

# Traitement aveugle des signaux EEG de surface : débruitage des signaux épileptiques critiques

AMAR KACHENOURA<sup>1,2</sup>, LAURENT ALBERA<sup>1,2</sup>, DOHA SAFIEDDINE<sup>1,2</sup>, FABRICE WENDLING<sup>1,2</sup>,  
LOTFI SENHADJI<sup>1,2</sup>, ISABELLE MERLET<sup>1,2</sup>

<sup>1</sup> Inserm, U1099, Rennes F-35000, France

<sup>2</sup> Université de Rennes 1, LTSI, Rennes, F-35000, France

[amar.kachenoura@univ-rennes1.fr](mailto:amar.kachenoura@univ-rennes1.fr)

**Résumé** – Dans cette communication, l'Analyse en Composantes Indépendantes (ICA), et la Séparation Aveugle de Sources basée sur l'Analyse de Corrélation Canonique (CCA), ont été exploitées pour l'élimination des artéfacts musculaires affectant les tracés électroencéphalographiques (EEG) de surface des patients épileptiques. Plus précisément, on s'intéresse ici au débruitage des activités rapides de bas voltage survenant au début des décharges critiques. Ces méthodes ont été évaluées, à la fois, sur des données simulées et sur des données réelles enregistrées chez un patient épileptique. Les résultats obtenus montrent qu'un algorithme ICA, à savoir CoM2 et la CCA sont des outils efficaces dans le contexte étudié.

**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 the contamination epileptic signals (rapid ictal discharges) by muscle artifact. In this context our aim was to compare the ability of Independent Component Analysis (ICA) (ten classical ICA methods) 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 the noisy data are best corrected with one ICA algorithm, namely CoM2 and CCA.

## 1 Introduction

L'EEG de surface a été le premier outil employé pour le diagnostic de l'épilepsie. C'est grâce à cette méthode que l'on peut visualiser les paroxysmes épileptiques qui sont le reflet d'une hyperexcitabilité neuronale et traduisent la survenue de décharges électriques neuronales brèves, soudaines et excessives au sein d'une ou de plusieurs régions cérébrales. Comme ces régions sont variables d'un patient à un autre, les manifestations cliniques associées aux crises sont donc variables et dépendent de la localisation des régions épileptiques. L'examen EEG, couplé à l'enregistrement vidéo des crises (Vidéo-EEG), est réalisé en continu pendant plusieurs jours au cours de la première phase du bilan préchirurgical des épilepsies partielles réfractaires, a pour objectif de recueillir les activités EEG critiques (décharges critiques) qui surviennent pendant les crises. Elles se caractérisent par des activités rythmiques dont la morphologie varie d'une part en fonction du type d'épilepsie (partielle vs. généralisée par exemple) et d'autre part dans le temps à mesure que l'activité critique se propage. Dans cette communication, on s'intéresse plus particulièrement à l'extraction de l'activité rapide et de bas voltage qui survient au tout début des décharges critiques. Contrairement aux autres activités intercritiques qui sont pour la plupart du temps un reflet d'un réseau de structure plus étendu que la zone épileptogène, cette activité rapide a été considérée comme la signature EEG de la région d'où la crise

démarre [5], et dont la résection est corrélée au bon pronostic postchirurgical [6]. Néanmoins, cette activité est difficilement observable en EEG de scalp du fait de sa faible amplitude et donc de son rapport signal à bruit peu favorable. En effet, les signaux EEG de surface peuvent être considérablement affectés par des artéfacts variés, tels que, les mouvements oculaires, l'électrocardiogramme, l'activité musculaire. Parmi tous ces artéfacts, l'élimination de l'activité musculaire est particulièrement ardue [3]. Ceci peut être essentiellement attribué au fait que : *i*) le muscle est largement distribué à la fois au niveau spatial et au niveau des bandes de fréquence, et *ii*) par rapport aux autres artéfacts, le muscle est moins stéréotypé.

Dans cette étude, nous proposons une analyse en deux étapes. La première étape cherche à comparer dix algorithmes ICA (parmi les plus connues dans la littérature) sur des signaux simulés. Cette première étape a pour but de choisir l'algorithme ICA le plus performant et le moins coûteux pour les signaux considérés. La deuxième étape présente les résultats quantitatifs et qualitatifs de la comparaison entre le meilleur algorithme ICA et la méthode CCA.

## 2 Méthodologie

### 2.1 Modèle de mélange

Le problème traité dans cette étude peut être formulé comme un problème de Séparation Aveugle de Source

(SAS). Le signal EEG enregistré à partir de  $N$  électrodes de surface peut alors être modélisé comme un mélange linéaire instantané de  $P$  sources élémentaires [7] :

$$\mathbf{x}[k] = \mathbf{A}\mathbf{s}[k] + \mathbf{v}[k]$$

où  $\mathbf{x}[k]$  est le vecteur des observations, représentant les  $N$  activités électriques enregistrées, à l'instant  $k$ , au niveau des électrodes de surface,  $\mathbf{s}[k]$  est associé à l'ensemble des sources du mélange (activités épileptiques, activités musculaires et EEG de fond),  $\mathbf{A}$  est la matrice de mélange globale, de taille  $(N \times P)$  et  $\mathbf{v}[k]$  est le vecteur décrivant le bruit d'instrumentation, de longueur  $N$  et supposé Gaussien. L'objectif de la SAS est d'estimer le vecteur source  $\mathbf{s}[k]$  et/ou d'identifier la matrice de mélange  $\mathbf{A}$  sans aucune information *a priori* sur les sources et le mélange. La méthode d'analyse du signal exploitée pour résoudre le problème de SAS dépend, essentiellement, des hypothèses qui seront admises sur les signaux sources.

## 2.2 Analyse en Composantes Indépendantes

L'ICA est une analyse multivariée dont le concept a émergé dans les années 1980-90[1]. L'ICA d'un  $N$ -vecteur aléatoire  $\mathbf{x}[k]$  se définit comme la recherche de  $P$  ( $P \leq N$ ) composantes linéaires  $s_p[k]$  qui soient "aussi indépendantes que possible". En d'autres termes, l'ICA cherche à trouver une matrice  $\mathbf{W}$  (dite de séparation), de telle sorte que le signal obtenu  $\mathbf{y}[k] = \mathbf{W}\mathbf{x}[k]$  soit une estimation du vecteur source  $\mathbf{s}[k]$ . Plusieurs algorithmes d'ICA ont été proposés dans la littérature [8, 9]. Notre choix, non exhaustif, s'est porté sur dix algorithmes d'ICA [8, 9] : SOBI, CoM2, JADE deux version de FastICA, InfoMax, PICA, ICAR, SIMBEC et 4-CANDHAP. Notons que les algorithmes InfoMax, ICAR, SIMBEC et 4-CANDHAP peuvent être appliqués sans étape de blanchiment préalable des observations. Cependant, les tests préliminaires que nous avons effectués montrent qu'un blanchiment spatial des observations améliore sensiblement les performances de ces méthodes dans le contexte de débruitage de l'EEG.

## 2.3 Analyse de Corrélation Canonique

Initialement proposée dans [10], la CCA est une méthode statistique permettant d'exploiter les relations qui peuvent exister entre deux vecteurs aléatoires, telles que : *i*) une décorrélation des composantes d'un même vecteur, et *ii*) une corrélation deux à deux des composantes des deux vecteurs. Friman et al. [2] ont montré que la CCA peut être exploitée pour résoudre le problème de SAS en prenant par exemple comme premier vecteur aléatoire, le vecteur source et comme second vecteur aléatoire, une version retardée du vecteur source. La CCA vise à séparer les sources d'intérêt du mélange présenté ci-dessus en minimisant leur corrélation spatiale et en maximisant leur autocorrélation temporelle.

## 2.4 Génération de données

Les données EEG simulées, sont générées en utilisant un modèle réaliste de genèse de l'EEG décrit dans [4]. Ce modèle combine une description réaliste des sources (maillage du cortex cérébral et champ dipolaire associé) et une description réaliste de l'activité des dipôles (modèle de populations de neurones). Les données EEG ont été simulées au niveau de 32 électrodes de scalp à partir d'une source unique ou « patch » de 5cm<sup>2</sup> situé dans le gyrus temporal supérieur gauche. Des activités épileptiques critiques, générées par le modèle de populations de neurones, sont attribuées aux dipôles du patch, L'EEG est ensuite généré en résolvant le problème direct (boundary element method, BEM) dans un modèle réaliste de la tête constitué d'un maillage du cerveau, du crâne et du scalp. 50 réalisations de ces EEG simulés sont générées pour l'étude. A chaque réalisation on ajoute une activité musculaire différente issue de tracés EEG de scalp réels. Par ailleurs, les algorithmes ont été testés sur un exemple de données critiques réelles.

## 2.5 Critère de performance

La performance des méthodes étudiées est évaluée en calculant l'Erreur Quadratique Moyenne Normalisée (EQMN), pour chacune des 32 électrodes, entre l'observation EEG initiale (EEG sans activité musculaire) est l'EEG de surface reconstruit après débruitage.

## 2.6 Sélection des sources d'intérêt

Avant de décrire les résultats obtenus, il est important de préciser que la sélection des sources d'intérêts est basée sur une inspection visuelle des composantes extraites pas chaque algorithme, dans le domaine temps-fréquence. En effet, la signature de l'activité rapide critique est bande étroite (type « chirp »), alors que celle de l'activité musculaire est large bande. Nous avons donc sélectionné les composantes d'intérêt visuellement en nous basant sur l'existence de la signature bande étroite. Le signal est ensuite reconstruit en ne gardant que les composantes correspondant aux activités critiques.

## 3 Résultats

### 3.1 Performances des dix algorithmes d'ICA

Cette première étude a pour but de choisir le meilleur algorithme d'ICA en termes de compromis entre performance et complexité numérique. Rappelons qu'un blanchiment spatial a été utilisé comme prétraitement pour les 10 algorithmes, bien qu'il ne soit pas nécessaire dans tous les cas ; ainsi nous définissons deux catégories : *i*) les algorithmes JADE, CoM2, SOBI, PICA, FastICA<sub>sym</sub> et FastICA<sub>def</sub> qui comportent intrinsèquement un blanchiment spatial, et *ii*) les algorithmes ICAR, 4-CANDHAP, InfoMax et SIMBEC pour lesquels un blanchiment sans réduction de

dimension a été utilisé afin d'améliorer les résultats, bien qu'il ne soit pas nécessaire.

### 3.1.1 Evaluation sur simulations réalistes

Pour cette étude, le nombre d'échantillons est fixé à 8192 (ce qui correspond à un signal de 32s à 256 Hz), les valeurs de RSB varient entre -30 dB et -5 dB avec un pas de 5 dB. La figure 1 illustre les valeurs d'EQMN obtenues pour les dix algorithmes d'ICA en fonction du RSB. Concernant le sous-groupe des algorithmes qui exigent un blanchiment spatial (figure 1.A), toutes les méthodes montrent des performances équivalentes pour des valeurs de RSB supérieures ou égales à -20 dB. Lorsque les signaux sont plus bruités (-25 et -30 dB) CoM2, FastICA<sub>sym</sub> et SOBI offrent les meilleures performances, et JADE, FastICA<sub>def</sub> et PICA les moins bonnes. Dans le cas des algorithmes qui utilisent un blanchiment sans réduction de dimension (figure 1.B) InfoMax et SIMBEC sont les méthodes les plus performantes, SIMBEC surpassant légèrement InfoMax pour des valeurs très faibles de RSB, alors que les moins bonnes performances sont obtenues pour ICAR. De plus, une évaluation de la complexité numérique de ces 10 algorithmes (non détaillée ici par manque de place) montre que CoM2 et InfoMax offrent le meilleur compromis entre performance et complexité numérique.

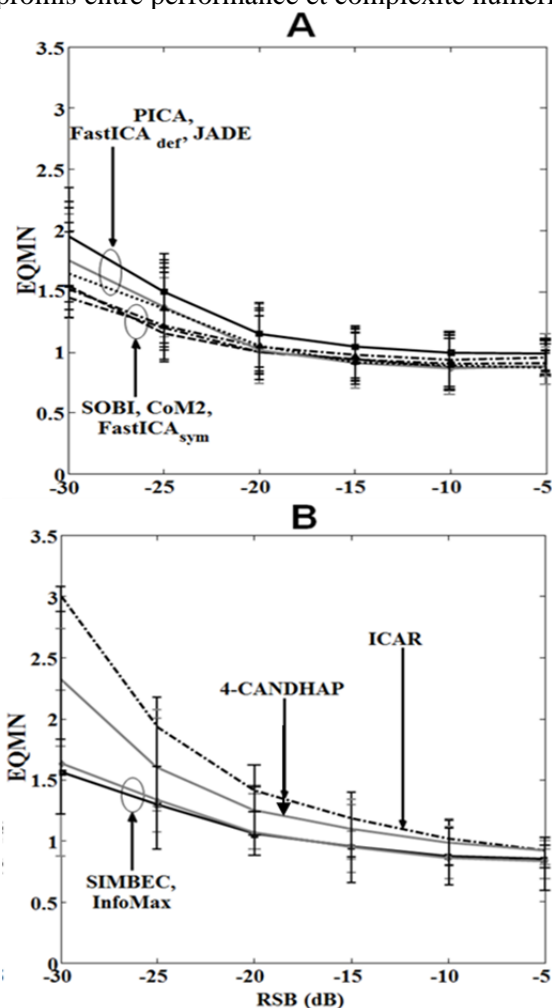


Figure 1 : EQMN calculées, à la sortie des 10 algorithmes d'ICA, pour : (A) méthodes qui comportent intrinsèquement un blanchiment spatial, et (B) méthodes pour lesquels un blanchiment sans réduction de dimension a été utilisé.

## 3.2 Comparaison entre CoM2, InfoMax et CCA

Sur la base des résultats obtenus dans la section précédente, nous confrontons les performances des deux meilleurs algorithmes d'ICA (CoM2 et InfoMax) précédemment identifiés avec celles de la CCA.

### 3.2.1 Evaluation sur simulations réalistes

Le calcul de l'EQMN est fait d'une part pour toutes les électrodes, et d'autre part pour l'électrode sur laquelle l'activité critique d'intérêt apparaît avec une énergie maximale (i.e. électrode T3 qui se trouve en face du patch épileptique). Les figures 2.A et 2.B, montrent que la performance de CCA reste sensiblement stable pour tous les RSB considérés alors que l'efficacité des deux autres méthodes s'améliore lorsque le RSB augmente. CCA présente des performances significativement supérieures à celle de CoM2 et InfoMax pour des RSB faibles (-30 dB à -20 dB), supérieures à InfoMax pour un RSB de -15 dB et comparables à CoM2 et InfoMax pour des RSB forts. Notons que CoM2 est significativement plus performant qu'InfoMax pour les RSB de -25 dB et -20 dB.

### 3.2.2 Application aux signaux réels

Afin d'évaluer les performances de CoM2, InfoMax et CCA sur des données réelles, nous avons identifié sur un patient épileptique une première période critique, peu contaminée par les artefacts musculaires ; cette activité critique sera utilisée comme référence. Une seconde crise a également été enregistrée chez ce patient. Cette fois, l'EEG au cours de cette seconde crise est contaminé par de l'activité musculaire. Au cours de la période critique de référence non bruitée (figure 3.A) on distingue la survenue de pointes rythmiques qui culminent à gauche au niveau de l'électrode C3. Cette activité est suivie par une décharge rapide visible au niveau des mêmes électrodes. Puis cette activité devient ensuite plus irrégulière et d'amplitude plus ample. Au cours de la période critique « bruitée » (figure 3.B) on distingue encore une décharge rythmique de pointes au niveau de l'électrode C3, en revanche la suite du tracé critique est entièrement masquée par les artefacts musculaires.

La figure 3.C montre qu'on restitue bien la signature temps fréquence de la décharge rapide après débruitage par CCA. Avec CoM2 (figure 3.D), la signature est bien visible également, mais il reste un peu d'activité musculaire pendant la décharge de pointes rythmiques précédant la décharge rapide. Enfin, après débruitage par InfoMax, l'activité musculaire résiduelle est encore plus importante (figure 3.D).

## 4 Conclusion

Les résultats préliminaires montrent que la méthode CCA est certaines méthodes d'ICA (CoM2) sont des outils efficaces dans le contexte de débruitage de l'activité rapide et de bas voltage survenant au début des décharges critiques. Les résultats sur signaux réels

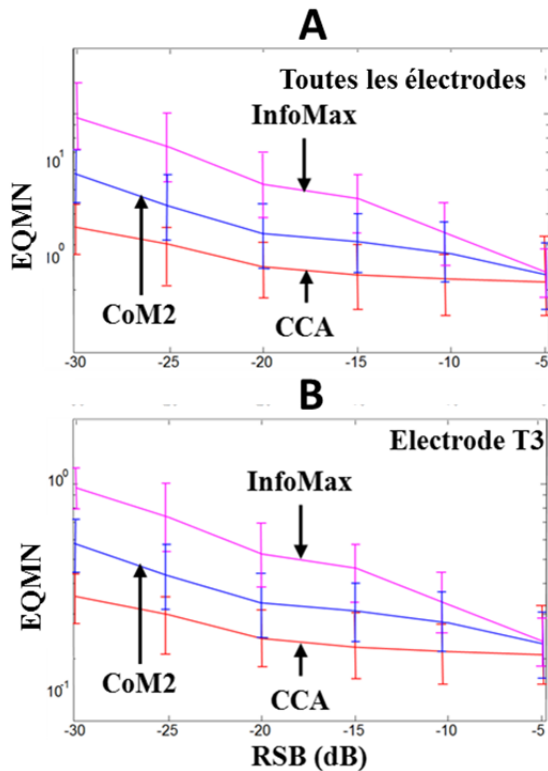


Figure 2 : EQMN calculées, à la sortie de CoM2, InfoMax et CCA, pour : (A) l'électrode T3 sur laquelle apparaît l'activité rapide avec une énergie maximale, et (B) toutes les 32 électrodes

indiquent la capacité des méthodes SAS à éliminer les artefacts musculaires sans affecter les activités critiques. Ces résultats prometteurs, devraient permettre de proposer une méthode de prétraitement des données EEG de scalp visant à améliorer *i)* l'analyse visuelle des tracés EEG critiques et, de fait, leur interprétation en routine clinique, et *ii)* le résultat issu de méthodes quantitatives de traitement de ces signaux comme par exemple la localisation de source ou l'analyse des relations statistiques entre voies.

### Références

[1] P. Comon, "Independent component analysis, A new concept?," *Signal Processing*, vol. 36, pp. 287-314, 1994.

[2] O. Friman, M. Borga, P. Lundberg, H. Knutsson, Exploratory fMRI analysis by autocorrelation maximization, *NeuroImage*, vol. 16, no. 2, pp. 454-464, 2002.

[3] B. W. McMenamin, A. J. Shackman, J. S. Maxwell, D. R. W. Bachhuber, A. M. Koppenhaver, L. L. Greischar, and R. J. Davidson, "Validation of ICA-based myogenic artifact correction for scalp and source-localized EEG," *Neuroimage*, vol. 49, pp. 2416-2432, 2011.

[4] D. Cosandier-Rimele, J. M. Badier, P. Chauvel, F. Wendling, "A physiologically plausible spatio-temporal model for EEG signals recorded with intracerebral electrodes in human partial epilepsy," *IEEE Trans Biomed Eng.*, vol. 54, pp. 380-8, Mar 2007.

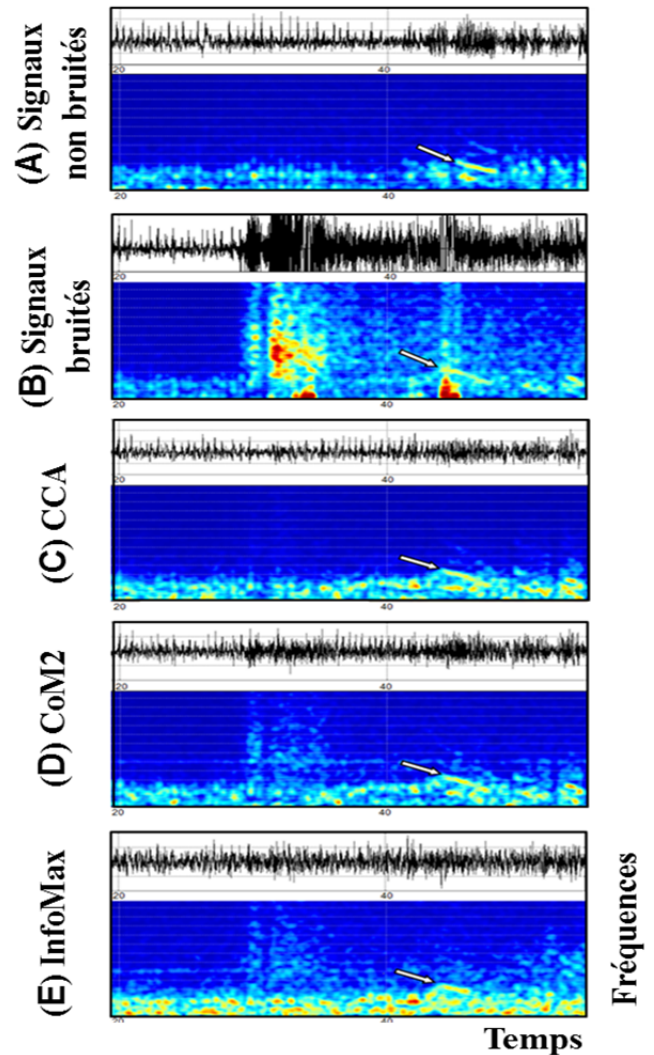


Figure 3 : Exemple de débruitage des données EEG réelles enregistrées chez un patient épileptique pendant la période intercritique-critique avec les cartes temps-fréquence correspondantes.

[5] J. Bancaud, J. Talairach, C. Schaub, G. Szikla, A. Bonis, "Stereotactic functional exploration of the epilepsies of the supplementary area of the mesial surfaces of the hemispheres," *Electroencephalogr Clin Neurophysiol*, 14: 788, 1962.

[6] G. Alarcon, C. Binnie, R. Elwes C. Polkey, "Power spectrum and intracranial EEG patterns at seizure onset in partial epilepsy." *Electroencephalogr Clin Neurophysiol*, vol. 94, no. 5, pp. 326-337, 1995.

[7] J. Sarvas, "Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem," *Phys Med Biol*, vol. 32, no. 1, pp. 11-22, 1987.

[8] A. Hyvarinen, J. Karhunen, P. Oja, *Independent component analysis*, ser. Wiley interscience. John Wiley and Sons, 2001.

[9] 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.

[10] H. Hotelling, Relations between two sets of variates, *Biometrika*, vol. 28, pp. 321-377, 1936.