



Muscle artifact removal in ictal scalp-EEG based on blind source separation

Ahmad Karfoul, Amar Kachenoura, Laurent Albera, Doha Safieddine, Anca Pasnicu, Fabrice Wendling, Lotfi Senhadji, Isabelle Merlet

► To cite this version:

Ahmad Karfoul, Amar Kachenoura, Laurent Albera, Doha Safieddine, Anca Pasnicu, et al.. Muscle artifact removal in ictal scalp-EEG based on blind source separation. 6th European Conference of the International Federation for Medical and Biological Engineering, Sep 2014, Dubrovnik, Croatia. 4 p. hal-01015798

HAL Id: hal-01015798

<https://hal.archives-ouvertes.fr/hal-01015798>

Submitted on 27 Jun 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.

Muscle artifact removal in ictal scalp-EEG based on blind source separation

A. Karfoul¹, A. Kachenoura², L. Albera^{2,3}, D. Safieddine², A. Pasnicu⁴, F. Wendling², L. Senhadji² and I. Merlet²

¹ Faculty of mechanical and electrical engineering, Al-Baath University, BP 77, Home, Syria

² INSERM, U1099, Rennes, 35042, France and the LTSI, Université de Rennes 1, Rennes, 35042, France

³ Centre INRIA Rennes-Bretagne Atlantique, Rennes 35042, France

⁴ Unité d'Epileptologie, Service de Neurologie, CHU, Rennes, 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. This paper focuses on the particular context of extraction of low-voltage rapid ictal discharges from ictal scalp-EEG activity contaminated by muscle artifact. In this context our aim was to evaluate the ability of Independent Component Analysis (ICA) and Canonical Correlation Analysis (CCA), to remove muscle artefacts from surface EEG signals. The efficiency of ICA and CCA to correct the muscular artifact was evaluated both on simulated data and on real data recorded in an epileptic patient. The obtained results show that some ICA methods and CCA removed successfully the muscle artifact without altering the recorded underlying ictal activity.

Keywords— EEG, rapid ictal discharges, ICA, CCA, muscle artifact.

I. INTRODUCTION

Electroencephalography (EEG), based on the recording of electrical brain activity with scalp electrodes, is an essential diagnostic tool in patients with epilepsy. In that context, EEG is used to record abnormal activities (or paroxysms) during epileptic seizures that reflect the occurrence of sudden, hypersynchronous discharges within one or several cerebral regions. During presurgical evaluation of patients with refractory partial epilepsy, scalp EEG and video recordings are performed simultaneously (video-EEG) during several days in order to record the ictal discharges occurring during seizures. These discharges are characterized on EEG by a rhythmic activity which morphology can change as the ictal activity speeds. In this paper, we will focus on low voltage rapid discharges that occur at the beginning of seizures in partial epilepsies. Indeed, rapid ictal discharges are considered as electrophysiological signatures of the brain region(s) from where the seizure starts (i.e. the epileptogenic zone) [1]. The surgical resection of regions generating rapid ictal discharges has been correlated with a good surgical out-

come [2]. Tracking these rapid discharges on scalp EEG is crucial in the presurgical evaluation of patients who are candidates for selective surgery. Unfortunately, this activity is of low amplitude and therefore difficult to detect on scalp EEG even more so when signals are contaminated with various kinds of artifacts. Among them, muscular activity arising from the contraction of head muscles is often associated with clinical symptoms occurring at the onset of seizures, and is particularly difficult to remove due to its broad distribution in frequency and its unsteretyped spatial distribution. Therefore, denoising of EEG is a challenging preprocessing step prior to qualitative or quantitative EEG analysis. Minimizing the disturbances due to muscular activity in EEG signals can be considered as a Blind Source Separation (BSS) problem [3], which consists in estimating the original sources underlying the multichannel EEG signals, without a priori (or very little) knowledge about the sources themselves and about the mixing process. BSS techniques were successfully applied to remove muscle artifacts from epileptic scalp-EEG [4–7]. However to date, no study addresses the challenging problem of extracting the low voltage rapid ictal discharges occurring during seizures. In this communication, this problem is tackled, by evaluating the ability of ICA [8, 9] and CCA [10] to remove muscle artifacts in particular context of the noisy ictal scalp-EEG signals in order to enhance the Signal to Noise Ratio (SNR) of low voltage rapid discharges. We first discuss the EEG denoising problem and the assumptions required by the used BSS methods. Then, the ICA and CCA concepts are briefly described and the numerical complexity of each used method is also calculated to evaluate its computational cost. Finally, quantitative results are obtained on simulated epileptic data generated with a physiologically-plausible model [11]. These results are also illustrated on real data recorded in a patient with epilepsy.

II. PROBLEM FORMULATION AND HYPOTHESIS

In the case of EEG, BSS assumes that electrical activities recorded at the level of surface electrodes can be considered

as an instantaneous linear mixture of elementary sources [3, 6, 7] so that the linear observation model below holds:

$$x[k] = A^{(e)}s^{(e)}[k] + A^{(b)}s^{(b)}[k] + A^{(m)}s^{(m)}[k] + n[k] \quad (1)$$

where $\{x[k]\}$ is the vector of observations, representing the electrical activities recorded at the level of scalp electrodes at instant k . $\{s^{(e)}[k]\}$, $\{s^{(b)}[k]\}$, $\{s^{(m)}[k]\}$, $\{n[k]\}$ are random vector processes representing the P_e epileptic activity sources, the P_b background activity sources, the P_m muscular activity sources and the N -dimensional instrument noise, respectively. The mixing matrices $A^{(e)}$, $A^{(b)}$ and $A^{(m)}$ model the transfer from all possible sources of activity within the brain to scalp electrodes. In order to estimate the source of interest, namely $\{s^{(e)}[k]\}$ in our case, the following assumption are made: **i**) $\{s^{(e)}[k]\}$, $\{s^{(b)}[k]\}$, $\{s^{(m)}[k]\}$ and $\{n[k]\}$ are assumed independent, as they correspond to different physiological/physical phenomena, **ii**) the $\{n[k]\}$ vector random process is Gaussian as most of instrument noises, and **iii**) in case of CCA, sources are temporally coloured (i.e. statistical dependence between the p -th component of $\{x[k]\}$ and the p -th component of $\{x[k-1]\}$). It is important to note that, in our context, the identification of three vector subspaces corresponding to the epileptic sources, the muscular sources and the background sources, respectively, but not exactly the $P = P_e + P_b + P_m$ sources involved in equation (1), is sufficient. Indeed, it is not necessary to exactly extract the P_e epileptic sources; we just aim to remove the contribution of the muscular and background activities from the scalp-EEG data.

III. METHODS

A. Independent component analysis

The concept of ICA was introduced by Herault and Jutten [12], especially in order to solve the BSS problem. In the mid 90s, Comon presented a mathematical formulation of ICA [8]. During the past 25 years a wealth of algorithms have been proposed [3] and ICA-based methods have now been extensively and successfully applied to solve many practical real-life problems (see [3] for details). More precisely, assuming the linear observation model of equation (1), ICA consists in looking for an overdetermined ($N \times P$) mixing matrix A (i.e. P is smaller than or equal to N) and a P -dimensional sources $s[k]$ 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, W , such that the output signal $y[k] = Wx[k]$ is an estimate of the source vector $s[k]$. Our previous works [7] showed that, compared

to 15 other ICA methods, CoM2 [8] and InfoMax [9] algorithms offer the best performance/complexity compromise for the denoising of epileptic spike-like EEG signals activity. This justifies the use of these two methods in this communication. In few words, CoM2 [8] explicitly maximizes a contrast function based on the Forth Order cumulants (FO) of the data by rooting successive polynomials. Regarding InfoMax method [9], the ICA problem is solved by maximizing the Differential Entropy (DE) of the output of an invertible non-linear transform of the expected sources using the natural gradient algorithm.

B. Canonical correlation analysis

Originally proposed by Hotelling, CCA is a method that measures the linear relationship between two multidimensional random variables [10]. Friman et al. [13] showed that CCA can be used to solve the BSS problem by taking the source vector as the first multi-dimensional random variable and the temporally delayed version of the source vector as the second multi-dimensional random variable. Indeed, assuming the instantaneous linear observation model of equation (1), the spatial statistical decorrelation of sources and the temporal coloration of sources, CCA aims at extracting a P -dimensional vector of sources $s[k]$ by both minimizing the spatial correlation between sources and maximizing their temporal correlation. In other words, CCA algorithm finds the linear combination of $x[k]$ that is maximally correlated with the linear combination of $z[k] = x[k-1]$. Let construct two new linear combinations $u[k]$ and $v[k]$ such that: $u[k] = w_x^T x[k]$ and $v[k] = w_z^T z[k] = x[k-1]$. The problem is to find the weight vectors w_x and w_z that maximize the correlation, ρ , between $u[k]$ and $v[k]$: $\rho(u, v) = \frac{w_x^T R_{xz} w_z}{(w_x^T R_{xx} w_x)(w_z^T R_{zz} w_z)}$, with R_{xx} and R_{zz} are autocovariance matrices of x and z , and R_{xz} represents the crosscovariance matrix of x and z . After some manipulations, the vector w_x , representing the columns of the separator matrix W , can be easily estimated by solving the eigen-matrix equation $R_{xx}^{-1} R_{xz} R_{zz}^{-1} R_{xz} w_x = \rho w_x$.

C. Numerical complexity of InfoMax, CoM2 and CCA

Although the ultimate goal of comparing denoising approaches is to evaluate the quality of methods as reflected by the reconstructed signals, it is also interesting to assess the numerical complexity of these methods. Indeed, in practice, knowing whether an algorithm is computationally heavy is as important as knowing its performances in terms of SNR. Numerical complexity is defined here as the number of floating point operations required to execute an algorithm (flops).

A flop corresponds to a multiplication followed by an addition. But, in practice, only the number of multiplications is considered since, most of the time, there are about as many (and slightly more) multiplications as additions. Let N be the number of EEG electrodes, P be the number of sources and K be the data length. For the CoM2 algorithm, let $f4(P) = P(P+1)(P+2)(P+3)/24$ be the number of free entries in an FO cumulant tensor of dimension P enjoying all symmetries, I the number of sweeps executed by a joint diagonalization process and Q the complexity required to compute the roots of a real fourth degree polynomial by Ferrari's techniques. it is the number of iterations required to ensure the convergence of InfoMax. The numerical complexity of InfoMax, CoM2 and CCA is reported in table 1 (See [3, 7] for more explications).

IV. DATASETS

A. Generation of simulated data

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 rapid ictal discharges. The simulated EEG data are generated using a realistic head and source model as described in [11]. 32-Channels EEG data were simulated from a single distributed source of $5cm^2$, referred to as "patch" in the following, located in the left superior temporal gyrus. Rapid ictal-like activities generated by a neural mass model are assigned to the patch. 50 realizations of rapid ictal discharges simulations were generated. These signals corresponded to "clean data". In order to generate noisy EEG simulations, 50 epochs of EEG muscle activity were extracted from real 32-channel EEG data. Each trial of EEG muscle activity was then normalized with respect to the channel showing the maximal power. Then, different levels of amplitude of noisy background and muscular activities were added to the simulated rapid ictal discharges to get noisy simulated signals with different SNR values.

B. Real data

In order to test the feasibility of the denoising algorithms on real data, CoM2, InfoMax and CCA are applied to the denoising of rapid ictal discharges in subject (referred as to "Patient P" in the following) suffering from drug-resistant partial epilepsy. During video-EEG monitoring, scalp-EEG data were acquired from 32 electrodes (19–20 standard 1020 electrodes plus additional electrodes at FC1, FC2, FC5, FC6, CP1, CP2, CP5, FT9, FT10, P9, and P10). These data were reviewed in order to isolate two different epochs: **i**) the first

one considered as the reference signal (see figure 2 line 1), contains a clean rapid ictal discharges (especially on electrode C3), and **ii**) the second epoch (illustrated in figure 2 line 2) includes rapid ictal discharges hidden in muscle activity.

V. RESULTS

The quantitative performance of two ICA methods and CCA algorithm has been evaluated by computing the Normalized Mean-Squared Error ($NMSE$) between the original EEG observed on the surface electrodes (EEG without muscle activity) and the reconstructed surface EEG after denoising using one of the BSS method, as described in [6, 7]. Note that the selection of sources of interest is based on a visual inspection of components extracted by each algorithm in the time-frequency domain. Indeed, the rapid ictal discharges are narrowband (type "chirp"), while the muscle activity is broadband. Then, the denoised surface EEG is reconstructed by keeping only the components accounting for the sources of interest (rapid ictal discharges).

A. Quantitative performance of CoM2, InfoMax and CCA

In this experiment the behavior of CoM2, InfoMax and CCA are evaluated in the context of synthetic epileptic signals (rapid ictal discharges generated by realistic biomathematical models) corrupted by muscle artifacts. The data length, K , is fixed to 8192 samples, the SNR values are equal to -30 , -25 , -20 , -15 , -10 and -5 dB. Figure 1 shows the variations of $NMSE$ of the three algorithms as a function of SNR. More precisely, the $NMSE$ of CoM2, InfoMax and CCA, is calculated either for all electrodes or for the single electrode (T3) facing the cortical patch. For both cases CCA and CoM2 outperform InfoMax, especially for low SNR values and CCA is slightly more effective than CoM2.

B. Application to real data

Figure 2 shows the results obtained when applying CoM2, InfoMax and CCA on real data described in section (IV.B). We can remark that the rapid ictal discharge, which is clearly visible in the clean data (line one), is hidden in the noisy data (line two) due to the muscular artefact. Clearly, CCA and CoM2 methods enhance the rapid ictal discharge. Indeed, line three and four of figure 2 illustrate that the rapid ictal discharge is visible at the denoised electrode (C3), where a narrowband (type "chirp") signal is clearly identified (see the white arrow on times frequency representation). We also remark that, in this real example, InfoMax algorithm is not denoising data as well as the two other methods.

VI. CONCLUSION

Muscle artifacts are a major source of contamination of scalp EEG data. As a result a rapid and reliable denoising of these data constitutes an essential issue particularly when these signals are used for diagnosis, which is the case in patients with epilepsy. In this communication, we evaluate the ability of ICA and CCA to remove muscle artifacts from EEG signals in particular context of low voltage rapid discharges that occur at the beginning of seizures in partial epilepsies. The obtained results show that, by taking into account both numerical complexity and performance, CoM2 and CCA seem to be the best methods in our epileptic simulation. These promising finding should allow us to propose a method of EEG scalp analysis in order: **i**) to improve visual analysis of ictal EEG and their interpretation in clinical routine, and **ii**) to enhance the performance of the methods dealing with the analysis of statistical relationships between EEG channels.

Table 1: Numerical complexity of the three BSS algorithms

InfoMax	$\min(KN^2/2 + 4N^3/3 + PNK, 2KN^2) + (P^2 + P^3 + 4P + 5KP)it$
CoM2	$\min(KN^2/2 + 4N^3/3 + PNK, 2KN^2) + IP^2Q/2 + \min(12If_4(P)P^2 + 2IP^3 + 3Kf_4(P) + KP^2, 13IKP^2/2)$
CCA	$5KN^2 + 5KN + 19N^3/3$

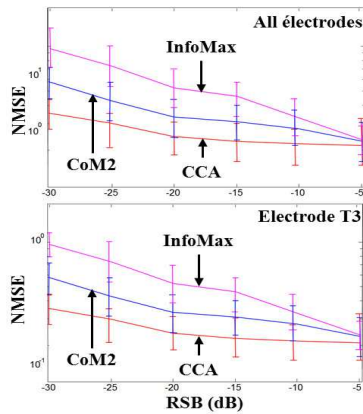


Fig. 1: NMSE of the CoM2, InfoMax and CCA, as a function of SNR

REFERENCES

1. Bancaud J, Talairach J, Schaub C, Szikla G, Bonis A. Stereotactic functional exploration of the epilepsies of the supplementary area of the mesial surfaces of the hemispheres *Electroencephalogr Clin Neurophysiol.* 1962;14–788.
2. Alarcon G, Binnie C, Polkey R Elwes C. Power spectrum and intracranial EEG patterns at seizure onset in partial epilepsy *Electroencephalogr Clin Neurophysiol.* 1995;94:326–337.

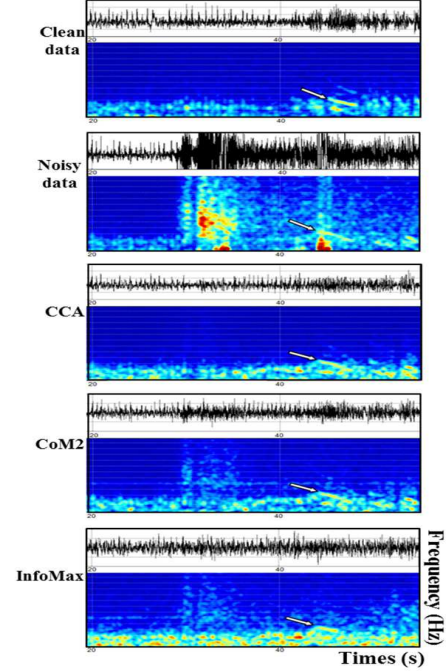


Fig. 2: Denoising of real rapid ictal discharge data using CCA, CoM2 and InfoMax.

3. Albera L, Comon P, Parra L, Karfoul A, Kachenoura A, Senhadji L. *Biomedical applications, in Handbook of blind source separation.* Eds. Academic Press 2010.
4. Iriarte J, Urrestarazu E, Artieda J, et al. Independent component analysis in the study of focal seizures *Clin. Neurophysiol.* 2006;23:551–558.
5. Vergult A, Clercq W De, et al. Improving the interpretation of ictal scalp EEG: BSSCCA algorithm for muscle artifact removal *Epilepsia.* 2007;48:950–958.
6. Saffeddine D, Kachenoura A, Albera L., et al. Removal of muscle artifact from EEG data: comparison between stochastic (ICA and CCA) and deterministic (EMD and Wavelet-based) approaches in *EURASIP Journal on Advances in Signal Processing.* 2012;article ID 2012:127.
7. Albera L., Kachenoura A, Comon P, et al. ICA-based EEG denoising: a comparative analysis of fifteen methods in *Special Issue of the Bulletin of the Polish Academy of Sciences - Technical sciences.* 2012;60:407–418.
8. Comon P. Independent component analysis: a new concept? *Signal Processing.* 1994;36:287–314.
9. Lee T W, Girolami M, Sejnowski T J. Independent component analysis using an extended infomax algorithm for mixed sub-gaussian and super-gaussian sources *Neural Computation.* 1999;11:417–441.
10. Hotelling H. Relations between two sets of variates *Biometrika.* 1936;28:321–377.
11. Cosandier-Rimele D, Badier J M, Chauvel P, Wendling F. A physiologically plausible spatio-temporal model for EEG signals recorded with intracerebral electrodes in human partial epilepsy *IEEE Trans Biomed Eng.* 2007;54:380–388.
12. Herault J, Jutten C. Space or time adaptative signal processing by neural networks models in *Int. Conf. on Neural Networks for Computing* (Snowbird) 1986.
13. Friman O, Borga M, Knutsson P Lundberg. Exploratory fMRI analysis by autocorrelation maximization *NeuroImage.* 2002;16:454–464.