



HAL
open science

Blind Source Separation Methods Applied to Muscle Artefacts Removing from Epileptic Eeg Recording: A Comparative Study.

Amar Kachenoura, Doha Safieddine, Laurent Albera, Gwénaél Birot, Fabrice Wendling, Lotfi Senhadji, Isabelle Merlet

► **To cite this version:**

Amar Kachenoura, Doha Safieddine, Laurent Albera, Gwénaél Birot, Fabrice Wendling, et al.. Blind Source Separation Methods Applied to Muscle Artefacts Removing from Epileptic Eeg Recording: A Comparative Study.. RITS 2011 (Colloque National Recherche en Imagerie et Technologies pour la Santé), Apr 2011, Rennes, France. pp.1708.1-1708.3. hal-00908765

HAL Id: hal-00908765

<https://hal.science/hal-00908765>

Submitted on 5 Mar 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Blind source separation methods applied to muscle artefacts removing from epileptic EEG recording: A comparative study

Amar kachenoura^(1,2), Doha Safieddine^(1,2), Laurent Albera^(1,2), Gwénaél Birot^(1,2), Fabrice Wendling^(1,2), Lotfi Senhadji^(1,2), Isabelle Merlet^(1,2)

¹ INSERM, U642, Rennes, F-35000, France;

² Université de Rennes 1, LTSI, Rennes, F-35000, France.

Abstract: *Electroencephalogram (EEG) recordings are often contaminated with muscle artifacts. These artifacts obscure the EEG and complicate its interpretation or even make the interpretation unfeasible. In this paper, realistic spike EEG signals are simulated from the activation of a 5 cm² epileptic patch in the left superior temporal gyrus. Background activities and real muscle artifacts are then added to the simulated data. We compare the efficiency of Empirical Mode Decomposition (EMD), Independent Component Analysis (ICA) and Blind Source Separation based on Canonical Correlation Analysis (BSS-CCA) to remove muscle artifacts from the EEG signals. The quantitative comparison indicates that the EMD approach exhibits a better performance than ICA and BSS-CCA, especially in the case of very low Signal to Noise Ratio (SNR).*

Keys words: *ICA, BSS-CCA, EMD, interictal spikes, epilepsy, muscle artefacts removing.*

I. INTRODUCTION

Epilepsy is common brain disorder characterized by abnormal neuronal firing in the brain. The activities of these neurons can be monitored by electroencephalogram (EEG). Interictal spikes recorded during the periods between seizures are the great importance for diagnostic purposes, since it can help the clinicians both to guide further explorations such as intracerebral implantation or to assist surgery. Nevertheless, the surface EEG signals have, typically, very low amplitude and hence they are severely contaminated by various noises and artifacts, such as broken wire contacts, ocular movements, electrocardiogram, muscular activity (electromyogram, EMG), etc. Thus, one of the challenging tasks in EEG signal processing, is how to extract very weak, brain signals hidden by noise and artifacts usually of higher amplitude. Among all these artifacts, the removing of muscular activity seems to be particularly difficult [1]. This can be essentially attributed to the fact that: i) EMG is spatially and spectrally distributed, ii) some muscular activity, such as facial EMG is sensitive to cognitive process, and iii) in comparison to other artefacts, EMG exhibits less stereotyped response.

Minimizing muscular artefacts from the acquired multi-channel EEG can be considered as a blind source separation problem (BSS), which can be defined as the

estimation of the original sources that generate the multi-channel EEG signals, without *a priori* (or very little) knowledge about the sources and the mixing process. According to the assumptions made on the sources, the problems of blind estimating the primary sources from a set of observations can be solved using different signal processing methods [2]. We focus, in this paper, on three different techniques, namely, Independent Component Analysis (ICA) [3], BSS using Canonical Correlation Analysis (BSS-CCA) [4-6], and Empirical Mode Decomposition (EMD) [7-8]. More precisely, we aim at showing how EMD performs, to remove muscle artifact from the real life interictal spikes EEG simulations, as compared to ICA and BSS-CCA.

II. MATERIALS AND METHODS

II.1- Problem formulation: In the case of EEG, BSS assumes that electrical activities recorded at the level of surface electrodes can be considered as a static linear mixture of elementary sources [9] so that the linear observation model below holds:

$$\mathbf{x}[m] = \sum_{p=1}^P \mathbf{a}_p s_p[m] + \mathbf{v}[m] = \mathbf{A}s[m] + \mathbf{v}[m] \quad (1)$$

where $s[m]$ is a P -dimensional vector called source vector, \mathbf{A} is $(N \times P)$ mixing matrix and $\mathbf{v}[m]$ is a N -dimensional additive noise. The objective of BSS is to recover the primary source vector $s[m]$ and/or to identify the mixing matrix \mathbf{A} from the observed vector $\mathbf{x}[m]$ without *a priori* knowledge of the sources and the mixing process, by assuming a few assumptions on the source signals.

II.2- Independent component analysis: The concept of ICA was introduced by Herault and Jutten [10], especially in order to solve the BSS problem. In the mid 90s, Comon presented a mathematical formulation of ICA [3]. During the past 25 years a wealth of algorithms have been proposed [2] and ICA-based methods have now been extensively and successfully applied to solve many practical real-life problems (see [2] for details). More precisely, assuming the linear observation model of equation (1), ICA consists in looking for an overdetermined $(N \times P)$ mixing matrix \mathbf{A} (i.e. P is smaller than or equal to N) and a P -dimensional sources $s[m]$ which components are the most statistically independent as possible. In other words, the goal of ICA

is to find a $(P \times N)$, full rank, separator matrix, \mathbf{W} , such that the output signal $\mathbf{y}[m] = \mathbf{W}\mathbf{x}[m]$ is an estimate of the source vector $\mathbf{s}[m]$. Note that in this work, we use the COM2 algorithm [2] which is based on a maximization of a contrast functions derived from Higher Order (HO) cumulants of the data.

II.2- Canonical correlation analysis: Originally proposed by [4], CCA is a method searching for a basis vectors for two sets of variables such that the projections of the two variables onto this basis are maximally correlated. Borga et al. [5] showed that CCA can be applied in BSS, and proposed a new algorithm (BSS-CCA) to solve the problem presented in equation (1) by forcing the sources $\mathbf{s}[m]$ to be mutually uncorrelated and maximally autocorrelated. To do so, let's consider the observed vector $\mathbf{x}[m]$ and its temporally delayed version $\mathbf{z}[m] = \mathbf{x}[m-1]$. The BSS-CCA algorithm finds the linear combination of $\mathbf{x}[m]$ that is maximally correlated with the linear combination of $\mathbf{x}[m-1]$. Now, construct two new linear combinations $b[m]$ and $c[m]$ such that:

$$\begin{aligned} b[m] &= w_{x_1} x_1[m] + \dots + w_{x_N} x_N[m] = \mathbf{w}_x^T \mathbf{x}[m] \\ c[m] &= q_{z_1} z_1[m] + \dots + q_{z_N} z_N[m] = \mathbf{q}_z^T \mathbf{z}[m] \end{aligned} \quad (3)$$

The problem is to find the weight vectors \mathbf{w}_x and \mathbf{q}_z that maximize the correlation, ρ , between $b[m]$ and $c[m]$ by solving the following maximization problem:

$$\max_{\mathbf{w}_x, \mathbf{q}_z} \rho(\mathbf{b}, \mathbf{c}) = \frac{\mathbf{w}_x^T \mathbf{R}_{xz} \mathbf{q}_z}{\sqrt{(\mathbf{w}_x^T \mathbf{R}_{xx} \mathbf{w}_x)(\mathbf{q}_z^T \mathbf{R}_{zz} \mathbf{q}_z)}} \quad (4)$$

with \mathbf{R}_{xx} and \mathbf{R}_{zz} the autocovariance matrices of \mathbf{x} and \mathbf{z} respectively, and \mathbf{R}_{xz} the crosscovariance matrix of \mathbf{x} and \mathbf{z} . After some manipulations, demixing matrix, \mathbf{W} , such that $\mathbf{y}[m] = \mathbf{W}\mathbf{x}[m]$ is an estimate of source vector $\mathbf{s}[m]$ can be easily calculated by solving the eigen-matrix equation $\mathbf{R}_{xx}^{-1} \mathbf{R}_{xz} \mathbf{R}_{zz}^{-1} \mathbf{R}_{zx} \mathbf{w}_x = \rho^2 \mathbf{w}_x$, where the eigen-values ρ^2 are the squared canonical correlation and the eigen-vectors \mathbf{w}_x are the columns of the demixing matrix \mathbf{W} .

II.3- Empirical mode decomposition: Empirical Mode Decomposition (EMD) was originally introduced in the late 90s to study water surface wave evolution [7]. EMD can be considered as an emerging technique in signal processing with a very important research topics and development in various fields [7, 8], Hurst exponent estimation [3], etc. EMD can be viewed as a method which splits a given signal into a set of AM-FM components. The latter components, referred to as the Intrinsic Mode Functions (IMFs), are estimated using an iterative procedure called *sifting process*. Let

$\{x[m]\} = \{a_0[m]\}$ be a signal to decompose. The sifting process to estimate the k^{th} IMF ($1 \leq k \leq K$) can be summarized as follows:

1. Initialization taking $d_{k,0}[m] = a_{k-1}[m]$ for every m .
2. Computation of the mean envelope $\mathcal{M}\{d_{k,j}[m]\}$.
3. Extraction of the details $\{d_{k,j+1}[m]\} = \{d_{k,j}[m]\} - \{\mathcal{M}\{d_{k,j}[m]\}\}$.
4. Incrementation of j and go to step 2, repeat until $\{d_{k,j+1}[m]\}$ is designated as an IMF.

Several EMD methods have been proposed in the literature, depending on the way the mean envelope $\mathcal{M}\{d_{k,j}[m]\}$ is calculated. The 2T-EMD method proposed in [8] is used hereafter. This method differs from the other approaches by its computational lightness and its algorithmic simplicity. It also enables both mono and multivariate decompositions without any change in the core of the algorithm.

III. PERFORMANCE COMPARISON

III.1- Data generation: The main goal of this subsection is to explain how we obtain synthetic but realistic data for comparing the above BSS methods in the particular context of epileptic interictal spikes. The simulated EEG data (one observation is displayed in figure 2), are generated using a realistic head and source model as described in [11]. 31-Channels EEG data were simulated from a single distributed source of 5cm², referred to as "patch" in the following, located in the left superior temporal gyrus. Hyper-synchronous spike-like activities generated by a neural mass model are assigned to the patch whereas uncorrelated background activities are attributed to the remaining part of the cortex. 50 trials of these simulated EEG are generated for the study. A different muscle activity issued from real EEG data is added to each trial.

III.2- Performance criterion: The performance of the three BSS methods, namely COM2, BSS-CCA and 2T-EMD, are evaluated by computing the following normalized mean-squared error (NMSE):

$$NMSE_{x_i} = \frac{\sum_{j=1}^J \sum_{m=1}^M \left[(x_i[m] - \hat{x}_i^{(j)}[m])^2 \right]}{J \sum_{m=1}^M (x_i[m])^2} \quad (5)$$

where $x_i[m]$ is the original EEG observation on the i^{th} electrode (EEG without muscle activity), $\hat{x}_i^{(j)}[m]$ is the reconstructed surface EEG after denoising from the j^{th} run, J is the number of Monte Carlo runs and M is the data length.

III.3- Simulation results: Before discussing the obtained performance, it is important to note that: i) the selection of

the ICA components accounting for the artifact removing is based on visual inspection and the signal was reconstructed, excluding the components related to the artifact, ii) for the BSS-CCA method, the selection of artifactual sources was based on thresholding criteria on the autocorrelation curve of the canonical components [6], and iii) as far as the EMD method is concerned, the IMFs of interest are selected using EMD-WAV procedure presented in [12] that was reported to exhibit a high performance on synthetic and real signals, and has the advantage of not needing either a reference signal or an a priori information about the noise.

Figure 1 shows the results obtained at the surface electrode T3, (facing the patch) in the case of SNR=-10dB. Clearly, the 2T-EMD has a better behavior than COM2 and BSS-CCA. This result is confirmed in Table 1 where the NMSE computed at the output of COM2, BSS-CCA and 2T-EMD methods, as a function of SNR, is depicted. Indeed the reconstruction error of 2T-EMD is lower than those of ICA and BSS-CCA, especially for lower SNR.

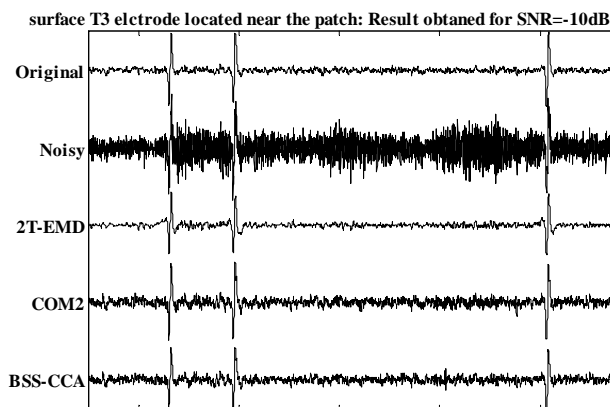


Fig.1 The obtained surface EEG on surface electrode T3 at the output of 2T-EMD, COM2 and BSS-CCA.

	-25dB	-20dB	-15dB	-10dB	-5 dB	0 dB
BSS-CCA	0.8661	0.5922	0.3664	0.2161	0.1245	0.0711
COM2	0.6186	0.4230	0.2617	0.1543	0.0889	0.0508
2T-EMD	0.3936	0.3080	0.2079	0.1217	0.07361	0.0528

TABLE 1: The NMSE, calculated from EEG of electrode T3, at the output BSS-CCA, COM2 and 2T-EMD methods.

IV. DISCUSSION - CONCLUSION

In this paper we present a study comparing the ability of EMD, ICA and BSS-CCA to remove muscle artifacts from EEG signals. Simulations of realistic epileptic EEG signals (interictal spikes) have been used to quantitatively compare the different algorithms. The results indicate that the EMD method as a performant tool to the remove muscle artifacts. Nevertheless, despite very promising

finding, a careful evaluation of the method on more realistic synthetic EEG data and the application of the method on clinical data must be performed in the future.

REFERENCE

- [1] B. W. McMenamin, A. J. Shackman, L. L. Greischar, R. J. Davidson, Electromyogenic artifacts and electroencephalographic inferences revisited, *NeuroImage*, vol. 54, pp. 4–9, 2011.
- [2] L. Albera, P. Comon, L. Parra, A. KarfoulL, A. Kachenoura, L. Senhadji, *Biomedical applications*, in *Handbook of blind source separation*, P. COMON and C. JUTTEN, Eds. Academic Press, 2010.
- [3] P. Comon, Independent Component Analysis, a new concept ? *Signal Processing, Elsevier*, vol. 36, no. 3, pp. 287–314, 1994.
- [4] H. Hotelling, Relations between two sets of variates, *Biometrika*, vol. 28, pp. 321–377, 1936.
- [5] O. Friman, M. Borga, P. Lundberg, H. Knutsson, Exploratory fMRI analysis by autocorrelation maximization, *NeuroImage*, vol. 16, no. 2, pp. 454–464, 2002.
- [6] W. De Clercq, A. Vergult, B. Vanrumste, W. Van Paesschen, S. Van Huffel, Canonical correlation analysis applied to remove muscle artifacts from the electroencephalogram, *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 12, pp. 2583 – 2587.
- [7] N. Huang, Z. Shen, S. Long, M. Wu, H. Shih, Q. Zheng, N.-C. Yen, C. Tung, H. Liu, The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis, *Proc. of the Royal Society A: Mathematical, Physical and Engineering Sciences*, pp. 903-995, 1998.
- [8] J. Fleureau, J. C. Nunes, A. Kachenoura, L. Albera, L. Senhadji, Turning Tangent Empirical Mode Decomposition: a framework for mono and multivariate signals, *IEEE Transactions on Signal Processing*, DOI: [10.1109/TSP.2010.2097254](https://doi.org/10.1109/TSP.2010.2097254), 2010(to appear).
- [9] C. J. James, C. W. Hesse, Independent component analysis for biomedical signals, *Physiological measurement*, vol. 26, no. 1, pp. 15–39, 2005.
- [10] B. Ans, J. Herault, C. Jutten, Adaptive neural architectures: Detection of primitives, in *COGNITIVA'85*, pp. 593–597, Paris, France, 1985.
- [11] D. Cosandier-Rim el e, J. Badier, P. Chauvel, F. Wendling, A physiologically plausible spatio-temporal model for depth-EEG signals recorded with intracerebral electrodes in human partial epilepsy, *IEEE Transactions On Biomedical Engineering*, vol. 3, no. 54, pp. 380–388, 2007.
- [12] Y. Kopsinis, E. Aboutanios, D. A. Waters, S. McLaughlin, Development of EMD-based Denoising Methods Inspired by Wavelet Thresholding, *IEEE Transaction. on Signal Processing*, vol. 57, no.4 ,pp. 1351-1362, 2009.