteratively estimating the sparse coefficients, the mixing matrix is used to extract the closest source to the original source from the mixtures. The method has been tested on simulated and real EEG data of epileptic discharges and the results are promising. In future work, keeping a part of dictionary pre-specified and updating the rest of it in order to better match the data will be considered.

## REFERENCES

[1] S. Sanei, *Adaptive Processing of Brain Signals*. John Wiley & Sons, 2013.
[2] A. Hyvärinen, J. Hurri, and P. O. Hoyer, “Independent component analysis,” in *Natural Image Statistics*. Springer, 2009, pp. 151–175.
[3] A. M. Bronstein, M. M. Bronstein, M. Zibulevsky, and Y. Y. Zeevi, “Sparse ica for blind separation of transmitted and reflected images,” *International Journal of Imaging Systems and Technology*, vol. 15, no. 1, pp. 84–91, 2005.
[4] V. Abolghasemi, S. Ferdowsi, and S. Sanei, “Blind separation of image sources via adaptive dictionary learning,” *IEEE Transactions on Image Processing*, vol. 21, no. 6, pp. 2921–2930, 2012.
[5] R. Rubinstein, A. M. Bruckstein, and M. Elad, “Dictionaries for sparse representation modeling,” *IEEE Proceedings*, vol. 98, no. 6, pp. 1045–1057, 2010.
[6] N. Kissani, G. Alarcon, M. Dad, C. Binnie, and C. Polkey, “Sensitivity of recordings at sphenoidal electrode site for detecting seizure onset: evidence from scalp, superficial and deep foramen ovale recordings,” *Clinical neurophysiology*, vol. 112, no. 2, pp. 232–240, 2001.
[7] S. G. Mallat and Z. Zhang, “Matching pursuits with time-frequency dictionaries,” *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, 1993.
[8] G. Davis, S. Mallat, and M. Avellaneda, “Adaptive greedy approximations,” *Constructive approximation*, vol. 13, no. 1, pp. 57–98, 1997.
[9] Y. C. Pati, R. Rezaiifar, and P. Krishnaprasad, “Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition,” in *IEEE Proceedings of 27th 1993 Asilomar Conference on Signals, Systems and Computers*. IEEE, 1993, pp. 40–44.
[10] J. A. Tropp, “Greed is good: Algorithmic results for sparse approximation,” *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2231–2242, 2004.
[11] D. Nayak, A. Valentın, G. Alarcón, J. J. Garcıa Seoane, F. Brunnhuber, J. Juler, C. E. Polkey, and C. D. Binnie, “Characteristics of scalp electrical fields associated with deep medial temporal epileptiform discharges,” *Clinical neurophysiology*, vol. 115, no. 6, pp. 1423–1435, 2004.