HAL
open science

Suppression du bruit de signaux EEG épileptiques

Sepideh Hajipour Sardouie

To cite this version:

Sepideh Hajipour Sardouie. Suppression du bruit de signaux EEG épileptiques. Signal and Image processing. Université de Rennes; Sharif University of Technology (Tehran), 2014. English. NNT : 2014REN1S185 . tel-01493660

HAL Id: tel-01493660

https://theses.hal.science/tel-01493660v1

Submitted on 21 Mar 2017

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.

N^o d'ordre : ANNÉE 2014

THÈSE / UNIVERSITÉ DE RENNES 1
sous le sceau de l'Université Européenne de Bretagne

En Cotutelle Internationale avec

UNIVERSITÉ TECHNOLOGIQUE SHARIF, IRAN

pour le grade de

DOCTEUR DE L'UNIVERSITÉ DE RENNES 1

Mention : Traitement du Signal et Télécommunications

École doctorale MATISSE

présentée par

Sepideh HAJIPOUR SARDOUIE

préparée au LTSI - INSERM UMR 1099
Laboratoire de Traitement du Signal et de l'Image
et au BiSIPL
Biomedical Signal and Image Processing Laboratory

Signal subspace identification for epileptic source localization from EEG data

Thèse soutenue à Rennes le 9 Octobre 2014

devant le jury composé de :

Christian Jutten
Professeur des universités / *rapporteur*

Vicente Zarzoso
Professeur des universités / *rapporteur*

Emad Fatemizadeh
Assistant Professor / *examinateur*

Laurent Albera
Maître de conférences / *directeur*

Mohammad B. Shamsollahi
Professeur des universités / *co-directeur*

Isabelle Merlet
Chargée de Recherche INSERM / *co-directeur*

Signal subspace identification for epileptic source localization from EEG data

A joint thesis submitted for the degree of

Doctor of Philosophy

From

University of Rennes 1

&

Sharif University of Technology

In

Signal Processing and Telecommunication

&

Biomedical Engineering

By

Sepideh HAJIPOUR SARDOUIE

JURY MEMBERS AND DIRECTORS

Reporters:	Prof. Christian JUTTEN	University of Joseph Fourier
	Prof. Vicente ZARZOSO	University of Nice Sophia-Antipolis
Evaluator:	Dr. Emad FATEMIZADEH	Sharif University of Technology
Directors:	Dr. Laurent ALBERA	University of Rennes 1
	Dr. Isabelle MERLET	University of Rennes 1
	Prof. Mohammad B. SHAMSOLLAHI	Sharif University of Technology

Dedicated to my dear parents

Acknowledgements

I would like to acknowledge with considerable gratitude the contributions and help received from:

Prof. *Mohammad Bagher Shamsollahi* for his continuous support and kind supervision, and for all the things he taught me during the past nine years. He has been far beyond a supervisor for me and I hope to carry the lessons that I learned from him in my personal life.

Dr. *Laurent Albera* for his constant guidance and inspiring discussions that let me develop my analytical knowledge in signal processing. I would also like to thank him for the great effort he put in revising my thesis report and papers.

Dr. *Isabelle Merlet* for her great help and supervision, especially for introducing me to the field of neurophysiology. I would also like to thank her for providing me with a great set of simulated and real epileptic data.

My committee members: Prof. Christian Jutten, Prof. Vicente Zarzoso and Dr. Emad Fatemizadeh for the evaluation of this work and their invaluable comments.

Prof. *Lotfi Senhadji* for giving me the opportunity to become a member of LTSI.

My dear *Father* for his great encouragement and support throughout my entire life. I wish he were among us to see my graduation.

My dear *Mother* for all the love, devotion and encouragement she gave me from the day I was born to this day.

My dear husband *Mahmoud* for his unconditional support. His encouragement, patience and unwavering love helped me to accomplish my research.

My dear brother *Alireza* for his never-ending cheerful encouragement.

All the teachers and professors that I had throughout my education, for everything that I learned from them.

My dear friends for all their help and supports and for all the good time we had together.

The *Embassy of France in Tehran* for providing a one-year research scholarship.

Abstract

In the process of recording electrical activity of the brain, the signal of interest is usually contaminated with different activities arising from various sources of noise and artifact such as muscle activity. This renders denoising as an important preprocessing stage in some ElectroEncephaloGraphy (EEG) applications such as source localization. In this thesis, we propose six methods for noise cancelation of epileptic signals.

The first two methods, which are based on Generalized EigenValue Decomposition (GEVD) and Denoising Source Separation (DSS) frameworks, are used to denoise interictal data. To extract *a priori* information required by GEVD and DSS, we propose a series of preprocessing stages including spike peak detection, extraction of exact time support of spikes and clustering of spikes involved in each source of interest. Two other methods, called Time Frequency (TF)-GEVD and TF-DSS, are also proposed in order to denoise ictal EEG signals for which the time-frequency signature is extracted using the Canonical Correlation Analysis method. We also propose a deflationary Independent Component Analysis (ICA) method, called JDICA, that is based on Jacobi-like iterations. Moreover, we propose a new direct algorithm, called SSD-CP, to compute the Canonical Polyadic (CP) decomposition of complex-valued multi-way arrays. The proposed algorithm is based on the Simultaneous Schur Decomposition (SSD) of particular matrices derived from the array to process. We also propose a new Jacobi-like algorithm to calculate the SSD of several complex-valued matrices. The last two algorithms are used to denoise both interictal and ictal data.

We evaluate the performance of the proposed methods to denoise both simulated and real epileptic EEG data with interictal or ictal activity contaminated with muscular activity. In the case of simulated data, the effectiveness of the proposed algorithms is evaluated in terms of relative root mean square error between the original noise-free signals and the denoised ones, number of required flops and the location of the original and denoised epileptic sources. For both interictal and ictal data, we present some examples on real data recorded in patients with a drug-resistant partial epilepsy.

Contents

Résumé **5**

Résumé long en français **7**

Acronyms and notations **13**

1 Introduction **17**
1.1 EEG and brain electrogenesis 17
1.2 Epilepsy and epileptic EEG paroxysms 19
1.2.1 Definition and epidemiological data 19
1.2.2 Epileptic EEG paroxysms 20
1.3 Sources of disturbance 21
1.4 Problem formulation 24
1.4.1 Statistical problem formulation 24
1.4.2 Deterministic problem formulation 25
1.5 Previous work 25
1.5.1 Blind methods 26
1.5.2 Semi-blind methods 37
1.5.3 Other methods 39
1.6 The outline of the thesis report 39
1.7 List of publications 39

2 Algorithms **45**
2.1 Semi-blind GEVD/DSS based methods 45
2.1.1 Generalized Eigenvalue Decomposition 45
2.1.2 The Denoising Source Separation framework 46
2.1.3 The GEVD and DSS based methods for denoising of interictal data 47
2.1.4 The TF-GEVD and TF-DSS methods for denoising of ictal data 55
2.2 The JDICA method 61
2.2.1 The JDICA algorithm: methodology 62
2.2.2 Denoising epileptic EEG signals using the JDICA algorithm . . 65
2.3 The SSD-CP method 65
2.3.1 Complex Simultaneous Schur Decomposition: the SSD technique 67
2.3.2 Direct CP decomposition: the SSD-CP method 70
2.3.3 Identifiability 76
2.3.4 Denoising epileptic EEG signals using the SSD-CP algorithm . . 78

2.4 The numerical complexity of the proposed algorithms 78
2.4.1 The GEVD/DSS methods to denoise interictal EEGs 79
2.4.2 The GEVD/DSS-based methods to denoise ictal EEGs 79
2.4.3 The JDICA algorithm 80
2.4.4 The SSD-CP algorithm 80
2.5 Summary 81

3 Data Generation 85
3.1 Simulated data 85
3.1.1 The source model 85
3.1.2 Construction of the intensity matrix 89
3.1.3 Genesis of simulated EEG signals 91
3.2 Real data 92
3.3 Performance criteria 93
3.4 Summary 97

4 Experimental Results 99
4.1 Common examples for testing the proposed SSD-CP method 99
4.1.1 A $(2 \times 2 \times 2)$ array of rank 2 100
4.1.2 Double bottlenecks in a $(30 \times 30 \times 30)$ array of rank 4 100
4.1.3 Triple bottlenecks in a $(30 \times 30 \times 30)$ array of rank 4 103
4.1.4 Multiway analysis of amino acids fluorescence data 104
4.2 EEG simulated data 106
4.2.1 Interictal data 106
4.2.2 Ictal data 115
4.3 Real data 130
4.3.1 Interictal data 130
4.3.2 Ictal data 136
4.4 Summary 136

5 Discussion, Conclusion and Perspectives 139
5.1 Discussion and Conclusion 139
5.1.1 SSD-CP and numerical experiments 139
5.1.2 Simulated interictal data 139
5.1.3 Simulated ictal data 141
5.1.4 Real interictal data 142
5.1.5 Real ictal data 142
5.2 Future work 143

A Deriving the DSS framework from the EM algorithm 145

B The wavelet-based spike peak detection algorithm 149

C Values of coefficients in the JDICA algorithm 151

D Values of coefficients in the SSD-CP algorithm 153

Résumé

Lorsque l'on enregistre l'activité cérébrale en électroencéphalographie (EEG) de surface, le signal d'intérêt est fréquemment bruité par des activités différentes provenant de différentes sources de bruit telles que l'activité musculaire. Le débruitage de l'EEG est donc une étape de pré-traitement important dans certaines applications, telles que la localisation de source. Dans cette thèse, nous proposons six méthodes permettant la suppression du bruit de signaux EEG dans le cas particulier des activités enregistrées chez les patients épileptiques soit en période intercritique (pointes) soit en période critique (décharges).

Les deux premières méthodes, qui sont fondées sur la décomposition généralisée en valeurs propres (GEVD) et sur le débruitage par séparation de sources (DSS), sont utilisées pour débruiter des signaux EEG épileptiques intercritiques. Pour extraire l'information a priori requise par GEVD et DSS, nous proposons une série d'étapes de prétraitement, comprenant la détection de pointes, l'extraction du support des pointes et le regroupement des pointes impliquées dans chaque source d'intérêt. Deux autres méthodes, appelées Temps Fréquence (TF) -GEVD et TF-DSS, sont également proposées afin de débruiter les signaux EEG critiques. Dans ce cas on extrait la signature temps-fréquence de la décharge critique par la méthode d'analyse de corrélation canonique. Nous proposons également une méthode d'Analyse en Composantes Indépendantes (ICA), appelé JDICA, basée sur une stratégie d'optimisation de type Jacobi. De plus, nous proposons un nouvel algorithme direct de décomposition canonique polyadique (CP), appelé SSD-CP, pour calculer la décomposition CP de tableaux à valeurs complexes. L'algorithme proposé est basé sur la décomposition de Schur simultanée (SSD) de matrices particulières dérivées du tableau à traiter. Nous proposons également un nouvel algorithme pour calculer la SSD de plusieurs matrices à valeurs complexes. Les deux derniers algorithmes sont utilisés pour débruiter des données intercritiques et critiques.

Nous évaluons la performance des méthodes proposées pour débruiter les signaux EEG (simulés ou réels) présentant des activités intercritiques et critiques épileptiques bruitées par des artéfacts musculaires. Dans le cas des données simulées, l'efficacité de chacune de ces méthodes est évaluée d'une part en calculant l'erreur quadratique moyenne normalisée entre les signaux originaux et débruités, et d'autre part en comparant les résultats de localisation de sources, obtenus à partir des signaux non bruités, bruités, et débruités. Pour les données intercritiques et critiques, nous présentons également quelques exemples sur données réelles enregistrées chez des patients souffrant d'épilepsie partielle.

Résumé long en français

Lorsque l'on enregistre l'activité cérébrale en électroencéphalographie (EEG) de surface, le signal d'intérêt est fréquemment bruité par des activités différentes provenant de différentes sources de bruit. Certains de ces bruits proviennent de sources externes au cerveau comme les interférences dues au réseau électrique ou le bruit lié à l'instrumentation (mauvais contact des électrodes par exemple). En outre, certains bruits peuvent aussi provenir de sources physiologiques, externes au cerveau, telles que le clignement des yeux, les mouvements oculaires lents, l'activité musculaire et les battements cardiaques. Dans certaines applications, une partie de l'activité cérébrale elle-même, comme l'EEG de fond, est considérée comme un bruit qui doit être supprimé. Le débruitage de l'EEG est donc une étape de pré-traitement importante à la fois pour permettre une meilleure interprétation qualitative des tracés par les cliniciens et pour améliorer la précision de certaines applications, telles que la localisation de source. Dans cette thèse, nous proposons six méthodes permettant la suppression du bruit de signaux EEG dans le cas particulier des activités enregistrées chez les patients épileptiques soit en période intercritique (pointes) soit en période critique (décharges).

EEG et Epilepsie

L'électroencéphalographie consiste à enregistrer l'activité électrique cérébrale au moyen d'électrodes posées à la surface de la tête au niveau du scalp (on parle d'EEG de surface ou d'EEG de scalp). Les potentiels électriques enregistrés en surface du scalp sont le reflet de l'activité des neurones corticaux sous-jacents. Pour que les potentiels puissent être enregistrés à la surface du scalp il faut qu'ils résultent de l'activation synchrone et simultanée d'un grand nombre de neurones.

L'épilepsie est une affection neurologique invalidante, responsable de la survenue répétitive d'épisodes paroxystiques brefs, ou "crises", qui peuvent comporter des manifestions motrices, sensorielles ou psychiques irrépressibles et peuvent s'accompagner d'une altération partielle ou totale de la conscience. Selon l'OMS, la prévalence de l'épilepsie (c'est à dire le nombre de cas à un moment donné et sur un groupe donné) est estimée entre 0,5 et 0,8Certains types d'épilepsies, dites pharmaco-résistantes ou réfractaires, sont totalement résistantes aux traitements médicamenteux. Elles sont souvent d'origine focale ou partielle, c'est-à-dire qu'elles prennent leur origine dans une région cérébrale relativement circonscrite du cerveau. On peut alors dans ce cas proposer au patient un traitement chirurgical visant à réséquer de façon sélective cette région, appelée " zone épileptogène " afin de mettre un terme aux crises, à condition évidemment que les déficits potentiels induits par la chirurgie restent acceptables. La

délimitation aussi précise que possible de la zone épileptogène constitue l'objectif ultime du bilan préchirurgical et repose sur la confrontation des résultats issus de diverses explorations cliniques, anatomiques et fonctionnelles plus ou moins invasives.

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 a pu 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. Divers paroxysmes EEG épileptiques peuvent être collectés:

i) Les paroxysmes EEG intercritiques surviennent en dehors des crises de façon inconstante et très variable selon les patients et les types d'épilepsie et sont en général d'aspect polymorphe. Il peut s'agir de pointes, pointes ondes, polypointes, polypointes-onde isolées ou en salves. Les régions d'où proviennent ces paroxysmes se regroupent sous le terme de "zone irritative". Ces paroxysmes sont sans conteste un marqueur de l'épilepsie mais les relations spatiales qu'ils entretiennent avec la zone épileptogène sont parfois mal définies et en tous cas variables d'un patient à l'autre.

ii) Les paroxysmes EEG épileptiques critiques, ou décharges critiques, 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. On peut par exemple enregistrer au cours des crises des décharges de pointes, des pointes ondes rythmiques, des activités rapides ou des ondes lentes rythmiques. La zone épileptogène est définie à l'origine comme la région cérébrale nécessaire au déclenchement des décharges critiques habituelles d'un patient.

L'objectif principal de la thèse

Dans le cadre de l'enregistrement de l'activité EEG épileptique, le signal enregistré peut être considéré comme une combinaison (i) du signal d'intérêt provenant des régions épileptiques du cerveau et (ii) du signal de non-intérêt composé de différents types de bruits, de l'activité de fond et des artefacts. L'objectif de ce rapport est de proposer des méthodes qui extraient le signal d'intérêt à partir du signal enregistré, c'est-à-dire qui débruitent les activités EEG épileptiques. Plus particulièrement, l'objectif principal de cette thèse est d'éliminer l'activité musculaire (parfois appelée activité myogénique) qui peut contaminer les activités épileptiques intercritiques et critiques enregistrées en EEG de surface.

Les algorithmes proposés

Dans cette thèse, les méthodes proposées pour le débruitage des signaux EEG peut être classées en trois familles.

I) Des méthodes semi-aveugles basées sur GEVD/DSS:

Des méthodes semi-aveugles qui exploitent la connaissance a priori d'informations sur les sources d'intérêt. Il s'agit par exemple du support temporel des pointes intercritiques ou du support temps-fréquence des décharges critiques. Essentiellement, la source est extraite en maximisant sa puissance relative sur le support estimé. Mathématiquement, les critères qui en découlent sont exprimés sous forme de quotients de Rayleigh, et maximisés de manière algébrique en calculant des vecteurs propres généralisés (generalized eigenvalue decomposition, GEVD). Le débruitage par séparation de sources (denoising source separation, DSS) incorpore l'information a priori sur le support de la source d'intérêt afin d'en améliorer l'estimation en utilisant des itérations du type expectation-maximization (EM).

Concernant l'activité intercritique, pour extraire l'information a priori requise par GEVD et DSS, nous proposons une série d'étapes de prétraitement, comprenant la détection de pointes, l'extraction du support temporal des pointes et le regroupement des pointes impliquées dans chaque source d'intérêt. Ces informations sont ensuite utilisées dans GEVD ou DSS pour extraire les sources de l'activité intercritique. Deux autres méthodes, appelées Temps Fréquence (TF)-GEVD et TF-DSS, sont également proposées afin de débruiter les signaux EEG critiques. Dans ce cas on extrait la signature temps-fréquence de la décharge critique par la méthode d'analyse de corrélation canonique (CCA).

II) JDICA :

Nous proposons également une méthode d'analyse en composantes indépendantes (ICA) à déflation, appelé JDICA, basée sur une stratégie d'optimisation de type Jacobi. Dans cette méthode, nous maximisons le critère de kurtosis en nous appuyant sur une paramétrisation particulière de la matrice séparatrice, unitaire après blanchiment. Cette paramétrisation, originalement proposée par Delfosse et Loubaton, a l'avantage de permettre la déflation de la source estimée (c'est-à-dire, la soustraction de la contribution de la source estimée au mélange) tout en réduisant la dimensionnalité du problème. Au lieu d'un algorithme de gradient utilisé dans l'implantation originale, la thèse propose un balayage de type Jacobi (d'où le nom de Jacobi-like deflationary ICA, JDICA) où chaque angle recherché est calculé comme l'une des racines d'un polynôme de degré 8. L'algorithme de débruitage proposé est testé sur données intercritiques et critiques.

III) SSD-CP :

La troisième famille méthodologique est basée sur l'exploitation du caractère trilinéaire du tableau espace x temps x fréquence construit a partir des observations spatio-temporelles. Cette approche nécessite de calculer une décomposition tensorielle du type canonique polyadique (CANDECOMP/PARAFAC, CP). Nous proposons une méthode directe, appelée SSD-CP, pour calculer la décomposition CP de tableaux à valeurs complexes. L'algorithme proposé est basé sur la décomposition simultanée de Schur (simultaneous Schur decomposition, SSD) de matrices particulières dérivées du tableau

d'ordre 3 à traiter. Nous proposons également un nouvel algorithme pour calculer la SSD de plusieurs matrices à valeurs complexes. L'algorithme SSD-CP est alors utilisé pour débruiter des données intercritiques et critiques.

Nous présentons également une étude détaillée du coût de calcul des différentes méthodes développées.

Donnée et résultats expérimentaux

Nous évaluons la performance des méthodes proposées pour débruiter les signaux EEG (simulés ou réels) présentant des activités intercritiques et critiques épileptiques bruitées par des artéfacts musculaires. Les signaux simulés sont issus d'un modèle réaliste de l'activité EEG de surface développé au LTSI, basé sur un modèle de populations de neurones et sur une description fine de la géométrie du néocortex. L'EEG est simulé en résolvant le problème direct à partir des sources synthétiques et un modèle réaliste de la tête humaine. Pour bruiter cet EEG simulé on lui ajoute de l'activité musculaire mêlée à de l'activité de fond issues d'un EEG réel. Les différents algorithmes proposés sont ensuite testés sur ces signaux synthétiques dans le cas de trois configurations de sources différentes : on considère une seule région (patch) neuronale d'activité épileptique, deux régions dont les activités sont décorrélées, et deux régions dont les activités sont corrélées. L'efficacité de chacune de ces méthodes est évaluée d'une part en calculant l'erreur quadratique moyenne normalisée entre les signaux originaux et débruités, et d'autre part en comparant les résultats de la localisation de sources, obtenus à partir des signaux non bruités, bruités, et débruités.

Parmi les nombreux résultats obtenus, on remarque la meilleure performance, dans la plupart des cas, des méthodes semi-aveugles basées sur la GEVD et la DSS, ce qui peut s'expliquer par le fait que ces méthodes exploitent explicitement l'information a priori sur les sources. Les performances de la SSD-CP sont relativement faibles si l'on considère l'estimation de la dynamique temporelle des sources, mais la localisation des sources épileptiques intercritiques est néanmoins satisfaisante pour cette méthode, probablement en raison d'une identification plus précise des signatures spatiales des activations neuronales obtenue par cette méthode.

Pour les données intercritiques et critiques, nous présentons également quelques exemples sur données réelles enregistrées chez des patients souffrant d'épilepsie partielle.

En plus des résultats mentionnés ci-dessus, la méthode de décomposition tensorielle directe SSD-CP est appliquée à des cas particuliers (présence de " bottleneck ", surestimation du rang du tenseur) qui posent problème aux algorithmes existants basés sur les moindres carrés alternés (ALS) ou sur l'itération de Levenverg-Marquardt. Dans tous ces cas difficiles, la SSD-CP montre des performances remarquables à un coût de calcul très raisonnable.

Perspectives

Pour la continuation du travail, nous proposons de quelques pistes telles que l'évaluation de performances sur une base de données des signaux réels, l'utilisation de techniques

plus performantes pour la détection et la classification de pics des pointes intercritiques, et le développement de modèles tensoriels alternatifs moins restrictifs que la décomposition CP.

Acronyms and notations

Acronyms

Models and methods

- **AF**: Adaptive Filtering
- **ALS**: Alternating Least Squares
- **BSS**: Blind Source Separation
- **BEMD**: Bivariate Empirical Mode Decomposition
- **CCA**: Canonical Correlation Analysis
- **CP**: CANDECOMP/PARAFAC model and Canonical Polyadic decomposition
- **DF**: Digital Filter
- **DSS**: Denoising Source Separation
- **ELS-ALS**: Enhanced Line Search-Alternating Least Squares
- **EM**: Expectation-Maximization
- **EMD**: Empirical Mode Decomposition
- **FLM**: Fast Levenberg-Marquardt
- **GEVD**: Generalized EigenValue Decomposition
- **ICA**: Independent Component Analysis
- **IMF**: Intrinsic Mode Function
- **JAD**: Joint Approximate Diagonalization
- **JDICA**: Jacobi-like Deflationary ICA
- **LM**: Levenberg-Marquardt
- **M-EMD**: Multivariate Empirical Mode Decomposition
- **MI**: Mutual Information

- **PCA**: Principal Components Analysis
- **RLS**: Recursive Least Square
- **SCICA**: Spatially-Constrained ICA
- **SOBI**: Second-Order Blind Identification
- **SSD**: Simultaneous Schur Decomposition
- **SVD**: Singular Value Decomposition
- **SVM**: Support Vector Machine
- **STF**: Space-Time-Frequency
- **STWV**: Space-Time-Wave-Vector
- **TF-DSS**: Time-Frequency DSS
- **TF-GEVD**: Time-Frequency GEVD
- **WD**: Wavelet Denoising
- **WPT**: Wavelet Packet Transform
- **WT**: Wavelet Transform

Signals

- **ECG**: ElectroCardioGram
- **EEG**: ElectroEncephaloGram
- **ECoG**: ElectroCorticoGram
- **EMG**: ElectroMyoGram
- **MCG**: MagnetoCardioGram
- **MEG**: MagnetoEncephaloGram

Notations

Typography

- vectors are denoted with lowercase letters ($\boldsymbol{a}, \boldsymbol{b}, \cdots$)
- matrices are denoted with uppercase letters ($\boldsymbol{A}, \boldsymbol{B}, \cdots$)
- HO arrays are denoted with calligraphic letters ($\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}}, \cdots$)
- the entry with row i and column j in a matrix $\boldsymbol{A}$ is symbolized by $A_{i,j}$

- the (i, j, k)-th component of a third order array $\boldsymbol{\mathcal{B}}$ is symbolized by $\mathcal{B}_{i,j,k}$
- $\hat{\boldsymbol{A}}$ denotes the estimate of $\boldsymbol{A}$

Variables

- M: number of observation channels
- N: number of time samples
- K: number of frequency samples
- P: number of all sources
- P': dimension of the source space
- P_e: number of epileptic sources
- P'_e: dimension of the epileptic subspace

Operators

- $\circ$: the outer product
- $\otimes$: the Kronecker product
- $^\sharp$: the Moore-Penrose pseudo-inverse
- $^\top$: the transpose operator
- $n(.)$: cardinality of a set
- E[.]: mathematical expectation
- $\langle . \rangle$: temporal mean

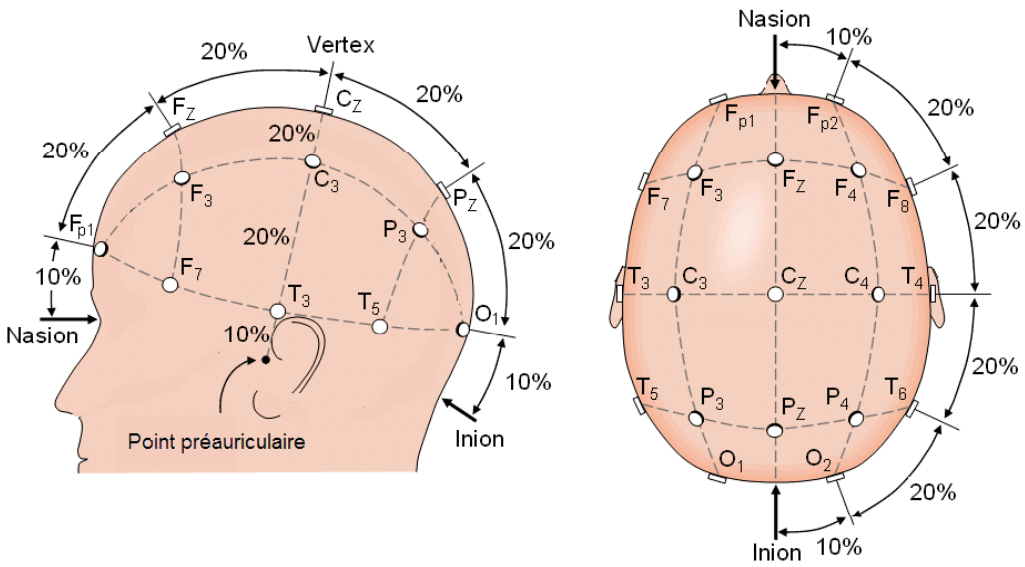
Chapter 1

Introduction

In recording brain activity, the signal of interest is often contaminated by different activities arising from various sources of noise and artifacts. Some of these artifacts are externally generated such as power line noise and instrumentation noise produced by inappropriate electrode connections. In addition to externally generated artifacts, noise can also arise from physiological sources, external to the brain, such as eye blinks, eye movements, muscle activity and heart pulse. In some applications, part of the brain activity itself such as brain background activity is considered as a noise that should be removed. This renders denoising as an important preprocessing stage in ElectroEncephaloGram (EEG) applications such as source localization or Brain Computer Interface (BCI). In this thesis, we focus on muscular noise cancelation of epileptic EEG data. Firstly, we introduce the EEG and brain electrogenesis. Then we specifically focus on Epilepsy and epileptic EEG paroxysms with an explanation of the sources of noise. Finally, the problem of denoising of epileptic EEG is formulated and previous work is examined.

1.1 EEG and brain electrogenesis

EEG consists of recording of brain activity using electrodes placed at the surface of the scalp (also called *surface EEG* or *scalp EEG*). The EEG represents the potential difference between the various electrodes taken in pairs (bipolar montage), between each electrode and a common reference electrode (unipolar or referential montage) or between each electrode and a virtual electrode representing for example the average activity on all recording channels (average reference montage). The number of electrodes varies widely (between 19 and 256 electrodes). They are symmetrically positioned on the two hemispheres, according to standard locations as shown in figure 1.1. These standard locations are defined by the 10-20 system [Jasper, 1958], the 10-10 system [Chatrian, 1985] or the 5-5 system [Oostenveld and Praamstra, 2001], with the number of electrodes increases from the first system to the third. There are other kinds of EEG recordings. For example, ElectroCorticoGraphy (ECoG) collects the intracranial electrical activity with electrodes positioned directly on the surface of the cerebral cortex, or on dura. Another example is Stereo-EEG (SEEG) that records intracerebral activity using multi-contact electrodes directly implanted in the brain tissue. In this work, we focus on the scalp EEG only.

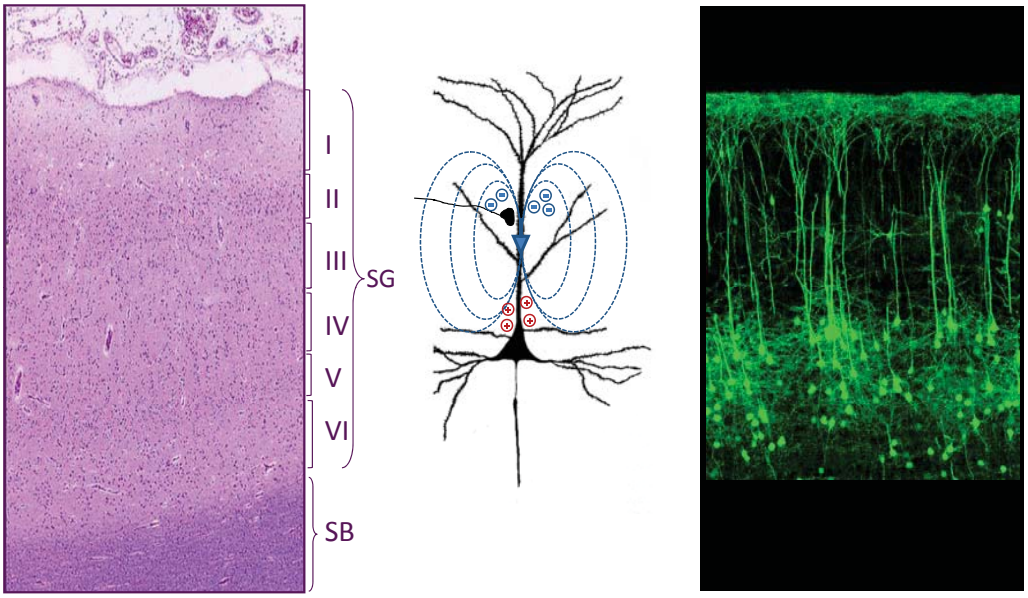


(a) (b)

Figure 1.1: The international 10-20 system seen from (a) left and (b) above the head. A=ear lobe, C=central, Pg=nasopharyngeal, P=parietal, F=frontal, Fp=frontal polar, O=occipital, T=temporal [Jasper, 1958].

Electrical potentials recorded at the scalp surface reflect the activity of underlying cortical neurons. In the cortical ribbon, postsynaptic potentials from the pyramidal neurons of layers III and V (figure 1.2(a)) are considered to be the main contributors to scalp EEG signals [Elul, 1972]. Schematically, the arrival of an action potential at a synapse of a pyramidal neuron causes a release of neurotransmitters at the presynaptic membrane. The neurotransmitters migrate to the postsynaptic membrane and bind to specific receptors. This causes opening of ion channels and change of ionic concentrations on either side of the cell membrane. The entry and exit of ions cause a potential difference, or postsynaptic potential, then can be recorded by microelectrodes in the extracellular space. It has been shown that these postsynaptic potentials and their summation are the origin of the electrical activity recorded at the scalp while action potentials, despite the fact that their amplitude is three times greater than that of a synaptic potential, contribute to a negligible part of the EEG activity [Elul, 1972]. By considering an excitatory synapse at dendrite level of a cortical pyramidal neuron (figure 1.2(b)), the synaptic activation causes a massive influx of ion Na^+ (depolarization) at the postsynaptic membrane which can be represented by a current input (or sink) and thus a deficit in positive charges in the extracellular medium. This massive influx of ions is balanced by current outputs (or sources) below of this point, along the membrane. An activated neuron can therefore be represented by a group of positive charges and a group of negative charges separated by a small distance, i.e. a dipole. Extracellular currents, and therefore the potential fields established between positive and negative regions, are the origin of EEG activity collected on the surface.

In order for this potential to be recorded at the scalp, it must result from the synchronous and quasi-simultaneous activation of a large number of neurons. In this



(a) (b) (c)

Figure 1.2: (a) Histological section of neocortex (somatosensory). The gray matter is formed by the molecular layer (I), the external granular layer (II), the outer pyramidal layer (III), the internal granular layer (IV), the ganglionic or internal pyramidal layer (V) and polymorphic layer (VI). The cell bodies of pyramidal neurons primarily located in the layers III and V. (b) Schematic of an excitatory synapse at a pyramidal neuron. (c) Slice of mouse neocortex (Dr. Fu-Ming Zhou, University of Tennessee). The pyramidal cell bodies of neurons of the layer V and their dendrites that are aligned parallel to the cortical surface are distinguished. This organization, called "lattice", is favorable to the summation of postsynaptic potentials.

case, and due to the lattice structure of pyramidal neurons in the cerebral neocortex (figure 1.2(c)), these unitary activities are summed to provide EEG signals with a sufficient amplitude to be detectable at the surface of the scalp.

1.2 Epilepsy and epileptic EEG paroxysms

1.2.1 Definition and epidemiological data

Epilepsy is a debilitating neurological disease, responsible for the repetitive occurrence of brief paroxysmal episodes or "seizures", which may include uncontrollable motor, sensory or mental demonstrations and may be accompanied by a partial or total impairment of consciousness. According to World Health Organization (WHO), the prevalence of epilepsy (i.e. the number of cases at a given time and a given group) is estimated between 0.5 and 0.8% of the population, including in France almost 1% of the population, making it the most common neurological disorder after migraine. The incidence of this disease is estimated to be two new cases per day in France with a population of one million. The occurrence of a single seizure may involve a higher population percentage

(up to 10% of people worldwide have at least one seizure during their lifetimes), but one seizure does not signal epilepsy.

Currently among the 500000 epileptic patients treated in France, approximately 50% of cases recover, while in 25% of patients medication controls only partially the seizures, and in the remaining 25% cases, patients are resistant to medication. These group of epilepsies, called drug-resistant or refractory, are among the most severe. They are often of focal or partial origin, i.e. seizure originate from a relatively circumscribed brain region. In this case a surgical treatment is possible to selectively resect this region called "epileptogenic zone" and to stop seizures, provided that the potential deficits induced by surgery are acceptable. The precise delimitation of the epileptogenic zone is the ultimate goal of presurgical evaluation. This delimitation is based on the results from different clinical, anatomical and functional explorations that are more or less invasive. The number of patients eligible for surgery is still low compared to the demand, which poses a real public health problem as the continuation of medication is more expensive than surgery.

1.2.2 Epileptic EEG paroxysms

The surface EEG is the first tool used in the diagnosis of epilepsy. This method can record epileptic paroxysms that reflect neuronal hyperexcitability and reflet the occurrence of brief, sudden and excessive neuronal discharges within one or more brain regions. As these regions vary from one patient to another, the clinical manifestations associated with seizures are variable and depend on the localization of epileptic regions.

The EEG examination, coupled with video recording of seizures (Video-EEG) is performed continuously for several days during the first phase of the presurgical evaluation of refractory partial epilepsy. In this examination, various epileptic EEG paroxysms are collected:

Interictal EEG paroxysms, also called "interictal spikes", occur inconsistently between seizures. Interictal spikes are highly variable among patients and types of epilepsy and they generally have polymorphic nature (figure 1.3). They may be spikes, spike-waves, polyspikes or polyspike-wave complexes. The regions where these paroxysms originate define the "irritative zone". These paroxysms are undoubtedly a marker for epilepsy but the spatial relationships that they have with the epileptogenic zone are sometimes unclear and can vary from one patient to another. For example in the case of specific lesions such as focal cortical dysplasia, interictal paroxysms are often frequent or continuous and arise from the vicinity of the epileptogenic region. In other cases such as in mesial temporal lobe epilepsy, interictal paroxysms are rather a reflection of a network of structures larger than the epileptogenic zone. The characterization of interictal paroxysms and their location are important aspects of the analyses conducted during the first phase of the presurgical evaluation.

Ictal EEG *paroxysms*, or ictal discharges occuring during seizures. They are characterized by rhythmic activities whose morphology varies depending on the type of epilepsy (for example partial vs generalized) and changes also with time as the ictal activity spreads. Figure 1.3(b) shows examples of ictal activity that can be recorded

during seizures such as spike discharges, spike-waves rhythms, fast activities or slow-wave rhythms. The epileptogenic zone is originally defined as the brain region necessary for triggering usual ictal discharges of a patient [Bancaud et al., 1962]. It is also defined as the smallest area whose resection (or disconnection) causes complete cessation of seizures.

In the particular case of partial epilepsies (which are considered in our study), intracerebral SEEG records a rapid and low voltage activity at the onset of seizures that is considered as the electroencephalographic signature of one (or more) region(s) from where seizures start [Bancaud et al., 1962]. These regions are also called "Onset zones" and their resection is correlated to good postsurgical prognosis [Alarcon et al., 1995]. The representation of this activity on the time-frequency map is specific [Shamsollahi et al., 1996, Wendling et al., 2003] and looks like a "chirp": it is a narrow band activity whose frequency gradually decreases and whose amplitude increases (figure 1.4). The maximum frequency of these discharges varies from one patient to another. For example, this rapid activity often peaks around 25 Hz when it comes from the mesial temporal lobe, whereas it can reach much higher frequencies (80 Hz or more) in some neocortical regions. Unlike SEEG, this activity is difficult to observe on scalp EEG, even in the absence of artifacts, because of its low amplitude and thus its unfavorable signal to noise ratio. When seizures are recorded during the first phase of presurgical evaluation, the clinician therefore tries to identify these rapid discharges on EEG traces. Depending on the size and topology of the region from which the discharge originates, it may have no reflection on the surface EEG or it may be associated with a focal flattening on EEG. In our study, we will consider on the one hand interictal spikes and on the other hand rapid activities at the onset of ictal discharges. These activities will constitute our signals of interest that we seek to denoise.

1.3 Sources of disturbance

Surface epileptic EEG signals are often contaminated with a variety of "noises" that can more or less hide the signal of interest. These noises are generally classified into two categories.

Instrumentation noise brings together non-physiological artifacts from the recording equipment itself or other close devices. For example, the 50 Hz noise is captured on scalp EEG in the presence of electronic devices in the recording room. Similarly, poor impedance recording electrode causes artifacts in the corresponding channel. However, this type of noise is directly attributable to the quality of the recording and it is theoretically managed when the recording system and methodology are of quality.

Physiological noise includes all the physiological activities that are not from cerebral origin and that can affect the quality of the EEG signal of interest. Thus, the eye movement (slow or fast) and blinking induce high amplitude variations in the potential most often seen in the frontal electrodes. Similarly, cardiac activity may generate an artifact on the electrodes placed in proximity to an artery. In addition, the activity of most of the muscles of the head, face and neck results in an often high amplitude and wide frequency band signal on EEG. Finally, in some applications, part of the brain

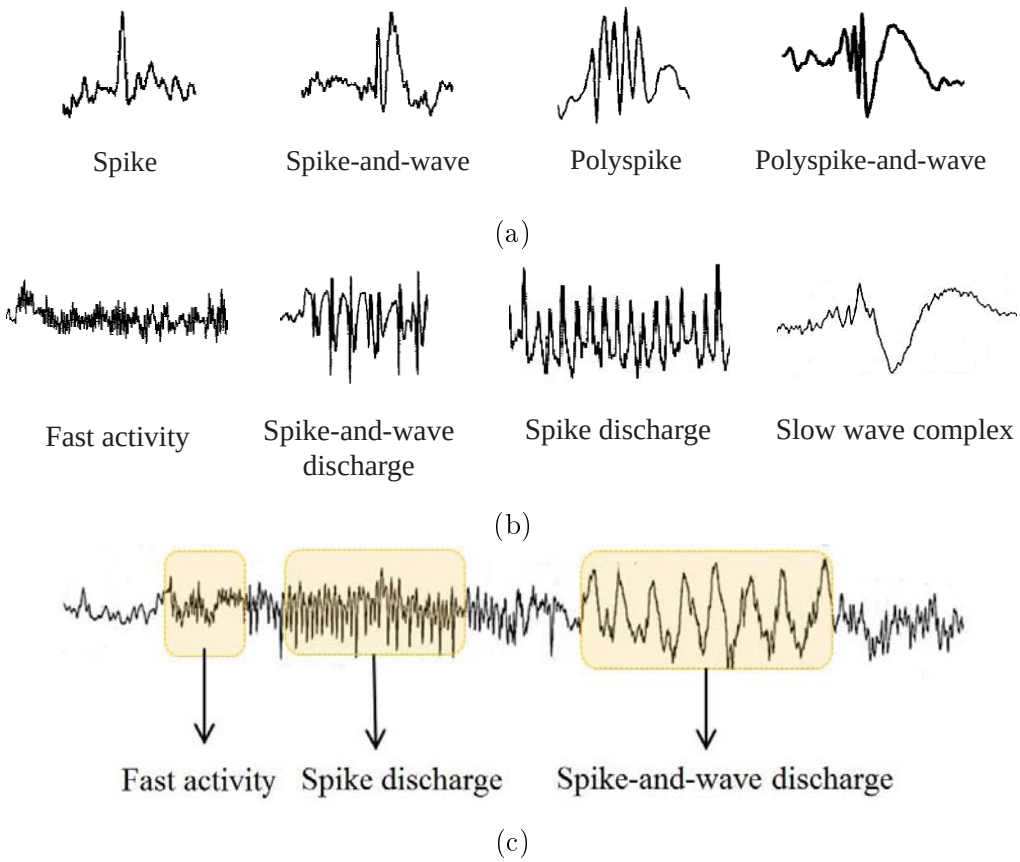


Figure 1.3: Epileptic paroxysms of scalp EEG. (a) Interictal paroxysms. (b) Ictal paroxysms. (c) Ictal discharge during which we identify different types of ictal activities [Crespel et al., 2006].

activity itself such as EEG background activity is considered as a noise that should be removed. The elimination of the muscle activity (sometimes called myogenic activity) which masks the EEG signal of interest is the main focus of our work.

The highly disruptive effect of muscle activity on the EEG is directly attributable to its properties. These have been studied in detail in the study of Goncharova and his colleagues [Goncharova et al., 2003], whose main conclusions are summarized as follows:

Muscular EEG signal has a wide frequency spectrum, generally ranging between 20 and 80 Hz. The activity of the frontal muscle has a frequency between 20 and 30 Hz, whereas the activity of the temporal muscle is rather between 40 and 80 Hz. For all muscle activity, the peak frequency increases as contraction increases. In addition, muscle activity may also contain low frequency component (0-5 Hz) or in the alpha band. Accordingly, all the frequency bands of EEG and therefore all rhythms (physiological or epileptic) are likely to be masked by muscle activity. The spatial distribution is not ubiquitous, contrarily to other artifacts such as eye movements. Muscle activity may involve EEG channels, based on the muscles involved. The anterior electrodes mainly record the activity of the facial muscles (frontal muscle, corrugator supercilii), the lateral electrodes rather record the activity of the chewing muscles (temporal muscle) and the posterior electrodes are rather sensitive to the activity of muscles of the neck or of base

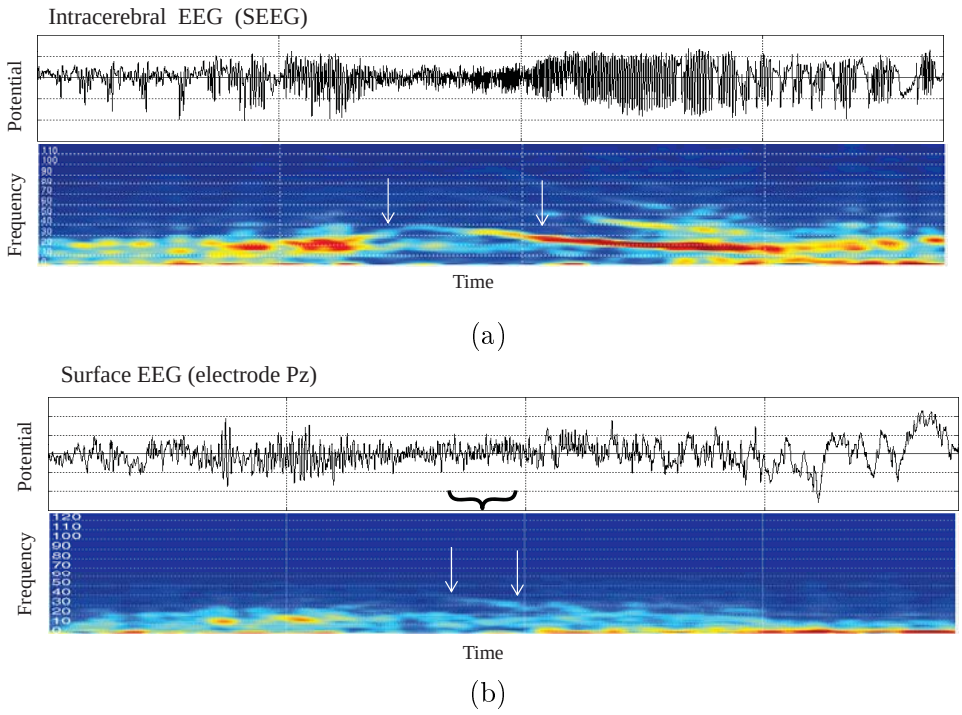


Figure 1.4: The fast activity at the beginning of the ictal discharge. (a) Ictal discharge SEEG recorded in a patient with a parieto-occipital epilepsy associated with dysplastic lesions. The intracerebral electrode contact placed in the lesion recorded a discharge that starts with a fast activity (40Hz). The time-frequency signature is illustrated between the two white arrows. (b) In the same patient, seizures have been recorded by scalp EEG. At the posterior electrode Pz, we guess the end of the fast discharge (white arrows).

of the skull (occipital muscle, trapezoid muscle). The co-contraction of certain muscles (frontal and temporal, for example) can induce myogenic activities at distant electrodes. Therefore, the myogenic activity is not only captured at the channels near the muscles that contract but also sometimes at more remote electrodes. Finally, myogenic activity can also induce an artifact with a more broad topography that varies according to the subjects. All the electrodes are potentially "perturbable" by the muscular activity.

As a matter of fact, muscle activity is often temporally confounded with seizures. Indeed, clinical signs of the ictal discharges induce muscle activity that is directly reflected in the EEG. For example, hypertonic contractions of facial muscles, jaw or chewing movements often occur at the beginning of seizures and directly perturb the ictal signal. Consequently, it is rare to record seizures on scalp EEG that are not partially or totally obscured by muscle activity. Pretreatment with denoising methods is essential to remove these contaminations before analyzing these signals.

1.4 Problem formulation

In the context of recording the epileptic EEG activity, the recorded signal can be considered as a combination of (i) the signal of interest arising from the epileptic brain regions and (ii) the signal of non-interest composed of different types of noise, background activity and artifacts. The aim of this thesis is to propose methods that extract the signal of interest from the recorded signal, i.e. that denoise epileptic EEG signals. Since the proposed methods use different approaches to denoise EEG signals, including both statistical and deterministic methods, we present the problem formulation both in statistical and deterministic formats.

1.4.1 Statistical problem formulation

In the statistical model, we model the EEG signal recorded from M electrodes as one realization $\{\boldsymbol{x}[n]\}$ of an M-dimensional random vector process $\{\tilde{\boldsymbol{x}}[n]\}$. We can represent the observation process $\{\tilde{\boldsymbol{x}}[n]\}$ as follows:

$$\begin{aligned}\tilde{\boldsymbol{x}}[n] &= \tilde{\boldsymbol{x}}^{(e)}[n] + \tilde{\boldsymbol{x}}^{(b)}[n] + \tilde{\boldsymbol{x}}^{(m)}[n] + \tilde{\boldsymbol{\nu}}[n] \\ \tilde{\boldsymbol{x}}[n] &= \boldsymbol{A}^{(e)}\tilde{\boldsymbol{s}}^{(e)}[n] + \boldsymbol{A}^{(b)}\tilde{\boldsymbol{s}}^{(b)}[n] + \boldsymbol{A}^{(m)}\tilde{\boldsymbol{s}}^{(m)}[n] + \tilde{\boldsymbol{\nu}}[n] \end{aligned} \tag{1.1}$$

where $\{\tilde{\boldsymbol{s}}^{(e)}[n]\}$, $\{\tilde{\boldsymbol{s}}^{(b)}[n]\}$, $\{\tilde{\boldsymbol{s}}^{(m)}[n]\}$, $\{\tilde{\boldsymbol{v}}[n]\}$ are the random vector processes representing the activity of P_e epileptic sources, P_b background sources, P_m muscular sources and the M-dimensional instrument noise, respectively. The mixing matrices $\boldsymbol{A}^{(e)}$, $\boldsymbol{A}^{(b)}$ and $\boldsymbol{A}^{(m)}$ of size $(M \times P_e)$, $(M \times P_b)$ and $(M \times P_m)$, model the transfer function from all possible sources of activity to scalp electrodes. All these mixing matrices can be combined in a global mixing matrix $\boldsymbol{A}$ of size $(M \times P)$ where $P = P_e + P_b + P_m$. Therefore, (1.1) can be rewritten as follows:

$$\tilde{\boldsymbol{x}}[n] = \boldsymbol{A}\tilde{\boldsymbol{s}}[n] + \tilde{\boldsymbol{\nu}}[n] \tag{1.2}$$

where $\{\tilde{\boldsymbol{s}}[n]\}$ is the random vector process representing all sources.

As far as the statistical properties of vector random processes $\{\tilde{\boldsymbol{s}}^{(e)}[n]\}$, $\{\tilde{\boldsymbol{s}}^{(b)}[n]\}$, $\{\tilde{\boldsymbol{s}}^{(m)}[n]\}$ and $\{\tilde{\boldsymbol{\nu}}[n]\}$ are concerned, we can assume that they are independent as they correspond to different physiological/physical phenomena. Nevertheless, such an assumption is not valid within each vector random process regarding its components. In particular, some of epileptic activity sources may be statistically mutually dependent. We will then assume that, for every time index, the three vectors $\tilde{\boldsymbol{s}}^{(e)}[n]$, $\tilde{\boldsymbol{s}}^{(b)}[n]$ and $\tilde{\boldsymbol{s}}^{(m)}[n]$ can be factorized as $\tilde{\boldsymbol{s}}^{(e)}[n] = \boldsymbol{B}^{(e)}\tilde{\boldsymbol{r}}^{(e)}[n]$, $\tilde{\boldsymbol{s}}^{(b)}[n] = \boldsymbol{B}^{(b)}\tilde{\boldsymbol{r}}^{(b)}[n]$ and $\tilde{\boldsymbol{s}}^{(m)}[n] = \boldsymbol{B}^{(m)}\tilde{\boldsymbol{r}}^{(m)}[n]$, respectively where $\tilde{\boldsymbol{r}}[n] = [\tilde{\boldsymbol{r}}^{(e)}[n]^{\mathsf{T}}, \tilde{\boldsymbol{r}}^{(b)}[n]^{\mathsf{T}}, \tilde{\boldsymbol{r}}^{(m)}[n]^{\mathsf{T}}]^{\mathsf{T}}$ is a P'-dimensional vector of mutually independent random variables such that $P' \leq M$ and where $\boldsymbol{B} = [\boldsymbol{B}^{(e)^{\mathsf{T}}}, \boldsymbol{B}^{(b)^{\mathsf{T}}}, \boldsymbol{B}^{(m)^{\mathsf{T}}}]^{\mathsf{T}}$ is a $(P \times P')$ matrix such as the product $\boldsymbol{G} = \boldsymbol{A}\,\boldsymbol{B}$ is a full column rank matrix. Eventually, the $\{\tilde{\boldsymbol{\nu}}[n]\}$ vector random process can be assumed to be Gaussian as most of instrument noises [Albera et al., 2012].

The aim of the proposed statistical approaches, is to extract the signal of interest, $\{\boldsymbol{x}^{(e)}[n]\}$, which represents one realization of the epileptic vector random process $\{\tilde{\boldsymbol{x}}^{(e)}[n]\}$, from $\{\boldsymbol{x}[n]\}$. Indeed, by computing from $\{\boldsymbol{x}[n]\}$ the corresponding realization, $\{\boldsymbol{r}^{(e)}[n]\}$, of $\{\tilde{\boldsymbol{r}}^{(e)}[n]\}$ and the corresponding mixing matrix $\boldsymbol{G}^{(e)} = \boldsymbol{A}^{(e)}\,\boldsymbol{B}^{(e)}$, the

M-dimensional signal $\{\boldsymbol{x}^{(e)}[n]\}$ given by $\boldsymbol{x}^{(e)}[n] = \boldsymbol{G}^{(e)}\boldsymbol{r}^{(e)}[n]$ will represent the denoised EEG data.

1.4.2 Deterministic problem formulation

Regarding the deterministic model, we assume that N samples of the M-dimensional EEG signal can then be modeled as follows:

$$\boldsymbol{x}[n] = \boldsymbol{A}^{(e)}\boldsymbol{s}^{(e)}[n] + \boldsymbol{A}^{(b)}\boldsymbol{s}^{(b)}[n] + \boldsymbol{A}^{(m)}\boldsymbol{s}^{(m)}[n] + \boldsymbol{\nu}[n] \tag{1.3}$$

where $\{\boldsymbol{s}^{(e)}[n]\}$, $\{\boldsymbol{s}^{(b)}[n]\}$, $\{\boldsymbol{s}^{(m)}[n]\}$ and $\{\boldsymbol{\nu}[n]\}$ are vector signals representing P_e epileptic activities, P_b background activities, P_m muscular activities and an M-dimensional instrument noise, respectively. The mixing matrices $\boldsymbol{A}^{(e)}$, $\boldsymbol{A}^{(b)}$ and $\boldsymbol{A}^{(m)}$ of size $(M \times P_e)$, $(M \times P_b)$ and $(M \times P_m)$, model the transfer function from all possible activities within the brain to scalp electrodes. All these mixing matrices can be combined in a global mixing matrix $\boldsymbol{A}$ of size $(M \times P)$ where $P = P_e + P_b + P_m$. Therefore, (1.3) can be rewritten as follows:

$$\boldsymbol{x}[n] = \boldsymbol{A}\boldsymbol{s}[n] + \boldsymbol{\nu}[n] \tag{1.4}$$

where $\{\boldsymbol{s}[n]\}$ is a vector signal representing all sources.

In addition, we assume that, for every time index n, the three vectors $\boldsymbol{s}^{(e)}[n]$, $\boldsymbol{s}^{(b)}[n]$ and $\boldsymbol{s}^{(m)}[n]$ can be factorized as $\boldsymbol{s}^{(e)}[n] = \boldsymbol{B}^{(e)}\boldsymbol{r}^{(e)}[n]$, $\boldsymbol{s}^{(b)}[n] = \boldsymbol{B}^{(b)}\boldsymbol{r}^{(b)}[n]$ and $\boldsymbol{s}^{(m)}[n] = \boldsymbol{B}^{(m)}\boldsymbol{r}^{(m)}[n]$, respectively, where the different signals of the P'-dimensional vector signal $\boldsymbol{r}[n]$ defined by $\boldsymbol{r}[n] = [\boldsymbol{r}^{(e)}[n]^{\mathsf{T}}, \boldsymbol{r}^{(b)}[n]^{\mathsf{T}}, \boldsymbol{r}^{(m)}[n]^{\mathsf{T}}]^{\mathsf{T}}$ have time constant frequency contents up to a multiplicative scalar. As in section 1.4.1, the $(M \times P')$ matrix $\boldsymbol{G}$ will denote the product $\boldsymbol{G} = \boldsymbol{A}\,\boldsymbol{B}$ where $\boldsymbol{B} = [\boldsymbol{B}^{(e)^{\mathsf{T}}}, \boldsymbol{B}^{(b)^{\mathsf{T}}}, \boldsymbol{B}^{(m)^{\mathsf{T}}}]^{\mathsf{T}}$. Nevertheless, contrarily to the statistical model, here P' can be strictly greater than M.

As a result, by computing the P'_e-dimensional vector signal $\{\boldsymbol{r}^{(e)}[n]\}$ and the corresponding $\boldsymbol{G}^{(e)} = \boldsymbol{A}^{(e)}\,\boldsymbol{B}^{(e)}$ matrix of coordinates from the measurements $\{\boldsymbol{x}[n]\}$, the M-dimensional signal $\{\boldsymbol{x}^{(e)}[n]\}$ given by $\boldsymbol{x}^{(e)}[n] = \boldsymbol{G}^{(e)}\boldsymbol{r}^{(e)}[n]$ will represent the denoised EEG data.

In the following chapters, the p-th epileptic source and the p-th source will correspond to the p-th entry of the epileptic signal $\{\boldsymbol{r}^{(e)}[n]\}$ and the p-th entry of the whole signal $\{\boldsymbol{r}[n]\}$, respectively.

1.5 Previous work

In this section, we study the previous work performed in the context of noise removal from EEG signals including epileptic EEG signal denoising. There are widely ranged researches in this context. Here, we study these researches in two main groups, namely blind and semi-blind methods. It is noteworthy that, despite their names, both blind and semi-blind methods use some knowledge about sources of interest. Although the blind methods blindly extract all sources, they use *a priori* knowledge to identify the sources of interest. On the other hand, the semi-blind methods use *a priori* knowledge in the stage of extracting the sources. In each group, we sort the studies by date. At the end of the section, other methods are briefly introduced.

1.5.1 Blind methods

Blind methods have been used in different fields such as audio processing, face recognition, biomedical signal processing and communications. In the rest of this section, we introduce some well-known results based on blind approaches in the context of EEG signal denoising.

1.5.1.1 ICA, DCCA and PCA-based methods

One of the widely used Blind Source Separation (BSS) approaches in the context of biomedical signal processing is Independent Component Analysis (ICA) with the fundamental assumption that the sources are mutually independent. Herault and Jutten are the first (around 1983) to use informally the concept of ICA, especially in order to solve the BSS problem [Ans et al., 1985]. A few years later, Comon presents a mathematical formulation of ICA and shows how Higher Order (HO) cumulants can be used to solve the problem of ICA: the HO contrast-based method CoM_2 arises from this work (see [Comon, 1994] and references therein). In parallel, Cardoso and Souloumiac develop the JADE algorithm [Cardoso and Souloumiac, 1993], based on a Joint Approximate Diagonalization (JAD). In the context of biomedical signal processing, ICA has been used in a wide range of studies such as ElectroCardioGram (ECG) [Zarzoso et al., 1997, De Lathauwer et al., 2000, Phlypo et al., 2010a], ElectroMyoGram (EMG) [Subasi and Kiymik, 2010, Naik et al., 2013], functional Magnetic Resonance Imaging (fMRI) [Akhbari et al., 2010, Griffanti et al., 2014], MagnetoEncephaloGram (MEG) [Fatima et al., 2013] and EEG [Albera et al., 2012].

Another approach is Decorrelated and Colored Component Analysis (DCCA), in which BSS-CCA and SOBI have been used in the context of biomedical signal processing. Blind Source Separation based on Canonical Correlation Analysis (BSS-CCA), briefly called CCA, has been used for processing of biomedical signals such as EEG [De Clercq et al., 2006, Lin et al., 2013], fMRI [Guccione et al., 2013] and ElectroHysteroGram [Hassan et al., 2011]. The Second-Order Blind Identification (SOBI) algorithm [Belouchrani et al., 1997] has been also used for processing of biomedical signals such as EEG [Tang et al., 2005, Wang et al., 2004] and ECG. Principal Components Analysis (PCA) is another BSS method used for processing of biomedical signals. In the rest of this section, we introduce some well-known ICA, DCCA and PCA-based researches in the context of EEG signal denoising.

ICA was used in [Vigário, 1997] to remove ocular artifacts from real EEG data recorded in a routine clinical 22-channel experiment from children aged 9-13 years. In another study [Jung et al., 2000], ICA was used to remove eye movement, muscle artifacts, line noise, blink and blink-related activities from three different EEG datasets collected from normal and autistic subjects. The results showed that ICA greatly surpassed regression and PCA methods.

In order to remove non-ictal noise and artifacts from ictal signals, ICA was applied to the ictal EEGs of patients with medial Temporal Lobe Epilepsy (TLE) in [Nam et al., 2002]. As illustrated in this paper, all ictal EEGs were successfully decomposed to epileptic and non-epileptic components, and after EEG reconstruction with exclusion of artifacts, the quality of the EEG was improved.

ICA, PCA and regression based methods were compared in [Wallstrom et al., 2004] for removing ElectroOculoGram (EOG) artifacts, using simulated and real (eyes open) EEG data. In contrary to the previous ICA-based studies, the obtained results supported the use of regression-based and PCA-based ocular artifact correction and suggested a need for further studies examining possible spectral distortion from ICA-based correction procedures.

A SOBI-based BSS algorithm was used in [Joyce et al., 2004] for automatic removal of ocular artifacts from EEG data. The BSS algorithm described herein could isolate correlated electroocular components with a high degree of accuracy. This method was studied on an example of real EEG data.

ICA and its combination with Digital Filter (DF) were used to denoise ictal EEGs in [Urrestarazu et al., 2004]. 20 seizures of nine patients with focal epilepsy were studied and the results showed that all the recordings except one improved after the use of ICA for the elimination of blinking and muscle artifacts. Moreover, the results showed that combination of ICA and DF improved the quality of results. As stated in this paper, the greatest difficulty for ICA was to separate the muscle artifact.

The CCA method was applied in [De Clercq et al., 2006] on synthetic and real ictal EEG recordings contaminated with muscle artifacts. For the synthetic data, an EEG epoch of 10 s containing mainly alpha (8-13 Hz) activity was selected and real muscle noise were superimposed. As shown in this paper, the CCA method outperformed a low-pass filter with different cutoff frequencies and an ICA technique, namely JADE [Cardoso and Souloumiac, 1993], for muscle artifact removal.

The goal of the study performed in [Iriarte et al., 2006] was to analyze focal seizures with ICA, i.e., decomposing the elements of the seizures to understand their genesis and propagation, and to differentiate between various types of focal seizures. In addition, ICA was used in order to denoise ictal EEGs from muscular activity. In this study, the JADE algorithm was used to separate independent components from 26 focal seizures of 12 different patients. The results showed that ICA demonstrated a high capacity to separate the ictal activity from the normal brain activity and artifacts.

Usual ICA methods of artifact removal require a tedious visual classification of the components. The proposed system in [LeVan et al., 2006] automated this process and removed simultaneously multiple types of artifacts, including EMG and EOG. This system was based on ICA and Bayesian classifier. The system was tested on real ictal EEG and compared with ICA method with manual classification with visual inspection. The results showed that the proposed Bayesian classifier had a performance that was only slightly worse than the errors that would be expected from human expert variability.

In another study [Delorme et al., 2007], three popular ICA algorithms, InfoMax [Lee et al., 1999, Makeig et al., 2000], SOBI, and FastICA [Comon and Jutten, 2010] were used to remove different types of noise and artifact, including muscle activity, eye blink, electrical shift artifacts, linear trends and white noise, from normal EEG data. To this end, event-related EEG data from a visual categorization task was recorded and different types of noise and simulated artifacts were added to noise-free signal with different SNRs from $-50dB$ to $0dB$. The results showed that except for muscle artifact

detection where no gain was found by using ICA, the ICA-decomposed data gave better results than the raw scalp data in the procedure of automated detection of small non-brain artifacts. Moreover, the results showed that the InfoMax algorithm performed better than the SOBI and FastICA.

A combination of BSS methods and Support Vector Machine (SVM) in [Halder et al., 2007] was used for online EOG and EMG noise removal from EEG data. In the first part of the algorithm, the BSS methods, such as JADE, InfoMax, FastICA and AMUSE [Tong et al., 1990], were used to decompose recorded signal to artifact and non-artifact components. SVM was then used for automatic classification of the components. EEG data were recorded while the subject was performing no particular task and were cleaned of blinks and other obvious artifacts. EMG and EOG artifact signals were recorded from forearm and electrode Fp1, respectively, and were added to noise-free signal. Four kinds of artifacts (jaw muscles, forehead movement, eye blinks, and eye movement), each recorded with four different subjects, were used to train the SVM to classify various EEG artifacts. To select the best BSS method, JADE, InfoMax, FastICA and AMUSE were examined. AMUSE and InfoMax gave the best results with EOG and EMG artifacts, respectively. This method was evaluated on three BCI datasets as a proof-of-concept of the method.

A study was performed [Fitzgibbon et al., 2007] to investigate and compare the performance of BSS algorithms for separating common types of contamination from EEG. The dataset was constructed with a known mixture of known real brain and contaminating signals. Five different BSS methods, namely PCA, SOBI, InfoMax, JADE and FastICA, were examined. The key finding was that although BSS is an effective and powerful tool for separating and removing contamination from EEG, the quality of the separation is highly dependent on the type of contamination, the degree of contamination, and the choice of BSS algorithm. BSS appeared to be more effective at separating muscle and blink contamination and less effective at saccadic and tracking contamination. Moreover, for all types of contamination, PCA was a strong performer when the contamination was greater in amplitude than the brain signal whereas other algorithms such as SOBI and InfoMax were generally better for specific types of contamination of lower amplitude.

The potential clinical relevance of BSS-CCA was investigated in [Vergult et al., 2007] to remove muscle artifacts in ictal scalp EEG. Applying BSS-CCA on the ictal data of thirty-seven patients with refractory partial epilepsy with a well-defined seizure onset zone showed that BSS-CCA significantly improved the sensitivity to localize the seizure onset from 62% to 81%, and gave the best results in ictal EEGs with moderate to severe muscle artifact contamination.

In the study performed in [Kachenoura et al., 2008], different ICA algorithms, namely SOBI, CoM_2 [Comon, 1994], JADE, ICAR, FastICA and InfoMax, were examined in the context of BCI systems based on the Mu rhythm. In this context, the surface observations can be considered as a noisy mixture of one source of interest, namely the Mu rhythm, and artifact sources such as the ocular and cardiac activities. A comparative study of these algorithms in terms of estimation accuracy and numerical complexity, conducted on simulated EEG data, showed that the use of an appropriate ICA algorithm may significantly improve the capabilities of BCI systems.

The study presented in [Crespo-Garcia et al., 2008] aimed at assessing the performance of four ICA algorithms, namely AMUSE, SOBI, InfoMax and JADE, to separate myogenic activity from EEG during sleep, in order to determine the optimal method. The algorithms were examined on both simulated and real sleep EEG data. The results showed that AMUSE, InfoMax, and SOBI performed significantly better than JADE at eliminating muscle artifacts over temporal regions, but AMUSE was independent of the signal-to-noise ratio over non-temporal regions and markedly faster than the remaining algorithms.

In [Congedo et al., 2008], different Second Order Statistics (SOS)-based and Higher Order Statistics (HOS)-based methods were compared for the extraction of spontaneous and induced EEG, and a method based on the approximate joint diagonalization of Fourier co-spectral matrices was proposed.

In the study performed in [Romero et al., 2009], regression analysis, adaptive filtering, and BSS were evaluated with simulated data in the context of EOG removal. In the first two methods, filtered versions were also taken into account by filtering EOG references in order to reduce the cancelation of cerebral high frequency components in EEG data. As BSS methods, PCA, SOBI, InfoMax and FastICA were considered. In general, filtered versions of time-domain regression and of adaptive filtering with Recursive Least Square (RLS) algorithm provided a very effective ocular reduction. However, BSS based on second order statistics, SOBI and PCA, showed the highest similarity indexes and the lowest errors in spectral variables.

In the study performed in [Safieddine et al., 2011, Safieddine et al., 2012], in addition to ICA algorithms, some other denoising algorithms, such as CCA, Empirical Mode Decomposition (EMD) and Wavelet Denoising (WD) were studied for removal of the muscular artifacts from surface EEG signals recorded in epileptic patients. This study showed that CCA and ICA approaches outperform EMD and WD in denoising of interictal signals with moderate noise levels.

A method by combining BSS, WD and discriminant analysis was proposed in [Romo Vázquez et al., 2012] in order to remove ocular, high frequency muscle and ECG artifacts from EEG. The method was examined on both simulated and real interictal and ictal signals.

As mentioned in [Albera et al., 2012], only few ICA algorithms such as InfoMax and FastICA are used nowadays to process biomedical signals. The authors in [Albera et al., 2012] examined fifteen ICA algorithms in terms of performance and computational complexity in the context of removal of muscle artifacts from interictal epileptiform activity. The results have shown that CoM_2 offers the best compromise between performance and numerical complexity.

An underdetermined blind source separation method and a novel MMSE beamformer were used in [Koldovsky et al., 2012] to remove artifacts from motor imagery signals in the context of BCI. They showed that by using the proposed method, the BCI classification accuracy was improved by up to 10%.

In another study [Zou et al., 2012], an artifact removal algorithm based on ICA and Hierarchal clustering was presented. This technique used general temporal and spectral features and particular information about target Event-Related Potentials (ERPs) to

separate ERPs and artifact activities. The EEG data was recorded from fifteen subjects in a Visual Evoked Potential (VEP) study. They showed that the proposed method could effectively enhance the ERPs for all subjects.

As presented in [Kang and Zhizeng, 2012], a combination of wavelet denoising, PCA and density estimation BSS (DEBSS) were used to remove EOG and ECG from EEG data. The DEBSS algorithm used in this method is a generalization of equivariant adaptive separation via independence (EASI) to the heterogeneous mixture signals.

In the study performed in [Qi, 2012], two different algorithms, RLS and BSS, were compared to investigate their performances on removing EOG artifacts from EEG signals. Results indicated that the performance of RLS algorithm was better than BSS algorithm no matter whether there were any EOG reference signals.

A novel method to enhance ERPs was proposed in [Kawaguchi et al., 2013] which was based on combining PCA with multivariate empirical mode decomposition (M-EMD). In the proposed method, PCA reduced the data dimensions, while M-EMD removed the relatively large background EEGs. The results of the method with simulated and real P300 ERP components showed the usefulness of the proposed method to improve the SNR of P300s.

Stone's method, which used two different linear filters that process the same set of sources, was used in [Chaozhu et al., 2013] to remove EOG artifact from EEG signals. Comparing the Stone's BSS results with those of FastICA and JADE showed the superiority of the proposed algorithm.

A new approach for joint BSS of datasets at multiple time lags using CCA was developed in [Shen et al., 2013] for removing muscular artifacts from EEG recordings. The proposed approach jointly extracted sources from each dataset in a decreasing order of between-set source correlations. Muscular artifact sources that typically have lowest between-set correlations can then be removed. The results on real data showed that the proposed method offered better performance in removing muscular artifacts than the conventional CCA.

A fully automatic system based on underdetermined blind source separation and using kurtosis of the sources to distinguish them was proposed in [Ge et al., 2014] to remove ocular artifacts from EEG. In this system, the fourth-order cumulant based method (FOOBI) [De Lathauwer et al., 2007] decomposed multiple EEG channels into a relative large number of source components and the kurtosis value was used to identify the ocular components in these source components. Results on real EEG data showed the superiority of the proposed method on FastICA.

1.5.1.2 Canonical Polyadic decomposition methods

The Canonical Polyadic or CANDECOMP/PARAFAC (CP) decomposition consists in decomposing a HO array as a linear combination of a minimal number of rank-1 terms. The CP model can then be seen as a generalization of the Singular Value Decomposition (SVD) of two-way data to multi-way data. But the main difference is that, under weak assumptions [Kruskal, 1977, Jiang and Sidiropoulos, 2004, Sidiropoulos and Bro, 2000, Stegman and Sidiropoulos, 2007], no orthogonality constraint is needed to ensure

uniqueness of the CP decomposition. This advantage makes the CP decomposition very useful in various applications. The CP decomposition was first proposed to solve psychometric problems [Carroll and Chang, 1970]. Afterwards, it was extended to various fields such as chemometrics [Bro, 1997], spectroscopy [Villringer and Chance, 1997] and biomedical engineering [Vos et al., 2007, Becker et al., 2014b]. In the rest of this section, we study previous work performed in the context of EEG signal processing.

The CP decomposition of time-varying EEG spectrum was used in [Miwakeichi et al., 2004] for the comparison of resting EEG to that recorded while the subject performed mental arithmetic. Mental arithmetic produces theta activity in the frontal area and a suppression of alpha activity in the occipital area, while the converse occurs when the eyes are closed in the resting condition. Real EEG data was used in this study and the HO array had time, space and frequency. For generating the frequency mode, a complex Morlet mother wavelet was used. As explained in this study, the CP decomposition can be used not only for extracting significant activities from EEG, but also for searching for the presence of atoms in a new data set, which were not used for estimating the loading matrices and can be either from the same or from a different subject. Moreover, if certain atoms obtained by the CP decomposition contain artifacts (e.g., eye movements, eye blinking, electromyogram), their space/frequency reconstructions can be used as templates for an artifact detector. The reconstruction, obtained by eliminating the component that corresponds to artifacts, will be an artifact removal method.

The CP decomposition was used in [Mørup et al., 2006] to decompose wavelet transformed (using Morlet mother wavelet) event-related EEG given by the inter-trial phase coherence encompassing ANOVA analysis of differences between conditions and 5-way analysis of channel, frequency, time, subject and condition. The CP decomposition of the 3-way array of ANOVA F-test values clearly showed the difference of regions of interest across modalities, while the 5-way analysis enabled visualization of both quantitative and qualitative differences.

Synthetic EEG data were used in [De Vos et al., 2007] in order to study the efficiency of the CP decomposition in localizing the ictal epileptic dipole sources. A three-way data array of size (space$\times$scale$\times$time) was obtained by wavelet-transforming every channel of the simulated EEG matrix. The EEG of the ictal activity was generated using a fixed dipole in a three-shell spherical head model. The different time courses generated by the dipole were a 4Hz sinusoidal waveform and a 4 Hz sawtooth (both of them representing seizure activity in patients with mesial temporal lobe epilepsy) at different noise levels. The influence of the frequency of the seizure signal on ictal scalp EEG source localization at a fixed noise level was also studied. In another simulation, the dipole localization error was examined when the frequency changed during the 2 seconds under investigation which does not give a trilinear signal after wavelet transformation. To this end, a chirp was used that linearly changed in frequency from 8Hz at the start to 4Hz at the end of the considered 2 seconds. Two rhythmical sources firing at the same frequency separated from each other by about 1 cm were considered in the other simulation. The effect of using dense array EEG (148 electrodes) was also studied. In all simulations, noise matrix was generated using 2 seconds of awake background EEG activity from a normal subject. It was shown that in a spherical head model with

realistically simulated EEG, the proposed algorithm correctly localized the seizure-related atom with an accuracy of about 5 mm, even at SNR ratios that were lower than $0dB$.

In another study of this group [Vos et al., 2007], both simulated and real ictal EEG data were used. The same algorithm as in the previous study was used and a three-way array of size (channels×time×scales) was generated. For the simulated data, a 5.7 Hz sine wave was used and muscle artifacts were superimposed with SNR values between $-15dB$ and $-2dB$. The results were compared with power CP decomposition (which used power of wavelet transform to create the three-way array) and SVD. The results showed that the CP decomposition based on pure wavelet-transformed EEG was better than SVD and the decomposition on the power of the wavelet-transformed EEG. Clinical validation on real ictal EEGs showed that the CP method correctly localized the seizure onset zone in 34 of 37 cases (92%) while a human reader, blinded to all other information, was able to localize 21 of 37 cases (57%).

The same approach to generate the three-way array from ictal epileptic EEG was used in [Acar et al., 2007]. Epileptic focus localization, artifact extraction and artifact removal were studied in this research. 10 real ictal EEGs from 7 patients were considered in this study. The results for 8 seizures matched with clinical observations in terms of seizure origin and extracted artifacts. On the other hand, for 2 of the seizures, seizure localization was not achieved using an initial trial of CP modeling. In these cases, first, they applied an artifact removal method and subsequently applied the CP model on the epilepsy tensor from which potential artifacts have been removed. This method successfully identified the seizure origin in both cases.

The two algorithms proposed in [Deburchgraeve et al., 2009] differ in the way the tensor was constructed and in the type of activity they were able to extract. While the first method extracted oscillatory seizure activity, the second extracted spike train activity. The first algorithm, called O-CP decomposition method in this study, was similar to the above-mentioned algorithms and was designed to localize the oscillatory type of neonatal seizures. The third dimension of the tensor was constructed using the continuous wavelet transform of the ongoing EEG recording and the tensor consisted of the dimensions channels, time and frequency. The suitable number of components was determined by performing the core consistency diagnostic (CorConDia). The second algorithm, called SP-CP decomposition, used the output of the spike train seizure detector to construct the three-way array. The first two dimensions of the tensor were EEG channels and time and the third dimension was built up by aligning all EEG segments (taking all channels) of the detected spikes one after the other in different slices of the tensor. The third dimension of the tensor thus consisted of an aligned set of EEG slices. After construction of the complete tensor, a rank-1 CP decomposition was computed. Simulations illustrate the robustness of the SP-CP method in comparison with spike triggered averaging. In this simulation, impact of noise (−15 dB-15 dB) on localization was studied. Moreover, the seizure localization results of 21 seizures from 6 neonates with post-asphyxial hypoxic ischemic encephalopathy were compared with those based on the visual analysis of the EEG by a clinical neurophysiologist. There was a good agreement between the two methods in the localization of seizure onset in all.

In the studies performed in [Becker et al., 2010, Becker et al., 2012, Becker et al., 2014b], a new multi-way approach based on Space-Time-Wave-Vector (STWV) data was proposed which was obtained by a 3D local Fourier transform over space accomplished on the measured data. This method can be seen as a preprocessing step that separates the sources, reduces noise as well as interference and extracts the source time signals. Most of the previously mentioned studies used frequency transform of data to construct the three-way (Space×Time×Frequency) (STF) arrays. The STF-based techniques depend on the source time signals, which are assumed to be oscillatory, and does not permit to separate several simultaneously active brain regions with correlated activities into more than one component, thus preventing the representation of such a scenario by an adequate number of equivalent dipoles. Contrary to the STF analysis, the STWV method allows also for the separation of correlated sources. Moreover, it permits to accurately localize one or several equivalent dipole sources and to extract at the same time a good estimate of the source time signals. In the simulations, epileptiform signals were generated with the help of the Jansen model. A noise matrix containing temporally white, but spatially correlated noise was computed in the same way as the data matrix for given noise dipoles with radial orientation emitting white Gaussian signals. The source localization error was calculated for the noise level of -8 dB to 10 dB. Compared to the STF analysis, the STWV method was particularly well suited if the data was only recorded for a small number of time samples. A potential application of the STWV technique is thus the tracing of the temporal evolution of sources. Moreover, the estimation error of the leadfield matrix was reduced, which improved the performance of the source localization. Moreover, a new algorithm was proposed in [Becker et al., 2014b], which shows how tensor-based approaches can be designed to perform an accurate source localization of spatially distributed sources. Note that realistic simulated and real EEG data were used to assess the good behavior of the proposed method.

In the study performed in [Phan et al., 2013a], the method Fast CP (FCP) was proposed to reduce the numerical complexity of CP decomposition by decomposing an unfolded tensor with lower order, e.g., order-3 tensor, instead of directly factorizing the high order data tensor. The method was applied on Event-Related EEGs stored in as a 4-way tensor of size (28 measurements $\times$ 61 frequency bins $\times$ 64 channels $\times$ 72 time frames). The aim of this paper was to compare the factorization time of ALS and FCP over various Ranks in the range of $5-72$ with and without a HOSVD-based compression prior to the CP decompositions. Moreover, another method proposed in [Phan et al., 2013b] reduced the extra memory requirements of CP algorithms and accelerates the standard alternating CP algorithms 20-30 times for order-5 and order-6 tensors. To study the factorization time of the proposed method, the same data as the previous study was used. In addition, another data set including EEG motor imagery data was examined by considering an order-5 tensor of size (2 dictionaries $\times$ 23 frequency bins (8-30 Hz) $\times$ 50 time frames $\times$ 62 channels $\times$ 200 trials).

The study performed in [Niknazar et al., 2014] studied blind source separation for underdetermined mixtures (i.e., more sources than sensors) of event-related sources that include quasi-periodic sources (e.g., ECG), sources with synchronized trials (e.g., ERP), and amplitude-variant sources. Their method consisted of two stages: (i) CP decom-

position for underdetermined source separation and (ii) signal extraction by Kalman filtering to recover the source dynamics. A tensor was constructed for each source by synchronizing on the event period of the corresponding signal and stacking different periods along the second dimension of the tensor. To cope with the interference from other sources that impede on the extraction of weak signals, two robust CP decomposition methods were proposed and compared in three different applications: ERP extraction, fetal ECG extraction and twin MagnetoCardioGram (MCG) extraction.

1.5.1.3 EMD-based methods

The EMD method [Huang et al., 1998] can decompose any complicated data set into a finite and often small number of components called intrinsic mode functions (IMF). An IMF represents a generally simple oscillatory mode as a counterpart to the simple harmonic function. Since the decomposition is based on the local characteristic time scale of the data, it can be applied to nonlinear and non-stationary processes. EMD was first used to study the laboratory made wave data and typical ocean wave data from the field stations [Huang et al., 1998]. After that, it has been used in different fields such as biomedical engineering [Bajaj and Pachori, 2012, Liang et al., 2005a], image processing [Hariharan et al., 2006] and speech processing [Huang and Pan, 2006].

In the biomedical signal processing, EMD has been widely used in order to analyze and denoise signals. It was shown in [Liang et al., 2000] that EMD could separate, identify and remove contamination from a wide variety of artifactual sources in ElectroGastroGram (EGG) recordings. EMD was also used in [Liang et al., 2005b] to analyze of esophageal manometric data in gastroesophageal reflux disease. Moreover, EMD was used in [Blanco-Velasco et al., 2008] to enhance ECG signals by removing both high-frequency noise and baseline wander with minimum signal distortion. A combination of the BSS-CCA and EMD methods, called CCA-EMD, was used in [Hassan et al., 2011] to denoise monopolar ElectroHysteroGram (EHG). A framework based on EMD was proposed in [Zhang and Zhou, 2013] to eliminate power line interference, white Gaussian noise and baseline wandering from EMG. A combination of EMD and ICA was used in [Lee and Lee, 2013] to remove ECG artifacts from EMG. An EMD method based on combination of instantaneous half period and soft-thresholding was used in [Samadi and Shamsollahi, 2014] for noise reduction of ECG signals.

EMD has been widely used in processing of EEG data, especially for removing noise and artifacts from EEG data. In this part, we introduce some new researches in this context. In the study performed in [Looney et al., 2008], EMD was used to remove ocular artifacts from real EEG. An EMD-based approach was used in [Rutkowski et al., 2009] to separate muscular interference from brain electrical activity. Indeed, the EMD components from each EEG channel have been compared in spectral domain and analyzed with a clustering technique in order to identify those similar across channels. The resulting reconstruction separated common muscular interferences from underlying brain activity.

To remove EOG from Steady-State Visual Evoked Potential (SSVEP) [Rutkowski et al., 2010], each of the recorded channels in multichannel EEG recording environment was decomposed into IMF components. The IMF components were further clustered

for their spectral similarity in order to identify only those carrying responses to present stimuli to the subjects. The results showed the superiority of this method in comparison with contemporary blind source separation algorithms.

The multivariate extension of EMD [Fleureau et al., 2011b, Fleureau et al., 2011a], called MEMD, was used in [Molla et al., 2012] to separate EOG artifacts from EEG data recorded during visual facial stimuli. A combination of EMD and Adaptive Filtering (AF) was used in [Navarro et al., 2012] to cancel ECG noise in a simplified EEG montage for preterm infants. In this study simulation signals were created from real artifact-free EEG and then the cardiac artifacts were added to these signals. As shown in this paper, cleaned signals improved up to 17% the correlation coefficient with original datasets in comparison with signals denoised solely with AF. A method based on combining EMD and ICA, called EMD-ICA, was proposed in [Mijovic et al., 2010b] in order to separate independent sources when the number of these sources was higher than the number of channels available. Two applications were considered for this method. In the first, the new method was used to separate the ECG artifact from the EMG recordings. And in the second one, EMD-ICA was applied to one of the EEG channels contaminated by muscle artifact, eye artifact and seizure activity. A method for a single-channel signal decomposition was proposed in [Mijovic et al., 2010a] which combined EMD with ICA. A two step extension to bivariate EMD (BEMD) was proposed in [Molla et al., 2010] to remove ocular artifacts from EEG. The BEMD is a generalized extension of the EMD to complex signals.

A real-time implementation of an EMD-based signal enhancement scheme was proposed in [Santillan-Guzman et al., 2013] for removing noise, suppressing muscle artifacts and detrending EEG signals in an automatic manner and in real-time. In order to test the performance of the EMD in real-time, semi-simulated EEG data were created: real EEG signals from the central and frontal parts of the brain were contaminated with simulated muscle artifacts. A real EEG dataset from an epileptic patient was also employed. In both real and simulated experiments, results indicated that the muscle artifacts were successfully suppressed.

A systematic approach to EEG signal enhancement was proposed in [Wang et al., 2013] which was based on EMD and threshold filtering/rejection. The results of applying the method on synthetic and real Auditory Evoked Potential (AEP) data showed the effectiveness of the method to improve the quality of AEPs. A method based on EMD and Orthogonal Matching Pursuit was proposed in [Mourad and Niazy, 2013] for automatic correction of eye blink artifact in single channel EEG recording. The effectiveness of the method was studied on real EEG data superimposed with the eye blink artifact recorded at the EOG electrode. A hybrid algorithm based on EMD and ICA was also proposed in [Soomro et al., 2013] that automatically removed the eye blink artifact from the EEG. The proposed algorithm was evaluated on both simulated and real EEGs with eye blink artifact. In another study [Mert and Akan, 2014], a method based on EMD and Mutual Information (MI) was proposed in order to remove white noise from EEG signals. In this method, MI was used as a metric to define the informative level of extracted intrinsic mode functions referencing to the noisy EEG signals. Two hybrid algorithms by combining the concept of Wavelet Packet Transform (WPT), EMD and ICA, called WPT-ICA and WPT-EMD, were proposed in [Bono et al., 2014].

The algorithms have been tested on multiple trials of four different artifact cases, eye-blinking and muscle artifacts including left and right hand movement and head-shaking. The results showed that the proposed methods were robust and holistic since they do not require any *a priori* information about the artifact.

1.5.1.4 Wavelet-based methods

The Wavelet Transform (WT) originated in 1980 with Morlet, a French research scientist working on seismic data analysis [Morlet, 1983]. The WT has been found to be particularly useful for analyzing signals which can be described as aperiodic, noisy, intermittent, transient and so on. Its ability to examine the signal simultaneously in both time and frequency in a distinctly different way from the traditional Short Time Fourier Transform (STFT) makes WT a widely used method in a variety of research areas such as biomedical signal processing [Senhadji et al., 1995, Shamsollahi et al., 1996], digital watermarking [Lai and Tsai, 2010], face recognition [Srinivasan and Ravichandran, 2013] and tracking of moving targets [Rui et al., 2013].

The WT has been widely used in the context of noise cancelation of biomedical signals such as ECG [Cherkassky and Kilts, 2001, Castillo et al., 2013], EMG [von Tscharner et al., 2011] and fMRI [Patel et al., 2014]. A lot of studies have been done in the context of EEG noise cancelation, including EOG and muscular noise cancelation. In this section, we study some of recent work in this context.

The BSS-CCA technique with wavelet filtering approach, called wavelet enhanced BSS-CCA, was proposed in [Raghavendra and Dutt, 2011] for minimizing both ocular and muscle artifacts simultaneously. The performance of the proposed wavelet enhanced BSS-CCA method was tested on real EEG recordings contaminated with ocular and muscle artifacts that suggested that the proposed hybrid approach minimizes ocular and muscle artifacts effectively, minimally affecting underlying cerebral activity in EEG recordings. The Automatic Wavelet Independent Component Analysis, AWICA, was proposed in [Mammone et al., 2012] for automatic EEG artifact removal. AWICA consists of a two-step procedure relying on the concepts of kurtosis and Renyi's entropy. The results on both simulated and real EEG data showed improved success in terms of suppression of artifact components while reducing the loss of residual informative data, since the components related to relevant EEG activity were mostly preserved.

A wavelet-based method was used in [Scolaro et al., 2012] to attenuate the background activity and high frequencies in real interictal epileptic EEG signals. In the study performed in [Mamun et al., 2013], four different discrete wavelet functions were used to remove noise from the EEG signal gotten from two different types of patients (healthy and epileptic) to examine the effectiveness of Discrete Wavelet Transform (DWT) on EEG noise removal. In another study [Zima et al., 2012], a method based on a combination of ICA and Wavelet, called Robust Artifact Removal (RAR), was proposed which was capable of removing short-duration, high-amplitude artifacts from long-term neonatal EEG recordings. Such artifacts are mainly caused by movement activity, and have an adverse effect on automatic processing of long-term sleep recordings. In order to gain robustness of the RAR method, the whole EEG recording was processed multiple times. The resulting tentative reconstructions were then combined.

The method was shown to perform better than either the wavelet-enhanced ICA or the simple artifact rejection method without the combination procedure.

A nonparametric Bayesian estimator for signal and image denoising in the wavelet domain, called MAP BKF, was presented in [Boubchir and Boashash, 2013]. The results obtained with real EEG data (with open eye and closed eye) corrupted by Gaussian noise indicated that the MAP BKF denoiser performed better than the soft and hard thresholding methods. In [Abtahi et al., 2014], wavelet coefficient thresholding was used to ECG removal from neonates EEG.

1.5.2 Semi-blind methods

In this section, we present semi-blind and/or constrained source separation methods that use some *a priori* knowledge about sources or signals of interest. Sometimes even rather weak data modeling priors exists, such as the general shape of the time curves or their frequency contents. It would be very useful if such prior knowledge could be incorporated into the separation algorithm directly. This kind of problem setting is often called semi-blind. It is noteworthy that finding a strict boundary between blind and semi-blind methods is somewhat difficult and in some researches, methods which used knowledge about sources of interest were presented as blind methods.

One semi-blind method is the Denoising Source Separation (DSS) [Sarela and Valpola, 2005], a powerful algorithmic framework built around the concept of a denoising function. In DSS, the independence criterion of ICA is replaced by the assumption that the sources should 1) be uncorrelated and 2) maximize some desired properties. Another approach is using Generalized EigenValue Decomposition (GEVD) with prior knowledge about sources of interest. Constrained ICA algorithms which use prior knowledge as temporally or spacially constraints are also used in denoising biomedical signals [Ille et al., 2002]. In the rest of this section, we study semi-blind source separation methods used in the context of noise removal from brain signals, including EEG, ECoG and MEG.

1.5.2.1 GEVD and DSS-based methods

DSS was used in [Gunduz et al., 2008] to remove the interictal activity from ECoG of patients who have lost their motor functions in order to improve linear prediction of hand trajectories. In this method, the high amplitude quasiperiodic nature of the observed interictal spikes was used as prior knowledge to present an iterative DSS method.

In the study recently performed in [Samadi et al., 2013], a fast method to extract the sources related to interictal epileptiform state was proposed. The method was based on GEVD using two correlation matrices during: 1) periods including Interictal Epileptiform Discharges (IED) as a reference activation model and 2) periods excluding IEDs or abnormal physiological signals as background activity. The IED and non-IED time intervals were manually identified by the epileptologist for each patient considering all the intracerebral EEG channels. The method was evaluated using both realistic simulated data and actual intracerebral EEGs of patients suffering from focal epilepsy.

A deflation algorithm based on GEVD was presented in [Sameni and Gouy-Pailler, 2014] for the automatic detection and removal of EOG artifacts from EEG. The EOG reference channel was used to identify time samples in which EOG artifact is high. The notion of effective number of identifiable dimensions was also used to estimate the number of dominant dimensions of the ocular subspace, which enabled the precise and fast convergence of the algorithm. The results of applying the proposed method on real and synthetic data showed the superiority of this method on some ICA-based approaches such as FastICA and JADE.

1.5.2.2 Constrained ICA-based methods

An spatial filter approach, called Spatially-Constrained ICA (SCICA), for artifact correction in EEG and MEG recordings was presented in [Ille et al., 2002]. SCICA used the prior knowledge about artifact topographies and combined it with the temporal-statistical strategy of ICA to estimate the brain signal topographies from the whole data segment to be corrected. The proposed method was used to remove EOG and ECG artifacts from EEG and MEG data.

Temporally constrained ICA was used in [James and Gibson, 2003] to remove EOG, ECG and line noise from EEG and EMG signals. To this end, for each type of artifacts one (or more) reference signal(s) was (were) extracted by using the knowledge available in the recorded signal and these reference signals were used to extract the corresponding artifact components. The results on synthetic and real data showed the effectiveness of this method.

A Jacobi-based constrained ICA technique was proposed in [De Vos et al., 2006] to remove eye artifacts from the EEG. The idea of this method was that there is strong prior information about the mixing vector of eye artifacts which can be used as a constraint in ICA model. The advantage of this method is that the components do not have to be inspected afterwards in order to detect the artifactual ones.

The concept and use of spatial constraints was introduced in [Hesse and James, 2006] in biomedical semi-blind source separation applications. To this end, different kinds of spatial constraint were defined and detailed general guidelines were provided on how these constraints could be incorporated into existing gradient-based BSS approaches. The effectiveness of using spatial constraints were studied by using normal and ictal EEG data in the context of artifact removal and source tracking.

BSS-CCA and SCICA were used in [Hallez et al., 2009] as a preprocessing step before the source localization by means of the Rap-MUSIC algorithm [Mosher and Leahy, 1999] during the ictal period. The methods were examined through the source localization procedure and showed that the accuracy of the source localization can be increased by preprocessing the ictal EEG segment by BSS-CCA and SCICA.

A framework was proposed in [Akhtar et al., 2012] based on SCICA and WD. In this framework, SCICA was used to extract artifact-only independent components from the given EEG data and WD was used to remove any cerebral activity from the extracted artifacts components. Finally the artifacts were projected back to be subtracted from EEG signals to get clean EEG data. Prior assumptions regarding the spatial topography

of some sources were used as spatial constraints on the mixing matrix. The proposed method was examined on both simulated EEG data with simulated eye-blink artifacts and real EEG data contaminated with ocular artifacts. It showed better results than the previous wavelet-based methods.

A constrained ICA approach [Shapoori et al., 2013] was used for separation of intracranial spikes from EEG. This method was based on creating a template from intracranial data, by using approximate entropy, and then this template was used in the form of a constraint in an ICA algorithm. The effectiveness of the algorithm was studied by using real intracranial EEG data.

1.5.3 Other methods

In addition to the above-mentioned methods, there exist other methods used in the context of noise removal from EEG signals. Among these methods we can mention methods based on Kalman Filters [Galka et al., 2011, Oikonomou et al., 2007b], adaptive and Volterra filters [Mateo et al., 2013], filtering and thresholding the analytic signal envelope [Melia et al., 2014] and Bayesian framework [Zumer et al., 2007]. In addition, several reviews and comparison studies in this context are available such as in [Kachenoura et al., 2011, Khatwani and Tiwari, 2013, Salis et al., 2013]. Table 1.1 summarizes the main advantages and disadvantages of the denoising methods.

1.6 The outline of the thesis report

The rest of this report is organized as follows. First, in chapter 2, we present the proposed methods to denoise epileptic EEG data. These methods are presented in three groups: the GEVD and DSS-based methods, the JDICA algorithm and the SSD-CP algorithm. For each proposed algorithm, we explain how it can be used in order to denoise interictal and/or ictal EEG data. At the end of this chapter, the numerical complexity of all proposed algorithms is studied. In the third chapter, we present the simulated and real data used in our various analyses. Then, we introduce the performance criteria used in the fourth chapter for comparison between the denoising methods. Chapter 4 shows the results obtained with interictal and ictal EEG signals, for both simulated signals and real data. Finally, the conclusion, discussion and perspectives of our work are presented in the last chapter.

1.7 List of publications

In this thesis, several methods are proposed for noise cancelation of epileptic signals. The first two methods, which are based on Generalized EigenValue Decomposition (GEVD) and Denoising Source Separation (DSS) approaches, are used to denoise interictal data. To extract a priori information required by GEVD and DSS, we propose a series of preprocessing stages including spike peak detection, extraction of exact time support of spikes and clustering of spikes involved in each source of interest. The first

version of the GEVD-based denoising method has been presented at a conference in 2012:

S. HAJIPOUR, M. B. SHAMSOLLAHI, L. ALBERA and I. MERLET, "Noise cancelation of epileptic interictal EEG data based on generalized eigenvalue decomposition," in TSP 12, 2012 IEEE Thirty-Fifth International Conference on Telecommunications and Signal Processing, Prague, Czech Republic, July 3-4 2012, pp. 591-595.

Then, we modify the preprocessing algorithms and add the DSS-based method. The presentation of the GEVD-based and DSS-based methods and results of comparing them with ICA and DCCA methods has been accepted to appear in the following journal:

S. HAJIPOUR, M. B. SHAMSOLLAHI, L. ALBERA and I. MERLET, "Interictal EEG noise cancelation: GEVD and DSS based approaches versus ICA and DCCA based methods," to appear in IRBM.

Two other methods, namely Time Frequency (TF)-GEVD and TF-DSS, are also proposed in order to denoise ictal EEG signals for which the time-frequency signature of ictal sources is extracted by using the CCA method. This work has been accepted to appear in the following journal:

S. HAJIPOUR SARDOUIE, M. B. SHAMSOLLAHI, L. ALBERA and I. MERLET, "Denoising of ictal EEG data using semi-blind source separation methods based on time-frequency priors," to appear in IEEE Journal of Biomedical and Health Informatics.

We also propose a deflationary ICA method, namely JDICA, that is based on Jacobi-like iterations. More particularly, while a projection-based deflation scheme inspired by Delfosse and Loubaton's ICA technique ($\mathrm{DelL}^{\mathbb{R}}$) is used, a Jakobi-like optimization strategy is proposed in order to maximize a fourth order cumulant-based contrast built from whitened observations. This work has been accepted to appear in the following journal:

S. HAJIPOUR, L. ALBERA, M. B. SHAMSOLLAHI and I. MERLET, "An efficient Jacobi-like deflationary ICA algorithm: application to EEG denoising," to appear in IEEE Signal Processing Letters.

Moreover, we propose a new direct algorithm, called SSD-CP, to compute the CP decomposition of complex-valued multi-way arrays. The proposed algorithm is based on the Simultaneous Schur Decomposition (SSD) of particular matrices derived from the array to process. We also propose a new Jacobi-like algorithm to calculate the SSD of several complex-valued matrices. This work has been presented in the following conference:

S. HAJIPOUR, L. ALBERA, M. B. SHAMSOLLAHI and I. MERLET, "Canonical polyadic decomposition of complex-valued multi-ways arrays based on simultaneous Schur decomposition," in ICASSP 13, 2013 IEEE International Conference on Acoustics Speech and Signal Processing, Vancouver, Canada, May 26-31 2013, pp. 4178-4182.

Finally, we plan to submit a journal paper from the above work very soon:

S. HAJIPOUR, L. ALBERA, M. B. SHAMSOLLAHI and I. MERLET, "From simultaneous Schur decomposition to canonical polyadic decomposition of complex-valued multi-way arrays," to be submitted in IEEE Transactions on Signal Processing.

Table 1.1: The main advantages and disadvantages of the denoising approaches.

	Advantages	Disadvantages
ICA, DCCA and PCA-based methods	• Used for denoising of different kinds of signals (Normal, interictal, ictal,...) and noises (EOG, EMG, ECG,...) • Various methods have been proposed in this category • Show better denoising performance in comparison with other methods for low SNR values	• Subspace of interest and subspace of non-interest should be independent or uncorrelated • Dimension of the source space should be estimated to achieve good results • The sources of interest and number of them should be chosen manually or with an extra step • High computational complexity in case of higher dimensions • The dimension of the source space should be less than the number of channels
CP decomposition based methods	• No independency assumption is needed • The dimension of the source space can be more than the number of channels	• Data should be represented as a higher order array • The array to be decomposed should fit trilinear (or higher) model • High computational complexity in case of higher dimensions • The sources of interest and number of them should be chosen manually or with an extra step
EMD-based methods	• Good performance for low SNR values	• Usually used in combination with other denoising methods • Choosing components of interest (or non-interest) is manual in most cases • High numerical complexity

Table 1.1: (cont.) The main advantages and disadvantages of the denoising approaches.

	Advantages	Disadvantages
Wavelet-based methods	• Low numerical complexity • Different mother wavelets can be used based on the type of signal and noise	• Usually used in combination with other denoising methods • Poor performance for low SNR values • Choosing components of interest (or non-interest) and thresholding is manual in most cases
GEVD and DSS-based methods	• Low numerical complexity • Good performance when the *a priori* knowledge is available about the subspace of interest • No need to choose the sources of interest	• Appropriate information about the subspace of interest is needed • Performance is greatly sensitive to the quality of the *a priori* information on subspace of interest • Poor performance for low SNR values
Constrained ICA-based methods	• Good performance when the *a priori* knowledge is available about the subspace of interest	• Appropriate information about the subspace of interest is needed • Performance is greatly sensitive to the quality of the *a priori* information on subspace of interest • Poor performance for low SNR values

Chapter 2

Algorithms

In this chapter, we present the proposed methods to denoise epileptic EEG data. These methods are presented in three groups. First, we study the methods that are based on GEVD and DSS frameworks. Then, we propose the JDICA algorithm that is a deflationary ICA method based on Jacobi-like iterations. The third method, called the SSD-CP algorithm, is based on the CP decomposition. For each proposed algorithm, we explain how it can be used in order to denoise interictal and/or ictal EEG data. At the end of the chapter, the numerical complexity of all proposed algorithms is studied.

2.1 Semi-blind GEVD/DSS based methods

GEVD/DSS based approaches can be used when we have *a priori* information about the subspace of interest. In this thesis, we separately propose two methods based on the GEVD and DSS [Sarela and Valpola, 2005] frameworks for denoising interictal and ictal data. Indeed, we want to test if *a priori* information on the subspace of interest can be used in the denoising process. To this end, first we introduce the GEVD and DSS frameworks and then we separately present the proposed methods to denoise interictal and ictal EEGs.

2.1.1 Generalized Eigenvalue Decomposition

The problem of GEVD of a pair of symmetric matrices $(\boldsymbol{C}_1, \boldsymbol{C}_2) \in \mathbb{R}^{M\times M} \times \mathbb{R}^{M\times M}$, is defined by finding the matrices $\boldsymbol{W}$ and $\boldsymbol{\Lambda}$ such that:

$$\boldsymbol{W}^{\mathsf{T}}\boldsymbol{C}_1\boldsymbol{W} = \boldsymbol{\Lambda} \tag{2.1}$$

$$\boldsymbol{W}^{\mathsf{T}}\boldsymbol{C}_2\boldsymbol{W} = \boldsymbol{I} \tag{2.2}$$

where $\boldsymbol{\Lambda}$ and $\boldsymbol{I}$ are a diagonal matrix and the identity matrix of size $(M \times M)$, respectively. After multiplying equations (2.1) and (2.2) with $\boldsymbol{W}^{-\mathsf{T}}$ and combining these equations, we obtain [Parra and Sajda, 2004]:

$$\boldsymbol{C}_1\boldsymbol{W} = \boldsymbol{C}_2\boldsymbol{W}\boldsymbol{\Lambda} \tag{2.3}$$

If $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$ are symmetric positive definite matrices, matrix $\boldsymbol{\Lambda}$ contains the positive generalized eigenvalues corresponding to the real generalized eigenmatrix $\boldsymbol{W} =$

$[\boldsymbol{w}_1, \ldots, \boldsymbol{w}_M]$. If the generalized eigenvalues are sorted in descending order on the diagonal of matrix $\boldsymbol{\Lambda}$, the first generalized eigenvector $\boldsymbol{w}_1$, corresponding to the largest generalized eigenvalue, maximizes the Rayleigh quotient given as follows:

$$J(\boldsymbol{w}) = \frac{\boldsymbol{w}^{\mathsf{T}} \boldsymbol{C}_1 \boldsymbol{w}}{\boldsymbol{w}^{\mathsf{T}} \boldsymbol{C}_2 \boldsymbol{w}} \tag{2.4}$$

Due to the assumption of linear mixtures, maximizing a Rayleigh quotient can be formulated in the time, frequency, wavelet, etc. in any domain obtained after a mapping reserving the linearity of the mixture [Sameni et al., 2010, Phlypo et al., 2010a, Phlypo et al., 2010b].

2.1.2 The Denoising Source Separation framework

To define the DSS framework, we use the statistical problem formulation given in section 1.4.1 and especially the following noisy linear static model:

$$\tilde{\boldsymbol{x}}[n] = \boldsymbol{G}\tilde{\boldsymbol{r}}[n] + \tilde{\boldsymbol{\nu}}[n] \tag{2.5}$$

To define the DSS algorithm, it is assumed that the observations $\{\tilde{\boldsymbol{x}}[n]\}$ are whitened as explained in [Comon, 1994], such that the covariance matrix of the noiseless observations becomes equal to the identity matrix. Consequently, with the whitening assumption, $M = P'$ and $\boldsymbol{G}$ is a $(P' \times P')$ orthogonal matrix. Based on the N observations, the problem consists in estimating (i) the mixing matrix $\boldsymbol{G}$ and (ii) the source signals.

The DSS framework is based on the Expectation-Maximization (EM) algorithm [Dempster et al., 1977]. For being self-contained, the detailed procedure to extract the DSS framework, based on [Sarela and Valpola, 2005, Belouchrani and Cardoso, 1995, Bermond and Cardoso, 1999], is given in appendix A. Here we give some brief explanations.

First, the likelihood $\mathcal{L}(\boldsymbol{\theta})$ of the observation matrix is calculated in terms of the vector of unknown parameters $\boldsymbol{\theta}$. To find the maximum likelihood estimate of the unknown parameters, all partial derivatives of $\mathcal{L}(\boldsymbol{\theta})$ should be equal to zero. These equations do not have a closed form solution. So, to solve this problem, an iterative EM-like algorithm is used. In this algorithm, in each step, the parameters are estimated again in such a way that a monotonic increase in the likelihood is guaranteed.

In the extracted EM algorithm, all the components are estimated simultaneously. However, by using the aforementioned assumptions, including the prewhitening of the observations, and simplifying the results, the DSS framework to extract the sources

one-by-one is obtained as follows [Sarela and Valpola, 2005]:

$$1) E - step:$$

$$a)\ Calculating\ the\ noisy\ estimate\ of\ one\ source:$$

$$\forall n,\ r_p^{(i)}[n] = \boldsymbol{w}_p^{(i)^\top} \boldsymbol{x}[n] \tag{2.6}$$

$$b)\ Denoising\ based\ on\ the\ a\ priori\ information\ about\ the\ sources:$$

$$\forall n,\ r_p^{(i)+}[n] = f(r_p^{(i)}[n]) \tag{2.7}$$

$$2)\ M - step:$$

$$a)\ Calculating\ the\ new\ ML\ estimate\ of\ \boldsymbol{w}_p^{(i)}:$$

$$\boldsymbol{w}_p^{(i)+} = \sum_{n=1}^{N} \boldsymbol{x}[n] r_p^{(i)+}[n] \tag{2.8}$$

$$b)\ Normalizing\ the\ mixing\ vector:$$

$$\boldsymbol{w}_p^{(i+1)} = \frac{\boldsymbol{w}_p^{(i)+}}{||\boldsymbol{w}_p^{(i)+}||} \tag{2.9}$$

$$Until\ convergence$$

where $\{\boldsymbol{x}[n]\}$ corresponds to the recorded realization of $\{\tilde{\boldsymbol{x}}[n]\}$ and $\{\boldsymbol{r}[n]\}$ is the corresponding realization of the source process $\{\tilde{\boldsymbol{r}}[n]\}$. In this framework, $\{r_p^{(i)}[n]\}$ and $\boldsymbol{w}_p^{(i)}$ represent the p-th source and its corresponding separator in the i-th iteration, respectively.

As stated in [Sarela and Valpola, 2005], this interpretation is not novel, but it allows for the development of new algorithms that are not derived starting from generative models. Therefore, the family of DSS algorithms is introduced as algorithms that have the form (2.6)-(2.9) and the equation $r_p^{(i)+}[n] = f(r_p^{(i)}[n])$ is interpreted as denoising.

2.1.3 The GEVD and DSS based methods for denoising of interictal data

For EEG signals, it has been previously shown that different assumptions about sources of interest, such as spatial constraints [Hesse and James, 2006], locations of known sources [Latif et al., 2006], shape and latency of the signal of interest [Ahmadian et al., 2013] and time support of spikes [Samadi et al., 2013], can be considered in semi-blind or constrained source separation methods. In this thesis, we use the timing information of the epileptic interictal sources (i.e. the time samples corresponding to each epileptic source) and propose a series of preprocessing stages to extract the useful information to be used in the GEVD and DSS frameworks. These preprocessing stages consist of the detection and clustering of the epileptic spikes involved in each source of interest. It should be mentioned that a GEVD-based method was previously proposed in [Samadi et al., 2013] in order to determine epileptic regions from epileptic intracerebral EEG signals. This method has a manual preprocessing stage to extract periods including interictal epileptiform discharges. On the contrary, in a GEVD-based

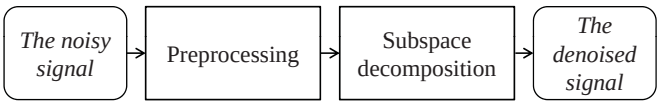


Figure 2.1: The flowchart of the GEVD and DSS-based denoising methods

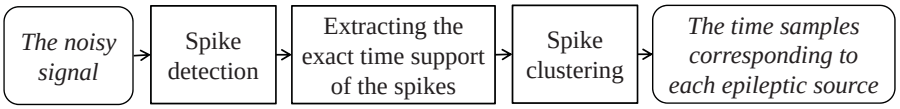


Figure 2.2: The flowchart of the preprocessing stage of the GEVD and DSS-based denoising methods

method used in this thesis to denoise interictal EEG data, all the preprocessing stages to extract the time samples corresponding to each epileptic source, including spike detection, extraction the exact time support of spikes and spike clustering, are fully automated. It should also be mentioned that both works have been done in parallel and independently. Some parts of our work was previously presented in [Hajipour et al., 2012]. In this section, we use the statistical model as presented in section 1.4.1.

As shown in figure 2.1, in order to denoise interictal EEG signals, we show how the GEVD/DSS approach can be used through two main stages: (1) a preprocessing stage and (2) the subspace decomposition stage. In the preprocessing stage, some *a priori* knowledge about the occurrence times of spikes (spike time samples involved in each epileptic source) are extracted. This information is used in the GEVD or DSS method to separate the epileptic subspace from the noise subspace. Finally, the estimated denoised signal is achieved by reconstructing the observation signals only from estimated epileptic sources. As shown in figure 2.2, the preprocessing stage has three parts. We first detect the spike peak times. Then the exact time support of spikes is identified. After that, we cluster the extracted spikes into P_e' groups corresponding to P_e' epileptic interictal sources. In the rest of this section, these three stages are explained in detail and the GEVD and DSS-based methods to denoise interictal data are then studied.

2.1.3.1 Spike detection

In the context of EEG signal processing, various methods have been proposed for spike detection [Senhadji et al., 1995, Senhadji et al., 1997, Wilson and Emerson, 2002, Acir et al., 2005, Bourien et al., 2007, Indiradevia et al., 2008, Ji et al., 2011]. These methods may use the information obtained from a single channel [Senhadji et al., 1995] or from several (multi-) channels [Ji et al., 2011, Acir et al., 2005]. They may also be categorized by the features they used: morphological features [Ji et al., 2011] or time-frequency ones [Senhadji et al., 1995]. Most of the spike detection methods have an enhancement stage that generates an output signal in which the distinction between the spikes and the noise is increased by some filtering methods such as WT [Senhadji and Wendling, 2002, Indiradevia et al., 2008], matched filters [Pfurtscheller and Fischer, 1978] or Kalman filter [Oikonomou et al., 2007a]. At this stage, the output signal is used in a decision procedure in order to extract the spike peak times. This decision procedure may be

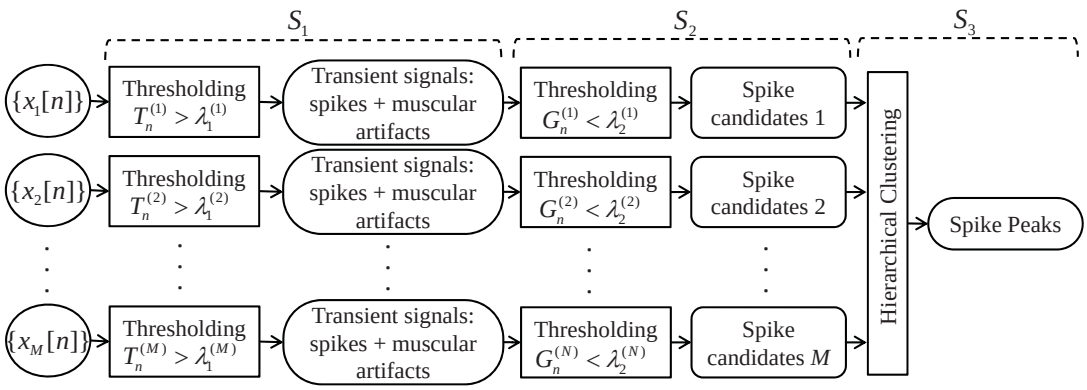


Figure 2.3: The spike peak detection algorithm

implemented by a simple thresholding algorithm [Bourien et al., 2007] or may have a more complex structure [Senhadji et al., 1995].

In this thesis, we use a three-stage multi-channel spike detection method that is a generalized version of the previously proposed single-channel spike detection method [Senhadji et al., 1995, Senhadji et al., 1997]. This single-channel wavelet-based spike detection method consists of two stages (S_1 and S_2) that detect the candidates for spike peak times of each channel of observation (figure 2.3). In [Senhadji et al., 1997], it was shown that this wavelet-based algorithm has a better performance than other spike detection algorithms. To generalize this method to multi-channel detection, we add a third stage (S_3) in which a clustering algorithm combines the extracted information from all channels to detect the spike peak times. Figure 2.3 shows the three stages of the algorithm.

The detailed procedure of this three-stage multi-channel spike detection method can be described as follows. First, note that the signal of interest, the spikes in our application, are contaminated by piece-wise stationary signals, such as background EEG and measurement noise as well as by transient signals such as muscular artifacts. Therefore, this algorithm eliminates the signals of non-interest in two successive stages: the first stage (S_1) separates transient signals (including the spikes) from the piece-wise stationary signals and the second stage (S_2) tries to separate the spikes from muscular artifacts (rejection of false alarms). To this end, two measures $T_n^{(m)}$ and $G_n^{(m)}$ are calculated for each channel of observation $\{x_m[n]\}$ and for each time sample n. Then by using the appropriate thresholding processes, the candidates for spike peak times are obtained for each channel. The detailed procedure of these two stages is presented in [Senhadji et al., 1995, Senhadji et al., 1997] and is also given in appendix B.

Up to this point, we estimate the candidates for the spike peak times for each observation channel. Since there may be false detections and since the peak time might be misestimated by a few samples, a decision algorithm has to be performed at a stage S_3 by using the information of all channels. We use a hierarchical clustering method [Ward, 1963] to group all possible candidates for spike peaks of all channels. The hierarchical grouping starts by defining a number N_s of clusters equal to the total number of candidates. In the first grouping step, the number of clusters is reduced to

$N_s - 1$ by combining the two nearest spike peak times. In successive stages the newly formed groups are linked to other spikes or spike groups with the lowest distance (i.e. the distance between the centers of the clusters). In each step we save the center and the maximum intra-distance, i.e. the distance between the individuals in one group, for each cluster. This procedure is repeated until the maximum intra-distance of all clusters is less than a given threshold Th_{spike} chosen as the length of a spike. After extracting all clusters, we select only the clusters for which the population is more than N_{eff} (i.e. the minimum number of channels which are affected by one spike) and accordingly the centers of these clusters are the estimated peak times.

2.1.3.2 Extraction of the exact time support of the spikes

Only a few studies have worked on extracting the time support of a spike. In most applications, such as extracting the morphological features, only the spike peak times and the distance between adjacent peaks are calculated. To extract the important morphological points, the curvature features based on the second derivative are usually used. In this thesis, we use a curvature measure proposed in [Rosenfeld and Johnston, 1973] to extract the time support of the spikes. First, in order to eliminate the ineffective channels, i.e. channels with little spike characteristics, we calculate a measure of spikiness based on the fourth standardized moment of each observation channel as follows:

$$Spikiness(\{\tilde{x}_m[n]\}) = \frac{\langle \mathrm{E}[(\tilde{x}_m[n] - \mathrm{E}[\tilde{x}_m[n]])^4] \rangle}{\langle \mathrm{E}[(\tilde{x}_m[n] - \mathrm{E}[\tilde{x}_m[n]])^2] \rangle^2} \tag{2.10}$$

where $\langle . \rangle$ stands for the temporal mean. We then eliminate the first N_{ue} channels with minimum Spikiness. The remaining channels are called "epileptic channels" in the following. Then we consider a symmetric window around each peak time (extracted in the previous step) and calculate the average value of the epileptic channels for all samples in the defined window. Then each signal is low pass filtered with a moving average filter of length $L_{ma,1}$. Up to this point, for each spike we have a smooth signal around the spike peak time. To find the start and end samples of each spike, we define a curvature measure as follows:

$$Curv_x[n] \;=\; \sum_{i=1}^{5} \cos(\theta_i^{(n)}) \tag{2.11}$$

where $\theta_i^{(n)}$ is the angle between two line segments that connect the points $(n-i, x[n-i])$ and $(n+i, x[n+i])$ to $(n, x[n])$ as shown in figure 2.4. Then we use Th_θ as a threshold to define the start and end points of the spike. Since the data are noisy, we must ensure that the curvature of at least N_θ consecutive points are greater than Th_θ. Thus, the start point of a spike is chosen as a first sample where its curvature and the curvature of the $N_\theta - 1$ following samples are greater than Th_θ. Similarly, the end point of a spike is chosen as a last sample where its curvature and the curvature of the $N_\theta - 1$ previous samples are greater than Th_θ.

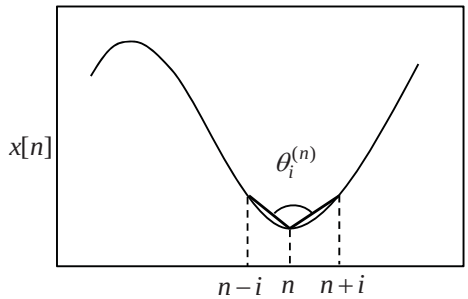


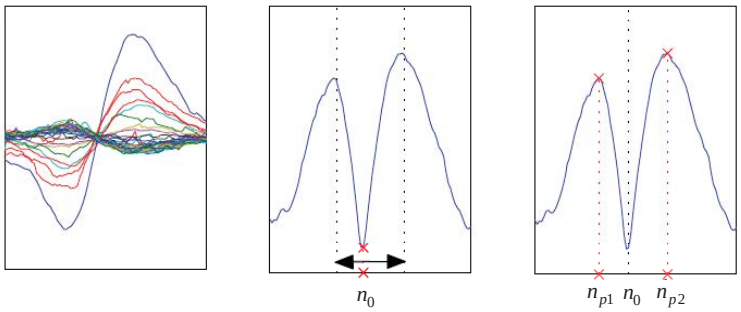
Figure 2.4: The angle $\theta_i^{(n)}$ used in the curvature definition

2.1.3.3 Spike clustering

In this stage, the spikes extracted in the previous stages are clustered into P_e' clusters corresponding to P_e' epileptic sources. The morphological features of spikes (including the peak location, the zero-crossing sample, slope and ...) originating from different epileptic sources are slightly different. These features may be used to cluster the spikes according to their corresponding epileptic sources. In this thesis, we use the ratio between the spike amplitudes on all electrodes at two main peak times. Since the location of epileptic sources in the brain has a direct impact on the spike amplitudes of different channels, this vector of features can effectively cluster the spikes. Since for the simulated data of this thesis we use biphasic spikes, we consider two peaks for each spike in this section. To extract this vector of features for each spike, we do the following procedure on each spike truncated at its extracted time support. These truncated spikes are represented by $\{\boldsymbol{x}^{(\ell)}[n_\ell]\}$ with $1 \leq \ell \leq N_s$ where N_s is the number of spikes. So, we should first detect the occurrence time of two main peaks as shown in figure 2.5. Note that the spike peak times extracted in the first preprocessing stage may match one of these peaks, but in this stage we determine all of the peak times. To this end, for each spike, we calculate the moving average low pass filtered signal (with the filter of length $L_{ma,2}$) of each channel as shown in figure 2.5(a). Then, we calculate the average of absolute values of amplitudes over all channels in each time sample, represented by $\{\overline{|x^{(\ell)}|}[n_\ell]\}$ as shown in figure 2.5(b). Then, the time of zero-crossing in approximately middle of the signal is determined. To find this point, the sample with minimum value in the middle third of the signal is extracted (n_0) as shown in figure 2.5(b). After that, for $\{\overline{|x^{(\ell)}|}[n_\ell]\}$, we consider two segments: one from the start point to n_0 and the other one from n_0 to end, and for each segment we find the point with maximum value as shown in figure 2.5(c). These two arguments of maximum, $n_{p1}^{(\ell)}$ and $n_{p2}^{(\ell)}$, correspond to the occurrence time of the negative and positive peaks of the spike. Then we use the values of observations of different channels in the time samples $n_{p1}^{(\ell)}$ and $n_{p2}^{(\ell)}$ to generate a normalized vector of features for each spike as follows:

$$\forall 1 \leq \ell \leq N_s, \qquad \boldsymbol{f}[\ell] = [\frac{\boldsymbol{x}[n_{p1}^{(\ell)}]^\intercal}{\sqrt{\sum_{m=1}^{M} x_m[n_{p1}^{(\ell)}]}}, \frac{\boldsymbol{x}[n_{p2}^{(\ell)}]^\intercal}{\sqrt{\sum_{m=1}^{M} x_m[n_{p2}^{(\ell)}]}}]^\intercal \tag{2.12}$$

Then by using the feature vectors of all spikes, $\boldsymbol{f}[\ell]$ for $\ell \in \{1, \ldots, N_s\}$, and the Fuzzy C-Means (FCM) clustering algorithm [Bezdek et al., 1984], the spikes are classified into



(a) (b) (c)

Figure 2.5: The procedure of feature extraction from one spike: (a) the low-pass filtered signals of all channels, (b) extracting the zero-crossing sample of $\overline{|x^{(\ell)}|}[k_\ell]$, (c) extracting the occurrence time of two negative and positive peaks

P_e groups corresponding to P_e epileptic sources. We represent the cluster containing all time samples corresponding to the p-th epileptic source by $cluster(p)$ and the sets $T_1^{(p)}$ and $T_2^{(p)}$ are defined as follows:

$$\begin{aligned} T_1^{(p)} &= \{n : n \in cluster(p)\} \\ T_2^{(p)} &= \{n : n \notin cluster(p)\} \end{aligned} \tag{2.13}$$

2.1.3.4 The GEVD-based method for denoising interictal EEGs

Let us assume that the vector random process $\{\tilde{\boldsymbol{x}}[n]\}$ can be approximately decomposed as a linear combination of decorrelated random processes, say as $\tilde{\boldsymbol{x}}[n] = \boldsymbol{G}\tilde{\boldsymbol{r}}[n]$ where P'_e components of $\{\tilde{\boldsymbol{r}}[n]\}$ correspond to the components of $\{\tilde{\boldsymbol{r}}^{(e)}[n]\}$, i.e. they span the epileptic source subspace, and the other $P' - P'_e$ components span the noise subspace. To extract these P'_e epileptic components as well as their corresponding columns in the mixing matrix $\boldsymbol{G}$, we use a GEVD-based subspace decomposition algorithm. In this algorithm, we extract each epileptic component separately and then reconstruct the denoised observation signal corresponding to epileptic sources. The whole procedure to extract the epileptic subspace is shown in figure 2.6. For the epileptic source $p \in \{1, \ldots, P'_e\}$ and for each time sample, we define two types of covariance matrices of the random vector $\{\tilde{\boldsymbol{x}}[n]\}$ as follows:

$$\begin{aligned} \forall n_1 \in T_1^{(p)}, \qquad \boldsymbol{C}_x^{(p)}[n_1] &= \mathrm{E}[(\tilde{\boldsymbol{x}}[n_1] - \mathrm{E}[\tilde{\boldsymbol{x}}[n_1]])(\tilde{\boldsymbol{x}}[n_1] - \mathrm{E}[\tilde{\boldsymbol{x}}[n_1]])^{\mathsf{T}}] \\ &= \boldsymbol{G}\boldsymbol{C}_r^{(p)}[n_1]\boldsymbol{G}^{\mathsf{T}} \end{aligned} \tag{2.14}$$

$$\begin{aligned} \forall n_2 \in T_2^{(p)}, \qquad \boldsymbol{C}_x^{(p)}[n_2] &= \mathrm{E}[(\tilde{\boldsymbol{x}}[n_2] - \mathrm{E}[\tilde{\boldsymbol{x}}[n_2]])(\tilde{\boldsymbol{x}}[n_2] - \mathrm{E}[\tilde{\boldsymbol{x}}[n_2]])^{\mathsf{T}}] \\ &= \boldsymbol{G}\boldsymbol{C}_r^{(p)}[n_2]\boldsymbol{G}^{\mathsf{T}} \end{aligned} \tag{2.15}$$

where the sets $T_1^{(p)}$ and $T_2^{(p)}$ are defined in (2.13) and the matrices $\boldsymbol{C}_{\tilde{r}}^{(p)}[n_1]$ and $\boldsymbol{C}_{\tilde{r}}^{(p)}[n_2]$ are the diagonal covariance matrices of the source vector $\{\tilde{\boldsymbol{r}}[n]\}$ at two time samples $n_1 \in T_1^{(p)}$ and $n_2 \in T_2^{(p)}$, respectively.

To identify the p-th epileptic component $\{r_p^{(e)}[k]\}$, we maximize the Rayleigh quotient defined as follows:

$$J^{(p)}(\boldsymbol{w}) = \frac{\boldsymbol{w}^{\mathsf{T}} \langle \boldsymbol{C}_x^{(p)}[n_1] \rangle_{T_1^{(p)}} \boldsymbol{w}}{\boldsymbol{w}^{\mathsf{T}} \langle \boldsymbol{C}_x^{(p)}[n_2] \rangle_{T_2^{(p)}} \boldsymbol{w}} \tag{2.16}$$

where $\langle \boldsymbol{C}_x^{(p)}[n_j] \rangle_{T_j^{(p)}}$ is the sample estimate (by time averaging) of the covariance matrices $\boldsymbol{C}_x^{(p)}[n_j]$ on the interval $T_j^{(p)}$ and $\boldsymbol{w}$ is the estimated separator. In practice and with ergodicity hypothesis, this average covariance matrix can be estimated from a realization $\{\boldsymbol{x}[n]\}_{n \in T_j^{(p)}}$ of the vector random process $\{\tilde{\boldsymbol{x}}[n]\}_{n \in T_j^{(p)}}$ as follows:

$$\forall j \in \{1,2\},\ \langle \boldsymbol{C}_x^{(p)}[n_j] \rangle_{T_j^{(p)}} \approx \frac{1}{n(T_j^{(p)})} \sum_{n \in T_j^{(p)}} \Big(\{\boldsymbol{x}[n]\} - \langle \{\boldsymbol{x}[n]\} \rangle_{T_j^{(p)}} \Big) \Big(\{\boldsymbol{x}[n]\} - \langle \{\boldsymbol{x}[n]\} \rangle_{T_j^{(p)}} \Big)^{\mathsf{T}} \tag{2.17}$$

where $n(T_j^{(p)})$ represents the number of samples in the set $T_j^{(p)}$ and $\langle \{\boldsymbol{x}[n]\} \rangle_{T_j^{(p)}} = \frac{1}{n(T_j^{(p)})} \sum_{n \in T_j^{(p)}} \boldsymbol{x}[n]$. Finding the argument of the maximum of (2.16) is equivalent to jointly diagonalizing the matrices $\langle \boldsymbol{C}_x^{(p)}[n_1] \rangle_{T_1^{(p)}}$ and $\langle \boldsymbol{C}_x^{(p)}[n_2] \rangle_{T_2^{(p)}}$. Consequently, the separator vector can be found by solving a problem of joint diagonalization by congruence. To this end, the GEVD of the pair of matrices $\langle \boldsymbol{C}_x^{(p)}[n_1] \rangle_{T_1^{(p)}}$ and $\langle \boldsymbol{C}_x^{(p)}[n_2] \rangle_{T_2^{(p)}}$ is solved in order to maximize (2.16). More particularly, the vector $\boldsymbol{w}_p$ maximizing $J^{(p)}$ (2.16) is computing as the eigenvector associated with the largest eigenvalue of matrix $\langle \boldsymbol{C}_x^{(p)}[n_2] \rangle_{T_2^{(p)}}^{-1} \langle \boldsymbol{C}_x^{(p)}[n_1] \rangle_{T_1^{(p)}}$. An estimate of $\{r_p^{(e)}[n]\}$ is then given by $\{\hat{r}_p^{(e)}[n]\}$ where $\hat{r}_p^{(e)}[n] = \boldsymbol{w}_p^{\mathsf{T}} \boldsymbol{x}[n]$ for any time index n. Once the estimate $\{\hat{r}_p^{(e)}[n]\}$ of the P'_e epileptic components $\{r_p^{(e)}[n]\}$ have been computed, the estimated denoised EEG signal $\{\hat{\boldsymbol{x}}^{(e)}[n]\}$ is achieved by reconstructing the observation signals only from the estimated epileptic subspace as follows:

$$\hat{\boldsymbol{x}}^{(e)}[n] = \hat{\boldsymbol{G}}^{(e)} \hat{\boldsymbol{r}}^{(e)}[n] \tag{2.18}$$

where $\hat{\boldsymbol{r}}^{(e)}[n] = [\hat{r}_1^{(e)}[n], ..., \hat{r}_{P'_e}^{(e)}[n]]^{\mathsf{T}}$, $\hat{\boldsymbol{G}}^{(e)} = \boldsymbol{X}(\hat{\boldsymbol{R}}^{(e)})^{\sharp}$ and where $\boldsymbol{X}$ and $\hat{\boldsymbol{R}}^{(e)}$ are the $(M \times N)$ and $(P'_e \times N)$ matrices standing for N samples of the M-dimensional signal $\{\boldsymbol{x}[n]\}$ and N samples of the P'_e-dimensional reconstructed signal $\{\hat{\boldsymbol{r}}^{(e)}[n]\}$, respectively, with the assumption that $N \geq M$. Therefore, $\{\hat{\boldsymbol{x}}^{(e)}[n]\}$ contains essentially the contribution of the epileptic activity on the scalp electrodes.

2.1.3.5 The DSS-based algorithm for denoising interictal EEGs

In this section, as mentioned before, we use the information extracted from the preprocessing stages of the algorithm to design an appropriate denoising method in the DSS framework. The model (2.5) for the observation is considered. The flowchart of the proposed DSS algorithm to extract one epileptic source is shown in figure 2.7. In this

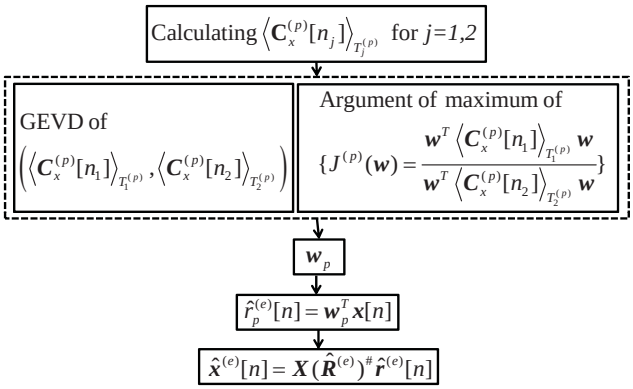


Figure 2.6: The procedure of extracting the epileptic subspace based on the GEVD method

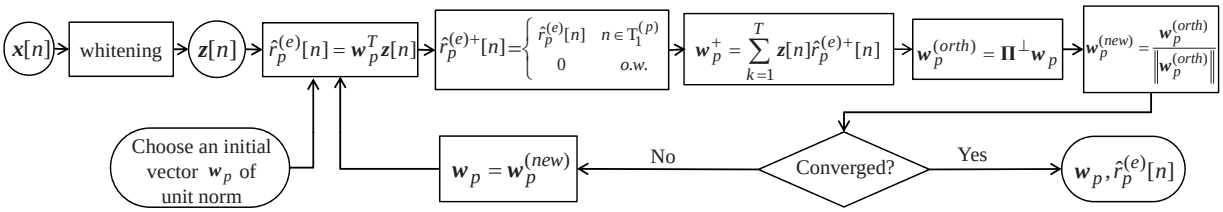


Figure 2.7: The DSS framework used to extract one epileptic source.

flowchart, $\{\hat{r}_p^{(e)}[n]\}$ and the vector $\boldsymbol{w}_p$ represent the p-th extracted epileptic source and its corresponding separator, respectively.

As shown in figure 2.7, the first step of the algorithm is a whitening of the observations $\{\boldsymbol{x}[n]\}$. The whitened signal $\{\boldsymbol{z}[n]\}$ is given by $\boldsymbol{z}[n] = \boldsymbol{\Theta}^{\sharp}\boldsymbol{x}[n]$ where $\boldsymbol{\Theta}$ is a square root of the covariance matrix of the noiseless observation vector process $\{\tilde{\boldsymbol{x}}[n] - \tilde{\boldsymbol{\nu}}[n]\}$. By means of this whitening procedure, the matrix $\boldsymbol{G}$ is transformed into an orthogonal matrix $\boldsymbol{W}$ of size $(P' \times P')$. The five sub-steps are the main steps of the DSS method which are repeated in an iterative procedure. First, by using an initial value of separator vector $\boldsymbol{w}_p$ of unit norm of size $(P' \times 1)$, a noisy estimate of the p-th source is calculated as follows:

$$\forall n, \quad \hat{r}_p^{(e)}[n] = \boldsymbol{w}_p^{\mathsf{T}} \boldsymbol{z}[n] \tag{2.19}$$

The next step, called the denoising step, is the main stage of the algorithm. In this stage, by using the pre-obtained information, the current estimated noisy source is modified or denoised. As stated in section 2.1.2, the general form of this stage is given by $\hat{r}_p^{(e)+}[n] = f(\hat{r}_p^{(e)}[n])$. In the DSS algorithm used in this thesis, the denoised source $\{\hat{r}_p^{(e)+}[n]\}$ is produced from the noisy source $\{\hat{r}_p^{(e)}[n]\}$ as follows:

$$\hat{r}_p^{(e)+}[n] = \begin{cases} \hat{r}_p^{(e)}[n] & if\ n \in T_1^{(p)} \\ 0 & otherwise \end{cases} \tag{2.20}$$

where $T_1^{(p)}$ is the set containing the time samples corresponding to the p-th source as defined in (2.13).

In the third, fourth and fifth steps, by using the denoised source $\{\hat{r}_p^{(e)+}[n]\}$, a new estimation of the separator vector is calculated and normalized as follows:

$$\boldsymbol{w}_p^+ = \sum_{n=1}^{N} \boldsymbol{z}[n]\hat{r}_p^{(e)+}[n] \tag{2.21}$$

$$\boldsymbol{w}_p^{(orth)} = \boldsymbol{\Pi}^{\perp}\boldsymbol{w}_p^+ \tag{2.22}$$

$$\boldsymbol{w}_p^{(new)} = \frac{\boldsymbol{w}_p^{(orth)}}{\|\boldsymbol{w}_p^{(orth)}\|} \tag{2.23}$$

where $\boldsymbol{\Pi}^{\perp}$ is defined by:

$$\boldsymbol{\Pi}^{\perp} = \boldsymbol{I} - \boldsymbol{W}^{(p-1)}(\boldsymbol{W}^{(p-1)^{\mathsf{T}}}\boldsymbol{W}^{(p-1)})^{-1}\boldsymbol{W}^{(p-1)^{\mathsf{T}}} = \boldsymbol{I} - \boldsymbol{W}^{(p-1)}\boldsymbol{W}^{(p-1)^{\mathsf{T}}} \tag{2.24}$$

$$\text{with} \quad \boldsymbol{W}^{(p-1)} = [\boldsymbol{w}_1, \ldots, \boldsymbol{w}_{p-1}] \tag{2.25}$$

It is clear that the matrix $\boldsymbol{\Pi}^{\perp}$ represents the orthogonal projector onto the subspace orthogonal to the previously extracted sources, so by multiplying this matrix by $\boldsymbol{w}_p^+$, the effect of previously extracted sources is removed. By using this deflation step in the DSS framework, the convergence to previously extracted sources is prevented [Sarela and Valpola, 2005]. Then, the present estimated separator vector $\boldsymbol{w}_p$ is passed to the first block and this procedure continues until convergence. The outputs of the algorithm are $\boldsymbol{w}_p$ and $\hat{r}_p^{(e)}[n] = \boldsymbol{w}_p^{\mathsf{T}}\boldsymbol{z}[n]$. By using the same procedure, all epileptic sources are extracted. The estimated denoised EEG signal can then be achieved as follows:

$$\forall k, \quad \hat{\boldsymbol{x}}^{(e)}[n] = \hat{\boldsymbol{G}}^{(e)}\hat{\boldsymbol{r}}^{(e)}[n] \tag{2.26}$$

where $\hat{\boldsymbol{G}}^{(e)} = \boldsymbol{\Theta}\hat{\boldsymbol{W}}^{(e)}$ with $\hat{\boldsymbol{W}}^{(e)} = [\boldsymbol{w}_1, ..., \boldsymbol{w}_{P_e'}]$.

2.1.3.6 Setting the parameters

In order to optimize the performance of the GEVD and DSS methods, we set the value of the parameters Th_{spike}, N_{ue}, L_w, N_θ, $L_{ma,1}$ and $L_{ma,2}$ equal to 80, 17, 81, 7, 9 and 19, respectively. The value of N_{eff} was set to be 4 and 6 for the first scenario and the other scenarios, respectively. Th_θ was set to be $0.4\overline{curv}_x$ where $\overline{curv_x}$ is the average of the curvature measure in the examined interval. Noted that these values were estimated from a few trials of the simulated data and then we used these constant values dealing with all data.

2.1.4 The TF-GEVD and TF-DSS methods for denoising of ictal data

As previously said, in this section we are interested in proposing GEVD/DSS-based source separation methods to denoise ictal EEGs. To this end, we look for some *a priori* information on the sources of interest to define the covariance matrices in the GEVD method and in denoising step of the DSS method. Accordingly, the proposed methods, namely TF-GEVD and TF-DSS, use the time-frequency signature of ictal

discharges to denoise ictal EEGs. To extract the time-frequency information on ictal sources, we use the CCA approach that has been shown to be the most efficient method to date, to denoise ictal EEG signals contaminated by muscular activity [De Clercq et al., 2006, Hallez et al., 2009, Safieddine, 2012, Karfoul et al., 2015].

Note that, in this section, we use the statistical model as presented in section 1.4.1. In addition, we denote by $\{T_{\tilde{x}_m}[n,k]\}$ the random process defined as the linear time-frequency transform of the random process $\{\tilde{x}_m[n]\}$, where n and k stand for the time and frequency indices, respectively. The M-dimensional random vector process $\{T_{\tilde{\boldsymbol{x}}}[n,k]\}$ then contains the linear time-frequency samples of the M channels. Therefore, according to (2.5), for each couple (n,k) of $\{1,...,N\}\times\{1,...,K\}$, we have:

$$T_{\tilde{\boldsymbol{x}}}[n,k] = \sum_{p=1}^{P} \boldsymbol{g}_p\, T_{\tilde{r}_p}[n,k] = \boldsymbol{G}\, T_{\tilde{\boldsymbol{r}}}[n,k] \tag{2.27}$$

where $\{T_{\tilde{r}_p}[n,k]\}$ and $\{T_{\tilde{\boldsymbol{r}}}[n,k]\}$ represent the linear time-frequency transform of the random processes $\{\tilde{r}_p[n]\}$ and $\{\tilde{r}[n]\}$, respectively. We also denote the linear time-frequency transform of a realization $\{\boldsymbol{x}[n]\}$ of the random vector process $\{\tilde{\boldsymbol{x}}[n]\}$ by $\{T_{\boldsymbol{x}}[n,k]\}$.

We can consider that the energy of the epileptic sources is local in the time-frequency domain. This hypothesis comes from the observation made by clinicians and epileptologists, that at the beginning of seizures, the fast ictal activity is very narrow band, and of low amplitude. According to this property, for the p-th ictal source we assume that the largest portion of its energy is reserved in a time-frequency subset $TF^{(p)}$ of $\{1,...,N\}\times\{1,...,K\}$. Thus, our aim is to compute the epileptic subspace $\{\boldsymbol{x}^{(e)}[n]\}$ using this *a priori* information on the ictal sources through specific time-frequency matrices. Given an L-dimensional random vector process $\{\tilde{\boldsymbol{u}}[n]\}$ and its linear time-frequency transform $\{T_{\tilde{\boldsymbol{u}}}[n,k]\}$, we define the $(L\times L)$ time-frequency covariance matrices $\boldsymbol{C}_{\boldsymbol{u}}^{(p)}[n_1,k_1]$ and $\boldsymbol{C}_{\boldsymbol{u}}^{(p)}[n_2,k_2]$ as follows:

$$\begin{aligned}
&\forall (n_1,k_1)\in TF^{(p)},\ \boldsymbol{C}_{\boldsymbol{u}}^{(p)}[n_1,k_1] = \mathrm{E}[(T_{\tilde{\boldsymbol{u}}}[n_1,k_1]-\mathrm{E}[T_{\tilde{\boldsymbol{u}}}[n_1,k_1]])(T_{\tilde{\boldsymbol{u}}}[n_1,k_1]-\mathrm{E}[T_{\tilde{\boldsymbol{u}}}[n_1,k_1]])^{\mathsf{T}}]\\
&\forall (n_2,k_2)\notin TF^{(p)},\ \boldsymbol{C}_{\boldsymbol{u}}^{(p)}[n_2,k_2] = \mathrm{E}[(T_{\tilde{\boldsymbol{u}}}[n_2,k_2]-\mathrm{E}[T_{\tilde{\boldsymbol{u}}}[n_2,k_2]])(T_{\tilde{\boldsymbol{u}}}[n_2,k_2]-\mathrm{E}[T_{\tilde{\boldsymbol{u}}}[n_2,k_2]])^{\mathsf{T}}]
\end{aligned} \tag{2.28}$$

We also represent the time-frequency average of the covariance matrices $\boldsymbol{C}_{\boldsymbol{u}}^{(p)}[n_1,k_1]$ and $\boldsymbol{C}_{\boldsymbol{u}}^{(p)}[n_2,k_2]$ on their time-frequency support by $\langle \boldsymbol{C}_{\boldsymbol{u}}^{(p)}[n_1,k_1]\rangle_{TF^{(p)}}$ and $\langle \boldsymbol{C}_{\boldsymbol{u}}^{(p)}[n_2,k_2]\rangle_{\overline{TF}^{(p)}}$, respectively. In practice, these average time-frequency covariance matrices can be estimated from a realization $\{T_{\boldsymbol{u}}[n,k]\}$ of the vector random process $\{T_{\tilde{\boldsymbol{u}}}[n,k]\}$ as follows:

$$\begin{aligned}
\langle \boldsymbol{C}_{\boldsymbol{u}}^{(p)}[n_1,k_1]\rangle_{TF^{(p)}} &\approx \frac{1}{n(TF^{(p)})}\sum_{(n_1,k_1)\in TF^{(p)}} (T_{\tilde{\boldsymbol{u}}}[n_1,k_1]-\langle T_{\tilde{\boldsymbol{u}}}[n_1,k_1]\rangle)(T_{\tilde{\boldsymbol{u}}}[n_1,k_1]-\langle T_{\tilde{\boldsymbol{u}}}[n_1,k_1]\rangle)^{\mathsf{T}}\\
\langle \boldsymbol{C}_{\boldsymbol{u}}^{(p)}[n_2,k_2]\rangle_{\overline{TF}^{(p)}} &\approx \frac{1}{NK-n(TF^{(p)})}\sum_{(n_2,k_2)\notin TF^{(p)}} (T_{\tilde{\boldsymbol{x}}}[n_2,k_2]-\langle T_{\tilde{\boldsymbol{u}}}[n_2,k_2]\rangle)(T_{\tilde{\boldsymbol{u}}}[n_2,k_2]-\langle T_{\tilde{\boldsymbol{u}}}[n_2,k_2]\rangle)^{\mathsf{T}}
\end{aligned} \tag{2.29}$$

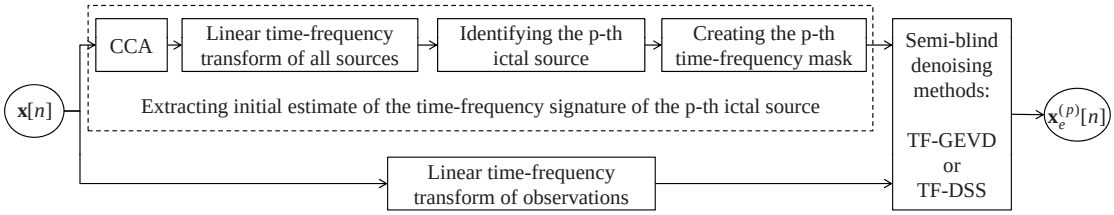


Figure 2.8: The flow chart of the denoising procedure to extract the p-th ictal subspace corresponding to the p-th ictal source by TF-GEVD and TF-DSS methods.

where $n(TF^{(p)})$ represents the number of samples in the set $TF^{(p)}$. Here, for simplicity, we represent matrices $\langle \boldsymbol{C}_{\boldsymbol{u}}^{(p)}[n_1,k_1]\rangle_{TF^{(p)}}$ and $\langle \boldsymbol{C}_{\boldsymbol{u}}^{(p)}[n_2,k_2]\rangle_{\overline{TF}^{(p)}}$ by $\boldsymbol{C}_{\boldsymbol{u}}^{(p)}$ and $\boldsymbol{C}_{\boldsymbol{u}}^{(\bar{p})}$, respectively. Consequently, the average time-frequency covariance matrix on all time-frequency samples, $\boldsymbol{C}_{\boldsymbol{u}}$, can be defined as follows:

$$\boldsymbol{C}_{\boldsymbol{u}} = \frac{1}{NK}\sum_{n=1}^{N}\sum_{k=1}^{K}(T_{\tilde{\boldsymbol{u}}}[n,k]-\langle T_{\tilde{\boldsymbol{u}}}[n,k]\rangle)(T_{\tilde{\boldsymbol{u}}}[n,k]-\langle T_{\tilde{\boldsymbol{u}}}[n,k]\rangle)^{\mathsf{T}} \tag{2.30}$$

The general flowchart of the proposed methods to extract the ictal subspace corresponding to the p-th ictal source is shown in Figure 2.8. As illustrated in this figure, the proposed methods consist of three main steps: 1) extracting the linear time-frequency signature of the p-th ictal source, 2) calculating the time-frequency transform of the observations and 3) applying a time-frequency-based denoising method to the linear time-frequency transform of the observations. The proposed methods only differ in the third block where either TF-GEVD or TF-DSS is used for the denoising procedure. These steps are explained in details in the rest of this section and then the relation between the two proposed methods is studied.

2.1.4.1 Extracting time-frequency mask corresponding to each ictal source

The CCA-based BSS method extracts the sources from observations with the main assumption that the sources are mutually uncorrelated as well as maximally autocorrelated [De Clercq et al., 2006]. As shown in previous studies [De Clercq et al., 2006, Safieddine, 2012, Karfoul et al., 2015], CCA surpasses other methods for the denoising of ictal signals. Nevertheless it can not achieve sufficiently good results in some instances such as low SNRs. In the method we propose, we use the CCA method to extract some *a priori* information on ictal sources in order to improve the denoising process. To this end, we apply the CCA method on the observed EEGs to obtain M sources. The next step consists in selecting the sources of ictal activity. A visual selection of ictal sources from their time course is difficult and sometimes impossible. Therefore different methods have been proposed to select ictal sources from all extracted ones, such as methods based on the time-frequency transform of each source [Safieddine, 2012] and methods based on spectral coherence of sources [Subramaniyam et al., 2013]. To select the sources corresponding to ictal activities, we use the linear time-frequency representation of each source and then we choose the sources of interest based on a

signature that consists of a narrow-band, high frequency signal at the beginning of the seizure, that decreases in frequency and increases in amplitude over time. Figures 2.9(a) and 2.9(b) show an example of the selected ictal source extracted by CCA and its linear time-frequency representation, respectively. In this thesis, we use the Continuous Wavelet Transform (CWT) with real Morlet mother wavelet which shows a good representation of ictal sources in the time-frequency domain. As explained later in this section, since the real-valued data should be used in the proposed algorithms, CWT with real Morlet mother wavelet is an appropriate choice. After selecting the sources of interest, we use an automated procedure to generate the time-frequency mask. To this end, we use an appropriate threshold to select the dominant time-frequency samples from the linear time-frequency representation of each source of interest. The threshold used for the source of interest $\{r_p[n]\}$ is obtained as $th_{r_p} = m_{|T_{r_p}|} + \sigma_{|T_{r_p}|}$ where $m_{|T_{r_p}|}$ and $\sigma_{|T_{r_p}|}$ are the mean value and standard deviation of the absolute values of $\{T_{r_p}[n,k]\}$ over all $(n,k) \in \{1,...,N\} \times \{1,...,K\}$, respectively. Then the absolute value of $\{T_{r_p}[n,k]\}$ is compared with th_{r_p} at each time-frequency sample (n,k) and the p-th time-frequency mask $S^{(p)}$ is obtained as follows:

$$S^{(p)}[n,k] = \begin{cases} 1 & \text{if } |T_{r_p}[n,k]| \geq th_{r_p} \\ 0 & \text{o.w.} \end{cases} \tag{2.31}$$

Then, we consider the obtained time-frequency mask $\{S^{(p)}[n,k]\}$ as a binary image and modify it by using morphological operators. To this end, we use the morphological "*closing*" operator [Gonzalez and Woods, 2002] with a disk structuring element of size 1. The "*closing*" operator performs morphological closing on the grayscale or binary image by applying a dilation followed by an erosion, using the same structuring element for both operations. By using this operator, the holes in the initial time-frequency mask are filled. Figure 2.9(c) shows the time-frequency mask created by the above-mentioned procedure. The time-frequency samples corresponding to the p-th ictal source are then stored in the set $TF^{(p)}$:

$$TF^{(p)} = \{(n^{(p)}, k^{(p)}) \in \text{the } p\text{-th ictal mask}\} \tag{2.32}$$

The obtained samples will then be used in the proposed TF-GEVD and TF-DSS methods to denoise ictal signals.

2.1.4.2 TF-GEVD

In this section, we propose a GEVD-based method to identify the p-th ictal source using its time-frequency support $TF^{(p)}$. More particularly, we aim at computing the M-dimensional vector $\boldsymbol{w}_p$ such that the signal $\{\hat{r}_p^{(e)}[n]\}$ defined by $\hat{r}_p^{(e)}[n] = \boldsymbol{w}_p{}^\top \boldsymbol{x}[n]$ is an estimate of the p-th ictal source $\{r_p^{(e)}[n]\}$. To this end, we maximize the following quotient of energies on and outside the time-frequency support $TF^{(p)}$:

$$J^{(p)}(\boldsymbol{w}) = \frac{\boldsymbol{w}^\top \boldsymbol{C}_{\boldsymbol{x}}^{(p)} \boldsymbol{w}}{\boldsymbol{w}^\top \boldsymbol{C}_{\boldsymbol{x}}^{(\bar{p})} \boldsymbol{w}} \tag{2.33}$$

which is a Rayleigh quotient. Finding the argument $\boldsymbol{w}_p$ of the maximum of (2.33) is equivalent to solving a specific GEVD problem, i.e. computing the eigenvector associated with the largest eigenvalue of the $(M \times M)$ matrix $\left(\boldsymbol{C}_{\boldsymbol{x}}^{(\bar{p})}\right)^{-1} \boldsymbol{C}_{\boldsymbol{x}}^{(p)}$.

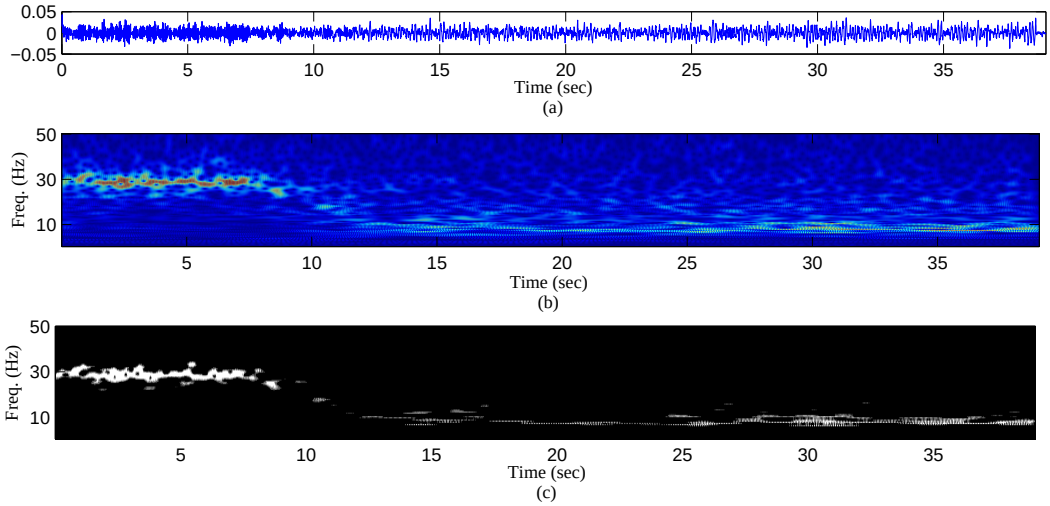


Figure 2.9: Procedure used to create the time-frequency mask corresponding to an ictal source, (a) selected ictal source generated by CCA, (b) time-frequency representation (absolute value of CWT) of the ictal source and (c) time-frequency mask corresponding to the ictal source of (a).

By executing the abovementioned procedure for each $p \in \{1, ..., P'_e\}$, P'_e vectors $\boldsymbol{w}_p$ are extracted and the matrix $\hat{\boldsymbol{G}}^{(e)}$ of coordinates can be estimated as $\boldsymbol{X}\left(\hat{\boldsymbol{R}}^{(e)}\right)^{\sharp}$ where $\boldsymbol{X}$ and $\hat{\boldsymbol{R}}^{(e)}$ are the $(M \times N)$ and $(P'_e \times N)$ matrices standing for the M-dimensional signal $\{\boldsymbol{x}[n]\}$ and the P'_e-dimensional signal $\{\hat{\boldsymbol{r}}^{(e)}[n]\}$, respectively, with $\hat{\boldsymbol{r}}^{(e)}[n] = [\hat{r}_1^{(e)}[n], \cdots, \hat{r}_{P'_e}^{(e)}[n]]^{\mathsf{T}}$.

2.1.4.3 TF-DSS

In this section, we propose a DSS-based method to denoise ictal EEG data in the time-frequency domain. The TF-DSS algorithm is based on a whitening of the linear time-frequency transform of observations, $\{T_{\boldsymbol{x}}[n,k]\}$, such that the time-frequency covariance matrix $\boldsymbol{C}_z$ of the output signal $\boldsymbol{z}[n] = \boldsymbol{\Theta}^{\sharp}\boldsymbol{x}[n]$ is equal to the identity matrix. In practice, $\boldsymbol{\Theta}$ is computed as a square root of the time-frequency matrix $\boldsymbol{C}_x$. By means of this whitening procedure, the matrix $\boldsymbol{G}$ is transformed into an orthogonal matrix of size $(P' \times P')$.

The four consequent steps are the main steps of the TF-DSS method which are repeated in an iterative procedure to extract the p-th ictal source. First, by using a P'-dimensional initial vector $\boldsymbol{w}_p$ of unit norm, a noisy estimate of the time-frequency transform of the p-th ictal source for each time-frequency sample (n,k) is calculated as follows:

$$T_{\hat{r}_p^{(e)}}[n,k] = \boldsymbol{w}_p{}^{\mathsf{T}}\, T_{\boldsymbol{z}}[n,k] \tag{2.34}$$

The next step, called the denoising step, is the main stage of the algorithm. In this stage, the current estimated time-frequency transform of the p-th ictal source, for each

time-frequency sample (n, k), is modified or denoised as follows:

$$T^{+}_{\hat{r}_p^{(e)}}[n,k] = \begin{cases} T_{\hat{r}_p^{(e)}}[n,k] & if\ (n,k) \in TF^{(p)} \\ 0 & otherwise \end{cases} \tag{2.35}$$

In the third and fourth steps, by using the denoised time-frequency transform of the p-th ictal source, namely $\{T^{+}_{\hat{r}_p^{(e)}}[n,k]\}$, a new estimation of the mixing vector is calculated and normalized as follows:

$$\boldsymbol{w}_p^{+} = \sum_{n=1}^{N}\sum_{k=1}^{K} T_{\boldsymbol{z}}[n,k] T^{+}_{r_p^{(e)}}[n,k] \tag{2.36}$$

$$\boldsymbol{w}_p = \frac{\boldsymbol{w}_p^{+}}{\|\boldsymbol{w}_p^{+}\|} \tag{2.37}$$

Then, the present estimated mixing vector $\boldsymbol{w}_p$ is passed to the first step given by (2.34) and this procedure continues until convergence. The global convergence of the iterative DSS scheme is proved in [Sarela and Valpola, 2005]. The outputs of the algorithm, $\boldsymbol{w}_p$ and the signal $\{\hat{r}_p^{(e)}[n]\}$, which is defined by $\hat{r}_p^{(e)}[n] = \boldsymbol{w}_p{}^{\mathsf{T}}\, \boldsymbol{z}[n]$, are the separator and an estimate of the p-th ictal source.

It should be noted that if we have extracted the first $p-1$ sources, to extract the p-th source, the fourth step given by (2.37) can be replaced by:

$$\boldsymbol{w}_p^{\perp} = \boldsymbol{\Pi}^{\perp}\, \boldsymbol{w}_p^{+} \tag{2.38}$$

$$\boldsymbol{w}_p = \frac{\boldsymbol{w}_p^{\perp}}{\|\boldsymbol{w}_p^{\perp}\|} \tag{2.39}$$

where the orthogonal projector $\boldsymbol{\Pi}^{\perp}$ is defined by:

$$\boldsymbol{\Pi}^{\perp} = \boldsymbol{I} - \boldsymbol{W}^{(p-1)}(\boldsymbol{W}^{(p-1)^{\mathsf{T}}}\boldsymbol{W}^{(p-1)})^{-1}\boldsymbol{W}^{(p-1)^{\mathsf{T}}} = \boldsymbol{I} - \boldsymbol{W}^{(p-1)}\boldsymbol{W}^{(p-1)^{\mathsf{T}}} \tag{2.40}$$

with:

$$\boldsymbol{W}^{(p-1)} = [\boldsymbol{w}_1, \ldots, \boldsymbol{w}_{p-1}] \tag{2.41}$$

By adding this deflation step to the DSS framework, the convergence to previously extracted ictal sources is prevented [Sarela and Valpola, 2005].

By executing the above-mentioned procedure for each $p \in \{1, ..., P'_e\}$, all separators $\boldsymbol{w}_p$ and consequently the corresponding ictal source are estimated. The matrix $\hat{\boldsymbol{G}}^{(e)}$ of coordinates can then be obtained similarly to the TF-GEVD method as described at the end of section 2.1.4.2.

2.1.4.4 Relation between TF-GEVD and TF-DSS methods

First, for the TF-DSS method, we do not consider the deflation step (2.38), which means that only one ictal source has to be estimated.

In the TF-GEVD method, the GEVD problem of the pair of matrices $\boldsymbol{C}_{\boldsymbol{x}}^{(p)}$ and $\boldsymbol{C}_{\boldsymbol{x}}^{(\bar{p})}$ can be reformulated as an exact joint diagonalization by congruence problem of these two matrices.

On the other hand, when the denoising stage in DSS algorithm is linear, the whole iterative algorithm can be expressed as a closed-form solution. In section 2.1.4.3, by inserting (2.34) and (2.35) into (2.36), we can show that the four steps of the TF-DSS algorithm can be summarized by the following procedure:

$$\boldsymbol{w}_p^+ = \boldsymbol{C}_{\boldsymbol{z}}^{(p)} \boldsymbol{w}_p \tag{2.42}$$

$$\boldsymbol{w}_p = \frac{\boldsymbol{w}_p^+}{\|\boldsymbol{w}_p^+\|} \tag{2.43}$$

which is actually the classical power method applied to matrix $\boldsymbol{C}_{\boldsymbol{z}}^{(p)}$ in order to compute its dominant eigenvector $\boldsymbol{w}_p$. Let $\boldsymbol{W}^{(p)}$ be the orthogonal matrix of the P'_e eigenvectors (including $\boldsymbol{w}_p$) of the symmetric matrix $\boldsymbol{C}_{\boldsymbol{z}}^{(p)}$. Consequently, the $(P'_e \times M)$ matrix $\boldsymbol{W}_{\text{DSS}}^{(p)\,\top}$ given by $\boldsymbol{W}_{\text{DSS}}^{(p)\,\top} = \boldsymbol{W}^{(p)\top} \boldsymbol{\Theta}^{\sharp}$ diagonalizes by congruence the time-frequency matrix $\boldsymbol{C}_{\boldsymbol{x}}^{(p)}$. Moreover, due to the orthogonality of $\boldsymbol{W}^{(p)}$ and since $\boldsymbol{\Theta}$ is a root square of $\boldsymbol{C}_{\boldsymbol{x}}$, we can easily show that the matrix $\boldsymbol{W}_{\text{DSS}}^{(p)\,\top}$ also diagonalizes by congruence the time-frequency matrix $\boldsymbol{C}_{\boldsymbol{x}}$. As a result, the TF-DSS algorithm estimates the p-th ictal source by means of a joint diagonalization by congruence of the matrices $\boldsymbol{C}_{\boldsymbol{x}}^{(p)}$ and $\boldsymbol{C}_{\boldsymbol{x}}$.

Now by comparing the two joint diagonalization problems corresponding to the TF-GEVD and TF-DSS methods, the former simultaneously diagonalizes the matrices $\boldsymbol{C}_{\boldsymbol{x}}^{(p)}$ and $\boldsymbol{C}_{\boldsymbol{x}}^{(\bar{p})}$, while the latter jointly diagonalizes the matrices $\boldsymbol{C}_{\boldsymbol{x}}^{(p)}$ and $\boldsymbol{C}_{\boldsymbol{x}}$.

Now, if we consider the deflation step (2.38) for the TF-DSS method, equations (2.42) and (2.43) will be changed to:

$$\boldsymbol{w}_p^{\perp} = \boldsymbol{\Pi}^{\perp} \boldsymbol{C}_{\boldsymbol{z}}^{(p)} \boldsymbol{w}_p \tag{2.44}$$

$$\boldsymbol{w}_p = \frac{\boldsymbol{w}_p^{\perp}}{\|\boldsymbol{w}_p^{\perp}\|} \tag{2.45}$$

which results in diagonalization of matrix $\boldsymbol{\Pi}^{\perp} \boldsymbol{C}_{\boldsymbol{z}}^{(p)}$. Consequently, the TF-DSS method to extract the p-th ictal source can be described as two serial diagonalization steps, say diagonalization of $\boldsymbol{C}_{\boldsymbol{x}}$ and diagonalization of $\boldsymbol{\Pi}^{\perp} \boldsymbol{C}_{\boldsymbol{z}}^{(p)}$. But contrarily to the previous case, it seems that these two diagonalizations cannot be simplified to a joint diagonalization problem preventing us to compare more precisely the TF-DSS and TF-GEVD algorithms when two or more ictal sources have to be estimated.

2.2 The JDICA method

ICA is a very useful tool to process biomedical signals such as EEG data. ICA consists of retrieving an unobserved realizations $\{\boldsymbol{r}[n]\}$ of a P'-dimensional random vector process $\tilde{\boldsymbol{r}}[n]$ from an observed realization $\{\boldsymbol{x}[n]\}$ of a N-dimensional random vector process $\{\tilde{\boldsymbol{x}}[n]\}$ that can linearly be modeled as follows:

$$\tilde{\boldsymbol{x}}[n] = \boldsymbol{H}\tilde{\boldsymbol{r}}[n] + \tilde{\boldsymbol{\nu}} \tag{2.46}$$

The fundamental assumption of ICA is that the P' unknown sources $\{\tilde{r}_p[n]\}$ are mutually independent.

ICA algorithms can be divided into two groups: i)"symmetric" approaches jointly extract the independent components while ii) "deflationary" techniques estimate sources one by one. Symmetric algorithms seem to converge to the expected solution in practice, but no theoretical result is available. On the other hand, the convergence of most of deflationary algorithms have been proved analytically [Hyvärinen and Oja, 1997, Delfosse and Loubaton, 1995, Zarzoso and Comon, 2010]. In addition, in deflationary algorithms, a penalty term can be added to the contrast function [Mi, 2014] to force the algorithm to extract the sources of interest during the early steps. As a consequence, when the number of all sources largely surpasses the number of sources of interest, the computational complexity of the deflationary algorithms is greatly reduced.

Here, we propose a new Jacobi-like Deflationary ICA algorithm, called JDICA, based on second and Fourth Order (FO) statistics. The deflation procedure of our algorithm is inspired by [Delfosse and Loubaton, 1995]. The gradient-based ICA algorithm (called DelL$^{\mathbb{R}}$ throughout this manuscript) proposed in [Delfosse and Loubaton, 1995], estimates the sources one by one using a smart projection-based deflation scheme. According to its gradient-based structure, the step size must be precisely chosen to guarantee acceptable results specially with noisy data. A multi-initialization procedure can even be necessary in some practical contexts. In order to overcome these drawbacks, we propose a Jacobi-like algorithm to maximize the contrast function computed from the FO cumulants of the whitened observations. In the following sections, we first introduce the proposed JDICA algorithm and then show how it can be used in the context of denoising of epileptic EEG signals.

2.2.1 The JDICA algorithm: methodology

We assume that we have one realization $\{\boldsymbol{x}[n]\}$ of length N of the real-valued random vector process $\{\tilde{\boldsymbol{x}}[n]\}$ (2.46). Since JDICA, like a large group of ICA algorithms, needs a prewhitening step [Comon and Jutten, 2010] without loss of generality we assume that the process $\{\tilde{\boldsymbol{x}}[n]\}$ denotes the prewhitened observation random vector process and matrix $\boldsymbol{H} = [\boldsymbol{h}_1, ..., \boldsymbol{h}_{P'}]$ is a $(P' \times P')$ real-valued orthogonal mixing matrix. The aim of our method is then to estimate the P' columns $\boldsymbol{h}_p$ of $\boldsymbol{H}$ and the P' corresponding sources such that $\hat{r}_p[n] = \boldsymbol{h}_p^{\mathsf{T}}\boldsymbol{x}[n]$. More particularly, vector $\boldsymbol{h}_p$ can be identified by maximizing the following contrast function:

$$\mathfrak{F}(\boldsymbol{g}^{(p)}) = \frac{1}{4}\langle C_4(y_p[n])\rangle^2 = \frac{1}{4}\langle C_4(\boldsymbol{g}^{(p)^{\mathsf{T}}}\tilde{\boldsymbol{x}}[n])\rangle^2 \tag{2.47}$$

with respect to $\boldsymbol{g}^{(p)}$ where $C_4(y_p[n])$ is the FO marginal cumulant of $y_p[n] = \boldsymbol{g}^{(p)^{\mathsf{T}}}\tilde{\boldsymbol{x}}[n]$. The advantage of defining such a contrast function is that the arguments of the local maxima of $\mathfrak{F}$ on the unit sphere are the vectors $\{\pm\boldsymbol{h}_p\}_{p=1,...,P'}$ [Delfosse and Loubaton, 1995]. This property ensures our maximization (2.47) to converge to one column of the matrix $\boldsymbol{H}$ and consequently one of the sources is extracted. Thus a projection deflation procedure is applied to subtract the contribution of the extracted source from the mixture. These two steps require a particular parametrization of the elements of the unit sphere which is given by:

Lemma 1 Each unit norm column vector $\boldsymbol{g}\in\mathbb{R}^{P'}$ whose last component $g_{P'}$ is strictly positive can be represented as the last column of an orthogonal matrix given by:

$$\boldsymbol{G}(\boldsymbol{t}) = \boldsymbol{G}_{P'-1}(t_{P'-1})\dots\boldsymbol{G}_2(t_2)\boldsymbol{G}_1(t_1) \tag{2.48}$$

where the $P'-1$ real-valued elements of $\boldsymbol{t} = [t_1, ..., t_{P'-1}]^\mathsf{T}$ correspond to tangents of uniquely defined angles belonging to $]-\pi/2, \pi/2[$ and $\boldsymbol{G}_p(t_p)$ is a Givens rotation of size $(P'\times P')$ derived from an identity matrix for which the (p,p)-th, (P',P')-th, (p,P')-th, (P',p)-th components are replaced with $(1+t_p^2)^{-1/2}$, $(1+t_p^2)^{-1/2}$, $t_p(1+t_p^2)^{-1/2}$ and $-t_p(1+t_p^2)^{-1/2}$, respectively.

Proof derives from [Delfosse and Loubaton, 1995, lemma 2.2] by expressing $\cos(\theta)$ and $\sin(\theta)$ as a function of $t = \tan(\theta)$ [Comon and Jutten, 2010]. This parametrization differs from that of [Delfosse and Loubaton, 1995] and allows us both to reformulate the contrast (2.47) as a rational function and to consider other optimization strategies such as a Jacobi-like procedure.

To extract the first source, we then propose to compute a matrix $\boldsymbol{G}^{(1)}(\boldsymbol{t})$ such that its last column, $\boldsymbol{g}^{(1)}(\boldsymbol{t})$, maximizes the contrast function (2.47) with respect to $\boldsymbol{t}$. Our Jacobi-like optimization procedure consists then of decomposing $\boldsymbol{G}^{(1)}(\boldsymbol{t})$ as a product of $(P'-1)$ elementary Givens rotations $\boldsymbol{G}_p^{(1)}(t_p)$ and sequentially identifying the $(P'-1)$ corresponding parameters t_p. The $(P'-1)$-dimensional optimization problem is thus replaced with $(P'-1)$ sequential mono-dimensional optimization problems. In practice, several sweeps of the $(P'-1)$ parameters are necessary to achieve convergence. More precisely, let us consider the p-th mono-dimensional maximization problem of a sweep of our Jacobi-like procedure. It consists in computing matrix $\boldsymbol{G}^{(1+)}(t_p)$ defined by $\boldsymbol{G}^{(1+)}(t_p) = \boldsymbol{G}_p^{(1)}(t_p)\boldsymbol{G}^{(1-)}$ such that its last column, $\boldsymbol{g}^{(1+)}(t_p)$, maximizes the contrast function (2.47), where $\boldsymbol{G}^{(1-)}$ is the product of all the elementary Givens rotations estimated previously.

Denoting the last column of $\boldsymbol{G}^{(1-)}$ by $\boldsymbol{g}^{(1-)} = [g_1^{(1-)}, \dots, g_{P'}^{(1-)}]^\mathsf{T}$, the last column of $\boldsymbol{G}^{(1+)}$ can be written as:

$$\begin{gathered}\boldsymbol{g}^{(1+)}(t_p) = [g_1^{(1+)}(t_p), \dots, g_{P'}^{(1+)}(t_p)]^\mathsf{T} = \\ [g_1^{(1-)}, \dots, g_{p-1}^{(1-)}, g_p^{(1+)}(t_p), g_{p+1}^{(1-)}, \dots, g_{P'}^{(1+)}(t_p)]^\mathsf{T}\end{gathered} \tag{2.49}$$

where:

$$g_p^{(1+)}(t_p) = \frac{1}{\sqrt{1+t_p^2}}g_p^{(1-)} + \frac{t_p}{\sqrt{1+t_p^2}}g_{P'}^{(1-)} \tag{2.50}$$

$$g_{P'}^{(1+)}(t_p) = \frac{-t_p}{\sqrt{1+t_p^2}}g_p^{(1-)} + \frac{1}{\sqrt{1+t_p^2}}g_{P'}^{(1-)} \tag{2.51}$$

It appears that only the p-th and P'-th components of $\boldsymbol{g}^{(1+)}(t_p)$ depend on t_p. Then, we set the derivative of the contrast function with respect to t_p equal to zero to find the appropriate t_p value:

$$\frac{\partial\langle C_4(y_1[n])\rangle^2/4}{\partial t_p} = \frac{1}{2}\langle C_4(y_1[n])\rangle\frac{\partial\langle C_4(y_1[n])\rangle}{\partial t_p} = 0 \tag{2.52}$$

which results in simply vanishing $\partial\langle C_4(y_1[n])\rangle/\partial t_p$.

Now using the multi-linearity property of cumulants, it can be shown that $\langle C_4(y_1[n])\rangle$ can be written as follows:

$$\begin{aligned}
&\langle C_4(y_1[n])\rangle = d_{40}(g_p^{(1+)}(t_p))^4 + d_{31}(g_p^{(1+)}(t_p))^3(g_{P'}^{(1+)}(t_p)) + \\
&d_{22}(g_p^{(1+)}(t_p))^2(g_{P'}^{(1+)}(t_p))^2 + d_{13}(g_p^{(1+)}(t_p))(g_{P'}^{(1+)}(t_p))^3 + \\
&d_{04}(g_{P'}^{(1+)}(t_p))^4 + d_{30}(g_p^{(1+)}(t_p))^3 + d_{21}(g_p^{(1+)}(t_p))^2(g_{P'}^{(1+)}(t_p)) \\
&+ d_{12}(g_p^{(1+)}(t_p))(g_{P'}^{(1+)}(t_p))^2 + d_{03}(g_{P'}^{(1+)}(t_p))^3 + d_{20}(g_p^{(1+)}(t_p))^2 \\
&+ d_{11}(g_p^{(1+)}(t_p))(g_{P'}^{(1+)}(t_p)) + d_{02}(g_{P'}^{(1+)})^2 + d_{10}(g_p^{(1+)}) \\
&+ d_{01}(g_{P'}^{(1+)}) + d_{00}
\end{aligned} \tag{2.53}$$

where the coefficients d_{ij} are given in appendix C.

Consequently, by computing the derivative of (2.53), we obtain:

$$\sum_{m=0}^{4} e_m t_p^m + \sqrt{1+t_p^2}\Big(\sum_{n=0}^{3} f_n t_p^n\Big) = 0 \tag{2.54}$$

where the coefficients e_m and f_n are given in appendix C.

Equation (2.54) can be simplified to an 8-th degree polynomial equation as follows:

$$\sum_{m=0}^{8}\sum_{n=0}^{m} e_n e_{m-n} t_p^m - (1+t_p^2)\sum_{m'=0}^{6}\sum_{n'=0}^{m'} f_{n'} f_{m'-n'} t_p^{m'} = 0 \tag{2.55}$$

By rooting (2.55), 8 solutions $\hat{t}_p$ are obtained. Then we calculate the contrast function (2.47) for all real-valued roots and we choose the root $\hat{t}_p^{(opt)}$ which maximizes it. Eventually, we calculate the matrices $\boldsymbol{G}_p^{(1)}(\hat{t}_p^{(opt)})$ and $\boldsymbol{G}^{(1+)}(\hat{t}_p^{(opt)})$. This procedure is performed iteratively for all $p \in \{1, ..., P'-1\}$ and for several sweeps until convergence. At this stage, the first column $\hat{\boldsymbol{h}}_1$ of the estimated mixing matrix is given by the last update of $\boldsymbol{g}^{(1+)}(\hat{t}_{P'-1}^{(opt)})$ and the realization of the first source is estimated by $\hat{r}_1[n] = \hat{\boldsymbol{h}}_1^{\mathsf{T}}\boldsymbol{x}[n]$.

After identifying the first source, we remove its contribution from the observations by projecting the observations onto the subspace orthogonal to that spanned by $\hat{\boldsymbol{h}}_1$ by computing $\boldsymbol{x}^{(1)[n]} = \Pi^{(1)}\boldsymbol{x}[n]$ where $\Pi^{(1)}$ is a $(P'-1 \times P')$ projection matrix built by stacking vertically the $(P'-1)$ first rows of the last update of $\boldsymbol{G}^{(1+)}(\hat{t}_p^{(opt)})^{\mathsf{T}}$. Now to estimate the other sources, the same procedure should be done by using equations (2.49) to (2.55). The only difference is that the vector of observations $\boldsymbol{x}[n]$ should be replaced by the observation $\boldsymbol{x}^{(p-1)}[n]$ of reduced dimension $(P'-p+1)$ in order to extract the p-th source.

Note that the estimation of FO cumulants is not required at each iteration of our Jacobi-like procedure. The M^4 FO averaged cumulants $C_{m_1,m_2,m_3,m_4,\boldsymbol{x}}$ of $\{\tilde{\boldsymbol{x}}[n]\}$ can be estimated at the beginning of the procedure and sorted in a $(M^2 \times M^2)$ matrix, $\boldsymbol{Q}_{\boldsymbol{x}}$, called quadricovariance [Albera et al., 2012]. The FO averaged cumulants

$C_{m_1,m_2,m_3,m_4,\boldsymbol{x}^{(p)}}$ of $\{\tilde{\boldsymbol{x}}^{(p)}[n]\}$ can then be derived using the following formula $\boldsymbol{Q}_{\boldsymbol{x}^{(p)}} = \triangle \boldsymbol{Q}_{\boldsymbol{x}} \triangle^{\mathsf{T}}$ where:

$$\triangle = \prod_{i=1}^{p-1} (\Pi^{(i)} \otimes \Pi^{(i)}) \tag{2.56}$$

2.2.2 Denoising epileptic EEG signals using the JDICA algorithm

The JDICA algorithm can be used to denoise both the interictal and ictal data. The procedure is almost similar and has three steps. First by using the JDICA algorithm, P' sources, including both sources of interest and non-interest, are extracted and the mixing matrix is estimated. Then, the sources of interest are identified and finally by using the estimated mixing matrix and the sources of interest, the denoised epileptic signal is obtained. The only problem that must be explained is how the sources of interest are identified. This procedure can be manual (visual) or automatic. In this thesis, we use the following procedure to select the sources of interest.

For interictal data, we use the spike detection algorithm (see section 2.1.3.1) to select the sources of interest. More particularly, by applying the two first steps of the spike detection algorithm (S_1 and S_2) on each extracted source, the spike candidates for each source are detected. Then by considering a window around each detected spike, the Energy Ratio (ER) of each source is calculated as follows:

$$ER_{r_p} = \frac{\frac{1}{n(\Gamma^{(p)}_{spike})} \sum_{n \in \Gamma^{(p)}} (r_p[n])^2}{\frac{1}{N - n(\Gamma^{(p)}_{spike})} \sum_{\notin \Gamma_{spike}} (r_p[n])^2} \tag{2.57}$$

where $\Gamma^{(p)}_{spike}$ is the set containing the time samples in the windows around all detected spikes for the extracted source $\{r_p[n]\}$ and $n(\Gamma^{(p)}_{spike})$ is the number of time samples in this set. The sources with the highest ERs are selected as interictal sources of interest.

For ictal data, as in section 2.1.4.1, sources of interest are extracted using the linear time-frequency representation of each source and then by choosing the sources of interest based on the existence of a narrow-band high frequency signature at the beginning of the seizure and of a decrease in the frequency in the following time samples.

2.3 The SSD-CP method

As mentioned in chapter 1, the CP decomposition consists in decomposing a HO array as a linear combination of a minimal number of rank-1 terms and can then be seen as a generalization of the SVD of two-way data to multi-way data. But the main difference is that, under weak assumptions [Kruskal, 1977, Jiang and Sidiropoulos, 2004, Sidiropoulos and Bro, 2000, Stegman and Sidiropoulos, 2007], no orthogonality constraint is needed to ensure uniqueness of the CP decomposition. This advantage makes the CP decomposition very useful in various applications. The CP decomposition was first proposed

to solve psychometric problems [Carroll and Chang, 1970]. Afterwards, it was extended to various fields such as chemometrics [Bro, 1997], spectroscopy [Villringer and Chance, 1997] and biomedical engineering [Acar et al., 2007, Vos et al., 2007, Becker et al., 2010, Becker et al., 2012].

Several algorithms were proposed to solve the CP problem. They can be classified in three main groups [Tomasi and Bro, 2006]: 1) alternating algorithms, which update only a subset of the parameters at each step; 2) derivative-based methods, seeking for an update of all the parameters simultaneously by successive approximations; and 3) direct procedures. For instance, the Alternating Least Squares (ALS) based techniques [Harshman and Lundy, 1994, Rajih et al., 2008, Nion and Lathauwer, 2008, Karfoul et al., 2011, Chen et al., 2011] belong to the first group. The ALS approach [Harshman and Lundy, 1994] is the most popular due its simplicity. Its main drawback is its slow convergence caused for instance by the presence of bottlenecks [Comon et al., 2009], say collinear factors. Recently, Enhanced Line Search (ELS) schemes [Rajih et al., 2008, Nion and Lathauwer, 2008, Karfoul et al., 2011, Chen et al., 2011] were proposed to cope with this drawback. But there are some simple cases such as decomposing the fully parameterizable ($2\times2\times2$) arrays given by Paatero [Paatero, 2000], where any alternating algorithm, including the ALS algorithm with ELS scheme (ELS-ALS), does not succeed in achieving the good decomposition [Comon et al., 2009]. The second main group of CP methods set is represented by the Levenberg-Marquardt (LM) [Comon et al., 2009, Tomasi and Bro, 2006] and nonlinear conjugate gradient approaches [Acar et al., 2011]. As the alternating algorithms, the derivative-based algorithms suffer from converging to local minima. To compensate this problem, different initial points must be used, which results to a time-consuming procedure. In order to overcome this drawback, a fast implementation of the LM algorithm, named Fast LM (FLM), has been recently proposed [Phan et al., 2013c, Tichavsky et al., 2013]. As far as the third group of CP algorithms is concerned, the most known implementations are the Generalized Rank Annihilation Method (GRAM) [Sanchez and Kowalski, 1986] and the Direct Tri-Linear Decomposition (DTLD) algorithm [Sanchez and Kowalski, 1990], both based on a generalized eigenvalue problem. In other words, the latter methods compute the CP decomposition by solving an alternative algebra problem of lower dimension, but they do not provide a solution in terms of least squares, contrarily to the alternating and derivative-based techniques. It is noteworthy that novel direct algorithms have appeared during the last decade. They reformulate the CP problem as a Joint EigenValue Decomposition (JEVD) [Roemer and Haardt, 2008, Luciani and Albera, 2010, Luciani and Albera, 2011, Roemer and Haardt, 2013, Luciani and Albera, 2014], Joint Diagonalization by Congruence (JDC) [De Lathauwer, 2006] or Simultaneous Generalized Schur Decomposition (SGSD) [De Lathauwer et al., 2004, Sørensen and De Lathauwer, 2010] problem. The latter reformulated problems are usually solved by means of a Jacobi-like procedure.

In this thesis, we propose a new direct algorithm to compute the CP decomposition of complex-valued multi-way arrays. The proposed algorithm is based on the Simultaneous Schur Decomposition (SSD) of particular matrices derived from the array to process. We also propose a new Jacobi-like algorithm to calculate the SSD of several complex-valued matrices. In the following sections, first the new Jacobi-like algorithm

for SSD of complex matrices is proposed. Then the proposed CP algorithm for decomposing a HO complex-valued array is introduced. After that, the identification constraints for the existence and uniqueness of the solution of the proposed algorithm are analyzed. Finally, we explain how to use the proposed SSD-CP algorithm in the context of denoising of epileptic EEG signals.

2.3.1 Complex Simultaneous Schur Decomposition: the SSD technique

In this section, we introduce a new Jacobi-like algorithm in order to calculate the SSD of several complex-valued matrices. It is a nontrivial extension of a method proposed in the nineties for real-valued matrices only [Haardt and Nossek, 1998]. The SSD problem can be expressed as the following simultaneous triangularization problem. Given R matrices $\boldsymbol{Y}^{(r)} \in \mathbb{C}^{d\times d}$ which have the following structure:

$$\boldsymbol{Y}^{(r)} = \boldsymbol{Q}^{(o)} \boldsymbol{X}^{(r)} \boldsymbol{Q}^{(o)^{\mathsf{H}}} \tag{2.58}$$

where $\boldsymbol{Q}^{(o)} \in \mathbb{C}^{d\times d}$ is a unitary matrix and where the R matrices $\boldsymbol{X}^{(r)} \in \mathbb{C}^{d\times d}$ are upper triangular matrices, our objective is the simultaneous triangularization of the R matrices $\boldsymbol{Y}^{(r)}$ by finding the appropriate unitary matrix $\boldsymbol{Q}^{(o)}$. We use the upper index (o) to denote the optimum value of the corresponding variable. In the presence of noise, the matrices $\boldsymbol{Y}^{(r)}$ do not exactly fit (2.58), so $\boldsymbol{Q}^{(o)}$ should be computed as an approximate solution to the problem of simultaneous upper triangularization. Indeed, we should find matrix $\boldsymbol{Q}^{(o)}$ which minimizes the following cost function with respect to the unitary matrix $\boldsymbol{Q}$:

$$\psi(\boldsymbol{Q}) = \sum_{r=1}^{R} \|\mathcal{L}(\boldsymbol{Q}^{\mathsf{H}} \boldsymbol{Y}^{(r)} \boldsymbol{Q})\|_F^2 \tag{2.59}$$

where $\mathcal{L}(\boldsymbol{Y}^{(r)})$ extracts the strictly lower triangular part of its matrix-valued argument by replacing the upper triangular part with zeros and where $\|.\|_F$ denotes the Frobenius norm.

Minimizing the function ψ in several variables is a hard optimization problem, except for small dimensions. The idea is to reformulate this multivariate optimization problem in a finite sequence of monovariate optimization problems using a Jacobi-like optimization scheme. Since any unitary matrix can be written as a product of elementary Givens rotation matrices, we can parameterize the unitary matrix $\boldsymbol{Q}$ as follows:

$$\boldsymbol{Q} = \prod_{sw=1}^{\#ofsweeps} \prod_{q=2}^{d} \prod_{p=1}^{q-1} \boldsymbol{Q}^{(p,q)} \tag{2.60}$$

where each elementary Givens rotation matrix $\boldsymbol{Q}^{(p,q)}$ is obtained from an identity matrix in which we have $Q_{p,p}^{(p,q)} = c$, $Q_{p,q}^{(p,q)} = s$, $Q_{q,p}^{(p,q)} = -s^*$ and $Q_{q,q}^{(p,q)} = c$, where $(c, s) \in \mathbb{R}\times\mathbb{C}$ such that $c^2+|s|^2 = 1$. An appropriate parametrization of c and s in the complex domain used in the proposed algorithm is:

$$s = \frac{t}{\sqrt{1+|t|^2}}, \quad c = \frac{1}{\sqrt{1+|t|^2}}, \quad t \in \mathbb{C} \tag{2.61}$$

In each iteration, a new elementary Givens rotation matrix $\boldsymbol{Q}^{(o,p,q)}$ which defines $\boldsymbol{Q}^{(o)}$ (2.60) is computed in order to minimize the cost function (2.59). So, for each couple (p,q), the R complex-valued matrices $\boldsymbol{Y}^{(r)}$ are transformed as follows:

$$\boldsymbol{Y}^{(r)\prime} = \boldsymbol{Q}^{(o,p,q)^{\mathsf{H}}} \boldsymbol{Y}^{(r)} \boldsymbol{Q}^{(o,p,q)} \tag{2.62}$$

Since this orthogonal similarity transformation changes only the components of $\boldsymbol{Y}^{(r)}$ which appear in rows and columns p and q, the components of $\boldsymbol{Y}^{(r)}$ on its strictly lower triangular part change as follows:

$$\begin{aligned}
&\forall k,\ 1 \le k < p, \quad Y_{p,k}^{(r)\prime} = (1+|t|^2)^{-\frac{1}{2}}(Y_{p,k}^{(r)} - tY_{q,k}^{(r)}) \\
&\forall k,\ 1 \le k < q, k \ne p, \quad Y_{q,k}^{(r)\prime} = (1+|t|^2)^{-\frac{1}{2}}(t^* Y_{p,k}^{(r)} + Y_{q,k}^{(r)}) \\
&Y_{q,p}^{(r)\prime} = (1+|t|^2)^{-1}(t^* Y_{p,p}^{(r)} - t^{*2} Y_{p,q}^{(r)} + Y_{q,p}^{(r)} - t^* Y_{q,q}^{(r)}) \\
&\forall k,\ p < k \le d, k \ne q, \quad Y_{k,p}^{(r)\prime} = (1+|t|^2)^{-\frac{1}{2}}(Y_{k,p}^{(r)} - t^* Y_{k,q}^{(r)}) \\
&\forall k,\ q < k \le d, \quad Y_{k,q}^{(r)\prime} = (1+|t|^2)^{-\frac{1}{2}}(tY_{k,p}^{(r)} + Y_{k,q}^{(r)})
\end{aligned} \tag{2.63}$$

and the other components do not change. So by applying the transformation (2.62), the change in the cost function can be expressed as follows:

$$\begin{aligned}
\Delta\psi(\boldsymbol{Q}^{(p,q)}) &= \sum_{r=1}^{R} \left(\|\mathcal{L}(\boldsymbol{Y}^{(r)\prime})\|_F^2 - \|\mathcal{L}(\boldsymbol{Y}^{(r)})\|_F^2 \right) \qquad (2.64) \\
&= \sum_{r=1}^{R} \left(|Y_{q,p}^{(r)\prime}|^2 + \sum_{k=p+1}^{q-1} \left(|Y_{k,p}^{(r)\prime}|^2 + |Y_{q,k}^{(r)\prime}|^2 \right) - |Y_{q,p}^{(r)}|^2 - \sum_{k=p+1}^{q-1} \left(|Y_{k,p}^{(r)}|^2 + |Y_{q,k}^{(r)}|^2 \right) \right)
\end{aligned}$$

where the equality $c^2 + |s|^2 = 1$ is used to simplify the equation. The equation (2.64) can also be presented as a function of t as follows:

$$\begin{aligned}
\Delta\psi(\boldsymbol{Q}^{(p,q)}) &= (1+|t|^2)^{-2} \left(v_8 + 2\mathrm{Re}\{tv_3^* - tv_9 - t^2 v_7^*\} \right. \\
&\left. + |t|^2(v_1 + v_{10}) + 2|t|^2 \mathrm{Re}\{t^* v_5 - t^* v_2^* - v_4\} + |t|^4 v_6 \right) \\
&+ (1+|t|^2)^{-1} \left(u_1 - 2\mathrm{Re}\{tu_2\} + |t|^2 u_3 + w_1 + 2\mathrm{Re}\{tw_2\} \right. \\
&\left. + |t|^2 w_3 \right) - \sum_{r=1}^{R} \left(|Y_{q,p}^{(r)}|^2 + \sum_{k=p+1}^{q-1} \left(|Y_{k,p}^{(r)}|^2 + |Y_{q,k}^{(r)}|^2 \right) \right) \qquad (2.65)
\end{aligned}$$

where the coefficients v_1-v_{10}, u_1-u_3 and w_1-w_3 are given in appendix D.

To calculate an appropriate complex value for t in each iteration, we use the parametrization $t = \rho e^{i\phi}$ and then differentiate (2.65) with respect to ρ and ϕ separately. Note that the roots of the following equation system are the stationary points of $\Delta\psi(\boldsymbol{Q}_{q,p})$.

$$\begin{cases} \frac{\partial \Delta\psi(\rho,\phi)}{\partial \rho} = 0 \\ \frac{\partial \Delta\psi(\rho,\phi)}{\partial \phi} = 0 \end{cases} \tag{2.66}$$

which results in:

$$\begin{aligned}\frac{\partial\Delta\psi}{\partial\rho}=&\big(\mathrm{i}\rho^4(e^{\mathrm{i}\phi}b_0-e^{-\mathrm{i}\phi}b_0^*)+\rho^3(b_3+\mathrm{i}(e^{\mathrm{i}2\phi}b_1+e^{-\mathrm{i}2\phi}b_1^*))\\&+\rho^2(-\mathrm{i}3e^{\mathrm{i}\phi}(b_0-b_2)+\mathrm{i}3e^{-\mathrm{i}\phi}(b_0^*-b_2^*))\\&+\rho(b_4-\mathrm{i}e^{\mathrm{i}2\phi}b_1+\mathrm{i}e^{-\mathrm{i}2\phi}b_1^*)\\&-\mathrm{i}e^{\mathrm{i}\phi}b_2+\mathrm{i}e^{-\mathrm{i}\phi}b_2^*\big)/(1+\rho^2)^3=0\end{aligned}\tag{2.67}$$

and:

$$\begin{aligned}\frac{\partial\Delta\psi}{\partial\phi}&=\big(\rho^3(e^{\mathrm{i}\phi}b_0+e^{-\mathrm{i}\phi}b_0^*)+\rho^2(e^{\mathrm{i}2\phi}b_1+e^{-\mathrm{i}2\phi}b_1^*)\\&\quad+\rho(e^{\mathrm{i}\phi}b_2+e^{-\mathrm{i}\phi}b_2^*)\big)/(1+\rho^2)^2=0\end{aligned}\tag{2.68}$$

where the coefficients b_0-b_4 are given in appendix D.

The equation (2.68) can then be converted into a second degree polynomial equation with regards to ρ:

$$p_2\rho^2+p_1\rho+p_0=0\tag{2.69}$$

with:

$$\begin{aligned}p_2&=e^{\mathrm{i}\phi}b_0+e^{-\mathrm{i}\phi}b_0^*=2\mathrm{Re}\{e^{\mathrm{i}\phi}b_0\}\\p_1&=e^{\mathrm{i}2\phi}b_1+e^{-\mathrm{i}2\phi}b_1^*=2\mathrm{Re}\{e^{\mathrm{i}2\phi}b_1\}\\p_0&=e^{\mathrm{i}\phi}b_2+e^{-\mathrm{i}\phi}b_2^*=2\mathrm{Re}\{e^{\mathrm{i}\phi}b_2\}\end{aligned}$$

So, we can compute the roots of (2.69) as follows:

$$\rho=\frac{-p_1\pm\sqrt{p_1^2-4p_2p_0}}{2p_2}\tag{2.70}$$

and insert them into (2.67). By replacing $e^{\mathrm{i}2\phi}$ with x and simplifying the equation, we can obtain a single 14-th degree polynomial P_A in variable x whose coefficients A_ℓ are given in appendix D.

By rooting this polynomial numerically, fourteen complex values $\hat{x}_k$ are extracted. Now, we should choose the optimum solution $\hat{x}^{(o)}$ which minimizes the cost function (2.65) with respect to all computed values $\hat{x}_k$. Since the equation (2.65) is a function of t, we have first to calculate the value t_k corresponding to each $\hat{x}_k$ in order to choose the best solution. To this end, we do the following computations. The values $\hat{\phi}_k$ associated with the fourteen values $\hat{x}_k$ can be computed as follows:

$$\hat{\phi}_k=\frac{\mathrm{angle}\{\hat{x}_k\}}{2}\tag{2.71}$$

Afterwards for each $\hat{\phi}_k$, two corresponding values $\hat{\rho}_k^{(+)}$ and $\hat{\rho}_k^{(-)}$ are calculated by using (2.70):

$$\hat{\rho}_k^{(+/-)}=\frac{-\mathrm{Re}\{b_1e^{\mathrm{i}2\hat{\phi}_k}\}\pm\sqrt{\left(\mathrm{Re}\{b_1e^{\mathrm{i}2\hat{\phi}_k}\}\right)^2-4\mathrm{Re}\{b_0e^{\mathrm{i}\hat{\phi}_k}\}\mathrm{Re}\{b_2e^{\mathrm{i}\hat{\phi}_k}\}}}{2\mathrm{Re}\{b_0e^{\mathrm{i}\hat{\phi}_k}\}}\tag{2.72}$$

Algorithm 2.1: Simultaneous triangularization of R complex-valued matrices

Inputs: R matrices $\boldsymbol{Y}^{(r)} \in \mathbb{C}^{d\times d}$.

Outputs: R upper triangular matrices $\boldsymbol{X}^{(r)} \in \mathbb{C}^{d\times d}$ and the unitary matrix $\boldsymbol{Q}^{(o)} \in \mathbb{C}^{d\times d}$ which satisfy (2.58).

```
Repeat
  For q = 2 : d do
    For p = 1 : q − 1 do
      Compute the fourteen roots x̂_k of P_A.
      For each root x̂_k, calculate φ̂_k = angle(x̂_k)/2.
      Calculate both corresponding values ρ̂_k^(+) and ρ̂_k^(−) by using (2.72).
      Compute t̂_k^(+/−) = ρ̂_k^(+/−) e^(iφ̂_k).
      Calculate the cost function (2.65) for the twenty-eight values t̂_k^(+/−) and find
      among them the best solution t̂^(o) that minimizes the negative-valued fitness.
      From the selected t̂^(o), generate matrix Q^(o,p,q) (2.61) and update matrices
      Y^(r) by using (2.62).
    End For
  End For
Until the algorithm converges
```

So, for each $\hat{x}_k$, two couples $(\hat{\phi}_k, \hat{\rho}_k^{(+)})$ and $(\hat{\phi}_k, \hat{\rho}_k^{(-)})$ are obtained. Then, for each couple $(\hat{\phi}_k\ , \hat{\rho}_k^{(+/-)})$, $\hat{t}_k^{(+/-)}$ is calculated as follows:

$$\hat{t}_k^{(+/-)} = \hat{\rho}_k^{(+/-)} e^{\mathrm{i}\hat{\phi}_k} \tag{2.73}$$

To select the optimum solution $\hat{t}^{(o)}$, the cost function (2.65) is computed for the twenty-eight values of $\hat{t}_k^{(+/-)}$ and the minimizer of (2.65) that makes it negative (to ensure that the cost function (2.59) is reduced in each step) is chosen as follows:

$$\hat{t}^{(o)} = \arg\min_{\hat{t}_k^{(+/-)}} \{\Delta\psi(\boldsymbol{Q}^{(p,q)})\} \tag{2.74}$$

By using the selected $\hat{t}^{(o)}$, $\boldsymbol{Q}^{(o,p,q)}$ is computed and the matrices $\boldsymbol{Y}^{(r)}$ are updated as described by equation 2.62. The summarized procedure to simultaneously triangularize complex-valued matrices is shown in algorithm (2.1).

2.3.2 Direct CP decomposition: the SSD-CP method

The CP decomposition of a q-th order complex-valued array $\boldsymbol{\mathcal{T}}$ of size $(N_1 \times N_2 \times \ldots \times N_q)$ is defined by the following minimal linear combination:

$$\boldsymbol{\mathcal{T}} = \sum_{p=1}^{P} \boldsymbol{u}_p^{(1)} \circ \boldsymbol{u}_p^{(2)} \circ \cdots \circ \boldsymbol{u}_p^{(q)} \tag{2.75}$$

where for each integer $i \in \{1, ..., q\}$ the P complex-valued vectors $\boldsymbol{u}_p^{(i)}$ are the column vectors of a $(N_i \times P)$ matrix $\boldsymbol{U}^{(i)}$ called i-th loading matrix. The aim of the CP problem

is to determine the q loading matrices $\boldsymbol{U}^{(i)}$ of $\boldsymbol{\mathcal{T}}$ up to a diagonal scaling matrix and a permutation which are the classical indeterminacies of the CP decomposition. It is noteworthy that the permutation indeterminacy is the same for the P loading matrices. Note also that the minimal number P of rank-1 terms allowing us to decompose $\boldsymbol{\mathcal{T}}$ (2.75) defines the rank of $\boldsymbol{\mathcal{T}}$.

Generally, people prefer to handle unfolding matrices of multi-way arrays. All possible matricizations $\boldsymbol{T}(\pi, L)$ of a q-way array $\boldsymbol{\mathcal{T}}$ can be defined by using two parameters, namely a permutation function π of $\{1, \ldots, q\}$ and an integer $L \in \{1, \ldots, q\}$. More precisely, the (k, ℓ)-th component of $\boldsymbol{T}(\pi, L)$ is given by:

$$T(\pi, L)_{k,\ell} = \mathcal{T}_{n_1, n_2, \ldots, n_q} \tag{2.76}$$

where:

$$k = n_{\pi(1)} + \sum_{j=2}^{L} (n_{\pi(j)} - 1) N^{(\pi(1), \pi(2), \ldots, \pi(j-1))}$$

and:

$$\ell = n_{\pi(L+1)} + \sum_{j=L+2}^{q} (n_{\pi(j)} - 1) N^{(\pi(L+1), \pi(L+2), \ldots, \pi(j-1))}$$

with $N^{(a_1, \ldots, a_K)} = N_{a_1} N_{a_2} \ldots N_{a_K}$. In the proposed algorithm, we use two different unfolding matrices, namely $\boldsymbol{T}^{(1)} = \boldsymbol{T}(\pi, L_1)$ and $\boldsymbol{T}^{(2)} = \boldsymbol{T}(\pi, L_2)$, where the permutation function π, the constants L_1 and L_2 ($L_1 < L_2$) are defined in such a way that the following assumptions hold:

1. The rank P of the array $\boldsymbol{\mathcal{T}}$ is lower than or equal to $N^{(\pi(1), \ldots, \pi(L_1))}$.
2. The entries of at least one row of the matrix $\boldsymbol{U}^{(\pi(q):\pi(L_2+1))}$ are non-zero.
3. The matrix $\boldsymbol{U}^{(\pi(L_2):\pi(L_1+1))}$ is full column rank.
4. The matrix $\boldsymbol{U}^{(\pi(q):\pi(L_2+1))}$ is full column rank.
5. The matrix $\boldsymbol{U}^{(\pi(L_1):\pi(1))}$ is full column rank.

where the $(N^{(n_1, \ldots, n_K)} \times P)$ matrix $\boldsymbol{U}^{(n_1:n_K)}$ is given by the following Khatri-Rao product of K matrices:

$$\boldsymbol{U}^{(n_1:n_K)} = \boldsymbol{U}^{(n_1)} \odot \boldsymbol{U}^{(n_2)} \odot \ldots \odot \boldsymbol{U}^{(n_K)} \tag{2.77}$$

These assumptions ensuring identifiability of the proposed CP algorithm will be discussed in section 2.3.3. For the sake of simplicity, we first explain the proposed method for a three-way complex-valued array $\boldsymbol{\mathcal{T}}$ of size $(N_1 \times N_2 \times N_3)$ defined by:

$$\boldsymbol{\mathcal{T}} = \sum_{p=1}^{P} \boldsymbol{u}_p^{(1)} \circ \boldsymbol{u}_p^{(2)} \circ \boldsymbol{u}_p^{(3)} \tag{2.78}$$

Next we will generalize our method to higher-order arrays. We also assume in the following of this section that π is the identity permutation. It can be easily shown that the unfolding matrix $\boldsymbol{T}^{(1)}$ of size $(N_1 \times N_2 N_3)$ (L_1 is equal to 1) can be expressed as follows:

$$\boldsymbol{T}^{(1)} = \boldsymbol{U}^{(1)} \left(\boldsymbol{U}^{(3)} \odot \boldsymbol{U}^{(2)}\right)^{\mathsf{T}} = \boldsymbol{U}^{(1)} \boldsymbol{U}^{(3:2)^{\mathsf{T}}} \tag{2.79}$$

Note that the $\boldsymbol{U}^{(3:2)}$ matrix can also be written as follows:

$$\boldsymbol{U}^{(3:2)} = \left[\boldsymbol{\Phi}^{(1)}\boldsymbol{U}^{(2)^{\mathsf{T}}}, \ldots, \boldsymbol{\Phi}^{(N_3)}\boldsymbol{U}^{(2)^{\mathsf{T}}}\right]^{\mathsf{T}} \tag{2.80}$$

where $\boldsymbol{\Phi}^{(i)}$ is a $(P \times P)$ diagonal matrix built from the i-th row of matrix $\boldsymbol{U}^{(3)}$. On the other hand, under assumptions **A1**), **A3**) and **A4**) we can compute the SVD truncated at order P to decompose $\boldsymbol{T}^{(1)}$ as follows:

$$\boldsymbol{T}^{(1)} = \boldsymbol{Z}\,\boldsymbol{\Sigma}\,\boldsymbol{V}^{\mathsf{H}} \tag{2.81}$$

where $\boldsymbol{Z}$, $\boldsymbol{\Sigma}$ and $\boldsymbol{V}$ are of size $(N_1 \times P)$, $(P \times P)$ and $(N_2 N_3 \times P)$, respectively. Under assumptions **A3**), **A4**) and **A5**) (note that here L_2 is necessarily equal to 2), by comparing (2.79) and (2.81), we can conclude that there exists a non-singular matrix $\boldsymbol{W}$ of size $(P \times P)$ such that:

$$\boldsymbol{V}^{\mathsf{H}} = \boldsymbol{W}\,\boldsymbol{U}^{(3:2)^{\mathsf{T}}} \tag{2.82}$$

Now by inserting (2.80) in (2.82), we can rewrite $\boldsymbol{V}$ as follows:

$$\boldsymbol{V} = \left[\boldsymbol{\Gamma}^{(1)}, \ldots, \boldsymbol{\Gamma}^{(N_3)}\right]^{\mathsf{H}} \tag{2.83}$$

where the N_3 matrices $\boldsymbol{\Gamma}^{(i)}$ of size $(P \times N_2)$ are given by $\boldsymbol{\Gamma}^{(i)} = \boldsymbol{W}\,\boldsymbol{\Phi}^{(i)}\,\boldsymbol{U}^{(2)^{\mathsf{T}}}$. Under assumptions **A2**) and **A3**), there is at least one matrix $\boldsymbol{\Gamma}^{(j)}$ admitting the Moore-Penrose matrix inverse $\Gamma^{(j)\sharp}$. Therefore, we can define R matrices $\boldsymbol{Y}^{(i,j)}$ as follows:

$$\begin{aligned} \boldsymbol{Y}^{(i,j)} &= \boldsymbol{\Gamma}^{(i)}\boldsymbol{\Gamma}^{(j)\sharp} \\ &= \boldsymbol{W}\boldsymbol{\Phi}^{(i)}\boldsymbol{U}^{(2)^{\mathsf{T}}}(\boldsymbol{U}^{(2)^{\mathsf{T}}})^{\sharp}(\boldsymbol{\Phi}^{(j)})^{-1}\boldsymbol{W}^{-1} \\ &= \boldsymbol{W}\boldsymbol{D}^{(i,j)}\boldsymbol{W}^{-1} \end{aligned} \tag{2.84}$$

where $N_3 - 1 \leq R \leq N_3(N_3 - 1)$ and for each couple (i, j), $\boldsymbol{D}^{(i,j)} = \boldsymbol{\Phi}^{(i)}(\boldsymbol{\Phi}^{(j)})^{-1}$ is a $(P \times P)$ diagonal matrix. Let us write $\boldsymbol{W} = \boldsymbol{Q}\boldsymbol{R}$, namely the QR decomposition of matrix $\boldsymbol{W}$ where $\boldsymbol{Q}\boldsymbol{Q}^{\mathsf{H}} = \boldsymbol{I}$ and $\boldsymbol{R}$ is an upper triangular matrix. Thus, we can rewrite $\boldsymbol{Y}^{(i,j)}$ as follows:

$$\begin{aligned} \boldsymbol{Y}^{(i,j)} &= \boldsymbol{W}\boldsymbol{D}^{(i,j)}\boldsymbol{W}^{-1} \\ &= \boldsymbol{Q}\boldsymbol{R}\boldsymbol{D}^{(i,j)}\boldsymbol{R}^{-1}\boldsymbol{Q}^{-1} \\ &= \boldsymbol{Q}\boldsymbol{X}^{(i,j)}\boldsymbol{Q}^{\mathsf{H}} \end{aligned} \tag{2.85}$$

Since $\boldsymbol{R}$ is an upper triangular matrix, $\boldsymbol{R}^{-1}$ and $\boldsymbol{X}^{(i,j)} = \boldsymbol{R}\boldsymbol{D}^{(i,j)}\boldsymbol{R}^{-1}$ are upper triangular matrices too. The elements on the diagonal of $\boldsymbol{X}^{(i,j)}$ are equal to the diagonal

components of $\boldsymbol{D}^{(i,j)}$. By applying the SSD algorithm proposed in section 2.3.1 to matrices $\boldsymbol{Y}^{(i,j)}$, the upper triangularized matrices $\boldsymbol{X}^{(i,j)}$ and consequently the diagonal matrices $\boldsymbol{D}^{(i,j)}$ can be extracted. Indeed, we have $D_{p,p}^{(i,j)} = X_{p,p}^{(i,j)}$ for any triplet (i,j,p).

In order to estimate the components of matrix $\boldsymbol{U}^{(3)}$, for each column of this matrix, we have R equations. If J is a set indexing the rows of matrix $\boldsymbol{U}^{(3)}$ whose elements are non-zero or equivalently the matrices $\Gamma^{(j)}$ which admit a Moore-Penrose matrix inverse, we have:

$$\forall j \in J, \forall i \in \{1,\ldots,N_3\}/\{j\}, \forall p \in \{1,\ldots,P\}, \quad \frac{U_{i,p}^{(3)}}{U_{j,p}^{(3)}} = D_{p,p}^{(i,j)} \tag{2.86}$$

Hence:

$$U_{i,p}^{(3)} - D_{p,p}^{(i,j)} U_{j,p}^{(3)} = 0 \tag{2.87}$$

These equations can be written in a matrix form:

$$\boldsymbol{M}^{(p)} \boldsymbol{u}_p^{(3)} = \boldsymbol{0}_{R\times 1} \tag{2.88}$$

where $\boldsymbol{u}_p^{(3)}$ is the p-th column of matrix $\boldsymbol{U}^{(3)}$, defined by $\boldsymbol{u}_p^{(3)} = [U_{1,p}^{(3)}, U_{2,p}^{(3)}, \ldots, U_{N_3,p}^{(3)}]^{\mathsf{T}}$. The matrix $\boldsymbol{M}^{(p)}$ of size $(R \times N_3)$ is defined as follows. For each ordered pair (i,j) with $j \in J$, $i \neq j$ and $1 \leq i,j \leq N_3$, we consider one row in matrix $\boldsymbol{M}^{(p)}$ such that the component in the i-th column is equal to 1, the component in the j-th column is equal to $-D_{p,p}^{(i,j)}$ and the other components are null. Note that the order of the rows is not important in the solution. For instance, assuming that all components of matrix $\boldsymbol{U}^{(3)}$ are non-zero leading to $R = N_3(N_3-1)$, we put the rows corresponding to $i=1$ in the first N_3-1 rows of matrix $\boldsymbol{M}^{(p)}$ and the rows corresponding to $i=N_3$ in its last N_3-1 rows as follows:

$$\boldsymbol{M}^{(p)} = \begin{array}{c} \left[\begin{array}{ccccc} 1 & -D_{p,p}^{(1,2)} & 0 & \ldots & 0 \\ \vdots & & & & \vdots \\ 1 & 0 & \ldots & 0 & -D_{p,p}^{(1,N_3)} \\ \vdots & & \ldots & & \vdots \\ -D_{p,p}^{(N_3,1)} & 0 & \ldots & 0 & 1 \\ \vdots & & & & \vdots \\ 0 & 0 & \ldots & -D_{p,p}^{(N_3,N_3-1)} & 1 \end{array}\right] \end{array} \begin{array}{l} \left.\begin{array}{c} \\ \\ \\ \end{array}\right\} i=1 \\ \vdots \\ \left.\begin{array}{c} \\ \\ \\ \end{array}\right\} i=N_3 \end{array} \tag{2.89}$$

Consequently, the loading factor $\boldsymbol{u}_p^{(3)}$ can be identified up to a scale factor by computing the right singular vector associated with the zero singular value of $\boldsymbol{M}^{(p)}$. By repeating this procedure P times, the P columns of $\boldsymbol{U}^{(3)}$ can be identified up to a right multiplicative diagonal matrix.

Next we derive $\boldsymbol{U}^{(2:1)}$ from matrix $\boldsymbol{U}^{(3)}$ by using the $(N_1N_2 \times N_3)$ unfolding matrix $\boldsymbol{T}^{(2)}$ (let us recall that L_2 is equal to 2). Indeed, it can be shown that $\boldsymbol{T}^{(2)}$ is equal to:

$$\boldsymbol{T}^{(2)} = \boldsymbol{U}^{(2:1)} \boldsymbol{U}^{(3)^{\mathsf{T}}} \tag{2.90}$$

where $\boldsymbol{U}^{(3)}$ was calculated in the previous stage. Therefore, under assumption **A4**), matrix $\boldsymbol{U}^{(2:1)}$ can be computed as follows:

$$\boldsymbol{U}^{(2:1)} = \boldsymbol{T}^{(2)}(\pi, L_2)\left(\boldsymbol{U}^{(3)^{\mathsf{T}}}\right)^{\sharp} \tag{2.91}$$

Since $\boldsymbol{U}^{(2:1)}$ is defined by the Khatri-Rao product of two loading matrices according to (2.77), we can use a procedure based on rank-1 SVDs to calculate $\boldsymbol{U}^{(1)}$ and $\boldsymbol{U}^{(2)}$. To this end, for each $p \in \{1, ..., P\}$, we build the rank-1 matrix $\boldsymbol{U}^{(2:1,p)}$ of size $(N_1 \times N_2)$ associated with the p-th column of $\boldsymbol{U}^{(2:1)}$ as follows:

$$\forall (i,j) \in \{1,\ldots,N_2\} \times \{1,\ldots,N_1\} \quad U_{i,j}^{(2:1,p)} = U_{(i+jN_2),p}^{(2:1)} \tag{2.92}$$

such that the (i,j)-th component, $U_{i,j}^{(2:1,p)}$, of $\boldsymbol{U}^{(2:1,p)}$ is equal to $U_{i,p}^{(2)} U_{j,p}^{(1)}$. Then the p-th column of the loading matrices $\boldsymbol{U}^{(1)}$ and $\boldsymbol{U}^{(2)}$ can be extracted by computing the left and right singular vectors associated with non-zero singular value of $\boldsymbol{U}^{(2:1,p)}$. Indeed, we have:

$$\boldsymbol{U}^{(2:1,p)} = \boldsymbol{z}\,\sigma\,\boldsymbol{v}^{\mathsf{H}} = \widetilde{\boldsymbol{z}}\,\boldsymbol{v}^{\mathsf{H}} \tag{2.93}$$

where $\widetilde{\boldsymbol{z}}$ and $\boldsymbol{v}$ are vectors of size $(N_2 \times 1)$ and $(N_1 \times 1)$, respectively. So, we have up to a scale factor:

$$\boldsymbol{u}_p^{(2)} = \widetilde{\boldsymbol{z}} \qquad \text{and} \quad \boldsymbol{u}_p^{(1)} = \boldsymbol{v}^* \tag{2.94}$$

By solving the SVD problem for all $p \in \{1,\ldots,P\}$, the two loading matrices $\boldsymbol{U}^{(1)}$ and $\boldsymbol{U}^{(2)}$ can be obtained up to a right multiplicative diagonal matrix.

We now explain in the following how to generalize the proposed SSD-CP method to the case of higher-order arrays. To calculate the loading matrices of a q-th order complex-valued array $\boldsymbol{\mathcal{T}}$ of size $(N_1 \times N_2 \times \ldots \times N_q)$, we do the following procedure. First, we define the unfolding matrix $\boldsymbol{T}^{(1)}$ of size $(N^{(1,\ldots,L_1)} \times N^{(L_1+1,\ldots,q)})$. Similarly to the third order case, we can express the matrix $\boldsymbol{T}^{(1)}$ as follows:

$$\boldsymbol{T}^{(1)} = \boldsymbol{U}^{(L_1:1)}\,\boldsymbol{U}^{(q:L_1+1)^{\mathsf{T}}} \tag{2.95}$$

where the matrix $\boldsymbol{U}^{(q:L_1+1)}$ given by (2.77) can also be written as follows:

$$\begin{aligned}\boldsymbol{U}^{(q:L_1+1)} &= \boldsymbol{U}^{(q:L_2+1)} \odot \boldsymbol{U}^{(L_2:L_1+1)} \\ &= \left[\boldsymbol{\Phi}^{(1)}\boldsymbol{U}^{(L_2:L_1+1)^{\mathsf{T}}}, \ldots, \boldsymbol{\Phi}^{(N^*)}\boldsymbol{U}^{(L_2:L_1+1)^{\mathsf{T}}}\right]^{\mathsf{T}}\end{aligned} \tag{2.96}$$

with $N^* = N^{(L_2+1,\ldots,q)}$ and where $\boldsymbol{\Phi}^{(i)}$ is a $(P \times P)$ diagonal matrix built from the i-th row of matrix $\boldsymbol{U}^{(q:L_2+1)}$. We can also compute the SVD of $\boldsymbol{T}^{(1)}$ truncated at order P, as described by (2.81), with $\boldsymbol{Z}$, $\boldsymbol{\Sigma}$ and $\boldsymbol{V}$ of size $(N^{(1,\ldots,L_1)} \times P)$, $(P \times P)$ and $(N^{(L_1+1,\ldots,q)} \times P)$, respectively. Then, we can conclude that there exists a non-singular matrix $\boldsymbol{W}$ of size $(P \times P)$ such that:

$$\boldsymbol{V}^{\mathsf{H}} = \boldsymbol{W}\,\boldsymbol{U}^{(q:L_1+1)^{\mathsf{T}}} \tag{2.97}$$

Thus by inserting (2.96) in (2.97), we can rewrite $\boldsymbol{V}$ as follows:

$$\boldsymbol{V} = \left[\boldsymbol{\Gamma}^{(1)}, \ldots, \boldsymbol{\Gamma}^{(N^*)}\right]^{\mathsf{H}} \tag{2.98}$$

where the N^* matrices $\boldsymbol{\Gamma}^{(i)}$ of size $(P \times N^{(L_1+1,\ldots,L_2)})$ are given by $\boldsymbol{\Gamma}^{(i)} = \boldsymbol{W}\,\boldsymbol{\Phi}^{(i)}\,\boldsymbol{U}^{(L_2:L_1+1)^\mathsf{T}}$. Similarly to the third order case, R matrices $\boldsymbol{Y}^{(i,j)}$ can be computed by using (2.84) where R is now a value between (N^*-1) and $N^*(N^*-1)$. Then, the same procedure as (2.85)-(2.89), i.e. applying the SSD algorithm and solving systems of linear equations, can be used to estimate the matrix $\boldsymbol{U}^{(q:L_2+1)}$.

After that, we define the unfolding matrix $\boldsymbol{T}^{(2)}$ of size $(N^{(1,\ldots,L_2)} \times N^{(L_2+1,\ldots,q)})$. It can be shown that $\boldsymbol{T}^{(2)}$ is equal to:

$$\boldsymbol{T}^{(2)} = \boldsymbol{U}^{(L_2:1)}\,\boldsymbol{U}^{(q:L_2+1)^\mathsf{T}} \tag{2.99}$$

where $\boldsymbol{U}^{(q:L_2+1)}$ was calculated in the previous stage. Therefore, matrix $\boldsymbol{U}^{(L_2:1)}$ can be computed as follows:

$$\boldsymbol{U}^{(L_2:1)} = \boldsymbol{T}^{(2)}(\pi, L_2)\left(\boldsymbol{U}^{(q:L_2+1)^\mathsf{T}}\right)^\sharp \tag{2.100}$$

Now, note that both matrices $\boldsymbol{U}^{(q:L_2+1)}$ and $\boldsymbol{U}^{(L_2:1)}$ are defined as the Khatri-Rao product of several loading matrices according to (2.77). So, we should apply the same procedure as (2.92)-(2.94) in order to calculate the corresponding loading matrices. Let's describe this procedure with the more general matrix $\boldsymbol{U}^{(n_1:n_K)}$ given by (2.77). According to the definition of the Khatri-Rao product, the p-th column of matrix $\boldsymbol{U}^{(n_1:n_K)}$ can be represented by the Kronecker product $\boldsymbol{u}_p^{(n_1)} \otimes \ldots \otimes \boldsymbol{u}_p^{(n_K)}$ where $\boldsymbol{u}_p^{(n)}$ denotes the p-th column vector of matrix $\boldsymbol{U}^{(n)}$. Then, let's build the $(N_{n_1} \times \ldots \times N_{n_K})$ rank-1 array $\boldsymbol{\mathcal{U}}^{(n_1:n_K,p)}$ associated with the p-th column of $\boldsymbol{U}^{(n_1:n_K)}$ as follows:

$$\forall (i_1, \ldots, i_K), \quad \mathcal{U}^{(n_1:n_K,p)}_{i_1,i_2,\ldots,i_K} = U^{(n_1:n_K)}_{i_1+\sum_{k=2}^{K}\left((i_k-1)N_{n_2}\ldots N_{n_k}\right),p} \tag{2.101}$$

such that the $(i_1, \ldots, i_K)$-th entry $\mathcal{U}^{(n_1:n_K,p)}_{i_1,i_2,\ldots,i_K}$ of $\boldsymbol{\mathcal{U}}^{(n_1:n_K,p)}$ is equal to $U^{(n_1)}_{i_1,p} U^{(n_2)}_{i_2,p} \ldots U^{(n_K)}_{i_K,p}$. Consequently, a simple rank-1 HOSVD [De Lathauwer et al., 2000] of the array $\boldsymbol{\mathcal{U}}^{(n_1:n_K,p)}$ allows us to extract the p-th column of the K loading matrices $\boldsymbol{U}^{(n)}$. More precisely, let us denote the mode-ℓ matricization of $\boldsymbol{\mathcal{U}}^{(n_1:n_K,p)}$ by $\boldsymbol{U}^{(n_1:n_K,p,\ell)}$ that arranges the mode-ℓ fibers to be the columns of the resulting matrix, such that:

$$\forall \ell \in \{1, \ldots, K\},\ U^{(n_1:n_K,p,\ell)}_{i_\ell,j} = \mathcal{U}^{(n_1:n_K,p)}_{i_1,i_2,\ldots,i_K} \tag{2.102}$$

with:

$$j = 1 + \sum_{k=1,k\neq\ell}^{K} (i_k - 1)J_k \text{ and } J_k = \prod_{m=1,m\neq\ell}^{k-1} N_{n_m} \tag{2.103}$$

Thus the p-th column $\boldsymbol{u}_p^{(n)}$ of the n-th loading matrix $\boldsymbol{U}^{(n)}$ is given by the left dominant singular vector of $\boldsymbol{U}^{(n_1:n_K,p,\ell)}$, i.e. the left singular vector associated with the non-zero singular value of $\boldsymbol{U}^{(n_1:n_K,p,\ell)}$. To extract all loading matrices, the aforementioned scheme should be executed for the N modes of $\boldsymbol{\mathcal{U}}^{(n_1:n_K,p)}$. So, by this way we can extract the q loading matrices $\boldsymbol{U}^{(i)}$ from $\boldsymbol{U}^{(q:L_2+1)}$ and $\boldsymbol{U}^{(L_2:1)}$ involving the computation of Pq SVDs.

The proposed algorithm of CP decomposition, named SSD-CP, is summarized in algorithm 2.2.

Algorithm 2.2: The SSD-CP technique

Inputs: $\boldsymbol{\mathcal{T}} \in \mathbb{C}^{N_1 \times \cdots \times N_q}$ and $P \in \mathbb{N}^*$.
Outputs: The q loading matrices $\boldsymbol{U}^{(i)}$ of $\boldsymbol{\mathcal{T}}$.

Choose the appropriate permutation Π and the couple (L_1, L_2) of integers.
Construct the unfolding matrices $\boldsymbol{T}^{(1)}$ and $\boldsymbol{T}^{(2)}$.
Compute the SVD of matrix $\boldsymbol{T}^{(1)}$ truncated at order P: $\boldsymbol{T}^{(1)} = \boldsymbol{Z}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathsf{H}}$
For $i = 1 : N^{(L_2+1,\ldots,q)}$ **do**
  Stack the i-th $(P \times N^{(L_1+1,\ldots,L_2)})$ matrix block of $\boldsymbol{V}^{\mathsf{H}}$ in $\boldsymbol{\Gamma}^{(i)}$.
  Calculate the Moore-Penrose inverse $\boldsymbol{\Gamma}^{(i)\sharp}$ of matrix $\boldsymbol{\Gamma}^{(i)}$.
End For
For $i = 1 : N^{(L_2+1,\ldots,q)}$ **do**
  For $j \in J/\{i\}$ **do**
    Calculate matrix $\boldsymbol{Y}^{(i,j)} = \boldsymbol{\Gamma}^{(i)}\boldsymbol{\Gamma}^{(j)\sharp}$.
  End For
End For
Simultaneously triangularize matrices $\boldsymbol{Y}^{(i,j)}$ by using algorithm 2.1.
For $p = 1 : P$ **do**
  Build matrix $\boldsymbol{M}^{(p)}$ from the triangularized from of matrices $\boldsymbol{Y}^{(i,j)}$.
  Calculate the p-th column of matrix $\boldsymbol{U}^{(q:L_2+1)}$ by computing the right singular vector associated with the lowest singular value of $\boldsymbol{M}^{(p)}$.
End For
Calculate matrix $\boldsymbol{U}^{(L_2:1)}$ from $\boldsymbol{T}^{(2)}$ and $\boldsymbol{U}^{(q:L_2+1)}$.
For $p = 1 : P$ **do**
  Reshape column p of matrix $\boldsymbol{U}^{(L_2:1)}$ into an L_2-th order rank-1 array $\boldsymbol{\mathcal{U}}^{(L_2:1,p)}$.
  Compute the rank-1 HOSVD of the array $\boldsymbol{\mathcal{U}}^{(L_2:1,p)}$ in order to identify the p-th column of the L_2 first loading matrices $\boldsymbol{U}^{(i)}$.
  Reshape column p of matrix $\boldsymbol{U}^{(q:L_2+1)}$ into a $(q - L_2)$-th order rank-1 array $\boldsymbol{\mathcal{U}}^{(q:L_2+1,p)}$.
  Compute the rank-1 HOSVD of the array $\boldsymbol{\mathcal{U}}^{(q:L_2+1,p)}$ in order to compute the p-th column of the $(q - L_2)$ other loading matrices $\boldsymbol{U}^{(i)}$.
End For

2.3.3 Identifiability

In this section we show that assumptions **A1**)-**A5**) ensure identifiability of the SSD-CP method proposed previously. To this end, we study the different steps of the proposed algorithm. Next, we show how the assumptions **A1**)-**A5**) can be simplified by considering the generic identifiability of the SSD-CP approach, which should help the user in practice to choose the π permutation and the couple (L_1, L_2) of integers. Recall that a property is called generic when it holds everywhere, except for a set with Lebesgue measure 0 [De Lathauwer, 2006].

First, in order to calculate the truncated SVD of the unfolding matrix $\boldsymbol{T}^{(1)}$, the dimensions of $\boldsymbol{T}^{(1)}$ must be greater than or equal to the rank P, i.e. $P \leq \min\{N^{(\pi(1),\ldots,\pi(L_1))}, N^{(\pi(L_1+1),\ldots,\pi(q))}\}$. Under assumption **A1**), we have $P \leq N^{(\pi(1),\ldots,\pi(L_1))}$. In addition, under assumption **A3**), matrix $\boldsymbol{U}^{(\pi(L_2):\pi(L_1+1))}$ is full column rank and consequently we

get $P \leq N^{(\pi(L_1+1),\ldots,\pi(L_2))}$. Similarly, under assumption **A4**), matrix $\boldsymbol{U}^{(\pi(q):\pi(L_2+1))}$ is full column rank and we have $P \leq N^{(\pi(L_2+1),\ldots,\pi(q))}$. Thus we get $P \leq \min\{N^{(\pi(1),\ldots,\pi(L_1))}, N^{(\pi(L_1+1),\ldots,\pi(L_2))}, N^{(\pi(L_2+1),\ldots,\pi(q))}\}$ which leads to $P \leq \min\{N^{(\pi(1),\ldots,\pi(L_1))}, N^{(\pi(L_1+1),\ldots,\pi(q))}\}$.

In the next step, in order to define matrices $\boldsymbol{Y}^{(i,j)}$ we compute the Moore-Penrose inverse of at least one matrix $\boldsymbol{\Gamma}^{(j)}$. Assumption **A2**) implies that it exists at least one index j such the diagonal matrix $\boldsymbol{\Phi}^{(i)}$ is invertible. So, under assumptions **A2**) and **A3**), since $\boldsymbol{W}$ is non-singular, at least one matrix $\boldsymbol{\Gamma}^{(j)}$ is full row rank and admits a Moore-Penrose matrix inverse denoted by $\boldsymbol{\Gamma}^{(j)\sharp}$.

Now, let's study the conditions for which matrix $\boldsymbol{U}^{(q:L_2+1)}$ is unique up to a diagonal matrix. In order to maintain this uniqueness, for each p, matrix $\boldsymbol{M}^{(p)}$ should be unique and rank deficient only by one implying that $\boldsymbol{M}^{(p)}$ has one and only one zero singular value. The specific structure of matrix $\boldsymbol{M}^{(p)}$, with only two non-zero entries in each row where one of them is constant 1 (as presented in (2.89)), clearly shows that each $(N^{(\pi(L_2+1),\ldots,\pi(q))} - 1)$ columns of $\boldsymbol{M}^{(p)}$ are linearly independent. Moreover, since $\boldsymbol{U}^{(q:L_2+1)}$ is a solution of (2.88), it is clear that the $N^{(\pi(L_2+1),\ldots,\pi(q))}$ columns of $\boldsymbol{M}^{(p)}$ are linearly dependent. Consequently, $\boldsymbol{M}^{(p)}$ is rank deficient only by one. On the other hand, since the non-constant entries of matrices $\boldsymbol{M}^{(p)}$ are the diagonal entries of matrices $\boldsymbol{D}^{(i,j)}$, for each ordered pair (i,j), the eigen values of matrix $\boldsymbol{Y}^{(i,j)}$ as well as the arrangement of these eigenvalues with respect to those of other $(R-1)$ matrices $\boldsymbol{Y}^{(i',j')}$ must be unique. The former condition inherently holds, but to satisfy the latter, the joint diagonalizer $\boldsymbol{W}$ used in (2.84) must be unique. To this end, we use the following uniqueness theorem [De Lathauwer et al., 2004]:

Theorem 1 The JEVD of R matrices $\boldsymbol{Y}^{(r)}$ of size $(P \times P)$ by a matrix transform $\boldsymbol{W}$ is unique up to a permutation and a scaling of the columns of $\boldsymbol{W}$ if and only if all the columns of the $(R \times P)$ matrix $\boldsymbol{E}$, whose (r,p)-th component of $\boldsymbol{E}$ is equal to $\boldsymbol{D}^{(r)}_{p,p}$, are distinct.

Note that in our case the number R of matrices to be triangularized is a value between $N^{(\pi(L_2+1),\ldots,\pi(q))} - 1$ and $N^{(\pi(L_2+1),\ldots,\pi(q))}(N^{(\pi(L_2+1),\ldots,\pi(q))} - 1)$. Although this theorem is expressed in the real domain, it can easily be extended to the complex case. This theorem states that the uniqueness is guaranteed if for each couple (p_1,p_2) there is at least one pair (i,j) such that $D^{(i,j)}_{p_1,p_1} \neq D^{(i,j)}_{p_2,p_2}$, i.e. such that $U^*_{i,p_1}/U^*_{j,p_1} \neq U^*_{i,p_2}/U^*_{j,p_2}$ by definition of $\boldsymbol{D}^{(i,j)}$ and consequently $U^*_{i,p_1}/U^*_{i,p_2} \neq U^*_{j,p_1}/U^*_{j,p_2}$ where $\boldsymbol{U}^* = \boldsymbol{U}^{(\pi(q):\pi(L_2+1))}$. As a consequence, the result of theorem 1 is valid if the columns of $\boldsymbol{U}^{(\pi(q):\pi(L_2+1))}$ are non-colinear, which is ensured under assumption **A4**).

Finally, to compute the Moore-Penrose inverse of $\boldsymbol{U}^{(\pi(q):\pi(L_2+1))}$ in order to calculate matrix $\boldsymbol{U}^{(\pi(L_2):\pi(1))}$, matrix $\boldsymbol{U}^{(\pi(q):\pi(L_2+1))}$ should be full column rank, which is satisfied under assumption **A4**).

Now let's show how the assumptions **A1**)-**A5**) can be simplified by considering the generic case. To this end, we use the lemma defined in [De Lathauwer, 2006], which is expressed as follows:

Lemma 2 Consider $\boldsymbol{A} \in \mathbb{R}^{I\times P}$ and $\boldsymbol{B} \in \mathbb{R}^{J\times P}$. Generically we have $\text{rank}(\boldsymbol{A} \odot \boldsymbol{B}) = \min(IJ, P)$.

This lemma can easily be generalized to complex-valued matrices. Consequently, the following equality holds generically:

$$\text{rank}(\boldsymbol{U}^{(\pi(L_2):\pi(L_1+1))}) = \min(N^{(\pi(L_1+1),\ldots,\pi(L_2))}, P) \tag{2.104}$$

Assumption **A3**) is then generically equivalent to the inequality $P \leq N^{(\pi(L_1+1),\ldots,\pi(L_2))}$. Similarly, **A4**) and **A5**) are generically equivalent to $P \leq N^{(\pi(L_2+1),\ldots,\pi(q))}$ and to $P \leq N^{(\pi(1),\ldots,\pi(L_1))}$, respectively. On the other hand the assumption **A2**) is equivalent to the assumption that at least the entries of one row of the $(q - L_2)$ loading matrices $\boldsymbol{U}^{\pi(L_2+1)}$ to $\boldsymbol{U}^{\pi(q)}$ are non-zero. In summary, we conclude that if $P \leq \min\{N^{(\pi(1),\ldots,\pi(L_1))}, N^{(\pi(L_1+1),\ldots,\pi(L_2))}, N^{(\pi(L_2+1),\ldots,\pi(q))}\}$ and at least the entries of one row of the loading matrices $\boldsymbol{U}^{\pi(L_2+1)}$ to $\boldsymbol{U}^{\pi(q)}$ are non-zero , then the identifiability of the proposed SSD-CP algorithm is generically ensured. It is noteworthy that these simplified identifiability conditions of the SSD-CP method should be used in practice to choose the π permutation and the couple (L_1, L_2) of integers which give a maximal achievable rank.

The last point that we should state is that in solving the CP problem by using the proposed method, there are many $\Gamma^{(i)}$ matrices which can be used to define matrices $\boldsymbol{Y}^{(i,j)}$ as defined in (2.84). There is no need to use all $\Gamma^{(i)}$ matrices to solve the CP problem. We can choose only matrices with good condition number in order not to affect the conditioning of the problem by computing the pseudoinverse.

2.3.4 Denoising epileptic EEG signals using the SSD-CP algorithm

To use the SSD-CP algorithm to denoise interictal and ictal data, we should define appropriate higher order arrays. For both interictal and ictal data, the continuous wavelet transform is used to construct a third-order array. Consequently, the tensor dimensions are channels, time and frequency. Then, we applied the SSD-CP algorithm on the constructed array with the rank P'. The first loading matrix $\boldsymbol{U}^{(1)}$ is the estimate of the mixing matrix $\boldsymbol{G}$. Then we calculate all the P' sources as $\hat{\boldsymbol{r}}[n] = \boldsymbol{U}^{(1)\sharp}\boldsymbol{x}[n]$. For interictal data, we use the same spike detection algorithm as described in section 2.2.2 to select the sources of interest. For ictal data, we choose the epileptic sources by using the time-frequency plot of each source. Eventually, we denoise signal by using only the sources of interest.

2.4 The numerical complexity of the proposed algorithms

In this section, we analyze the numerical complexity of the proposed algorithms in terms of real-valued floating point operations (flops). A flop corresponds to a multiplication followed by an addition, but in practice the number of multiplications is computed because they are more expensive than additions.

Table 2.1: The numerical complexity of each step of the GEVD/DSS-based interictal denoising methods

	Multiplications	Comparisons
Spike detection	$MN(M_wF_w+12)+2(M-1)N_s$	$2N\log(N)+MN_slog(MN_s)+\frac{1}{2}(M-1)N_s(M(N_s+1)-3)+N_s$
Extraction of the exact time support	$M(2N+2)+44N_sL_w$	N_sL_w
Spike clustering	$M(L_{ma,2}T_s+2N_sP_e'^2)$	$\frac{4}{3}T_s$
The GEVD algorithm	$\frac{91}{3}M^3P_e'+M^2P_e'(N+T_s+1)$	
The DSS algorithm	$B+P_e'(M^2T+MN+M)+\sum_{p=1}^{P_e'}It_p(2MN+M)$	

In the following computations, M, N, P' and P_e' are the number of channels, time samples, all independent sources and independent epileptic sources, respectively. $B=\min\{NM^2/2+4M^3/3+P'MN, 2NM^2\}$ is the number of flops required to perform spatial whitening. Moreover, the parameters corresponding to each method are defined at the beginning of each subsection. Note that the results of this section will be used later in chapter 4 in order to compare the denoising algorithms.

2.4.1 The GEVD/DSS methods to denoise interictal EEGs

In this section, M_w and F_w are the number of wavelet decompositions and the length of discrete mother wavelet in spike detection algorithm. N_s is the total number of spikes in all epileptic sources and T_s is the total number of samples in the spike durations. L_w is the length of window used in extracting the exact time support of the spikes. The length of the moving average filters used in the sections "Extraction of the exact time support of the spikes" and "Spike clustering" are denoted by $L_{ma,1}$ and $L_{ma,2}$, respectively. It_p is the number of iterations in the DSS algorithm to extract the p-th epileptic source. The numerical complexity of each step of the semi-blind methods is shown in Table 2.1.

By using these calculations, the total numerical complexity of the GEVD-based and DSS-based methods are computed and shown in Table 2.3. Since the comparison is much simpler than multiplications, we consider only the number of multiplications for calculating the required flops. It should be noted that for these methods, we separately show the numerical complexities of two cases (1) $P_e'=1$ and (2) $P_e'\neq 1$, because for the case of $P_e'=1$ the spike clustering stage is discarded.

2.4.2 The GEVD/DSS-based methods to denoise ictal EEGs

In the TF-GEVD and TF-DSS denoising methods, K and TF_s are the number of frequency samples and total number of time-frequency samples involved in epileptic sources. It_p is the number of iterations in the DSS algorithm to extract the p-th epileptic source. The CCA method which is the preprocessing stage of these methods needs $5NM^2+5NM+19M^3/3$ flops. By considering the complexity of the GEVD and

DSS algorithms (table 2.1), the numerical complexity of the TF-GEVD and TF-DSS methods to denoise ictal signals are calculated as shown in table 2.4.

2.4.3 The JDICA algorithm

In the following computations, $f_4(P') = P'(P'+1)(P'+2)(P'+3)/24$ is equal to the number of free entries in a fourth order cumulant tensor of dimension P enjoying all symmetries. R is the complexity required to compute the roots of a real 8-th degree polynomial by using the companion matrix technique (we may take $T = 972$ flops). As a result the proposed JDICA algorithm requires C_{JDICA} as follows:

$$\begin{aligned} C_{JDICA} &= B + 2N + 2P' + MP'^2 + 3Nf_4(P') + \sum_{p=2}^{P'-1}(2p^2(p-1)^2(p^2-p+1)) \\ &+ \sum_{p=2}^{P'}(p^2N + pP' + (p-1)P'^2) + \sum_{p=2}^{P'} It_{P'-p+1}(T + 4p^3/3 - 7p^2 + 62p/3 \\ &+ 195 + \min((p-2)N + 8(4N+8), 4p + 8(2p^4 + p^2 + 7))) \end{aligned} \tag{2.105}$$

As explained in section 2.2.2, in the JDICA-based method to denoise interictal data we need to select the sources of interest. The numerical complexity of this step is $P'N^2$. Therefore, by using this value and the numerical complexity of the JDICA algorithm, we can calculate the numerical complexity of the JDICA-based methods used to denoise interictal and ictal signals. The numerical complexity of these methods are shown in tables 2.3 and 2.4.

2.4.4 The SSD-CP algorithm

In the following computations, N_1 to N_q are the dimensions of the array. Symbols q, P, L_1 and L_2 are the order of the array, the rank of the array and the integers used to define the unfolding matrices $\boldsymbol{T}^{(1)}$ and $\boldsymbol{T}^{(2)}$, respectively. For the sake of simplicity, we assume that π is the identity permutation. R is the number of matrices to be jointly triangularized. The number of required iterations in the SSD step is equal to It. We also define $N^- = N^{(1,\ldots,L_1)}$, $N^+ = N^{(L_1+1,\ldots,L_2)}$ and $N^* = N^{(L_2+1,\ldots,q)}$. The number of required flops for each step of the algorithm is given in table 2.2.

To calculate the numerical complexity of each step, we use the following facts. First, for a matrix $\boldsymbol{H}$ of size $(M \times N)$ with the rank r $(M \geq N \geq r)$, the numerical complexity of calculating singular values, r dominant left singular vectors and r dominant right singular vectors are $(2MN^2 - 2N^3/3)$, $(5Mr^2 - r^3/3)$ and $(5Nr^2 - r^3/3)$ flops, respectively. The numerical complexity of the Moore-Penrose pseudoinverse of matrix $\boldsymbol{H}$ is equal to $(8MN^2 + 11N/3)$ flops. Calculating the roots of a p-th degree polynomial by using the companion matrix method is equal to $(4p^3/3 + p)$ flops. As presented in table 2.2, since in each iteration, the simultaneous triangularization step should be done for each pair of (p_i, q_i), numerical complexity of this step should be summed up for all pairs of (p_i, q_i), $1 \leq p_i, q_i \leq P$ and $p_i < q_i$. This value should also be multiplied by the number of iterations It. By summing up the total complexity of this stage with the

complexity of the other steps, the total numerical complexity of the SSD-CP algorithm to calculate the CP decomposition of a q-th order complex array as follows:

$$
\begin{aligned}
&C_{SSD-CP} = It\left(\frac{P(P-1)}{2}\left(\frac{4}{3}14^3+1073+47R+\frac{200}{3}PR\right)\right)+P^3\left(\frac{10}{3}\right)+\\
&P^2\big(13N^+N^*+8N^*+RN^++1\big)+P\left(\frac{11}{3}N^*+2(N^+)^2R+\right.\\
&4(N^+)^3+N^-N^+N^*+2N^-N^+(\sum_{i=1}^{L_2}N_i)\\
&\left.+\,2N^*(\sum_{i=L_2+1}^{q}N_i)-\frac{2}{3}(\sum_{i=1}^{q}N_i^3)+5(\sum_{i=1}^{q}N_i)-\frac{q}{3}+1\right)+\\
&2(N^-)^2N^+N^*-\frac{2}{3}(N^-)^3
\end{aligned}
\tag{2.106}
$$

By using this numerical complexity, we can calculate the complexity of the SSD-CP based methods to denoise interictal and ictal signals as explained in section 2.3.4. The numerical complexity of these methods are shown in tables 2.3 and 2.4. In these tables, N_t and N_f are respectively the number of time and frequency samples in the time-frequency transform.

2.5 Summary

In this chapter, we presented the proposed methods to denoise epileptic EEG data, including i) the GEVD and DSS-based methods for denoising of interictal data, ii) the TF-GEVD and TF-DSS methods for denoising ictal data, iii) the JDICA algorithm for denoising both interictal and ictal data and iv) the SSD-CP algortihm for denoising both interictal and ictal data. We also studied the numerical complexity of all proposed algorithms. In the next chapter, we present the simulated and real data used in various analysis.

Table 2.2: Numerical complexity of different steps of the SSD-CP algorithm.

SVD of $\boldsymbol{T}^{(1)}$	
Calculation of $\boldsymbol{\Sigma}$	$2(N^-)^2N^+N^* - \frac{2}{3}(N^-)^3$
Computation of $\boldsymbol{V}$ (only the first P dominant right singular vectors)	$5N^+N^*P^2 - \frac{P^3}{3}$
Computation of the R matrices $\boldsymbol{Y}^{(i,j)}$	
Calculation of the pseudo-inverse of the N^* matrices $\boldsymbol{\Gamma}^{(i)}$	$N^*(8N^+P^2 + \frac{11}{3}P)$
Computation of the R matrices $\boldsymbol{Y}^{(i,j)}$	RN^+P^2
Simultaneous Triangularization (for p_i and q_i)	
Calculation of the polynomial coefficients	$6R(q_i - p_i - 1) + 10R + 745$
Computation of the fourteen roots	$14 + \frac{4}{3}14^3$
Calculation of the corresponding $\hat{\rho}$ values	140
Calculation of the cost function for all couples $(\hat{\rho}, \hat{\phi})$	$170R(q_i - p_i - 1) + 197R + 168$
Update of the matrices $\boldsymbol{Y}^{(i,j)}$	$8PR + 16R + 6$
Calculation of $\boldsymbol{U}^{(q:L_2+1)}$	
Computation of the right singular vector (corresponding to the zero singular value)	$P(2R(N^+)^2 + 4(N^+)^3)$
Estimating the loading matrices	
Calculation of $\boldsymbol{U}^{(L_2:1)}$	$PN^-N^+N^* + 8P^2N^* + \frac{11}{3}P^3 + P^2 + P$
HOSVD of the P arrays $\boldsymbol{\mathcal{U}}^{(L_2:1,p)}$	$P(2N^-N^+(N_1+...+N_{L_2}) - \frac{2}{3}(N_1^3+...+N_{L_2}^3) + 5(N_1 + ... + N_{L_2}) - \frac{L_2}{3})$
HOSVD of the P arrays $\boldsymbol{\mathcal{U}}^{(q:L_2+1,p)}$	$P(2N^*(N_{L_2+1} + ... + N_q) - \frac{2}{3}(N_{L_2+1}^3 + ... + N_q^3) + 5(N_{L_2+1} + ... + N_q) - \frac{q-L_2}{3})$

Table 2.3: The numerical complexity of the interictal denoising methods

	Numerical Complexity
GEVD ($P'_e = 1$)	$91M^3/3 + M^2(N + T_S + 1) + 44N_sL_w + M(14N + M_wF_wN + 2 + 2N_s) - 2N_s$
GEVD ($P'_e \neq 1$)	$91M^3P'_e/3 + M^2P'_e(N + T_S + 1) + 44N_sL_w + M(14N + MF_wN + L_{ma,2}T_s + 2 + 2N_s(1 + P'^2_e)) - 2N_s$
DSS ($P'_e = 1$)	$B + It_1(2MN + M) + N(16N + M_wF_wN + 4 + 2N_s) + 44N_sL_w - 2N_s$
DSS ($P'_e \neq 1$)	$B + M^2NP'_e + \sum_{p=1}^{P'_e} It_p(2MN + M) + 44N_sL_w - 2N_s + M(15N + M_wF_wN + P'_e(N + 1) + L_{ma,2}T_s + 3 + 2N_s(1 + P'^2_e))$
JDICA	$P'N^2 + B + 2N + 2P' + MP'^2 + 3Nf_4(P) + \sum_{p=2}^{P'-1}(2p^2(p-1)^2(p^2 - p + 1)) + \sum_{p=2}^{P'}(p^2N + pP' + (p-1)P'^2) + \sum_{p=2}^{P'} It_{P'-p+1}(T + 4p^3/3 - 7p^2 + 62p/3 + 195 + \min((p-2)N + 8(4N + 8), 4p + 8(2p^4 + p^2 + 7)))$
SSD-CP	$P'N^2 + 2M^2N_tN_f - 2M^3/3 + 5N_tN_fP'^2 - P'^3/3 + N_f(8P'^2N_t + 11/3P'^3 + P'^2 + P) + RN_tP'^2 + It((13R + 8RP' + 325/3)(P'(P' - 1)/2) + (20RP'(P' + 1)(P' - 1)/6)) + P'(7R(N_f - 1)^2 + 11/3(N_f - 1)^3 + (N_f - 1)(R + 1) + 2(N_f - 1)^2) + P'(2MN_t^2 - 2N_f^3/3 + 6M + 5N_t - 2/3) + MN_tN_fP' + 8P'^2N_f + 11P'/3$

Table 2.4: The numerical complexity of the ictal denoising methods

	Numerical Complexity
TF-GEVD	$5NM^2 + 5NM + 19M^3/3 + 91/3M^3P'_e + M^2P'_e(NK + TF_s + 1)$
TF-DSS	$5NM^2 + 5NM + 19M^3/3 + B + P'_e(M^2NK + MNK + M) + \sum_{p=1}^{P'_e} It_p(2MNK + M)$
JDICA	$B + 2N + 2P' + MP'^2 + 3Nf_4(P') + \sum_{p=2}^{P'-1}(2p^2(p-1)^2(p^2 - p + 1)) + \sum_{p=2}^{P'}(p^2N + pP' + (p-1)P'^2) + \sum_{p=2}^{P'} It_{P'-p+1}(T + 4p^3/3 - 7p^2 + 62p/3 + 195 + \min((p-2)N + 8(4N + 8), 4p + 8(2p^4 + p^2 + 7)))$
SSD-CP	$2M^2N_tN_f - 2M^3/3 + 5N_tN_fP'^2 - P'^3/3 + N_f(8P'^2N_t + 11/3P'^3 + P'^2 + P) + RN_tP'^2 + It((13R + 8RP' + 325/3)(P'(P' - 1)/2) + (20RP'(P' + 1)(P' - 1)/6)) + P'(7R(N_f - 1)^2 + 11/3(N_f - 1)^3 + (N_f - 1)(R + 1) + 2(N_f - 1)^2) + P'(2MN_t^2 - 2N_f^3/3 + 6M + 5N_t - 2/3) + MN_tN_fP' + 8P'^2N_f + 11P'/3$

Chapter 3

Data Generation

In this chapter, we present the simulated and real data used in various analysis. We start by introducing how interictal and ictal signals are generated from a realistic model of coupled neuronal population and then we describe the real signals recorded in a patient with partial epilepsy. Then, we introduce the performance criteria used in the next chapter for the comparison between the denoising methods.

3.1 Simulated data

To quantitatively evaluate the performance of the proposed methods presented in Chapter 2, simulations of 32 channels of EEG data are generated with a spatio-temporal model developed in our team [Cosandier-Rimélé et al., 2007, Cosandier-Rimélé et al., 2008, Cosandier-Rimélé et al., 2010]. This model combines a biophysical model of the sources of brain activity, represented by a dipole field at the cortical surface, and a biomathematical model of coupled neuronal populations. The first model is used to describe the spatial characteristics of sources and the second one describes the time course of sources (see figure 3.1).

3.1.1 The source model

3.1.1.1 Spatial characteristics of sources

To realistically reproduce the complex geometry of the cerebral neocortex, a mesh of the cortical surface at the interface of gray and white matter is built from the segmentation of a 3D T1 MRI image (BrainVisa, SHFJ, Orsay , France) [Mangin et al., 1995]. This mesh consists of 19626 triangles with an average area of 4.8 mm^2. Each triangle in the mesh is associated with an elementary current dipole placed at the barycenter of each triangle and oriented perpendicular to its surface. The norm of the dipole moment is proportional to the area of the corresponding triangle and is obtained by multiplying the area of the triangle by the dipole moment density per unit area. This density is defined by the product between the thickness of the neocortex, estimated to be averagely 3 mm in this study, and the cortical current density, estimated to be 100 nA/mm^2 [Hämäläinen et al., 1993]. The average value of density of the dipole moment

per unit area is equal to 300 nA/mm ($nA.m/m^2$), and the norm q_i of the i-th dipole moment is given by $q_i = 300\Delta_i\, nA/mm$ where Δ_i denotes surface of the i-th triangle.

The source or sources of epileptic activity (interictal or ictal) are defined manually on the mesh as a set of contiguous triangles (patch). Here, each patch consists of 100 triangles which represents a source of approximately 5 cm^2. Different locations were considered. In cases where one epileptic patch is considered, this patch is delineated in the left superior temporal gyrus. In the case where two epileptic patches are considered, the first one is placed in the left superior temporal gyrus, while the second one is located either in the left inferior parietal region (for simulation of interictal signals) or in the left inferior frontal region (for simulation of ictal signals).

3.1.1.2 Temporal dynamics of sources

The temporal dynamics of activities of each population of neurons associated with each dipole of the epileptic patch are generated by a macroscopic model of neural populations [Jansen et al., 1993, Jansen and Rit, 1995, Wendling et al., 2000, Wendling et al., 2001, Wendling et al., 2002]. This model considers the average activity of interconnected neuronal sub-populations without an explicit representation of the unit cells or of mechanisms associated with transmembrane exchanges. Although macroscopic, this type of model is based on neurophysiological data and has two essential features. First, the model parameters reflect the inhibitory and excitatory processes of the considered neuronal tissues (here the neocortex). Second, the temporal dynamics generated at the output of the model are similar to those observed in a local potential field. The above-mentioned work has shown that, in fact, they are comparable to real signals recorded by electrodes implanted in the cerebral cortex for an appropriate set of parameters in the model. In other words, we consider that the local field potentials are mainly a reflection of the dynamic processes associated with synaptic interactions within sub-populations of cells (pyramidal neurons and interneurons). The macroscopic model that we used is based on neurophysiological data of neuronal organization and connectivity within the cerebral cortex. In the model illustrated in figure 3.1, each neuronal population is formed by three subassemblies of cells: the main cells (the pyramidal cells which represent about 70 percent of the cortical neurons) and two types of interneurons (slow and fast inhibition). The synaptic organization of the neocortex is represented in the model by excitatory projections between pyramidal cells and interneurons. In turn, these interneurons project to the pyramidal cells via inhibitory connections (GABAergic).

In each subset of cells, the input/output relationship is represented generally by means of two functions: i) a linear transfer function, which transforms the presynaptic information (the average presynaptic pulse density of afferent action potentials) into postsynaptic information (an average postsynaptic membrane potential) and ii) a static non-linear function, which relates the average postsynaptic potential of a given subset to an average pulse density of potentials fired by the neurons. The linear transfer function models the effects associated with synaptic kinetics and passive conduction in dendritic trees. It may be represented by a low-pass filter of order 2, whose impulse response is given by $h(t) = Wtwe^{-wt}$ where (W, w) are the pairs (amplitude, time constant) of the postsynaptic potential mediated by a specific receptor-type. These pairs are marked

(A, a) in the excitatory case (AMPA-type receptors) and (B, b) and (G, g), in the three modes of inhibition dependent on postsynaptic receptors $\text{GABA}_{A,slow}$, $\text{GABA}_{A,fast}$.

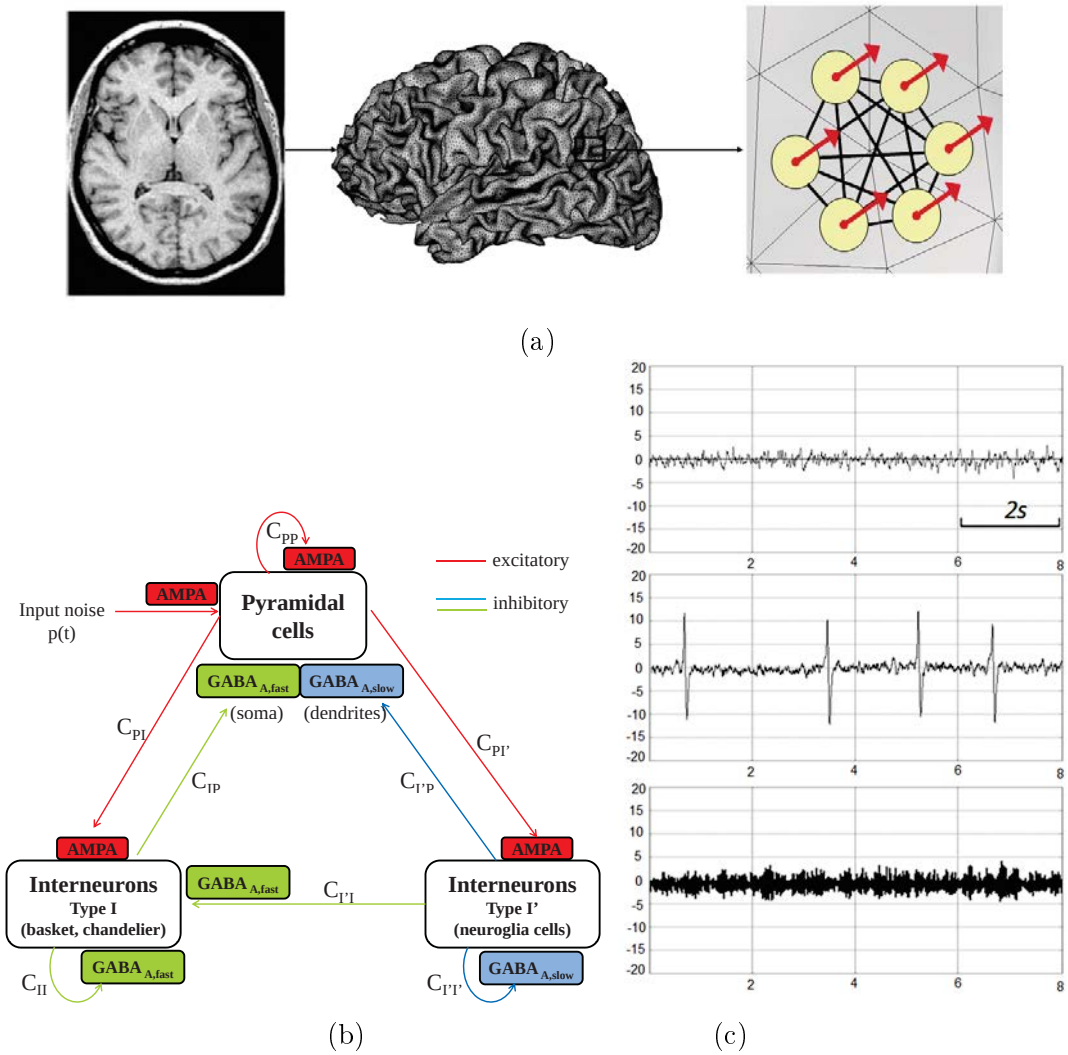


Figure 3.1: Simulation of cortical activities. (a) realistic mesh of the cortical surface (gray matter/white matter interface), built from the segmentation of a 3D anatomical T1 MRI. A current dipole is located at the barycenter of each triangle and oriented orthogonally to the surface. The intensity of each dipole is proportional to the area of the corresponding triangle. (b) Structure of the model used to simulate the activity of the cerebral cortex. Three subpopulations of neurons are represented: i) pyramidal cells, ii) somatic-projecting and dendritic-projecting interneurons (type I: basket and chandelier cells involving currents $\text{GABA}_{A,fast}$), and iii) dendritic interneurons (type I': neurogliaformes cells involving currents $\text{GABA}_{A,slow}$ and GABA_B). The collateral excitation between pyramidal cells and excitatory input on interneurons are represented in the model by the currents mediated by AMPA receptors. The parameters (C_{PP}, C_{PI}, C_{IP}, C_{II}, $C_{PI'}$, $C_{I'P}$, $C_{I'I'}$, $C_{I'I}$) are constants of local connectivity between subpopulations. (c) Signals obtained at the output of the model. From up to down: the signal produced by the model representing a background activity, a signal representing an interictal spike activity and a signal representing a fast ictal activity.

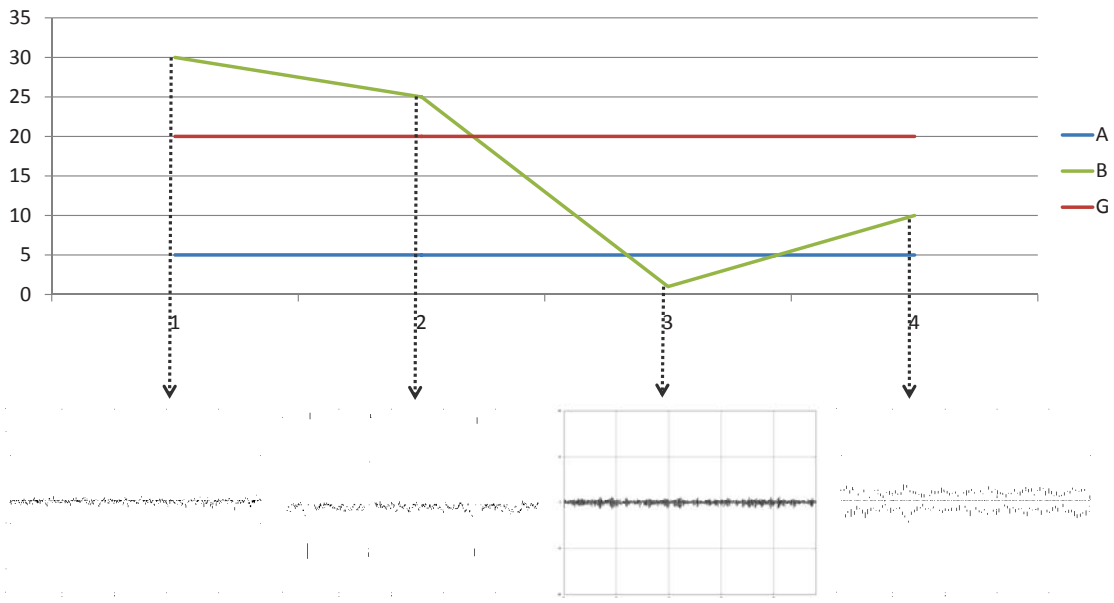
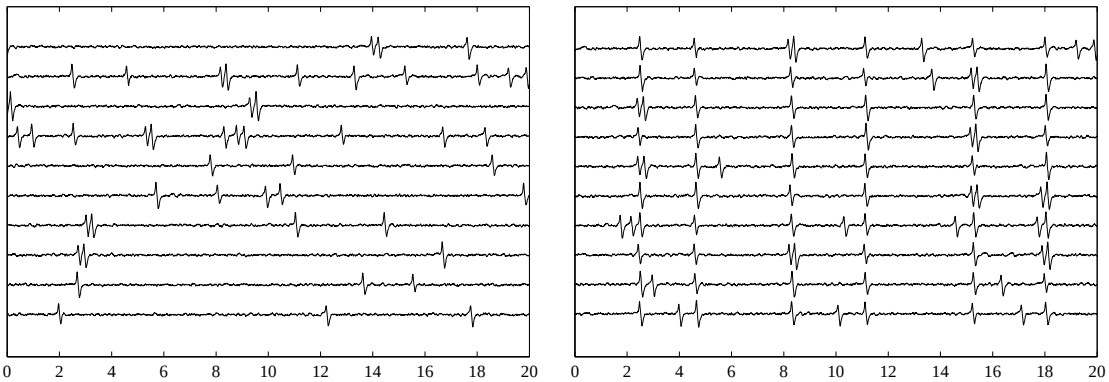


Figure 3.2: In the model, the amplitudes of excitatory and inhibitory parameters A, B, and G are adjusted such that the model generates normal and epileptic activity. Signals obtained at the output of the model. From left to right, a background activity with the basic values used in the model, an interictal spikes activity obtained after a slight decrease of the slow inhibition (B), a fast activity obtained after a sharp decline in the slow inhibition, a slower rhythmic activity after increase in slow inhibition.

The synaptic interactions between neuronal subpopulations are also characterized in the model by connectivity constants (C_{PP}, C_{PI}, C_{IP}, C_{II}, $C_{PI'}$, $C_{I'P}$, $C_{I'I'}$, $C_{I'I}$), which represent the average number of synaptic contacts between the considered subpopulations. Finally, the nonspecific influence of neighbors or more distant populations is modeled by a Gaussian input noise $\boldsymbol{v}(t)$ (input of type excitatory) that describes generally the average density of the related action potentials. Finally, in this model, the inter-population couplings can be defined using the average density of action potentials generated by the subset of principal cells of a given population as related excitation on the subset of principal cells of another population. These couplings are resolved through constant connectivity (K_{ij}), which defines the degree of coupling of a population i to population j.

In this model, we adjusted the amplitudes of A, B and G to obtain an activity of type sporadic spikes or an activity of type seizure discharge (figure 3.2). Compared with baseline values used in the model for the genesis of a background activity, sporadic spikes are obtained by slightly decreasing of slow inhibition. If B is greatly reduced, we obtain a fast rhythmic activity (25-30 Hz) that resembles the tonic rapid discharge recorded at the beginning of the seizure in intracerebral recordings. When rapid inhibition decreases in turn, we get a slower rhythmic activity (8-10 Hz) that may be related to the slowdown of the ictal discharge [Wendling et al., 2005, Cosandier-Rimélé et al., 2012]. To generate a temporal dynamics of type ictal discharge, we apply the settings to get a quick activity for the first 10 seconds, then the parameters are set to obtain a slower rhythmic activity



(a) (b)

Figure 3.3: Examples of activities obtained by the model in several populations. (a) Asynchronous sporadic spikes activities in a group of 10 uncoupled populations. (b) Sporadic spikes activities in the same group of populations, after introduction of a constant and uniform coupling value K_{ij} across populations. Strong synchronization between the temporal dynamics of populations are obtained.

during the last 15 seconds. Between the two activities, we consider a 2 sec period of linear transition in which the parameter B is varied linearly from one value to another.

3.1.2 Construction of the intensity matrix

In the case of interictal activity, the dynamic of the dipoles related to an epileptic patch made of 100 triangles are obtained using a model of 100 coupled populations. The connectivity constant K_{ij} is adjusted to obtain a series of temporal dynamic of strongly correlated interictal spikes (figure 3.3).

For the genesis of interictal spikes signals, we consider that the activities of dipoles within a patch are highly synchronized. In the case of two patches, two scenarios are considered whether activities of patch 1 are: i) decorrelated from those of the patch 2 (two independent sources) or ii) correlated to activities of the patch 2 by considering a few milliseconds time delay (12-16ms) (propagation of interictal activity from patch 1 to patch 2).

In the case of ictal activities, to simulate the interaction between two brain regions during a seizure, we consider a model of two coupled populations. Indeed, the coupling between a large number of populations is difficult to control in the model in the case of fast activities. The parameters of the first population are set to generate a fast ictal activity, followed by a slower ictal activity. In the case of the second population, the excitation parameter A is slightly decreased, in order to obtain a "normal" activity in the absence of coupling. The introduction of the coupling between the two populations then drives the second population into the ictal activity (figure 3.4).

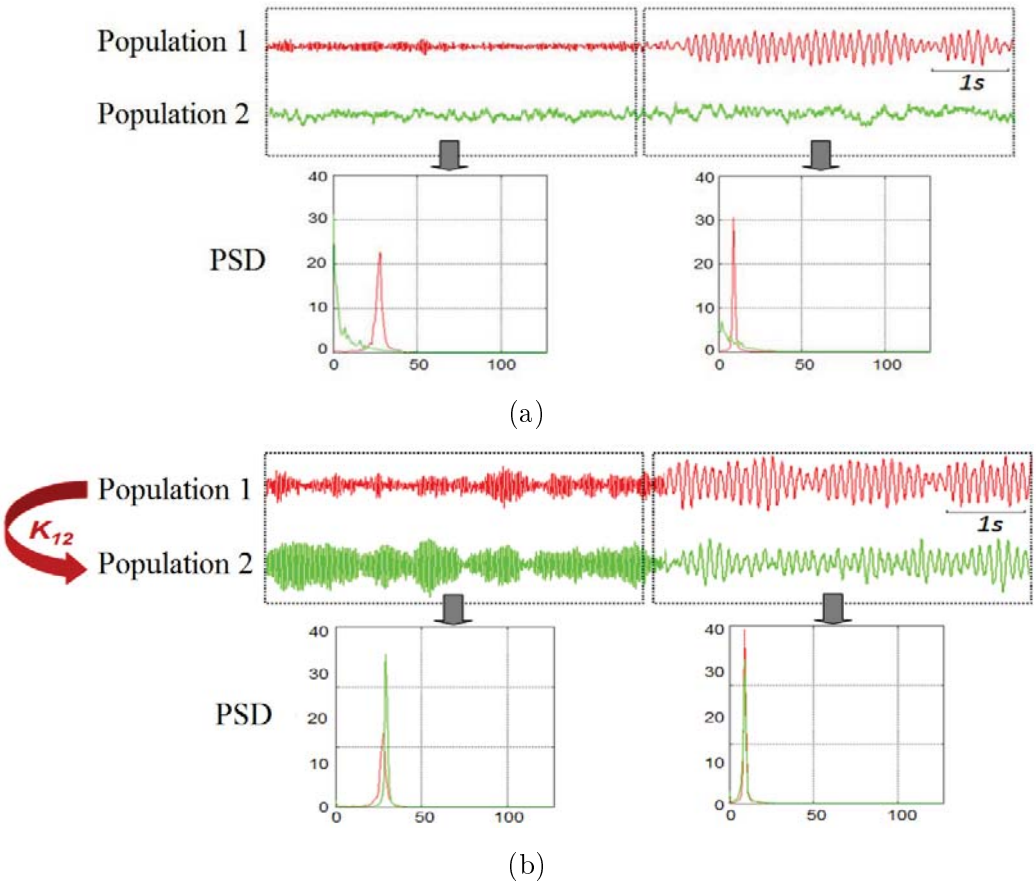
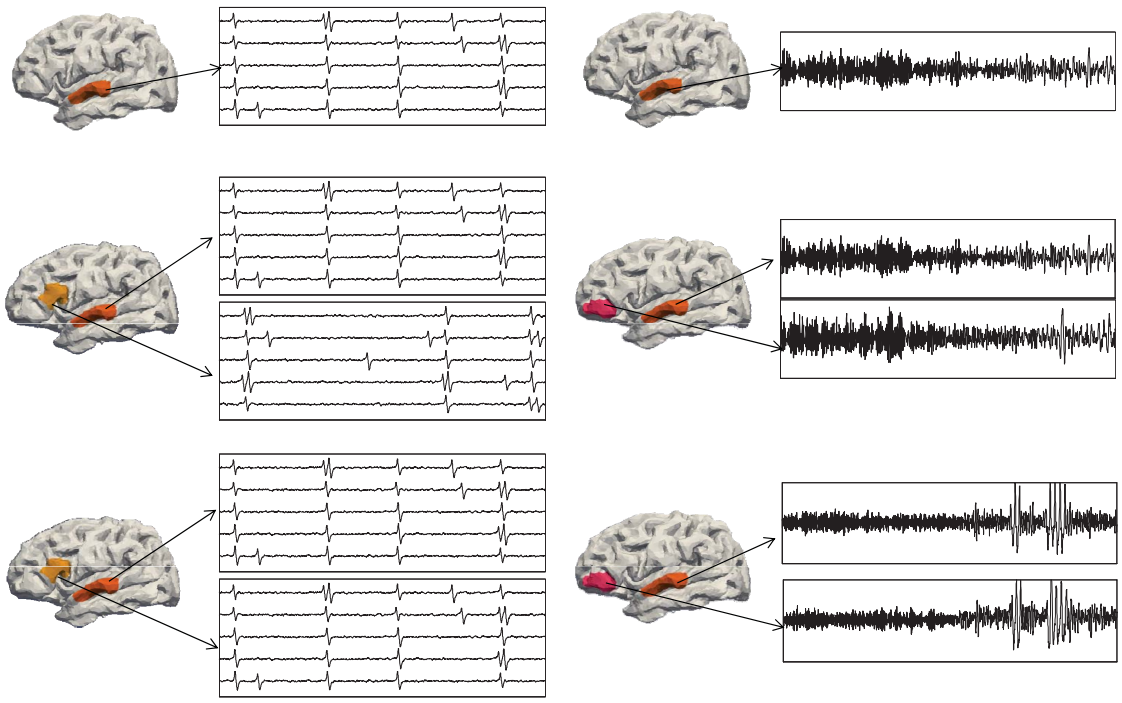


Figure 3.4: Model of two populations in the case of ictal activities. (a) In the first population, the parameters are set to generate a fast ictal activity, followed by a slower rhythmic ictal activity. In the second population, the excitation parameter A is slightly decreased to obtain a "normal" activity. The coupling between the two populations is zero. The power spectral density (PSD) of the population 1 shows the peak of fast activity at 25 Hz and the slower peak activity at 8 Hz. (b) The introduction of the coupling between the two populations drives the second population into the ictal activity. The PSD of the population 1 shows the same activity peaks at the same frequencies.

Moreover, as interictal spikes, we also considered a scenario in which the ictal activity in the patch 1 was independent of the ictal activity in the patch 2. The intensity matrix is then constructed by assigning the first temporal ictal dynamics to all dipoles of patch 1 and the second temporal ictal dynamics to all the dipoles of patch 2. Finally the dipoles outside patches have zero activity. This produces a matrix of intensities $\boldsymbol{S}$ whose columns and rows represent the patch dipoles and temporal samples, respectively (figure 3.5).

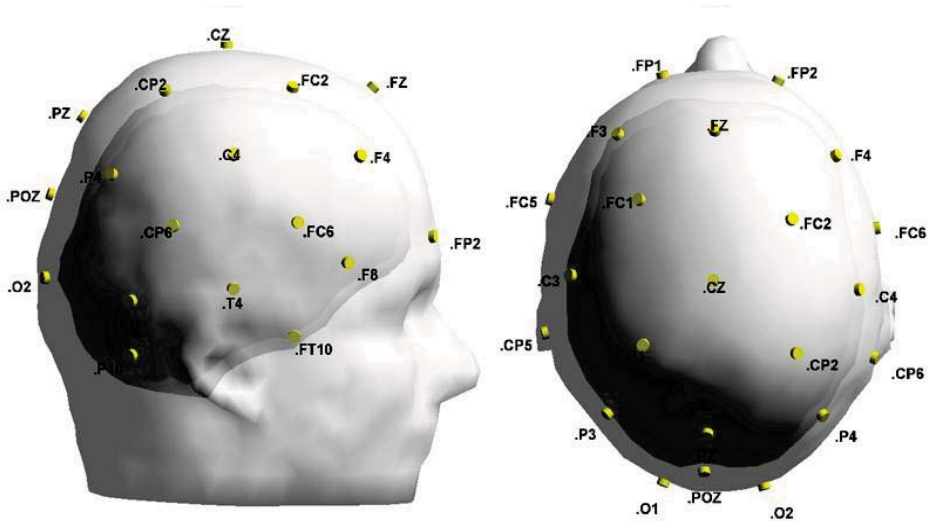


(a) (b)

Figure 3.5: (a) Configurations of sources for the genesis of the interictal spikes activities. From top to bottom, the dynamics of dipoles associated with a single patch located in the left superior temporal region (orange patch), or two patches, the second one being located at the left inferior parietal region (yellow patch). Within a given patch, the spikes are very synchronous. Between patches, activities are either asynchronous (uncorrelated) or synchronous (correlated). In the latter case, a time delay of a few milliseconds (16 ms) is added to mimic the activity of the spread between the two patches. (b) Configurations of sources for the genesis of ictal activities. From top to bottom, the dynamics of dipoles associated with a single patch located in the left superior temporal region (orange patch), or from two patches, the second one being located in the left inferior frontal region (red patch). Ictal activities are generated either by two independent populations (uncorrelated activities) or by two coupled populations (correlated activities).

3.1.3 Genesis of simulated EEG signals

In our study, we solve the direct problem in a realistic head model. This model consists of three nested homogeneous conducting media representing the brain, skull and scalp. The conductivities of these areas are set at $0.33S/m$, $0.0082S/m$ and $0.33S/m$ [Gonçalves et al., 2003]. The surfaces that define each environment are extracted from the segmentation of the 3D T1 MRI as the source model, and are meshed by 2440 triangles (ANT, ASA, Enschede, Netherlands [Zanow and Peters, 1995]). In this realistic head model, the direct problem can only be solved numerically. We use the Boundary Element Method (BEM) implemented in ASA, based on [Hamalainen



(a) (b)

Figure 3.6: The position of 32 scalp electrodes: (a) profile view, (b) top view.

and Sarvas, 1989, Meijs et al., 1989]. We get a Leadfield matrix $\boldsymbol{A} = [\boldsymbol{a}(\rho_1), ..., \boldsymbol{a}(\rho_M)]$ where ρ_m is the position of the barycenter of the triangle m and $\boldsymbol{a}_m$ is the leadfield vector corresponding to the dipole m. The space-time matrix $\boldsymbol{X}$ of the signals at the 32 surface electrodes, defined by the montage illustrated in figure 3.6, is then obtained directly by applying the formula $\boldsymbol{X} = \boldsymbol{AS}$ where $\boldsymbol{S}$ is the intensity matrix.

For each source configuration, 50 realizations were generated. These signals correspond to the noise-free data. In order to generate simulations of noisy EEG, 50 periods of muscle activity were extracted from real scalp EEG (32 electrodes, figure 3.6). These data contain a mixture of background activity and muscle activity; we simply call it muscle activity in the rest of this report. Each muscle activity was normalized to the channel containing the maximum signal strength. Then, different levels of amplitude of muscle activity were added to epileptic EEG signals to obtain the simulated noisy EEGs with different values of Signal to Noise Ratio (SNR). Figure 3.7 shows an example of original noise-free interictal signals and corresponding noisy signals for two different SNRs (-10 dB and -20 dB). Figure 3.8 shows an example of noise-free ictal signals and corresponding noisy signals for two different SNRs (-10 dB and -20 dB).

3.2 Real data

To test the feasibility of denoising algorithms on real data, different methods have been applied to denoise interictal spikes recorded from a 40 years old patient ("patient P") suffering from a drug-resistant partial epilepsy since age of 26. As part of the preoperative evaluation, patient P underwent two sessions of video-EEG recording, a brain MRI and ictal and interictal SPECT examinations. During video-EEG recording, EEG data were acquired from 32 electrodes (standard 10-20 electrodes system and additional electrodes FC1, FC2, FC5, FC6, CP1, CP2, CP5, FT9, FT10, P9 and P10), at a sampling frequency of 256 Hz with a bandpass filter (0.3 - 100 Hz). These data

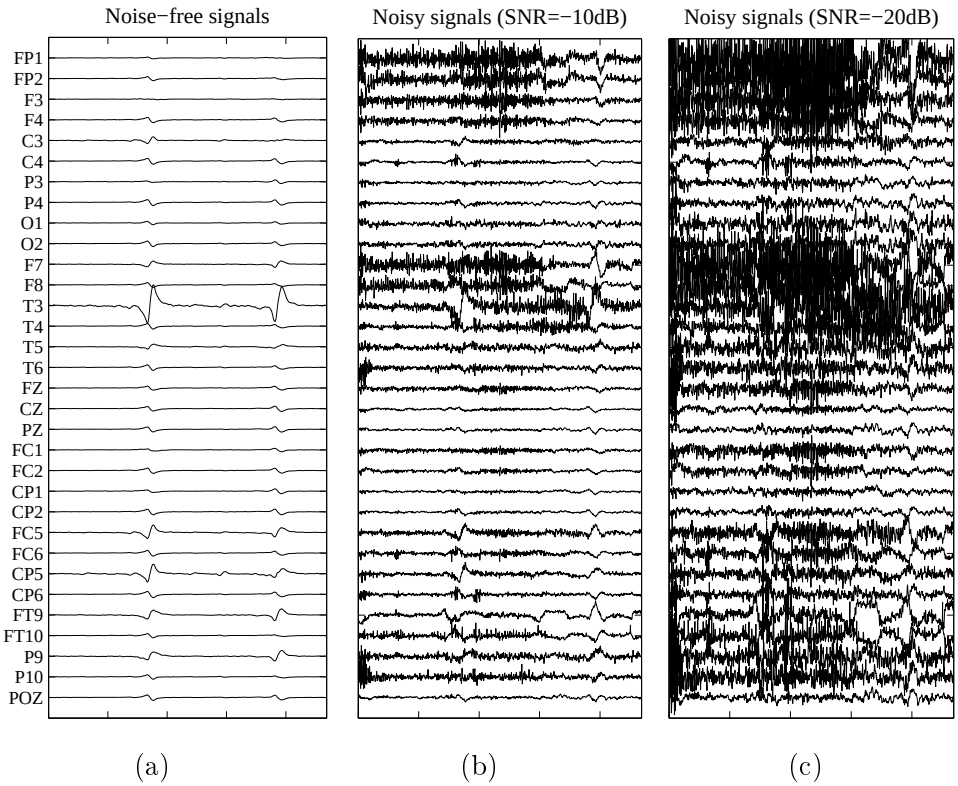


Figure 3.7: An example of simulated interictal signals. (a) An example of a noise-free simulation. These data were generated from a single source (5 cm^2) in the left superior temporal gyrus. The spikes has the maximal amplitude at the electrode T3 (facing the superior temporal region). (b) and (c) The same simulated data after adding a real muscle activity with SNR=-10 dB and SNR = -20dB, respectively.

were examined to isolate different periods of 2048 samples (8s) containing a noise-free period and the periods of 2048 samples containing spikes embedded in muscle activity, at two different qualitative levels of noise (high noisy and low noisy). An example of actual data of interictal spike is illustrated in figure 3.9.

Furthermore, the denoising algorithms have also been tested on real ictal data, recorded in a 23 years old patient ("patient N"). This patient was also recorded on long-term video-EEG (32 electrodes, 256 Hz, PB [0.3-100Hz]) examination during which several seizures were collected. One of these ictal discharges was almost noise-free and could be used as a reference signal. A second discharge was much more noisy and various denoising algorithms were applied on it. The reconstructed signal was then compared to the reference signal. An example of actual data of ictal signals is shown in figure 3.10.

3.3 Performance criteria

To compare the results of the denoising algorithms, we use six measures as follows:

Relative Root Mean Square Error (RRMSE): This measure quantifies the error in

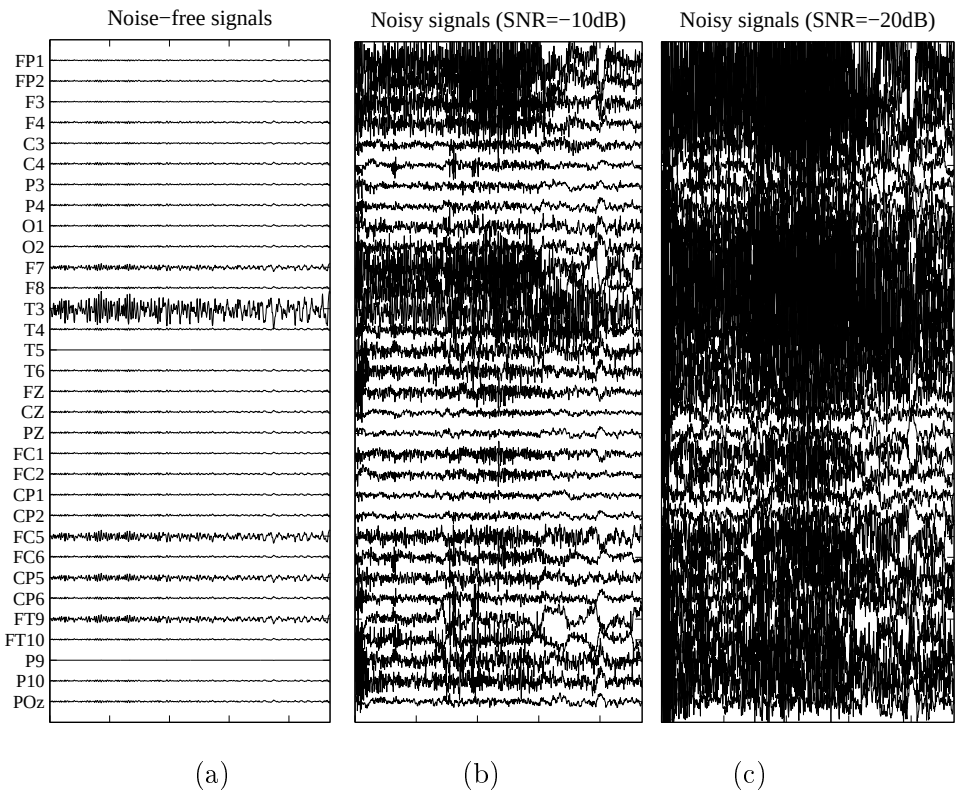


Figure 3.8: An example of simulated ictal signals. (a) An example of a noise-free simulation. These data were generated from a single source (5 cm^2) in the left superior temporal gyrus. The peak value is at the electrode T3 (next to the superior temporal region). (b) and (c) The same simulated data after adding a real muscle activity with SNR=-10 dB and SNR = -20dB, respectively.

separating the signal of interest, the epileptic interictal activity, from the signal of non-interest. The RRMSE is given by:

$$RRMSE = \frac{\sqrt{\sum_{m=1}^{M}\sum_{n=1}^{N}\left(x_m^{(e)}[n] - \hat{x}_m^{(e)}[n]\right)^2}}{\sqrt{\sum_{m=1}^{M}\sum_{n=1}^{N}\left(x_m^{(e)}[n]\right)^2}} \tag{3.1}$$

where $\{x_m^{(e)}[n]\}$ and $\{\hat{x}_m^{(e)}[n]\}$ are the actual and estimated epileptic subspaces on the m-th channel, respectively.

Local RRMSE: This measure quantifies the error in separating the epileptic interictal activity from the signal of non-interest at each electrode. The $\text{RRMSE}_{ElecName}$ at m-th electrode is given by:

$$RRMSE_m = \frac{\sqrt{\sum_{n=1}^{N}\left(x_m^{(e)}[n] - \hat{x}_m^{(e)}[n]\right)^2}}{\sqrt{\sum_{n=1}^{N}\left(x_m^{(e)}[n]\right)^2}} \tag{3.2}$$

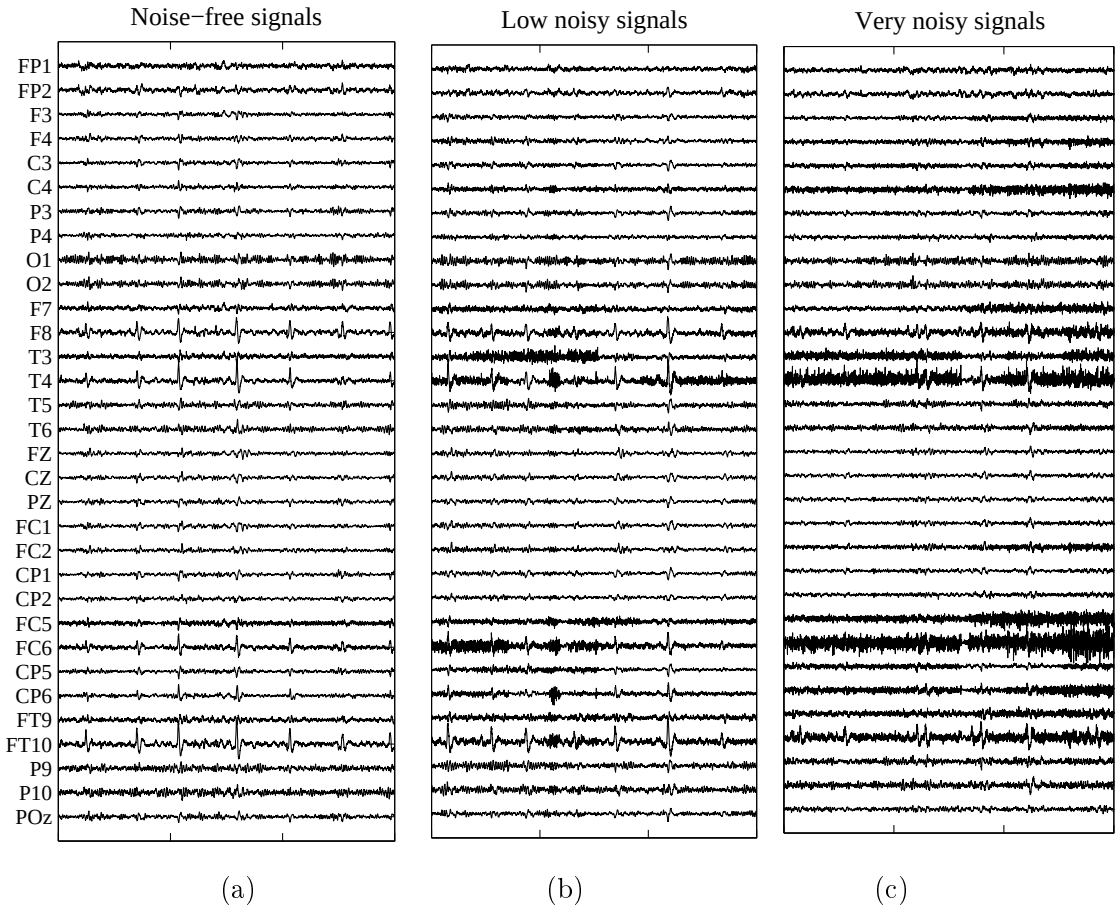


Figure 3.9: Examples of real interictal signals. (a) A typical example of noise-free interictal spikes. (b) An example of interictal spikes affected by muscle activity. (c) An example of interictal spikes greatly affected by muscle activity.

where m is the channel number corresponding to the examined electrode and $\{x_m^{(e)}[n]\}$ and $\{\hat{x}_m^{(e)}[n]\}$ are the actual and estimated epileptic subspaces at this channel, respectively.

Spike RRMSE: This measure quantifies the error in separating the signal of interest, the epileptic interictal activity, from the signal of non-interest, in spike intervals. The Spike RRMSE is given by:

$$RRMSE_{spikes} = \frac{\sqrt{\sum_{m=1}^{M} \sum_{n \in Spikes} \left(x_m^{(e)}[n] - \hat{x}_m^{(e)}[n] \right)^2}}{\sqrt{\sum_{m=1}^{M} \sum_{n \in Spikes}^{N} \left(x_m^{(e)}[n] \right)^2}} \tag{3.3}$$

Local Spike RRMSE: This measure quantifies the error in separating the signal of interest, the epileptic interictal activity, from the signal of non-interest, in spike intervals

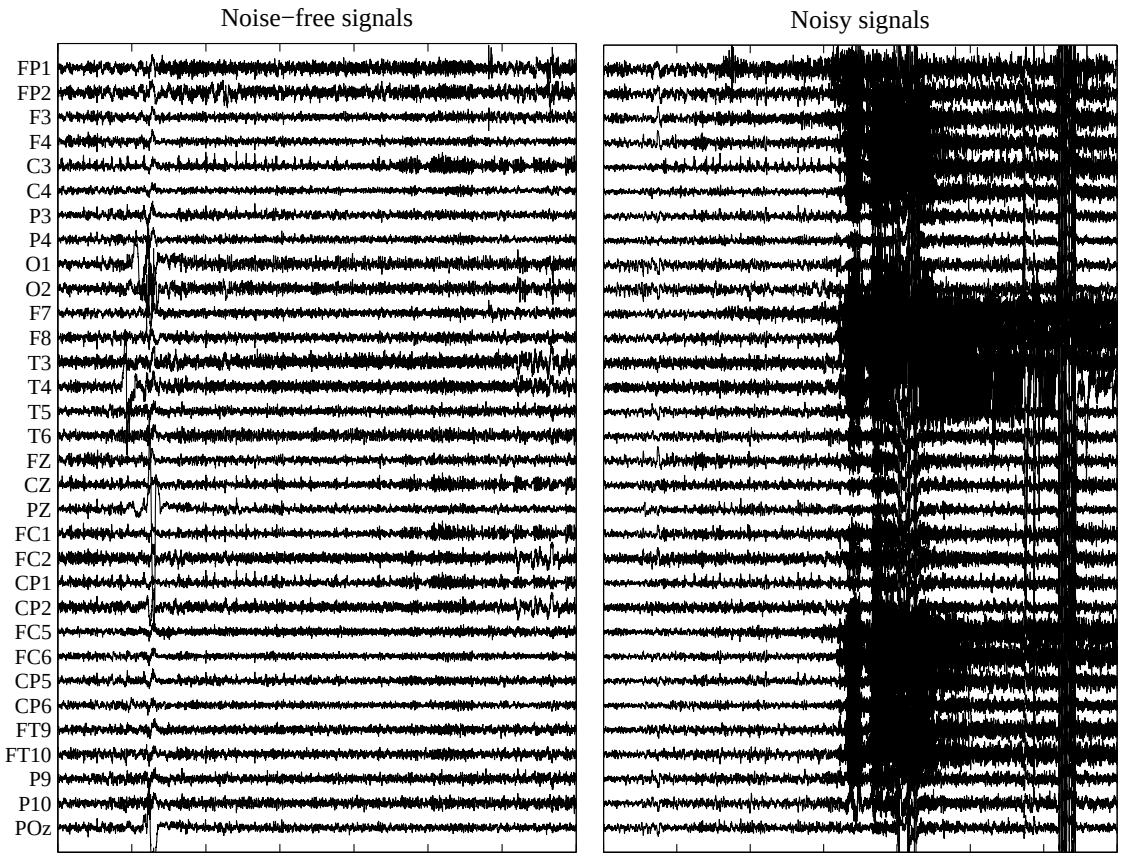


Figure 3.10: A real EEG recorded from an epileptic patient during two separate ictal periods. (a) An almost noise free ictal EEG. (b) Another ictal EEG contaminated by muscle activity.

at each electrode. The $\text{RRMSE}_{spikes,ElecName}$ at m-th electrode is given by:

$$RRMSE_{spikes,ElecName} = \frac{\sqrt{\sum_{n\in Spikes}\left(x_m^{(e)}[n] - \hat{x}_m^{(e)}[n]\right)^2}}{\sqrt{\sum_{n\in Spikes}^{N}\left(x_m^{(e)}[n]\right)^2}} \tag{3.4}$$

Flops: In order to compare the numerical complexity of the denoising algorithms, we calculate the number of flops required by each denoising algorithm. To this end, we use the results illustrated in tables 2.3 and 2.4.

ROC and AUC: We also compared source localization results obtained on noise-free, noisy and denoised data by the 4-ExSo-MUSIC algorithm using the Receiver Operating Characteristic (ROC) curve as a performance criterion. This criterion represents the mathematical expectation of the True Positive Fraction (TPF) as a function of the mathematical expectation of the False Positive Fraction (FPF). The TPF is the fraction between the area of the patch truly retrieved and the total patch area while the FPF is the fraction between the area falsely localized outside the patch and the total cortical

area minus the patch area. The Area Under the Curve (AUC) of each algorithm is also calculated.

3.4 Summary

In this chapter, we presented the simulated and real data used in various analysis. We first introduced the simulated interictal and ictal data. Secondly, the real interictal and ictal data were presented. Then, we introduced the performance criteria. In the next chapter, the performance of the proposed algorithms on both simulated and real data will be studied.

Chapter 4

Experimental Results

In this chapter, the performance of the proposed algorithms presented in chapter 2 is studied in two main parts.

In the first part, we examine the proposed SSD-CP algorithm in some common scenarios in the context of CP decomposition. The first three tests are well-known numerical examples extracted from [Comon et al., 2009]. In the last example of this section, we study the robustness of the proposed algorithm with respect to overfactoring by using amino acids fluorescence data.

In the second part, we study the performance of the proposed methods to denoise both simulated and real epileptic EEG data with interictal or ictal activity contaminated with muscular activity. In the case of simulated data, the effectiveness of the proposed algorithms is evaluated in terms of Relative Root Mean Square Error between the original noise-free signals and the denoised ones, number of required flops and the location of the original and denoised epileptic sources. For both interictal and ictal data, we present some examples on real data recorded in patients with a drug-resistant partial epilepsy. The results of this part are presented in four sections: i) simulated interictal data, ii) simulated ictal data, iii) real interictal data and iv) real ictal data.

4.1 Common examples for testing the proposed SSD-CP method

In this section, we present the results of applying the SSD-CP algorithm to some well-known examples. To compare these results with those obtained from other algorithms, we define an error criterion as follows. Each algorithm gives, for each loading matrix, a normalized root mean squared estimation error whose median values are computed from 100 Monte-Carlo experiments and denoted by $r^{(i)}$ where $1 \leq i \leq q$ represents the index of the present loading matrix. Our estimation criterion is then obtained by $Error = \frac{1}{q}\sum_{i=1}^{q} r^{(i)}$. The first three examples of this section are extracted from [Comon et al., 2009] to examine the results of the proposed algorithm on some well-known numerical examples. In the last example, we use the Amino acids fluorescence data downloaded from [Ami, 1998] and described in [Ami, 1998] to examine the effect of overfactoring, i.e. overestimating the array rank, on the performance of different CP

methods.

4.1.1 A $(2 \times 2 \times 2)$ array of rank 2

A good example for studying the efficiency of our algorithm is the fully parameterizable $(2 \times 2 \times 2)$ array which was given by Paatero [Paatero, 2000]. The mode-2 unfolding matrix of this array is as follows:

$$\boldsymbol{T} = \left(\begin{array}{cc|cc} 0 & 1 & e & 0 \\ 1 & d & 0 & h \end{array} \right) \tag{4.1}$$

where e is regarded as a constant, typically $e = 30$. The parabola $h = -ed^2/4$ consists of degenerate border points. All points below the parabola have rank 3, while all points above have rank 2.

The CP decomposition of $\boldsymbol{T}$ defined by 4.1 is expressed in the following form:

$$[\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}] = \left[\begin{pmatrix} 1/\epsilon & -1/\epsilon \\ p & v \end{pmatrix}, \begin{pmatrix} 1/\epsilon & -1/\epsilon \\ r & x \end{pmatrix}, \begin{pmatrix} 1/\epsilon & -1/\epsilon \\ t & z \end{pmatrix} \right] \tag{4.2}$$

where parameters ϵ, p, v, r, x, t and z can be calculated as functions of e, d and h.

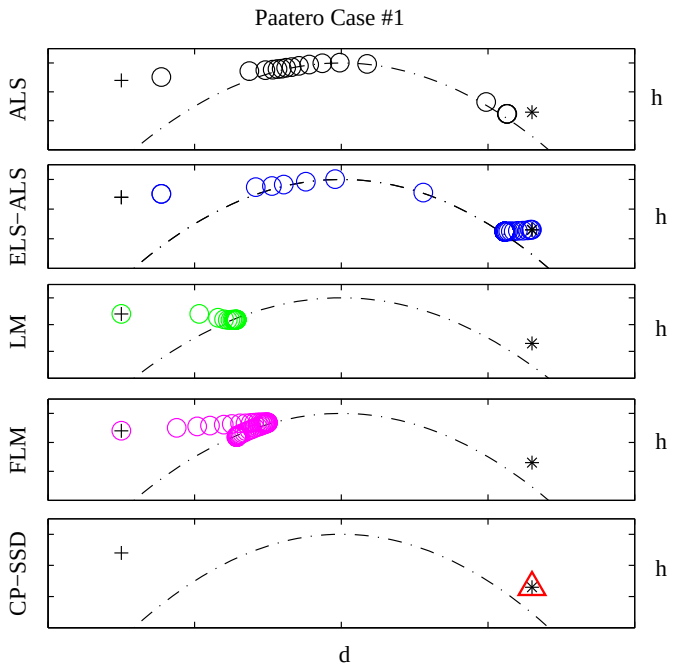
We study two different cases and in each case we use the parameters similar to [Comon et al., 2009]. In these cases, the tensors to be decomposed are chosen such that they have a rank 2, but they lie close to degenerate border points, represented by a parabola in the subsequent figures. In each case we compare the results of our algorithm with the ALS, ELS-ALS, LM and FLM algorithms.

In the first case, the array parameters are considered as $(e, d, h) = (30, 0.26, -0.34)$ and the initial value for non-direct algorithms is $(e, d, h) = (30, -0.3, -0.12)$. The ELS enhancement is executed in each iteration. As shown in figure 4.1, the alternating algorithms ALS and ELS-ALS are very sensitive to the initial value and their convergence speed are very low with the given initial value. With this initial value, the LM and FLM algorithms do not converge. Unlike these non-direct algorithms, the proposed SSD-CP algorithm finds the exact solution in a few flops.

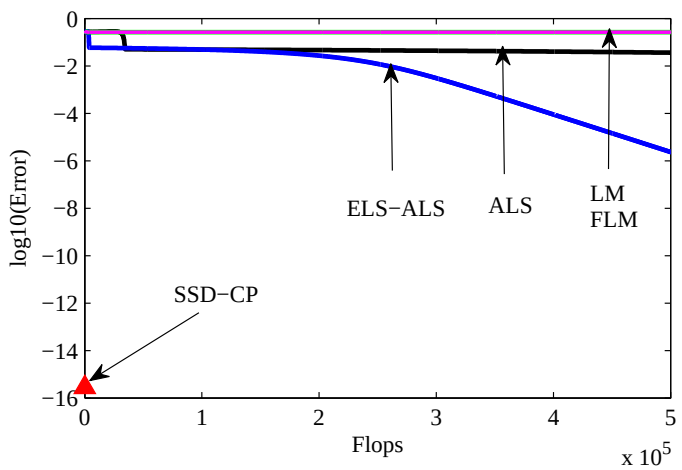
In the second case which is an easier case for non-direct algorithms, the array parameters are given by $(e, d, h) = (30, 0.26, -0.29)$ and the initial value is $(e, d, h) = (30, -0.26, -0.19)$. The results of using the ALS, ELS-ALS, LM, FLM and the proposed SSD-CP algorithms are shown in figure 4.2. In this case, the LM and FLM algorithms converge to the global minimum within a moderate number of iterations. The ALS algorithm works similarly to the previous case but the ELS-ALS converges to the solution in fewer flops. Similar to the first case, the proposed SSD-CP algorithm converges to the exact solution in a few flops.

4.1.2 Double bottlenecks in a $(30 \times 30 \times 30)$ array of rank 4

In this case, an array with two double bottlenecks is considered. This array is generated from three random matrices $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ of size (30×4). To create two bottlenecks in



(a)



(b)

Figure 4.1: Results obtained after applying five algorithms on the first Paatero case. (a) symbols $+$, $*$, $\circ$ and $\triangle$ denote, respectively the initial value, the global minimum, the successive iterations for the non-direct algorithms and the solution for the SSD-CP algorithm, in the (d, h) plane. The parabola shows degenerate border points. (b) Value of the error as a function of flops.

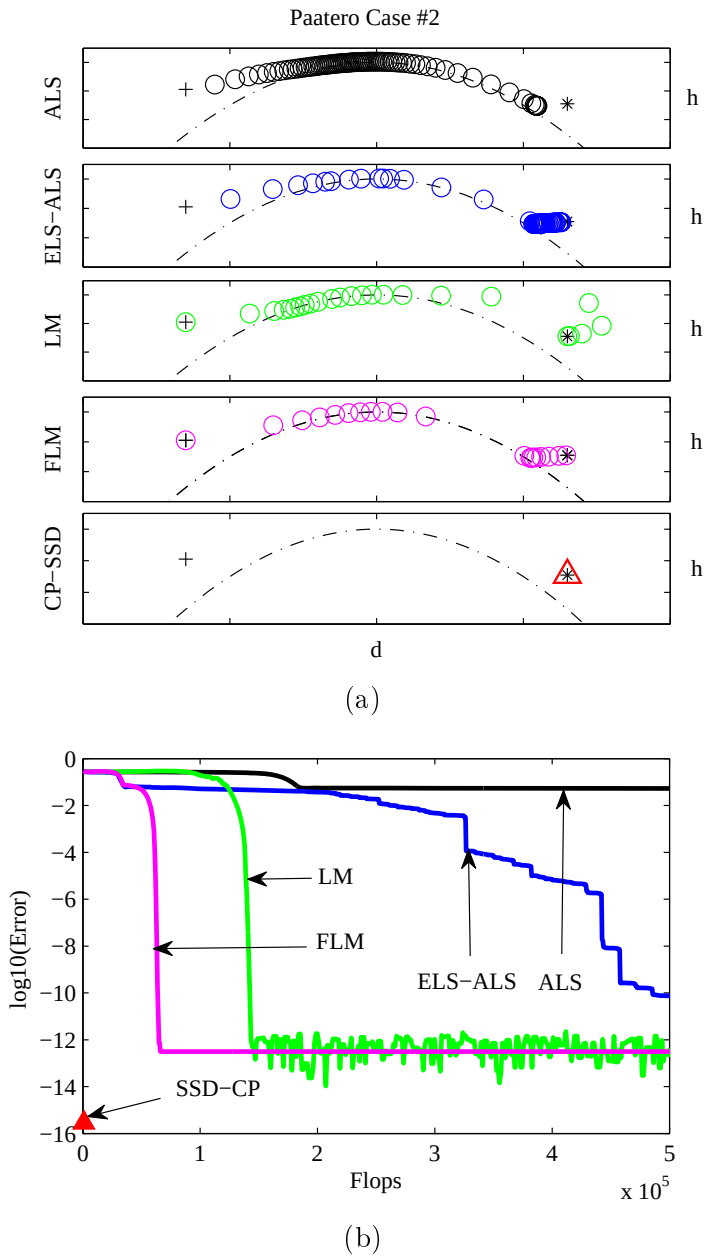


Figure 4.2: Results obtained after applying five algorithms on the second Paatero case. (a) symbols $+$, $*$, $\circ$ and $\triangle$ denote, respectively the initial value, the global minimum, the successive iterations for the non-direct algorithms and the solution for the SSD-CP algorithm, in the (d, h) plane. The parabola shows degenerate border points. (b) Value of the error as a function of flops.

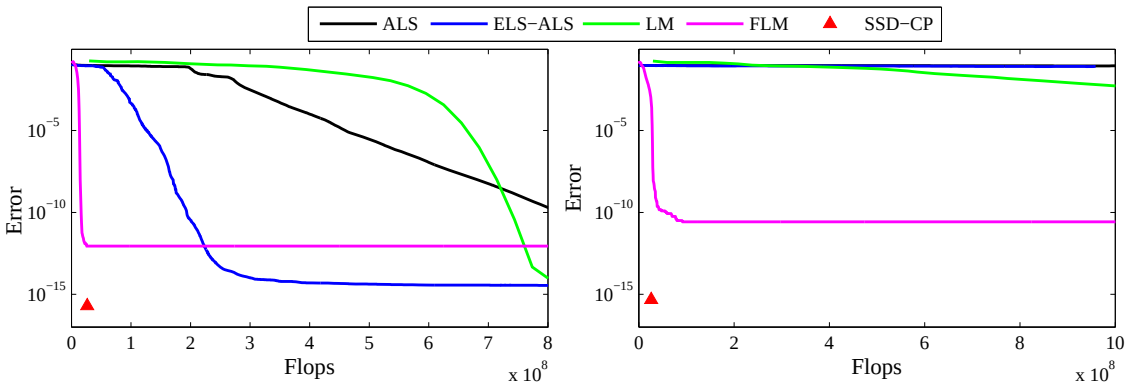


Figure 4.3: Value of error as a function of flops: the results of applying five algorithms on (a) A $30 \times 30 \times 30$ array of rank 4 with double bottlenecks, (b) A $30 \times 30 \times 30$ array of rank 4 with triple bottlenecks.

the two first modes as stated in [Comon et al., 2009], the first column $\boldsymbol{a}_1$ of $\boldsymbol{A}$ is drawn randomly from a zero-mean unit-variance Normal distribution and the second column is set to be $\boldsymbol{a}_2 = \boldsymbol{a}_1 + 0.5\boldsymbol{v}_2$ where $\boldsymbol{v}_2$ is drawn from a zero-mean unit-variance Normal distribution. The third and the fourth columns of $\boldsymbol{A}$ are generated in the same manner, respectively. Matrix $\boldsymbol{B}$ is independently generated similarly to $\boldsymbol{A}$ while matrix $\boldsymbol{C}$ is just drawn from a zero-mean unit-variance Normal distribution without bottleneck. For 100 Monte-Carlo runs, we compare the results of the ALS, ELS-ALS, LM, FLM and SSD-CP algorithms. For each non-direct algorithm, we consider the random initializing approach. In this experiment, the ALS, ELS-ALS, LM algorithms converge for 72, 84, and 98 trials, respectively. The FLM algorithm converges for all 100 trials. For each algorithm, the computed error with respect to number of flops is presented in figure 4.3(a). As shown in this figure, the SSD-CP algorithm outperforms the other methods in terms of performance. In terms of the number of flops, the SSD-CP and FLM algorithms surpass the other methods.

4.1.3 Triple bottlenecks in a $(30 \times 30 \times 30)$ array of rank 4

In this case, an array with two triple bottlenecks is randomly generated. Matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are generated in the same way as in section 4.1.2. But matrix $\boldsymbol{C}$ is generated similarly to $\boldsymbol{A}$ and $\boldsymbol{B}$. 100 Monte-Carlo runs are used to compare the performance of the methods. The ALS, ELS-ALS and LM algorithms converge for 0%, 34% and 51% of trials, respectively. The FLM and SSD-CP algorithms converge in all 100 trials. As shown in figure 4.3(b), the SSD-CP algorithm converges with an average error of 10^{-15} in 2.5×10^7 flops. Among the non-direct algorithms only FLM achieve acceptable errors in comparable number of flops. The other algorithms can not achieve good results in 10^9 flops.

4.1.4 Multiway analysis of amino acids fluorescence data

In this section, we examine the effect of overfactoring, i.e., overestimating the array rank, on the performance of different CP methods. To this end, we use the Amino acids fluorescence data downloaded from [Ami, 1998] and described in [Ami, 1998], [Bro, 1998], [Kiers, 1998]. This data set comprises five simple laboratory-made samples. Each sample contains different amounts of tyrosine, tryptophan and phenylalanine dissolved in phosphate buffered water. The samples were measured by fluorescence (excitation $250-300$ nm, emission $250-450$ nm, 1 nm intervals) on a PE LS50B spectrofluorometer with excitation slit-width of 2.5 nm, an emission slit-width of 10 nm and a scan-speed of 1500 nm/s [Ami, 1998]. Therefore, the array to be decomposed is of size $(5 \times 61 \times 201)$. Since each individual amino acid gives a rank-1 contribution to the data, we can ideally describe these data with three canonical factors [Ami, 1998]. We applied three classical CP methods, namely ALS [Harshman and Lundy, 1994], ELS-ALS [Comon et al., 2009], FLM [Phan et al., 2013c] and the SSD-CP method proposed in this paper, to canonically decompose the amino acids fluorescence data. Then we compared the emission-mode factors estimated by the four methods by varying the rank P of the CP decomposition from 3 to 5 as shown in figure 4.4.

For each algorithm, we scaled the p-th column of the emission-mode loading matrix by the norm of the rank-1 three-way array produced by the outer product of p-th columns of each of the three loading matrices. Note that, regarding the three non-direct methods, they are run several times with different random initial guess in order to use the best initialization.

As displayed in figure 4.4(a), for $P = 3$, all CP algorithms generate almost the same emission factors corresponding to three amino acids. As shown in figures 4.4(b-c), in the case of $P = 4$ and $P = 5$, the factors obtained from the ALS and ELS-ALS change, however the proposed SSD-CP technique and the FLM algorithm generate three factors that match with the true factors while the forth and the fifth estimated factors are very small. These results show that SSD-CP and FLM are robust with respect to overfactoring contrary to the other methods.

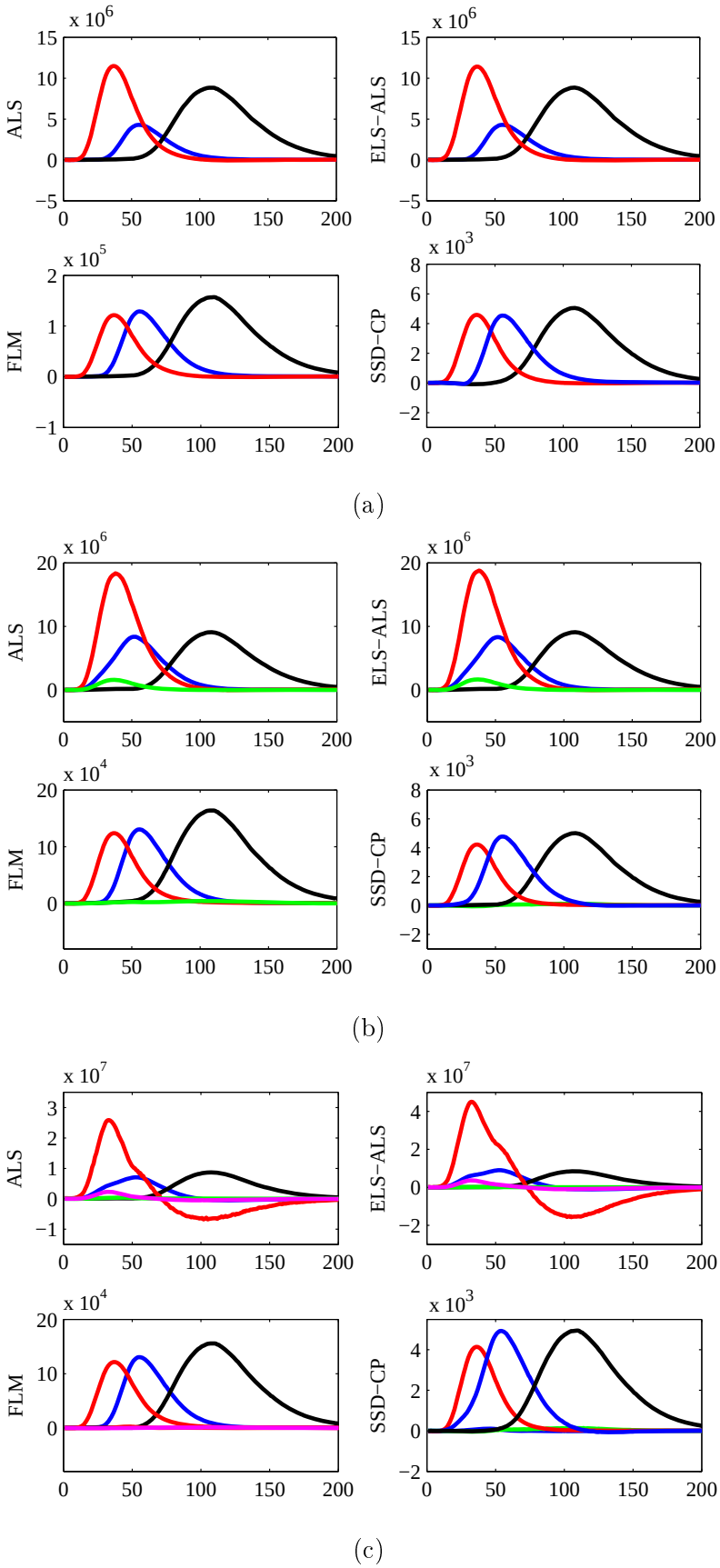


Figure 4.4: The emission mode factors with the estimated number of components equal to (a) $P = 3$, (b) $P = 4$ and (c) $P = 5$.

4.2 EEG simulated data

In this section, we study the performance and numerical complexity of the proposed methods in the case of interictal and ictal EEGs. For both interictal and ictal data, we examine the algorithms in three different scenarios: i) scenario 1: one epileptic source, ii) scenario 2: two epileptic sources with uncorrelated activities and iii) scenario 3: two epileptic sources with correlated activities. The location of epileptic sources and the procedure of generating epileptic signals in each scenario are explained in details in chapter 3.

4.2.1 Interictal data

In this section, in addition to the proposed methods, we study the efficiency of two well-known methods, namely CoM_2 [Comon, 1994] and SOBI [Belouchrani et al., 1997], on simulated interictal data. The reason for selecting these algorithms is that the CoM_2 algorithm offers the best compromise between performance and numerical complexity as shown in [Albera et al., 2012] and SOBI is one of the most common DCCA algorithms used in biomedical applications. In the following tests, the same procedure as explained for the JDICA algorithm (see section 2.2.2) is used by CoM_2 and SOBI to choose interictal sources from all extracted ones. In each scenario, the number of channels is 32, the number of samples is 10240 (which corresponds to a signal of length 40 seconds with sampling frequency of 256 Hz) and the SNR values change from $-20dB$ to $5dB$ with a step of $5dB$. It is noteworthy that for the ICA and DCCA algorithms, we changed the dimension of the source space P' in the range of $\{4, 8, 12, 16, 20, 24, 28, 32\}$ and, in each trial of each scenario, we generated the results by choosing the value of P' yielding the best result.

4.2.1.1 Scenario 1: one epileptic patch

In the first interictal scenario, we consider one patch in the left superior temporal gyrus. We study the effect of changing SNR on the performance and numerical complexity of each algorithm. Figure 4.5 shows an example of simulated noise-free, noisy and denoised data of the first scenario with SNR= $-10dB$. Electrode T3 shows clear spike-like activity on original simulated data. This activity is partially disturbed in the simulated noisy data, on which large muscular activity is seen at T3, T4, F7, F8, FC5, FC6, FT9 and P9 electrodes. For this example, a visual analysis of denoised data shows that the spike activity at T3 is well retrieved with all six algorithms. However, in the final samples of the signal at T3, a distortion is still present in almost all denoised signals, but appears smaller in the GEVD and DSS denoised signals. The other channels were acceptably denoised by all six algorithms. Regarding the source localization results (bottom of each column of figure 4.5), the epileptic source is correctly localized in the left superior temporal region for the noise-free data. For the noisy data, the epileptic source is incorrectly localized in the left inferior temporal region. The interictal data denoised by GEVD, DSS and SSD-CP (figure 4.5(c-d,f)) are localized as expected in the left superior temporal region. The interictal data denoised by JDICA, CoM_2 and SOBI (figure 4.5(e,g,h)) are mislocalized in the inferior temporal gyrus.

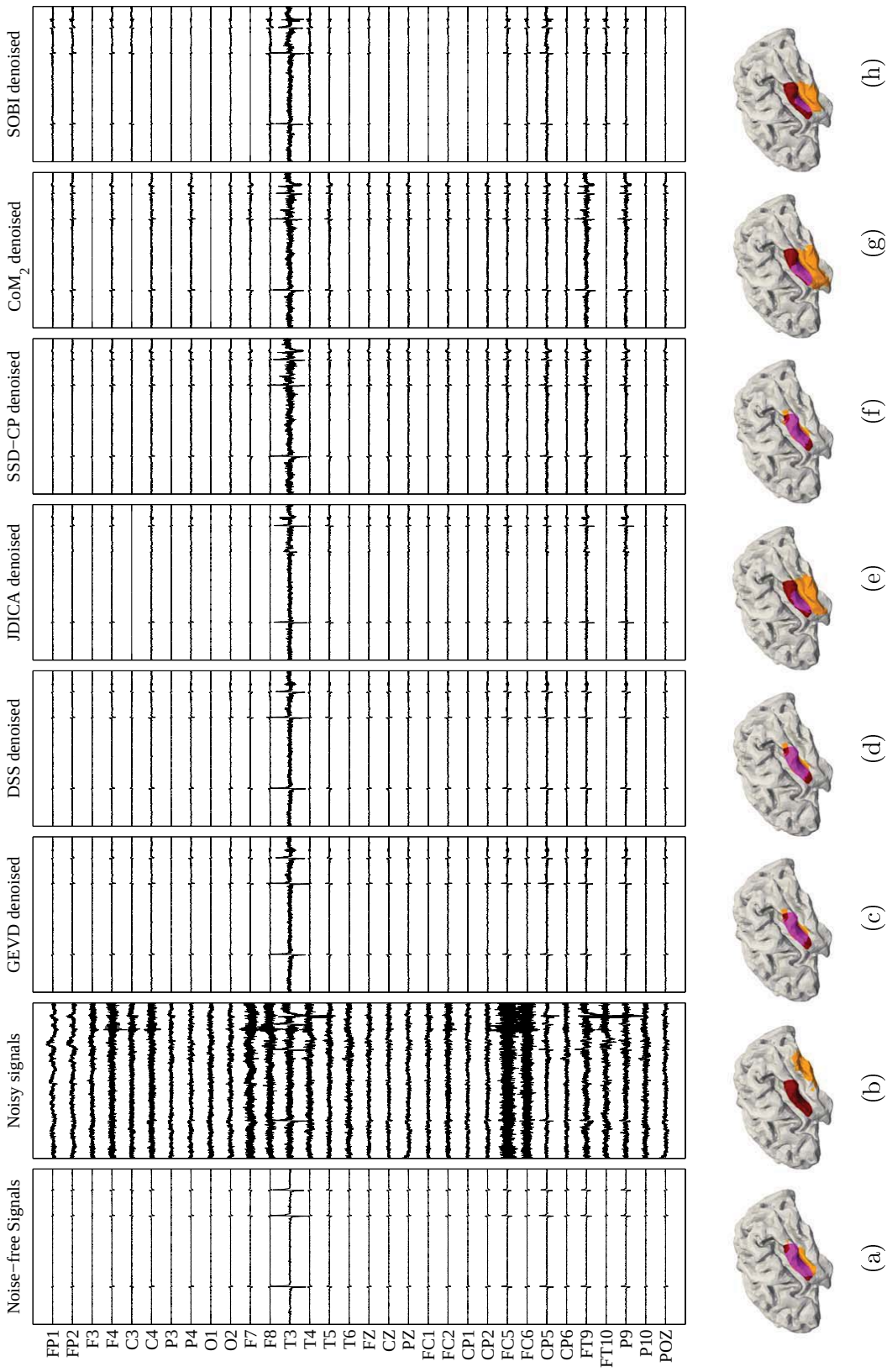


Figure 4.5: An example (trial ♯4) of denoising procedure in the case of simulated interictal data of scenario 1 generated from the activation of a single patch located in the superior temporal gyrus: (a) noise-free simulated EEG with interictal spike-like activity, (b) noisy EEG after adding real muscle activity and background EEG with SNR$= -10dB$, (c-h) EEG denoised by the GEVD, DSS, JDICA, SSD-CP, CoM_2 and SOBI algorithms, respectively. At the bottom of each column, the source localization results obtained from 4-ExSo-MUSIC are illustrated for this trial. Dark brown: real patch; violet: correctly estimated part of the patch; orange: incorrectly estimated part of the patch.

Figure 4.6 shows the results obtained with the first interictal scenario by using the six denoising algorithms for different SNRs. Figures 4.6(a) and 4.6(b) show the RRMSE at all electrodes and at channel T3, respectively. As illustrated in these figures, the SSD-CP denoising method does not work as well as the other methods. Among the other methods, the GEVD and DSS methods are better than the others for high SNRs ($-10dB$ and more) both on a global point of view (i.e. when all electrodes are considered) and on a local point of view (i.e. when only T3 electrode is considered). But for lower SNRs ($-20dB$ and $-15dB$), the ICA-based algorithms, JDICA, CoM_2 and SOBI give better results. Figures 4.6(c) and 4.6(d) show the Spike RRMSE at all electrodes and at channel T3, respectively. As shown in these figures, unlike RRMSE, the Spike RRMSE of SSD-CP is similar to or lower than other methods for SNR values higher than $-20dB$. Similarly to RRMSE results, GEVD and DSS are better than the other methods for high SNRs. It shows that GEVD, DSS and SSD-CP denoise signals in the intervals around spikes better than the other methods. In terms of numerical complexity (figure 4.6(e)), the GEVD and DSS algorithms have the lowest number of flops. Among the other methods, JDICA and SSD-CP are the most time-consuming. Note that, for CoM_2 and JDICA, the required flops greatly decreases for higher SNRs. This is due to the fact that less sources are used to optimize the performance of the CoM_2 and JDICA algorithms. Figure 4.6(f) shows the area under ROC curve (AUC) obtained after source localization of signals denoised by each method. As shown in this figure, for high SNRs ($-5dB$ and higher) all methods generate almost similar and high AUCs. For $-10dB$, the AUC after GEVD, DSS and SSD-CP are greater than those obtained after ICA-based denoised signals. For lower SNRs, CoM_2 and JDICA surpass GEVD and DSS, especially for $-20dB$. These AUC results confirm the Spike RRMSE results (4.6(c,d)).

Figure 4.7 shows the source localization ROC curves of the six denoising algorithms for different SNR values. We do not show the results for $0dB$ and $5dB$, since for these SNRs the curves are almost similar to $-5dB$. As illustrated in this figure, all algorithms can acceptably estimate the source location and their ROC curves have an acceptable distance from the curve corresponding to the Noisy EEG, especially for SNRs equal to or higher than $-15dB$.

4.2.1.2 Scenario 2: two uncorrelated epileptic patches

In the second scenario, two patches are considered: the first patch is located in the superior temporal gyrus and the second patch is located in the inferior frontal gyrus. In this scenario, these two patches have uncorrelated dynamics. Figure 4.8 shows an example of simulated noise-free, noisy and denoised data of the interictal scenario 2 with SNR= $-10dB$.

In the original noise-free data (figure 4.8(a)), electrodes T3 and FC5 show the maximum spike-like activity. This activity is completely hidden in the simulated noisy data and the other electrodes are considerably noisy (figure 4.8(b)). As shown in figure 4.8(c-h), the spike activity at T3 and FC5 is retrieved by all algorithms except by the SSD-CP method. Although SSD-CP can reduce noise in some channels, it can not succeed in denoising signals where the maximal activity is present. In addition, the amplitude of spikes is reduced in channel FC5 after denoising with SOBI. Regarding

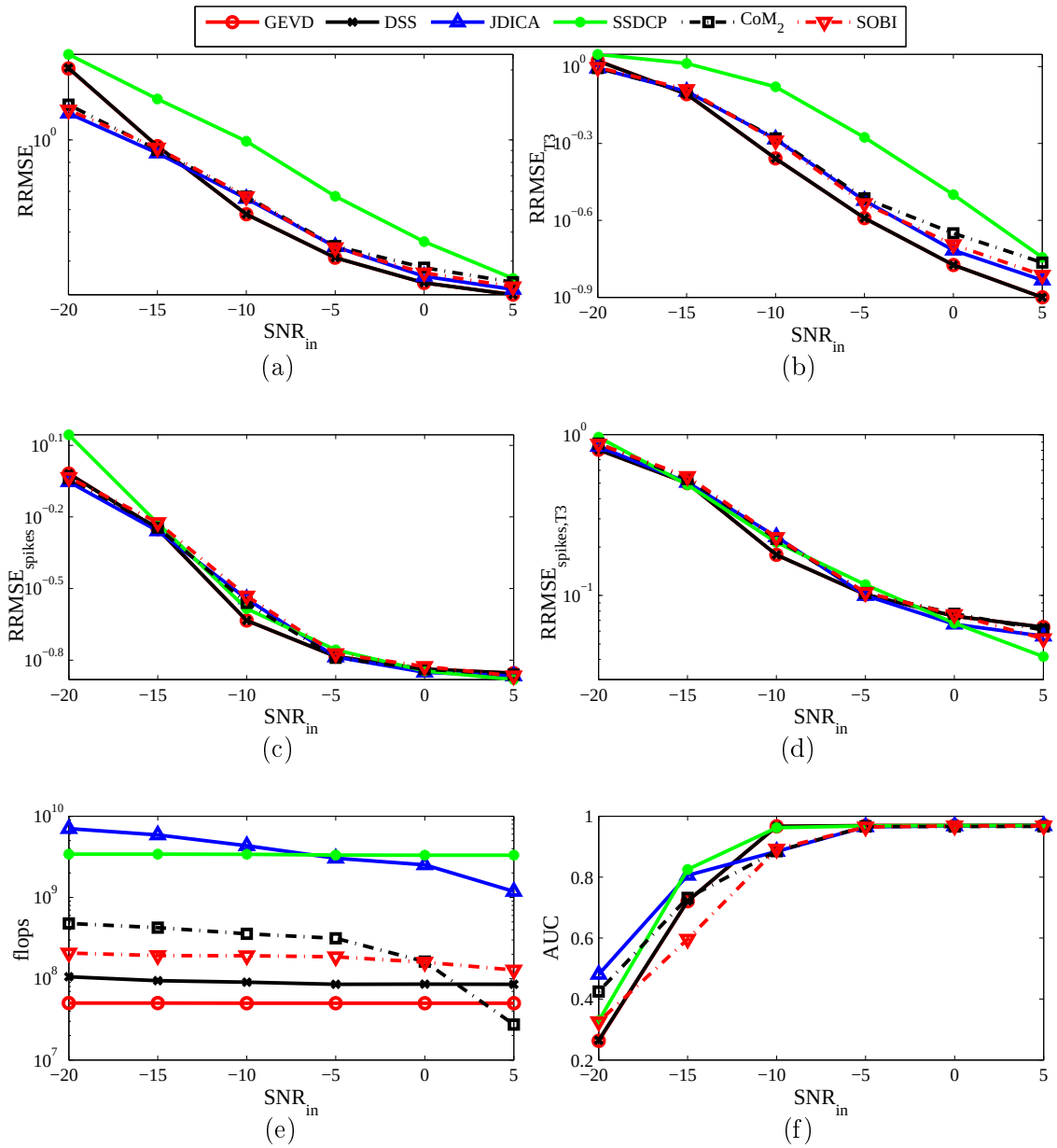


Figure 4.6: Denoising results of the six methods in the case of interictal scenario 1 for different SNR values between -20 and $5dB$: (a) the RRMSE denoising error, (b) the local RRMSE denoising error at channel T3, (c) the RRMSE denoising error in spike intervals, (b) the local RRMSE denoising error at channel T3 in spike intervals, (e) number of flops and (f) AUC of the source localization ROC curve.

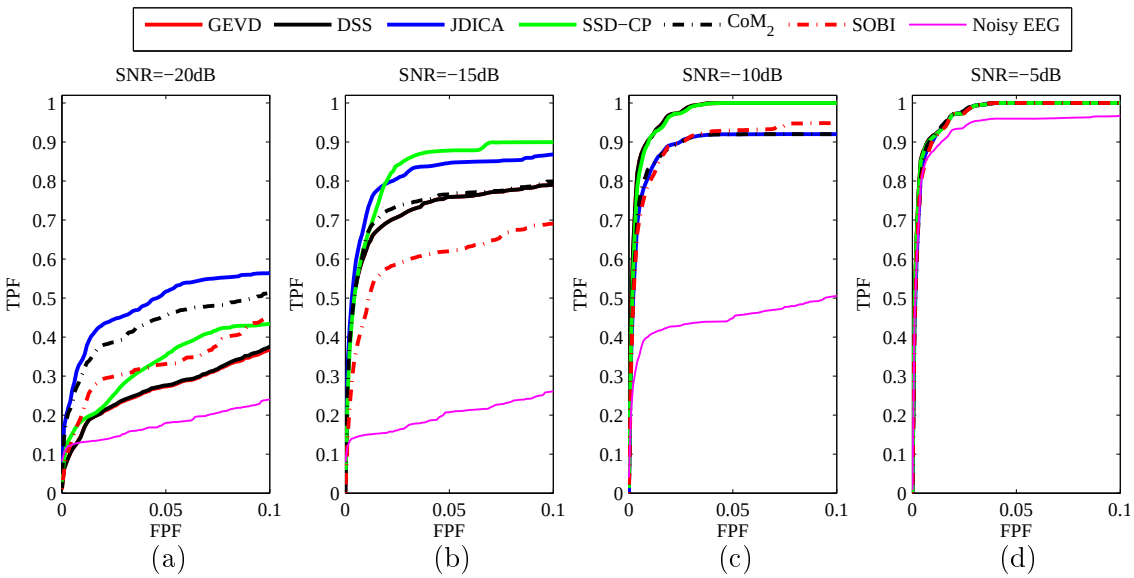


Figure 4.7: The source localization ROC curves of six denoising algorithms in the case of the interictal scenario 1 for different SNR values of (a) $-20dB$, (b) $-15dB$, (c) $-10dB$ and (d) $-5dB$

the source localization results (bottom of each column of figure 4.8), the two epileptic sources are correctly localized in the left superior temporal and left inferior frontal regions for the noise-free data. For the noisy data, only one epileptic source is localized in the left inferior region. The interictal data denoised by all methods are localized close to the expected regions, i.e. in the superior temporal and inferior frontal gyri. For the interictal data denoised by SSD-CP and CoM_2, the source localization results are not as accurate as those of the other methods.

Figure 4.9 shows the results obtained from the second interictal scenario by using the six denoising algorithms for different SNRs. Figures 4.9(a), (b) and (c) show the RRMSE at all electrodes, at channels T3 and FC5, respectively. Similar to the previous scenario, the SSD-CP method does not denoise as well as the other methods. On a global point of view (figure 4.9(a)), all methods (except SSD-CP) perform almost the same for SNRs equal to or higher than $-10dB$. But for lower SNRs, the ICA-based algorithms surpass the other methods. The results at the electrodes T3 and FC5 show that, for SNRs higher than $-15dB$, the GEVD and DSS methods denoise better the signal at electrode T3 than the other algorithms, whereas ICA-based methods denoise better the signal at channel FC5 than the other methods. For SNR$=-20dB$, the ICA-based methods are better than GEVD and DSS for these two electrodes. Figures 4.9(d), (e) and (f) show the Spike RRMSE at all electrodes, at channels T3 and FC5, respectively. Although the results of SSD-CP in terms of RRMSE and local RRMSE are not suitable, it has acceptable results in terms of Spike RRMSE and local Spike RRMSE, especially for SNR values equal to or higher than $-10dB$. In terms of numerical complexity (figure 4.9(g)), results are similar to the first scenario and the GEVD, DSS, SOBI, CoM_2, SSD-CP and JDICA denoising methods require from the least to greatest number of flops, respectively.

Figure 4.9(h) shows the AUC obtained after source localization of signals denoised by each method. As shown in this figure, for the SNRs of $0dB$ and $5dB$, similar AUCs are obtained. For $-5dB$ and $-10dB$, the GEVD and DSS denoised signals give the best AUCs, but for $-20dB$ CoM_2-denoised signals are better localized than the other ones. In average for all SNRs, SOBI gives the lowest AUCs.

Figure 4.10 show the source localization ROC curves obtained after denoising interictal signals with the six algorithms for different SNR values. According to these results, for SNR values equal to or higher than $0dB$, signals denoised with all algorithms were accurately localized. For $-15dB$, $-10dB$ and $-5dB$, acceptable results were obtained by all algorithms. For SNR value of $-20dB$, none of the denoised signals were accurately localized.

4.2.1.3 Scenario 3: two correlated epileptic patches

In the third interictal scenario, as in the second one, two patches are considered, the first patch is located in the superior temporal gyrus and the second patch is in the inferior frontal gyrus. However, in this scenario, these two patches have correlated activity. Figure 4.11 shows an example of simulated noise-free, noisy and denoised data of the interictal scenario 3 with SNR= $-15dB$.

Similar to the second scenario, in the original noise-free data (figure 4.11(a)), electrodes T3 and FC5 show clear spike-like activity. In the simulated noisy data (figure 4.11(b)), the spike-like activity is completely hidden and a large noise is seen in the other electrodes. By using all the six denoising methods, the spike activity at T3 and FC5 is retrieved (figure 4.11(c-h)). Among the denoising methods, SSD-CP leaves much greater noise on the denoised signals, especially at electrodes T3 and FC5. Regarding the source localization results (bottom of each column of figure 4.11), the two epileptic sources are correctly localized in the left superior temporal and left inferior frontal regions for the noise-free data. For the noisy data, two epileptic sources are localized in the right frontal region. The interictal data denoised by all methods are almost truly localized as far as the first patch is considered (temporal superior region). However the second patch is poorly localized by all the algorithms.

Figure 4.12 shows the results obtained with the third interictal scenario by using the six denoising algorithms for different SNRs. Figures 4.12(a), (b) and (c) show the RRMSE at all electrodes, at channels T3 and FC5, respectively. Similar to the previous scenarios, the SSD-CP denoising method does not genearate suitable results in terms of RRMSE and local RRMSE. On a global point of view (figure 4.12(a)) as well as on local results (figure 4.12(b,c)), GEVD and DSS surpass the ICA-based methods for SNRs equal to or higher than $-15dB$. But for $-20dB$, the ICA-based methods give better results. Figures 4.12(d), (e) and (f) show the spike RRMSE at all electrodes, at channels T3 and at FC5, respectively. As shown in these figures, GEVD and DSS are better than the other algorithms in terms of Spike RRMSE and local Spike RRMSE. The numerical complexity results (figure 4.12(g)) are similar to the previous scenarios.

Figure 4.12(h) and figure 4.13 respectively show the AUC obtained by each denoising method and the source localization ROC curves corresponding to each algorithm. As

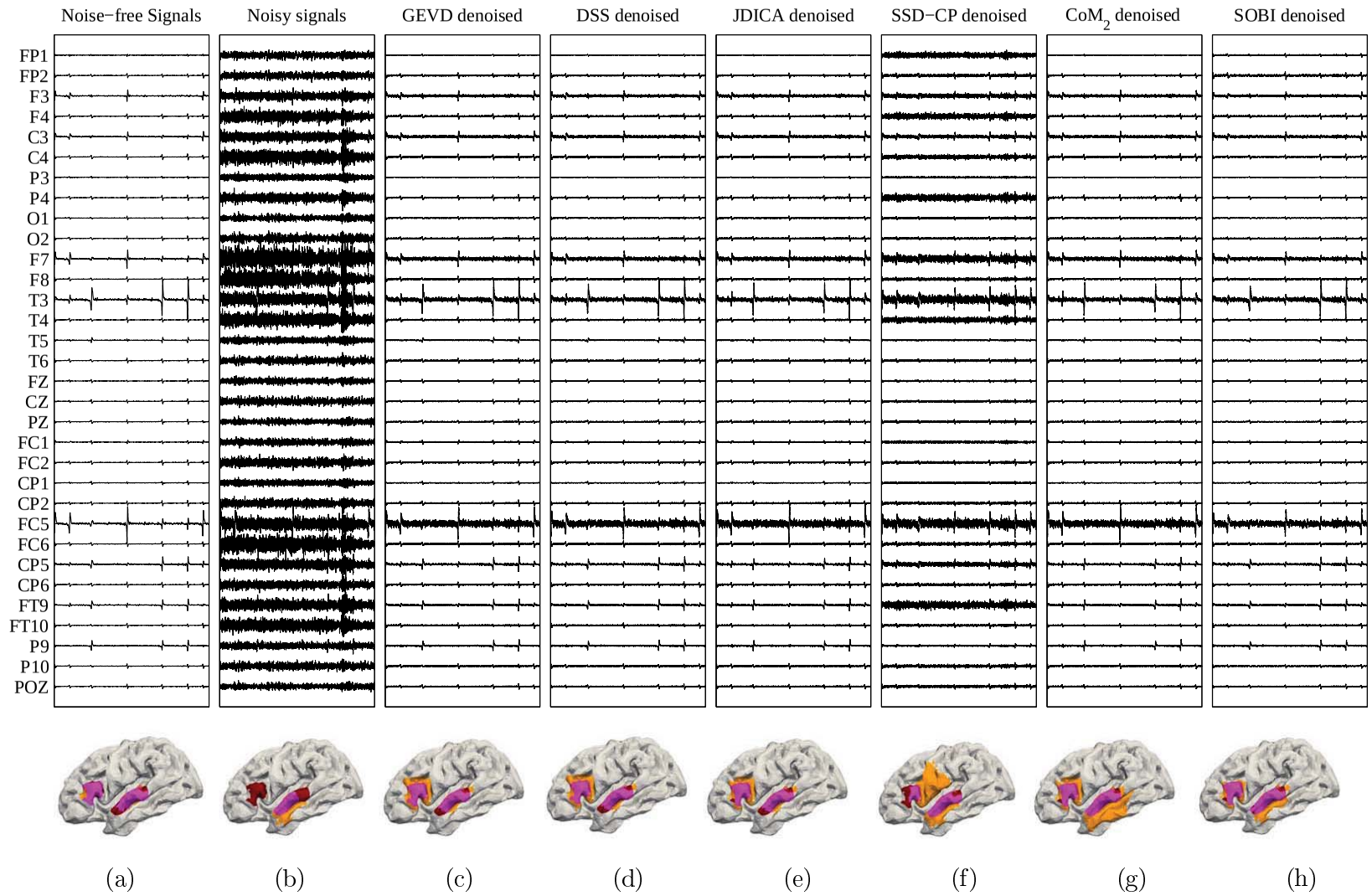


Figure 4.8: An example (trial ♯2) of denoising procedure in the case of simulated interictal data of scenario 2 generated from the activation of two patches with uncorrelated activities: (a) noise-free simulated EEG with interictal spike-like activity, (b) noisy EEG after adding real muscle activity and background EEG with SNR= $-10dB$, (c-h) EEG denoised by the GEVD, DSS, JDICA, SSD-CP, CoM_2 and SOBI methods, respectively. At the bottom of each column, the source localization results obtained from 4-ExSo-MUSIC are illustrated for this trial. Dark brown: real patch; violet: correctly estimated part of the patch; orange: incorrectly estimated part of the patch.

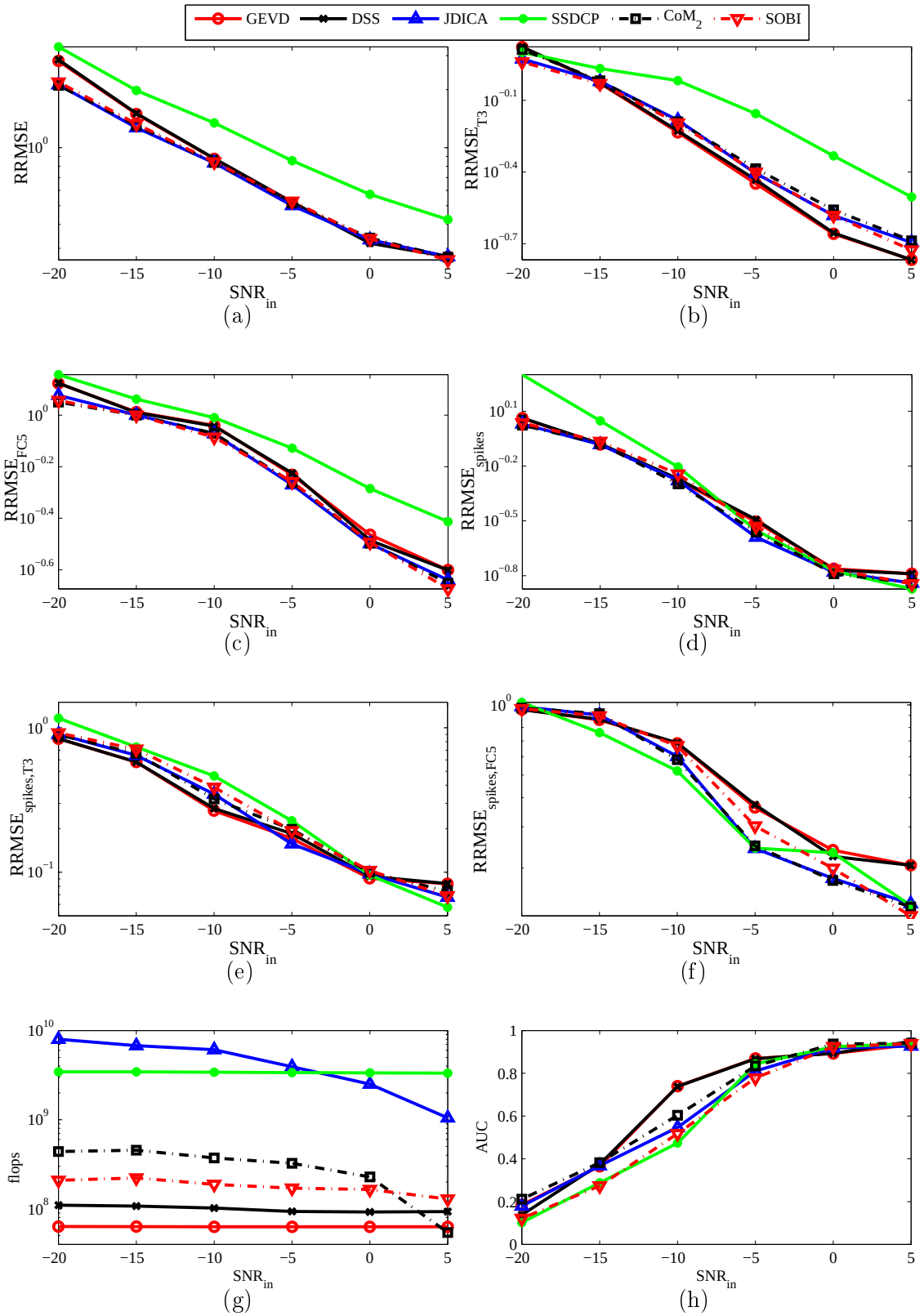


Figure 4.9: Denoising results of the six methods in the case of interictal scenario 2 for different SNR values between -20 and $5dB$: (a) the RRMSE denoising error, (b) the local RRMSE denoising error at channel T3, (c) the local RRMSE denoising error at channel C5, (d) the RRMSE denoising error in spike interval, (e) the local RRMSE denoising error at channel T3 in spike interval, (f) the local RRMSE denoising error at channel C5 in spike interval, (g) number of flops and (h) AUC of the source localization ROC curve.

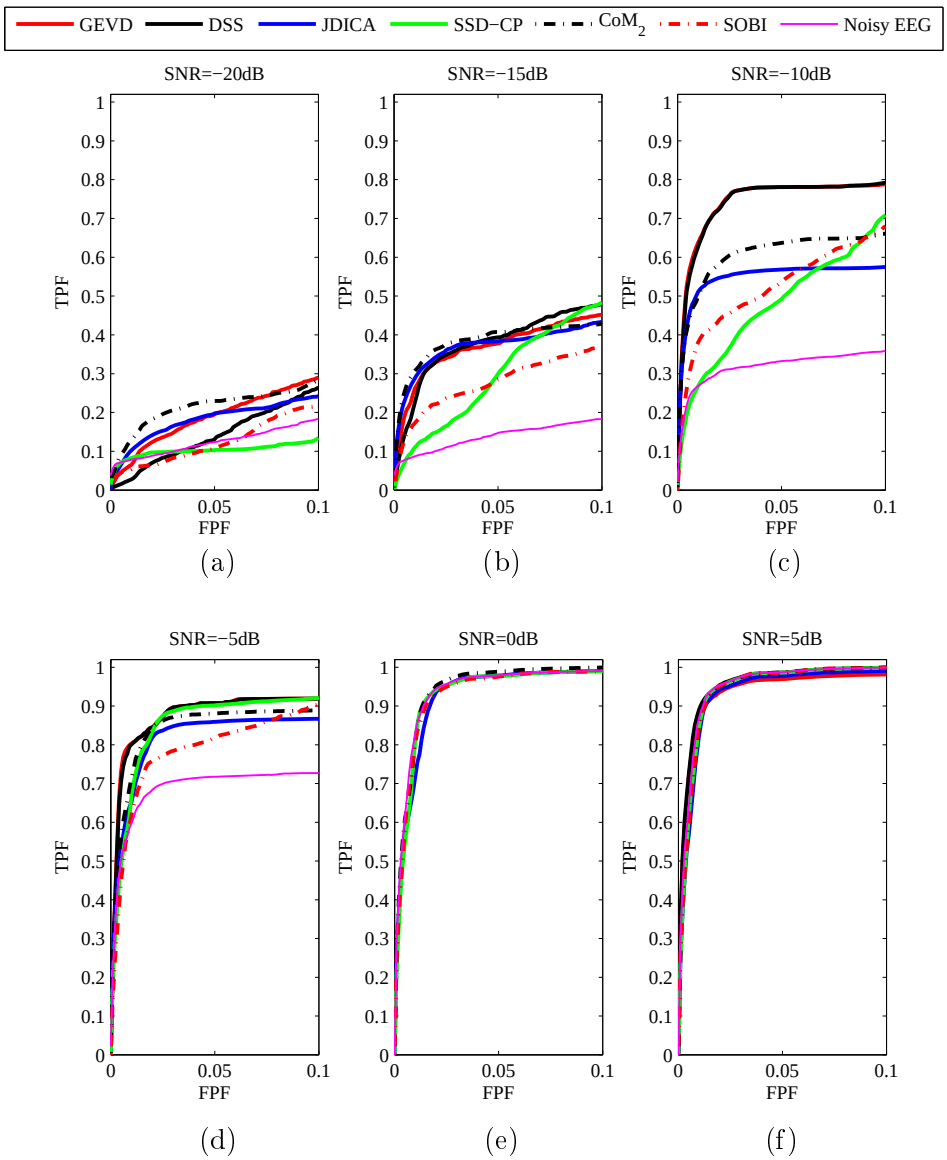


Figure 4.10: The source localization ROC curves of six denoising algorithms in the case of the interictal scenario 2 for different SNR values of (a) $-20dB$, (b) $-15dB$, (c) $-10dB$, (d) $-5dB$, (e) $0dB$ and (f) $5dB$.

shown in these figures, GEVD and DSS methods give the best source localization results for almost all SNRs except for $-20dB$ in which JDICA surpasses the other algorithms. These results confirm the Spike RRMSE and local Spike RRMSE results shown in figures 4.12(d), (e) and (f).

4.2.2 Ictal data

In this section, in addition to the proposed methods, we study the efficiency of the CoM_2 and CCA methods, for the denoising of simulated ictal data. The CCA approach has been shown to be the most efficient method to date for denoising ictal EEG signals contaminated by muscular activity [De Clercq et al., 2006,Hallez et al., 2009,Safieddine, 2012,Karfoul et al., 2015]. In the following tests, for CCA and CoM_2, we use the time-frequency plot of the extracted sources to choose the ictal sources. In each scenario, the number of channels is 32, the number of samples is 10000 and the SNR values change from $-30dB$ to $0dB$ with a step of $5dB$. Similar to the interictal tests, for the ICA algorithms, we change the number of sources P in the range of $\{4, 8, 12, 16, 20, 24, 28, 32\}$ and in each trial of each scenario, we generate the results by choosing the value of P yielding the best result.

4.2.2.1 Scenario 1: one epileptic patch

In the first ictal scenario, one patch in the left superior temporal gyrus is considered. Figure 4.14 shows an example of simulated noise-free, noisy and denoised data of the first scenario with SNR= $-20dB$. Electrode T3 shows a fast ictal activity followed by a slower rhythmic ictal activity on the original simulated data. Lower amplitude ictal activities are also seen in F7, FC5, CP5 and FT9. These activities are completely hidden in the simulated noisy data.

As shown in figure 4.14(c,d,h), TF-GEVD, TF-DSS and CCA can generate denoised signals almost similar to the noise-free signals. Although the JDICA and CoM_2 methods can retrieve a high amplitude activity at electrode T3, there is some differed ictal activity on other channels. The SSD-CP algorithm can not properly denoise the data of this example. Regarding the source localization results (bottom of each column of figure 4.14), the epileptic source is correctly localized in the left superior temporal region for the noise-free data. For the noisy data, the epileptic source is incorrectly localized in the right frontal region. The ictal data denoised by GEVD and DSS (figure 4.14(c,d)) are correctly localized in the left superior temporal region. By applying the source localization method on the CCA denoised data, two epileptic patches are obtained: a true one in the left superior temporal region and a false one in the left anterior temporal region. JDICA and CoM_2-denoised signals are not properly localized (left central region) and the SSD-CP denoised data is also incorrectly localized in the right hemisphere.

Figure 4.15 shows the time course and time-frequency representations of the channel T3 for the noise-free signal, noisy ictal data and signals denoised with the six denoising methods. The ictal fast activity is clearly observable in the time-frequency representation of the noise-free signal(figure 4.15(a)). This time-frequency signature is almost invisible in the time-frequency representation of the noisy data (figure 4.15(b)). Figures

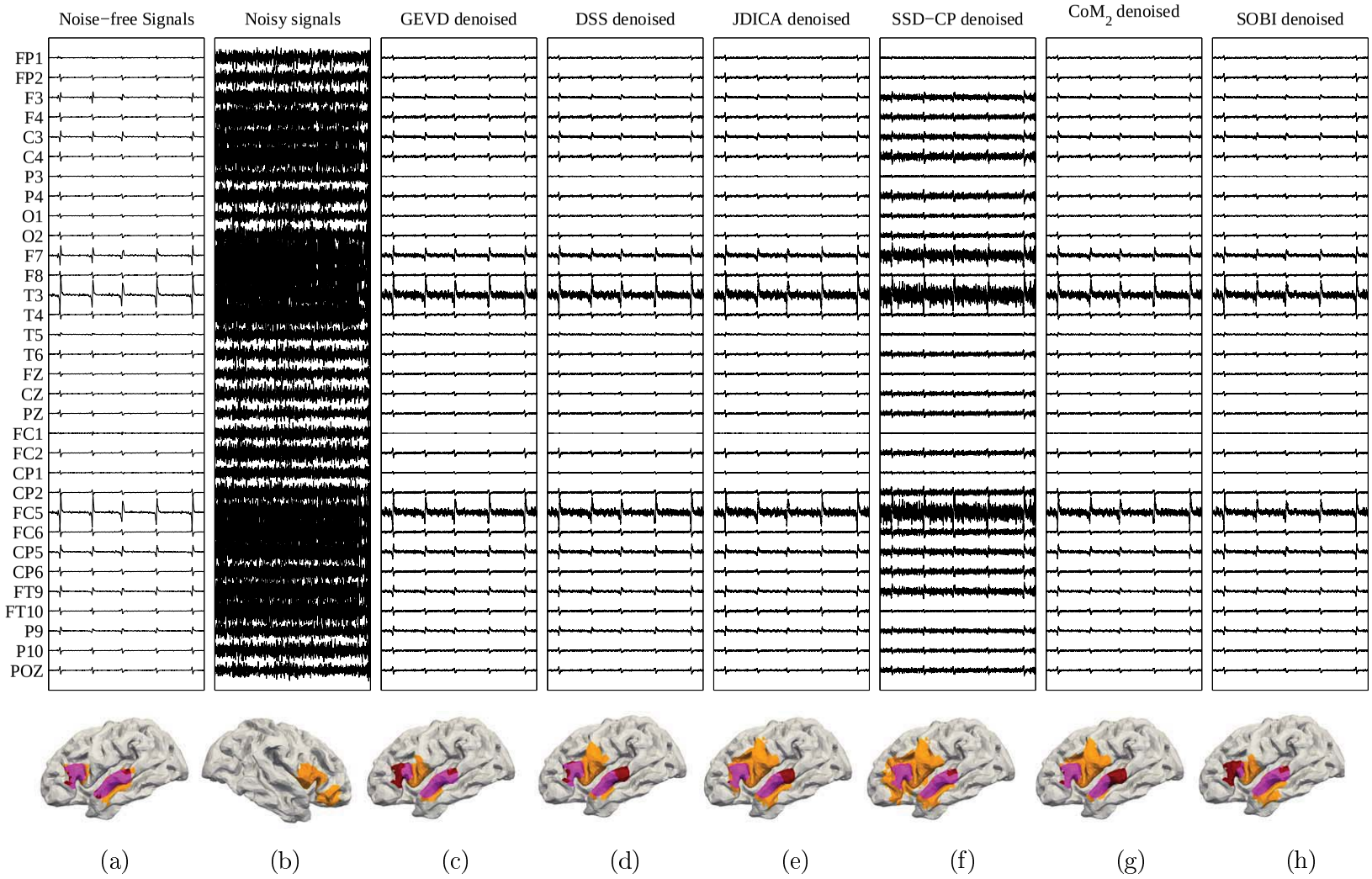


Figure 4.11: An example (trial ♯44) of denoising procedure in the case of simulated interictal data of scenario 3 generated from the activation of two patches with correlated activities: (a) noise-free simulated EEG with interictal spike-like activity, (b) noisy EEG after adding real muscle activity and background EEG with SNR= $-15dB$, (c-h) EEG denoised by the GEVD, DSS, JDICA, SSD-CP, CoM_2 and SOBI methods, respectively. At the bottom of each column, the source localization results obtained from 4-ExSo-MUSIC are illustrated for this trial. Dark brown: real patch; violet: correctly estimated part of the patch; orange: incorrectly estimated part of the patch.

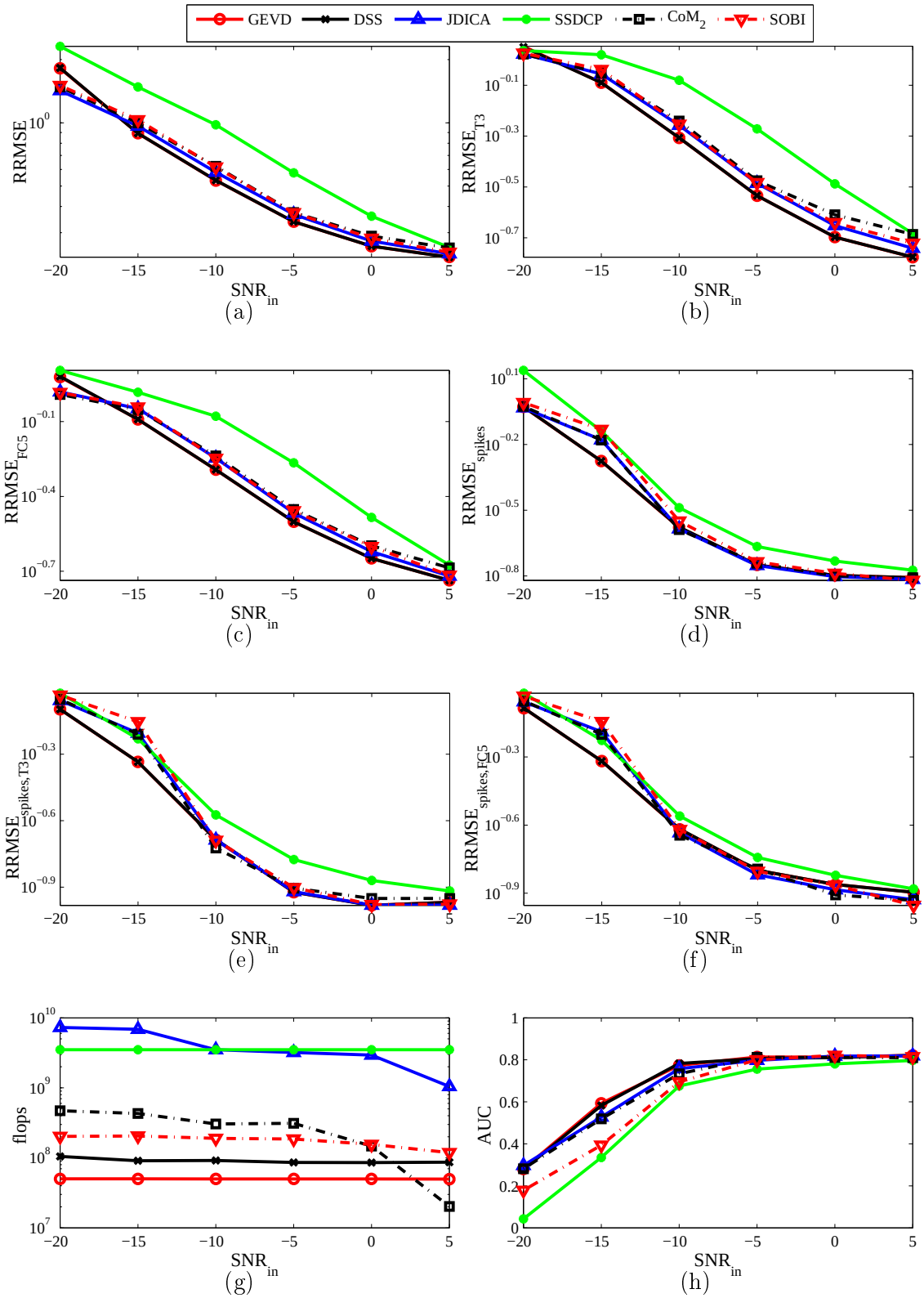


Figure 4.12: Denoising results of the six methods in the case of interictal scenario 3 for different SNR values between -20 and $5dB$: (a) the RRMSE denoising error, (b) the local RRMSE denoising error at channel T3, (c) the local RRMSE denoising error at channel FC5, (d) the RRMSE denoising error in spike interval, (e) the local RRMSE denoising error at channel T3 in spike interval, (f) the local RRMSE denoising error at channel FC5 in spike interval, (g) number of flops and (h) AUC of the source localization ROC curve.

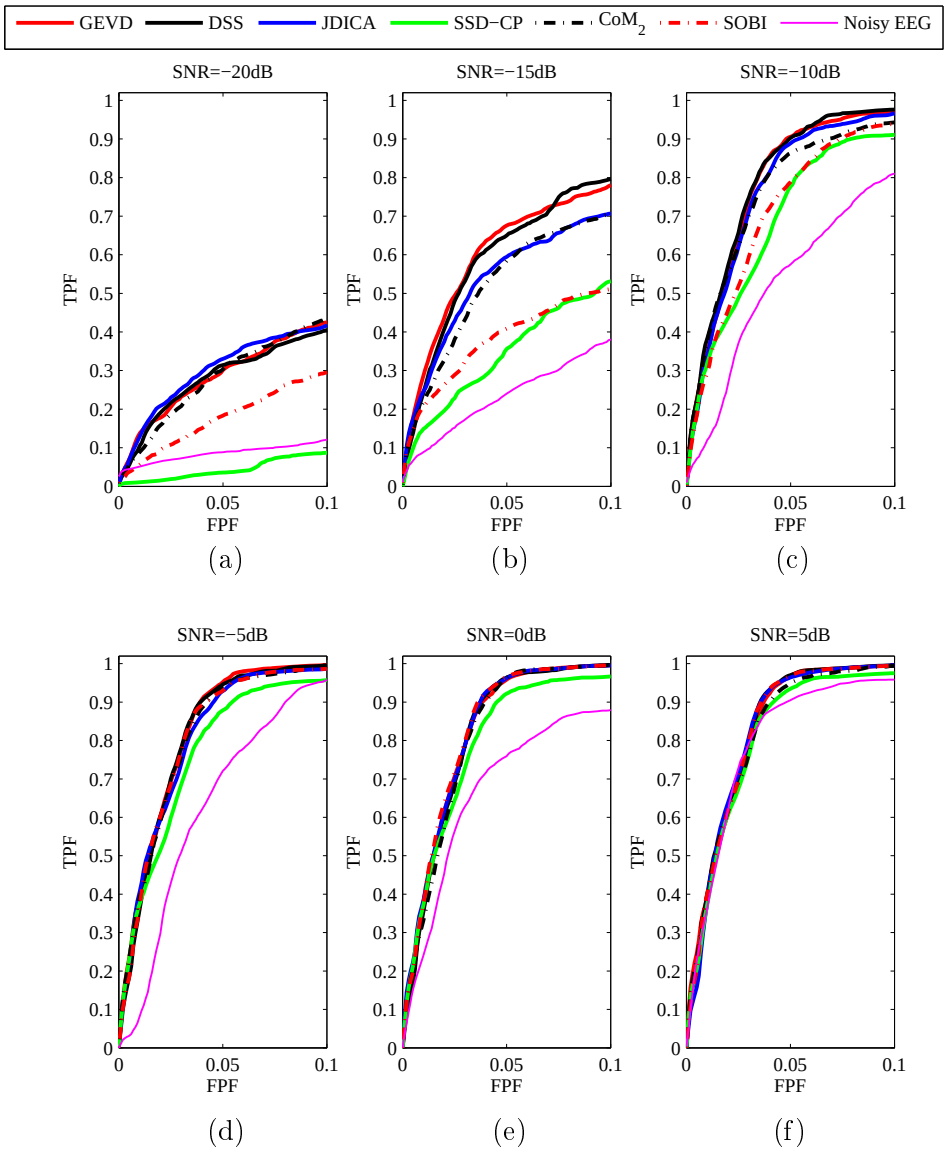


Figure 4.13: The source localization ROC curves of six denoising algorithms in the case of the interictal scenario 3 for different SNR values of (a) $-20dB$, (b) $-15dB$, (c) $-10dB$, (d) $-5dB$, (e) $0dB$ and (f) $5dB$.

4.15(c)-(h) show the time course and time-frequency representations for channel T3 of the ictal signals denoised by the TF-GEVD, TF-DSS, JDICA, SSD-CP, CoM_2 and CCA algorithms, respectively. As shown in these figures, the time-frequency signature of the ictal activity appears in the signals denoised by all methods except by SSD-CP. It is noteworthy that although the time course of signals denoised by JDICA and CoM_2 is not visually similar to that of the noise-free signal, their time-frequency representations show a signature of ictal activity in the beginning of the denoised signals.

Figure 4.16 shows the results obtained from the first ictal scenario by using the six denoising algorithms for different SNRs. Figures 4.16(a) and (b) show the RRMSE at all electrodes and at channel T3, respectively. As illustrated in these figures, the TF-GEVD and TF-DSS methods surpass the other methods for SNR values equal to or lower than $-10dB$. For SNR values of $-5dB$ and $0dB$, JDICA give better results than the other algorithms in terms of RRMSE, but local RRMSE at electrode T3 is not as good as that from TF-GEVD and TF-DSS methods. Among the denoising algorithms, SSD-CP gives the worst results almost for all SNRs.

In terms of numerical complexity (figure 4.16(c)), CCA and CoM_2 have the lowest number of flops and the TF-GEVD and TF-DSS algorithms have low number of flops. The SSD-CP and JDICA methods have the highest number of flops. Similar to interictal scenarios, for CoM_2 and JDICA, the required flops greatly decrease for higher SNRs. This is due to the fact that less sources are used to optimize the performance of the CoM_2 and JDICA algorithms.

Figure 4.16(d) and figure 4.17 show the AUC and the source localization ROC curves obtained by each denoising method, respectively. As shown in these figures, for high SNRs ($-10dB$ and higher) all methods give high AUCs except in the case of $-10dB$ with SSD-CP. For SNR values lower than $-15dB$, TF-GEVD and TF-DSS methods give the best source localization results. After these two methods, the CCA method precedes the ICA-based methods and SSD-CP in source localization of the first ictal scenario.

4.2.2.2 Scenario 2: two uncorrelated epileptic patches

In the second ictal scenario, two patches are considered, first patch in the left superior temporal gyrus and the second one in the left inferior frontal gyrus with uncorrelated ictal activities. Figure 4.18 shows an example of simulated noise-free, noisy and denoised data of the second scenario with SNR= $-20dB$.

As shown in 4.18(a), in the original noise-free data, electrodes T3 and F7 show high amplitude ictal activities. Lower amplitude ictal activities are also seen in electrodes FC5, CP5 and FT9. In the simulated noisy data, the signal of all channels is contaminated with high amplitude muscular noise (figure 4.18(b)). A visual inspection of figure 4.18(c-h) shows that the TF-GEVD and TF-DSS denoised signals are more similar to the noise-free data than the signals denoised by other methods. Signals are well denoised but the amplitude of ictal activities at T3 and F7 is smaller than that of noise-free data. The CCA method can somewhat denoise data of channels T3 and F7, but some small activities of non-interest are seen in the other electrodes such as C4. The JDICA and CoM_2 algorithms show high amplitude ictal activities at electrodes F7

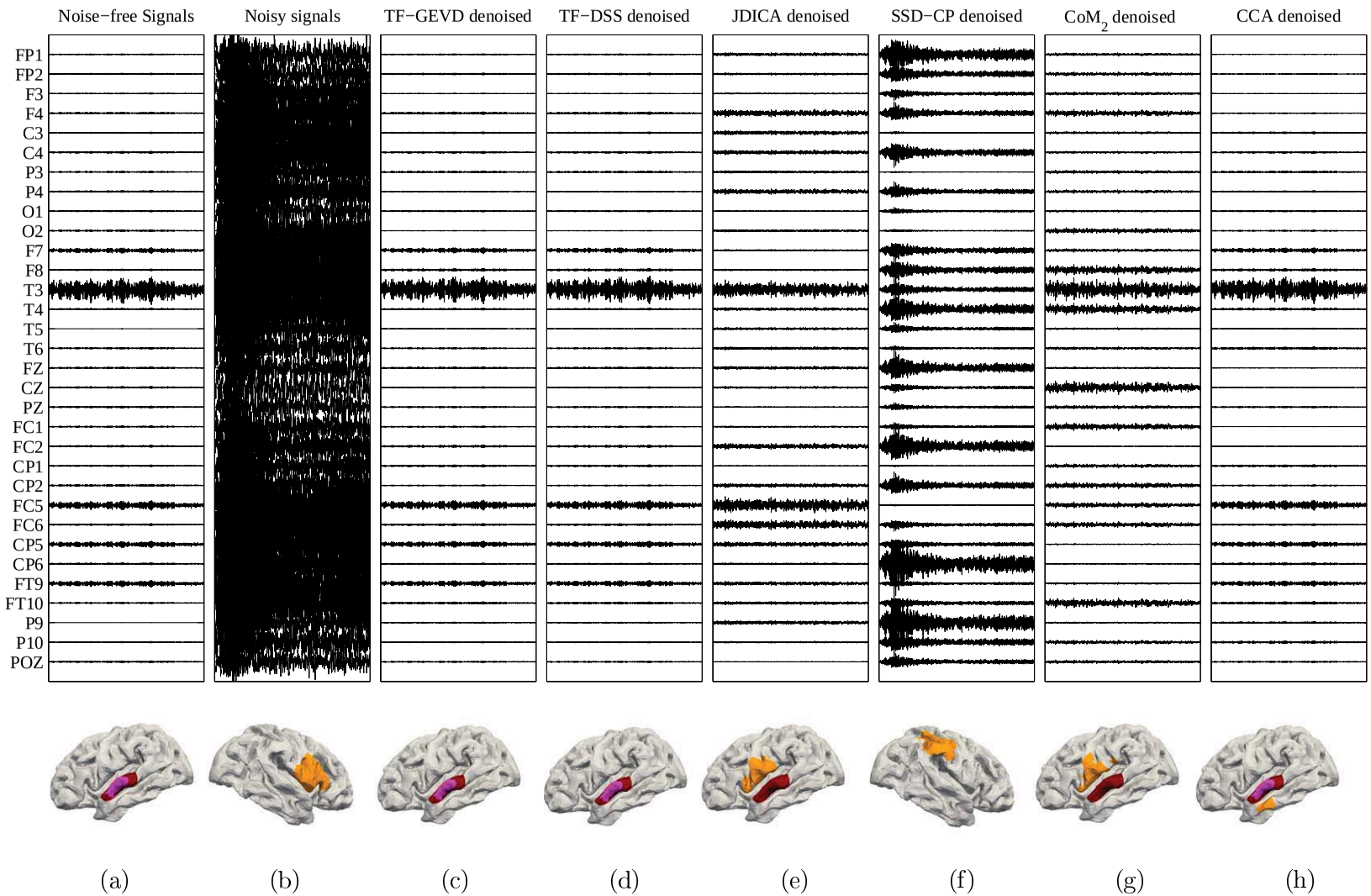


Figure 4.14: An example (trial ♯2) of denoising procedure in the case of simulated ictal data of scenario 1 generated from the activation of one patch: (a) noise-free simulated EEG, (b) noisy EEG after adding real muscle activity and background EEG with SNR= $-20dB$, (c-h) EEG denoised by the GEVD, DSS, JDICA, SSD-CP, CoM_2 and CCA methods, respectively. At the bottom of each column, the source localization results obtained from 4-ExSo-MUSIC are illustrated for this trial. Dark brown: real patch; violet: correctly estimated part of the patch; orange: incorrectly estimated part of the patch.

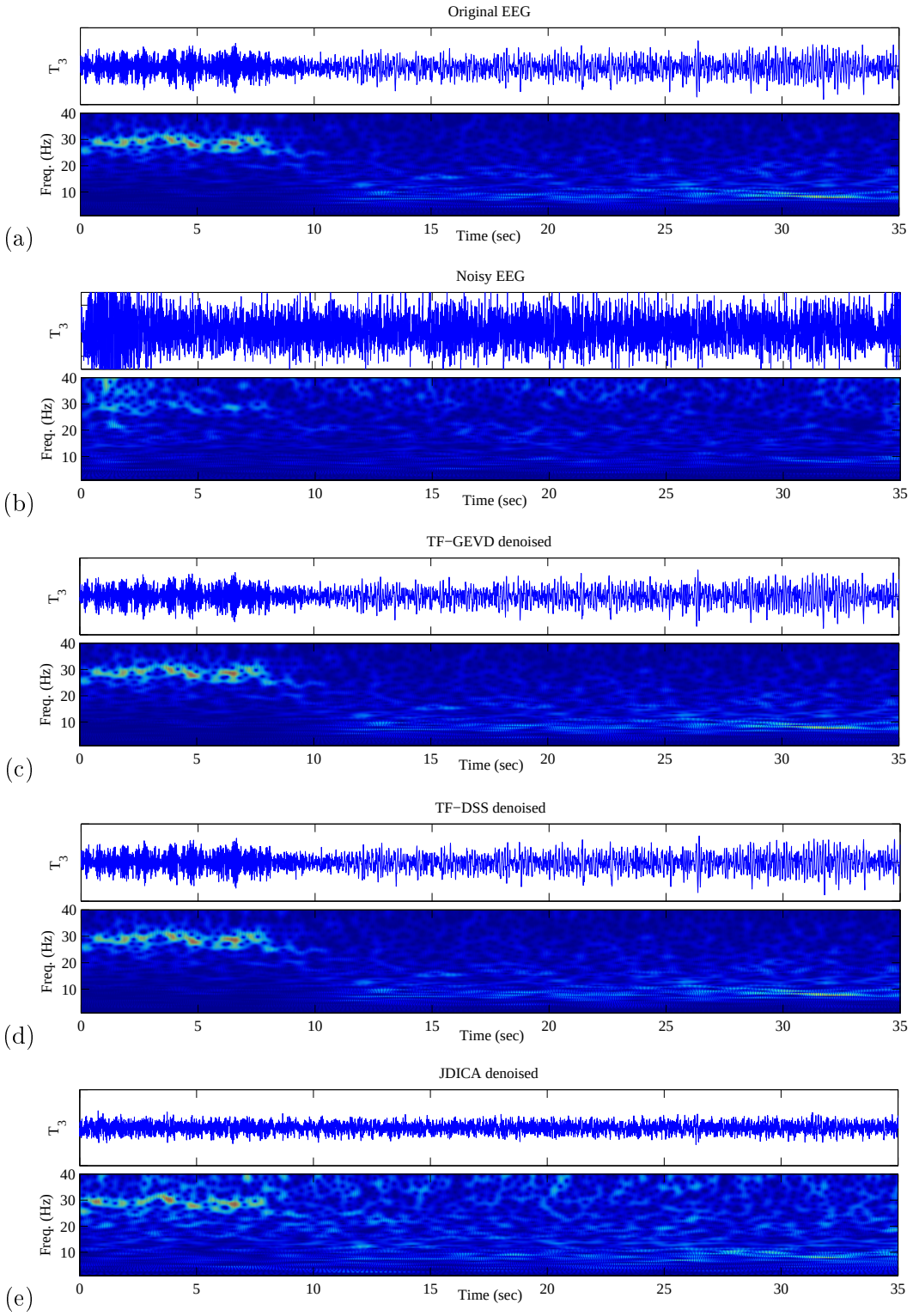


Figure 4.15: The time courses and time-frequency representations in the case of the ictal scenario 1 at the channel T3 of: (a) the noise-free ictal EEG, (b) the noisy ictal segment, (c) TF-GEVD denoised, (d) TF-DSS denoised, (e) JDICA denoised EEGs.

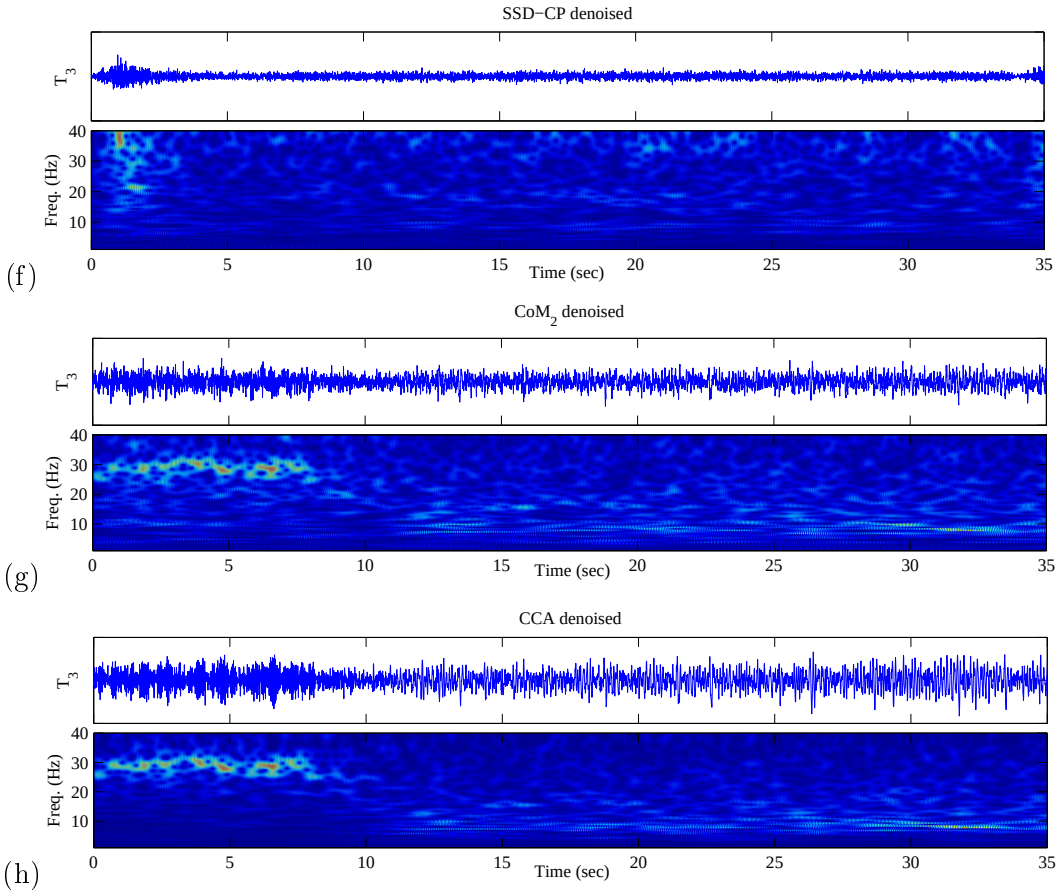


Figure 4.15: (cont.) The time courses and time-frequency representations in the case of the ictal scenario 1 at the channel T3 of: (f) SSD-CP denoised, (g) CoM_2 denoised and (h) CCA denoised EEGs.

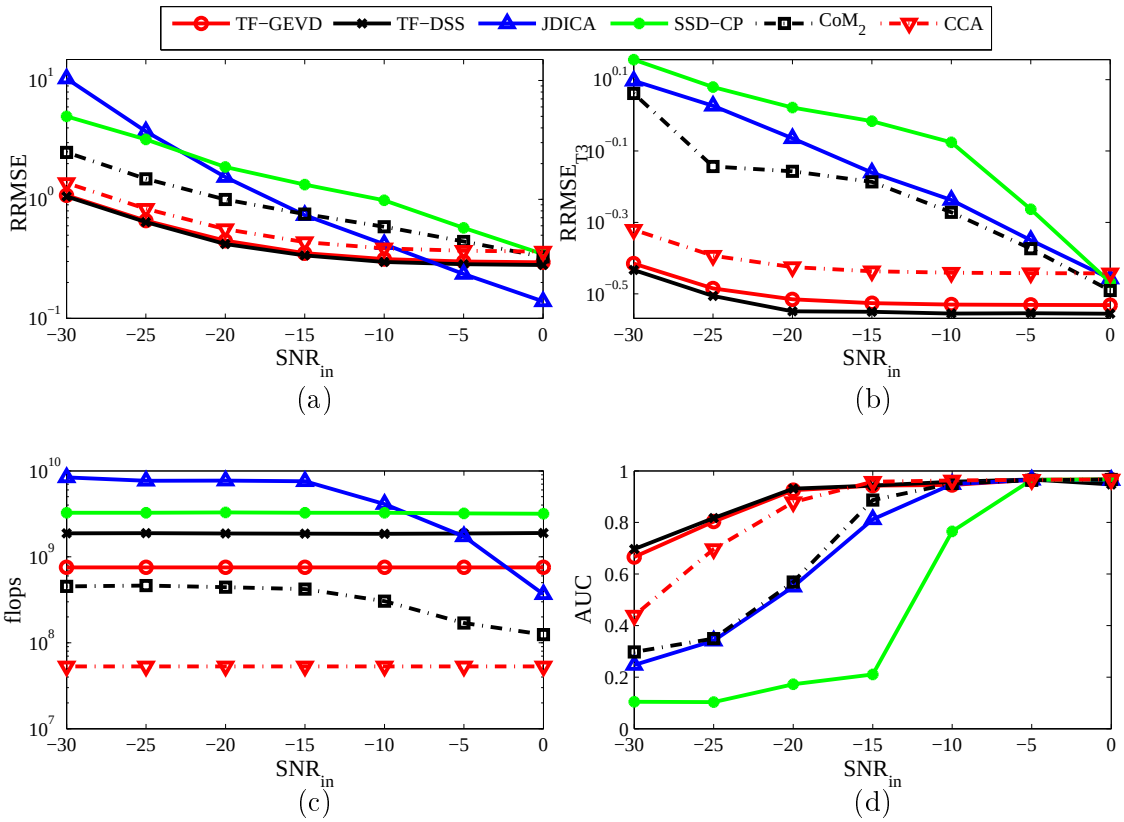


Figure 4.16: Denoising results of the six methods in the case of ictal scenario 1 for different SNR values of -30 to $0dB$: (a) the RRMSE denoising error, (b) the local RRMSE denoising error at channel T3, (c) number of flops and (d) AUC of the source localization ROC curve.

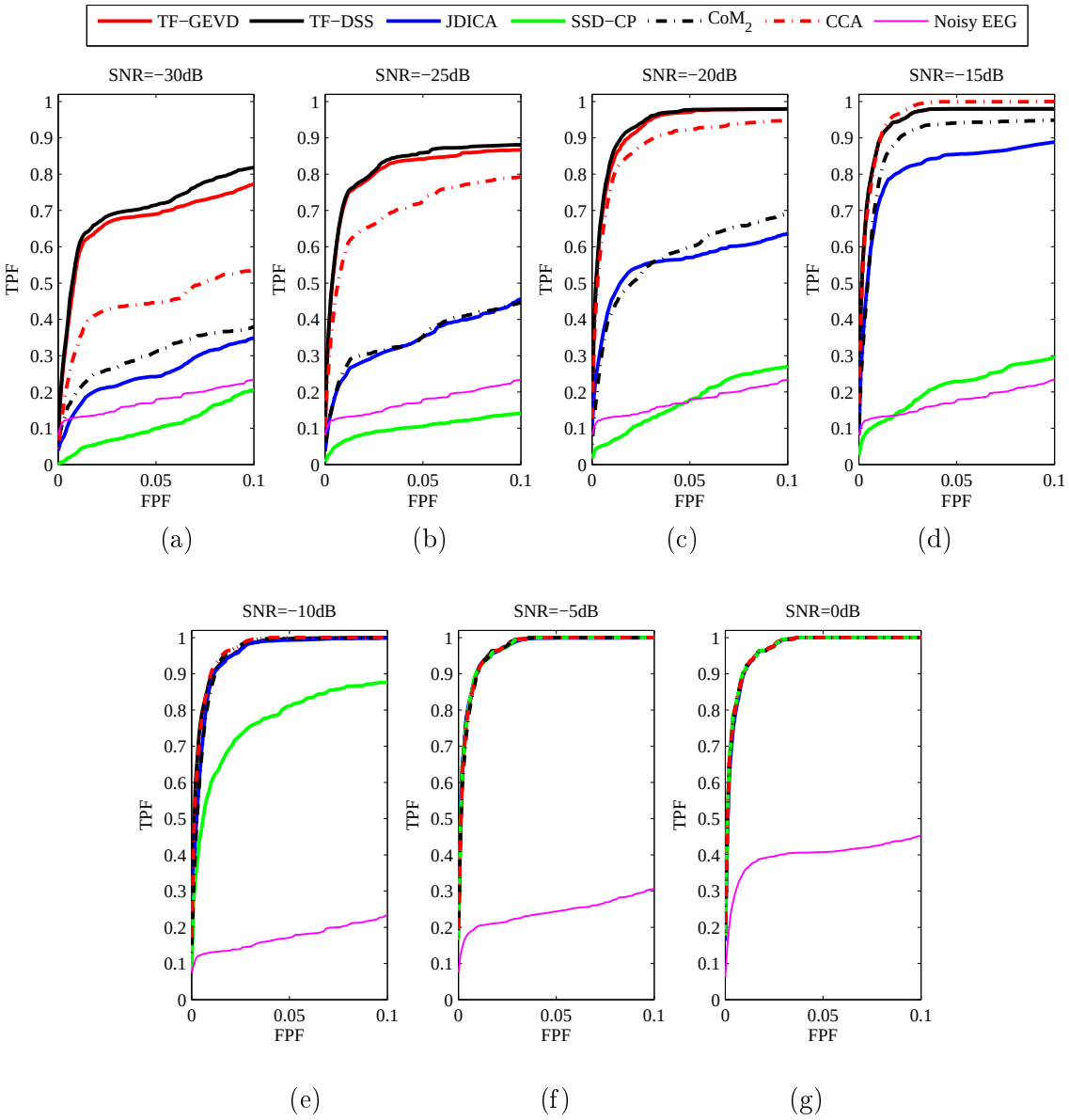


Figure 4.17: The source localization ROC curves of six denoising algorithms in the case of the ictal scenario 1 for different SNR values of (a) $-30dB$, (b) $-25dB$, (c) $-20dB$, (d) $-15dB$, (e) $-10dB$, (f) $5-dB$ and (g) $0dB$.

and T3, but high amplitude activities are also seen in other electrodes specially at F8. The SSD-CP method can not denoise the ictal data of this example.

Figure 4.19 shows the time course and time-frequency representations of the channel T3 for the noise-free signal, noisy ictal data and denoised signals by using the six denoising methods. The ictal fast activity can be seen clearly in the beginning of the time-frequency representation of the noise-free signal (figure 4.19(a)), while this signature is almost invisible in the time-frequency representation of the noisy data (figure 4.19(b)). Figures 4.19(c)-(h) show the time course and time-frequency representations of the channel T3 of the ictal signals denoised by the TF-GEVD, TF-DSS, JDICA, SSD-CP, CoM_2 and CCA algorithms, respectively. Similar to the previous scenario, the time-frequency signature of the ictal activity appears in the signals denoised with all methods except with SSD-CP. It is noteworthy that, although JDICA and CoM_2 do not show good time representation as shown in figure 4.18, the time-frequency representation of the channel T3 of their denoised signals show a clear ictal activity.

Figure 4.20 shows the results obtained with the second ictal scenario by using the six denoising algorithms for different SNRs. Figures 4.20(a), (b) and (c) show the RRMSE at all electrodes, at channel T3 and at channel F7, respectively. In this scenario, the TF-GEVD and TF-DSS methods surpass the other methods for all SNR values. After TF-GEVD and TF-DSS, the CCA method gives better results than the ICA-based algorithms and SSD-CP. For SNR values equal to or lower than $-20dB$, the local RRMSE of the CCA method at channel T3 and F7 is equal to that of TF-GEVD and TF-DSS.

The numerical complexity of the denoising methods in this scenario (figure 4.20(c)) is almost similar to the first ictal scenario. The difference is that the numerical complexity of TF-GEVD and TF-DSS increases for high SNR values and for some SNRs it is higher than that of JDICA and SSD-CP.

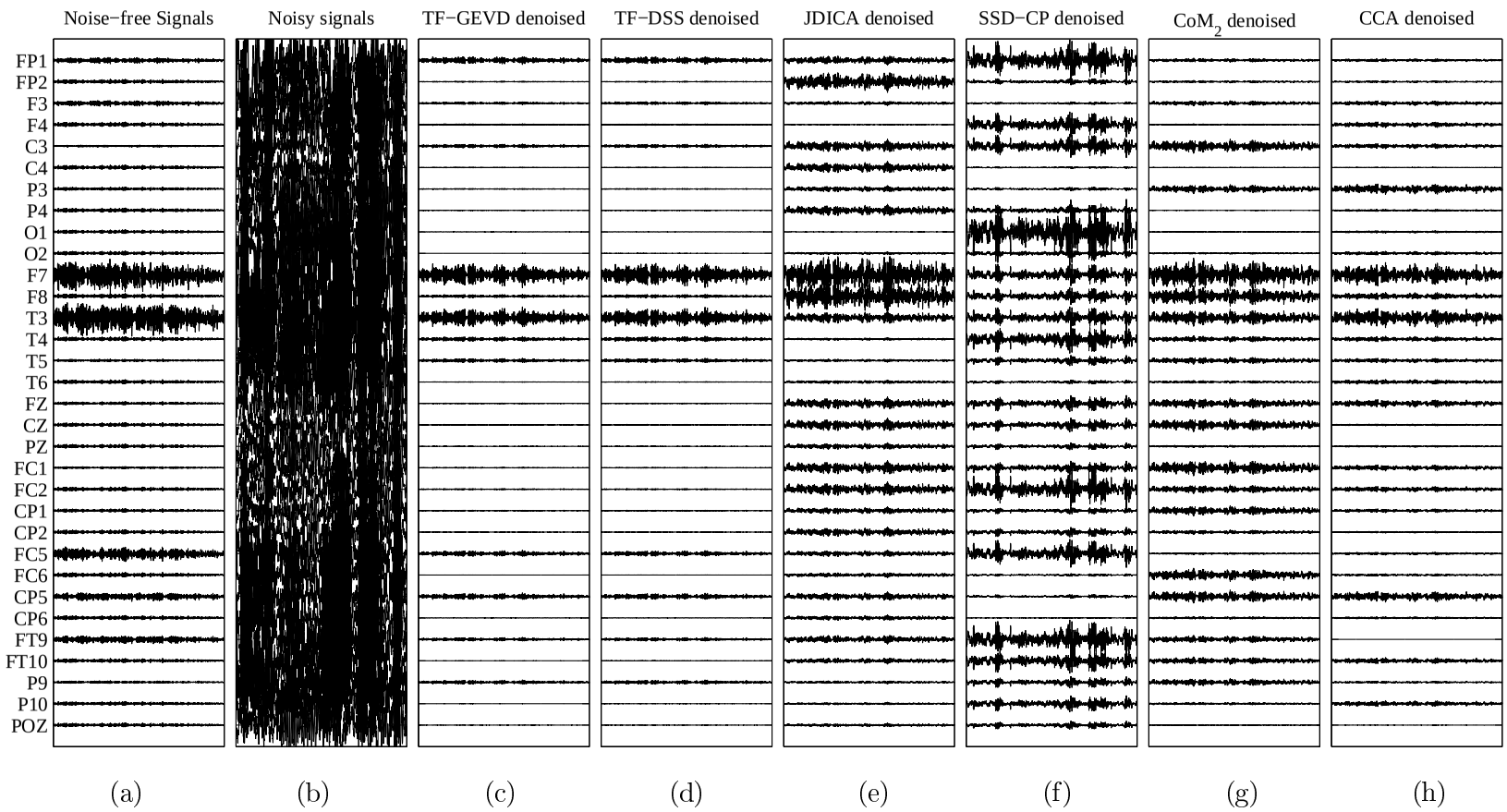


Figure 4.18: An example (trial ♯26) of denoising procedure in the case of simulated ictal data of scenario 2 generated from the activation of two patches with uncorrelated activities: (a) noise-free simulated EEG, (b) noisy EEG after adding real muscle activity and background EEG with SNR= $-20dB$, (c-h) EEG denoised by the GEVD, DSS, JDICA, SSD-CP, CoM_2 and CCA methods, respectively.

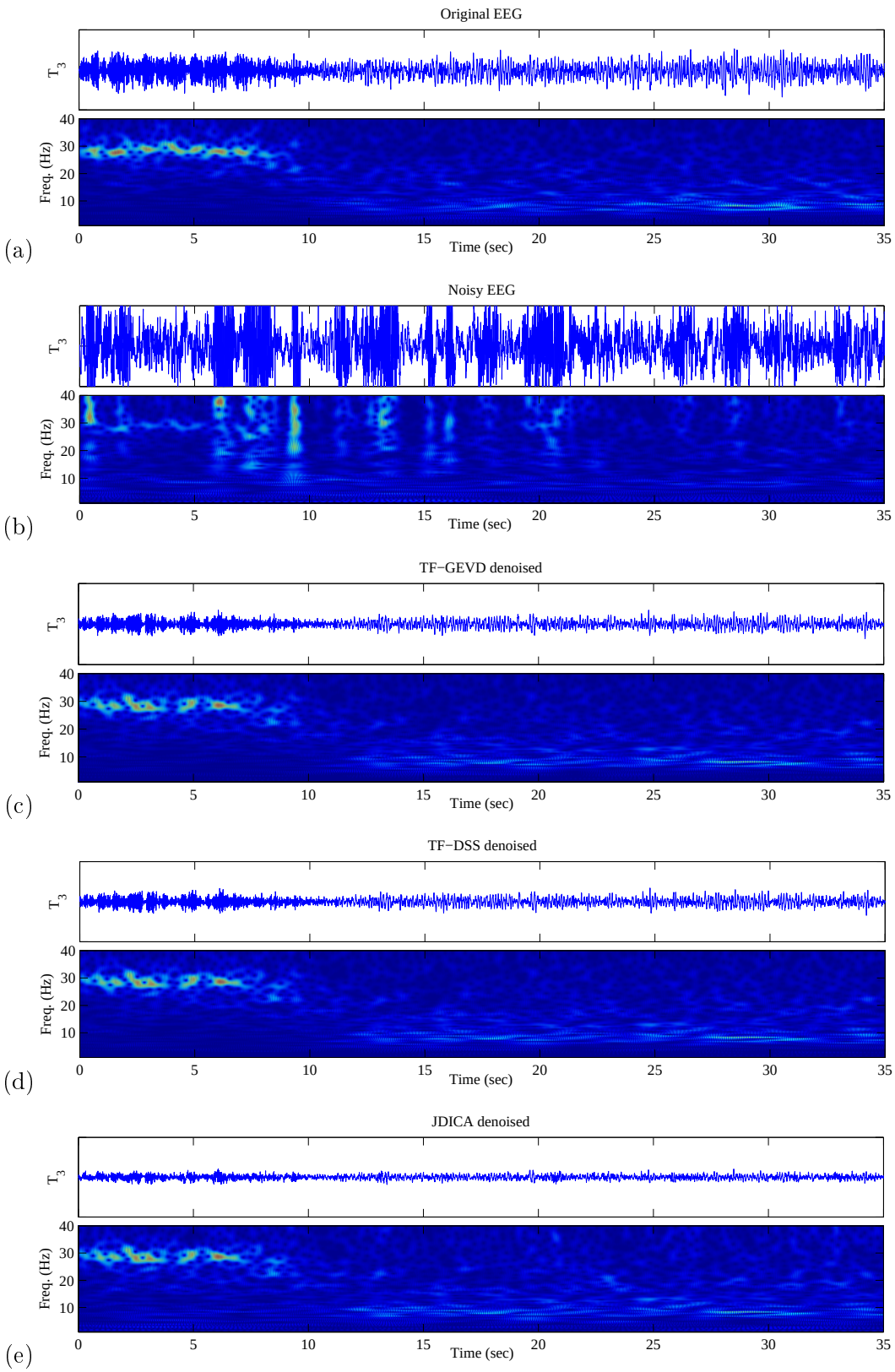


Figure 4.19: The time courses and time-frequency representations in the case of the ictal scenario 2 at the channel T3 of: (a) the noise-free ictal EEG, (b) the noisy ictal segment, (c) TF-GEVD denoised, (d) TF-DSS denoised, (e) JDICA denoised EEGs.

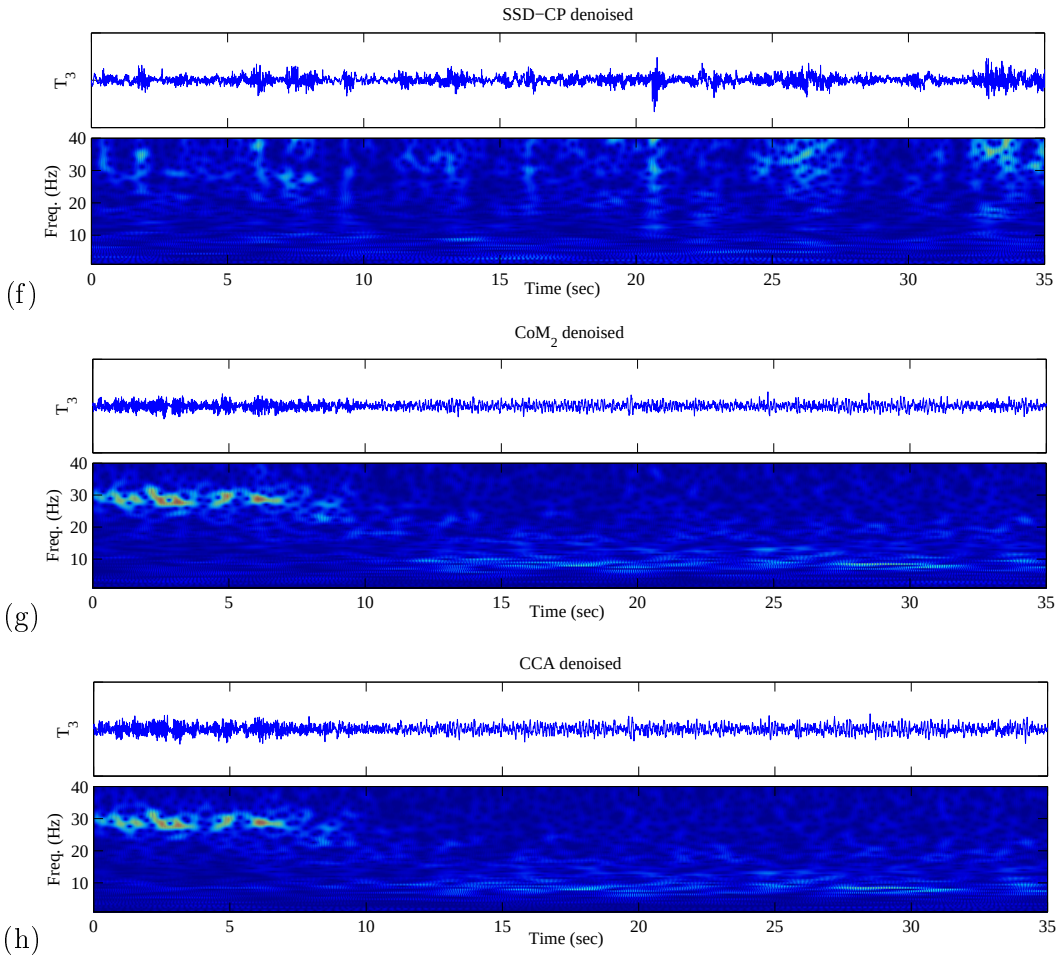


Figure 4.19: (cont.) The time courses and time-frequency representations in the case of the ictal scenario 2 at the channel T3 of: (f) SSD-CP denoised, (g) CoM_2 denoised and (h) CCA denoised EEGs.

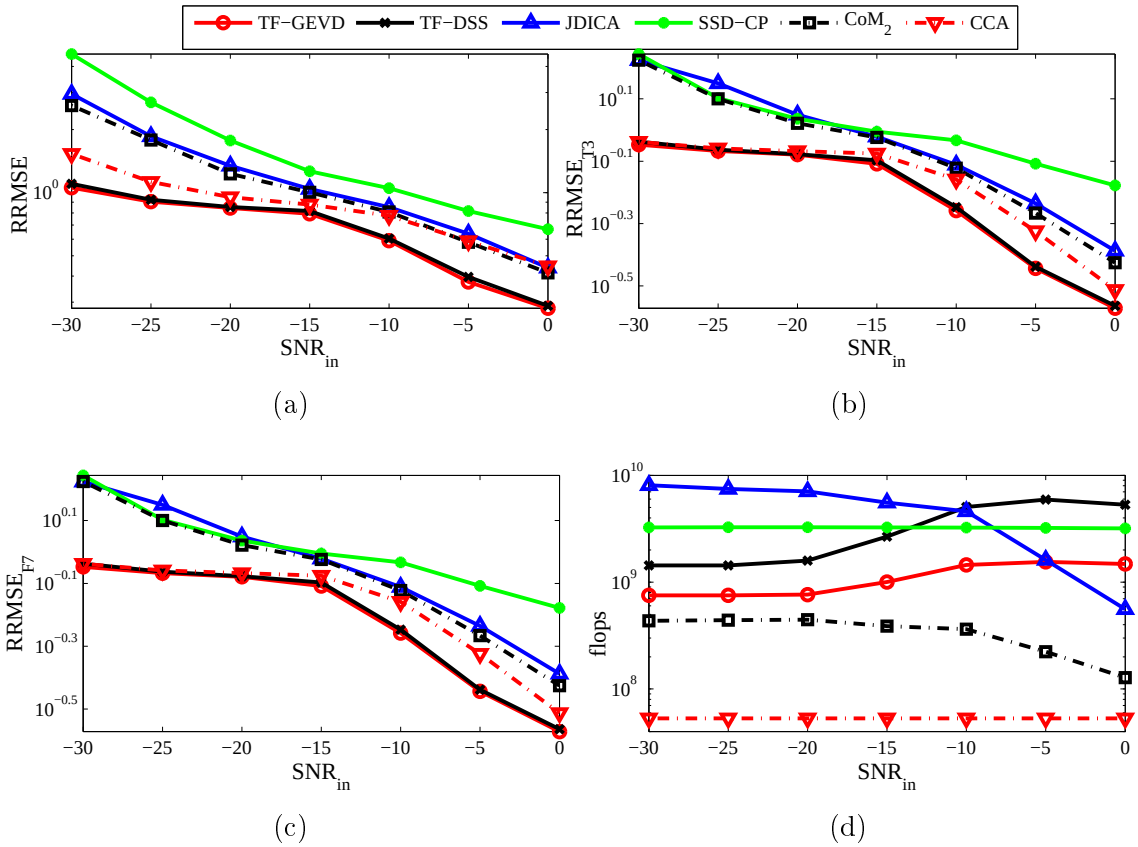


Figure 4.20: Denoising results of the six methods in the case of ictal scenario 2 for different SNR values between -30 and $0dB$: (a) the RRMSE denoising error, (b) the local RRMSE denoising error at channel T3, (c) the local RRMSE denoising error at channel F7 and (d) number of flops.

4.2.2.3 Scenario 3: two correlated epileptic patches

In the third ictal scenario, similar to the ictal scenario 2, two patches are considered, first patch in the left superior temporal gyrus and the second one in the left inferior frontal gyrus. In this scenario the patches are considered to have correlated ictal activities. Figure 4.21 shows an example of simulated noise-free, noisy and denoised data of the second scenario with SNR$= -15dB$.

As shown in 4.21(a), in the original noise-free data, electrodes T3 and F7 show ictal activities, but this activity is contaminated with high amplitude muscular noise in the simulated noisy data (figure 4.21(b)). A visual inspection of figure 4.21(c-h) shows that TF-GEVD, TF-DSS, JDICA, CoM_2 and CCA can somewhat extract the ictal activity at channels T3 and F7. Among these methods, TF-DSS and CCA highly decrease the amplitude of signal at channels F7 and T3, respectively. The SSD-CP method can not denoise the ictal data of this example.

Figure 4.22 shows the time course and time-frequency representations of the channel T3 for the noise-free signal, noisy ictal data and denoised signals by using the six denoising methods. The ictal fast activity can be seen clearly in the middle times (10 sec to 20 sec) of the time-frequency representation of the noise-free signal (figure

4.22(a)) and the denoised signals by using TF-GEVD, TF-DSS, JDICA, CoM_2 and CCA (figure 4.22(c-e,g-h)) while this signature is almost invisible in the time-frequency representation of the noisy data (figure 4.19(b)) and of the SSD-CP denoised one (figure 4.19(f)).

Figure 4.23 shows the results obtained with the third ictal scenario by using the six denoising algorithms for different SNRs. Figures 4.23(a), (b) and (c) show the RRMSE at all electrodes, at channel T3 and at channel F7, respectively. In this scenario, the CoM_2 algorithm surpass the other methods almost for all SNR values. After CoM_2, the TF-GEVD, JDICA and CCA methods give better results than the TF-DSS and SSD-CP algorithms. For some SNR values, the results of TF-GEVD, JDICA and CCA are as good as those of CoM_2.

The numerical complexity of the denoising methods in this scenario (figure 4.23(c)) is almost identical to the second ictal scenario.

4.3 Real data

In this section, we evaluate the denoising methods in the case of real data with both interictal and ictal activity.

4.3.1 Interictal data

In this section, the four proposed methods and two ICA algorithms, CoM_2 and SOBI, are applied to denoise interictal spikes obtained from a patient suffering from drug-resistant partial epilepsy as explained earlier in section 3.2. Figure 4.24(a) and (b) show two epochs of the real interictal data, one epoch containing clean spikes and one epoch including spikes partially hidden in muscle activity, respectively. The same procedure as for simulated data was applied to the noisy real EEG epoch to reconstruct the denoised EEG signals by using the six denoising algorithms as shown in figure 4.24(c-h). Since we do not know the ground truth to evaluate qualitatively the performance of the six methods, a source localization process was performed on the original clean signal (considered as a reference), on the noisy data, as well as on the data denoised by GEVD, DSS, JDICA, SSD-CP, CoM_2 and SOBI. As shown in figure 4.24(c-h), the interictal spikes are visible at electrodes F8, T4, FC6 and FT10 in the denoised data whereas they are partially hidden in the noisy data. It seems that among the denoising algorithms, SSD-CP leaves more muscular activity at all electrodes. Regarding the source localization results (bottom of each column of figure 4.24), the spike sources are localized in the right anterior temporal region for the clean epoch. For the noisy epoch, the spike source is not in agreement with the source obtained from clean data since it is localized in the right inferior frontal region. The interictal data denoised by JDICA, SSD-CP and CoM_2 are localized in the right temporal neocortex which is concordant with the source localization obtained from the clean epoch. The interictal data denoised by GEVD, DSS and SOBI are localized in the right temporal neocortex but also in the right insula. These results can indicate that JDICA, SSD-CP and CoM_2 denoise this example of real interictal data better than GEVD, DSS and SOBI.

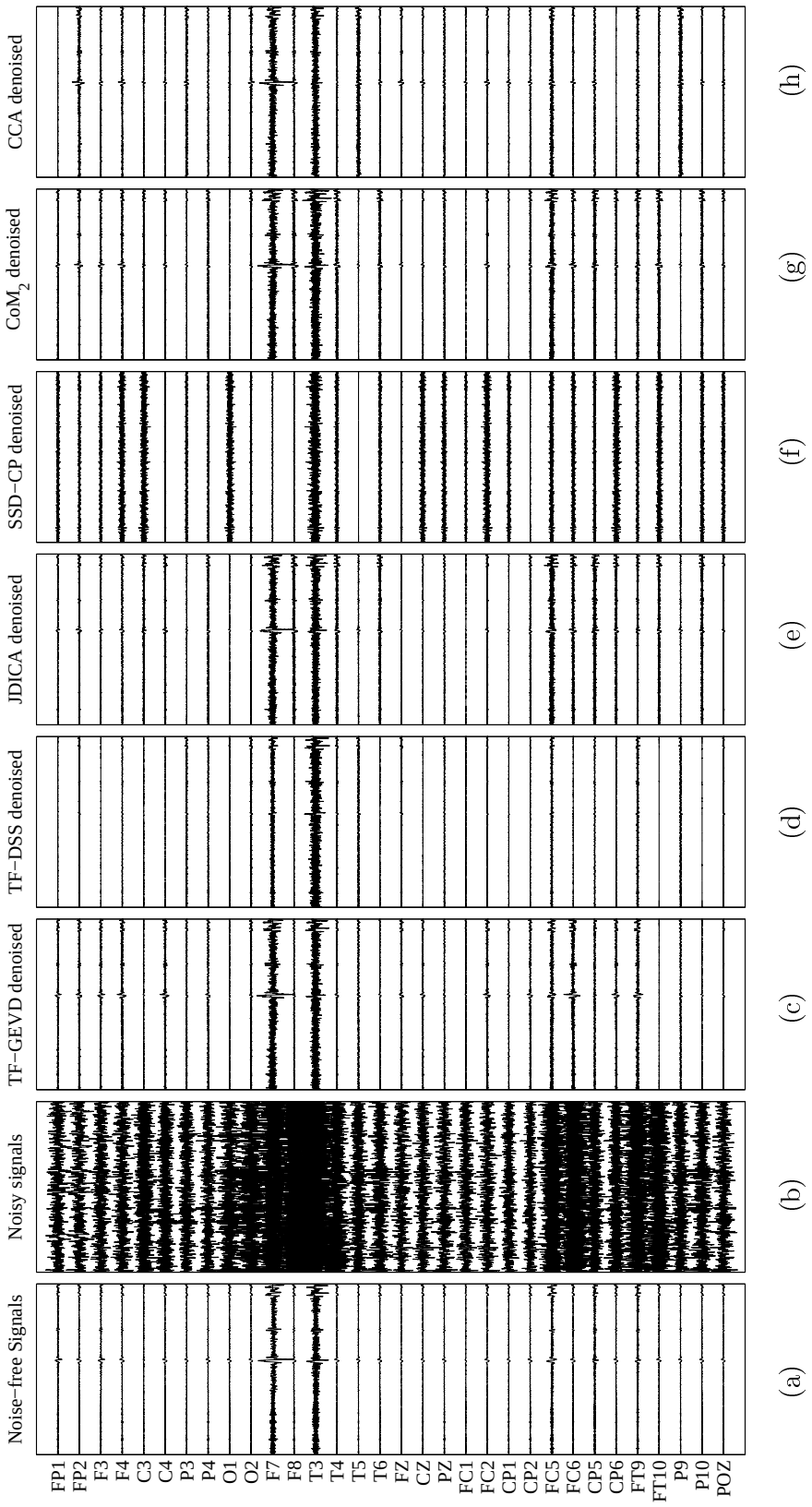


Figure 4.21: An example (trial ♯46) of denoising procedure in the case of simulated ictal data of scenario 3 generated from the activation of one patch: (a) noise-free simulated EEG, (b) noisy EEG after adding real muscle activity and background EEG with SNR$= -15dB$, (c-h) EEG denoised by the GEVD, DSS, JDICA, SSD-CP, CoM$_2$ and CCA methods, respectively.

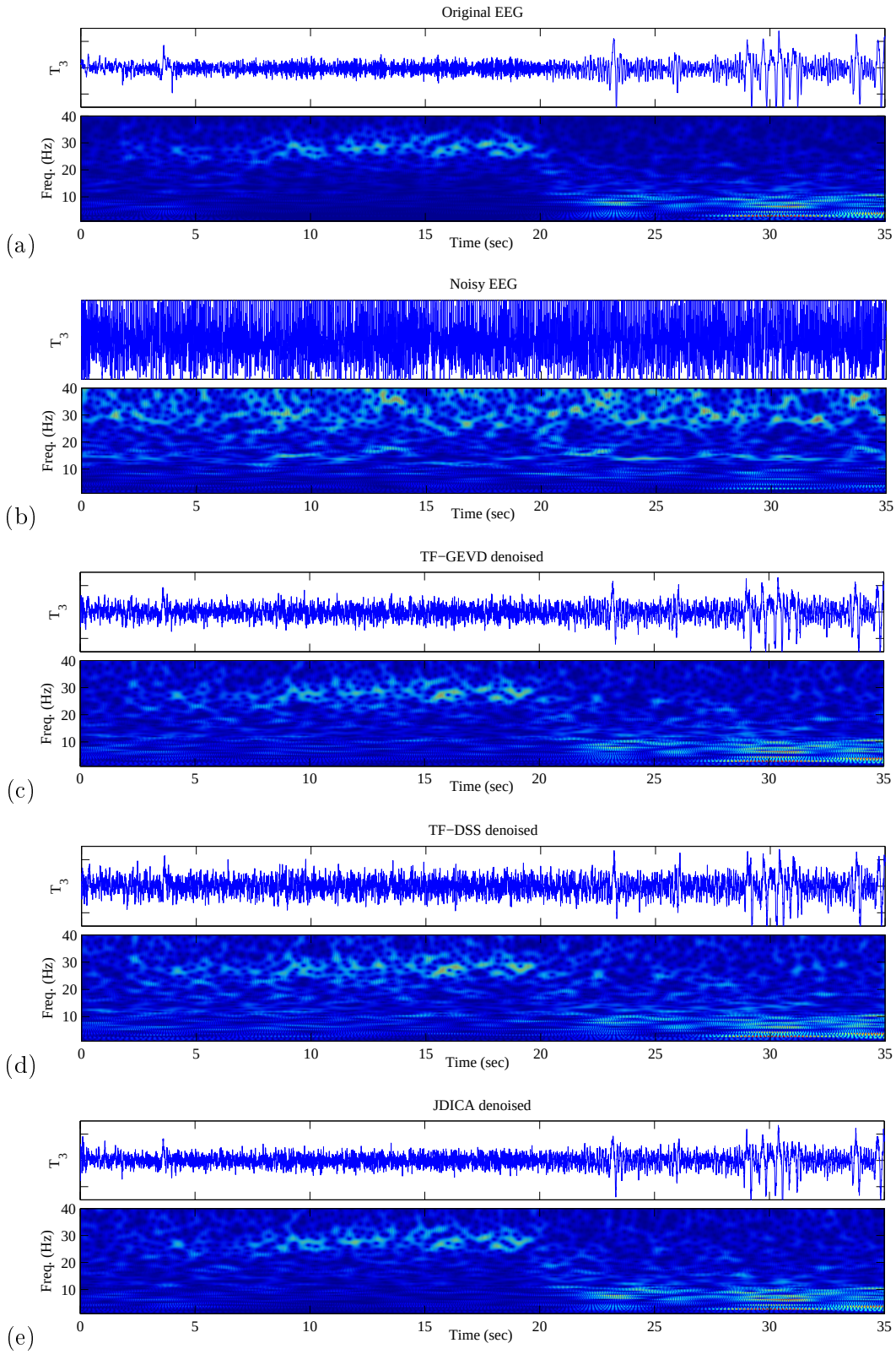


Figure 4.22: The time courses and time-frequency representations in the case of the ictal scenario 3 at the channel T3 of: (a) the noise-free ictal EEG, (b) the noisy ictal segment, (c) TF-GEVD denoised, (d) TF-DSS denoised, (e) JDICA denoised EEGs.

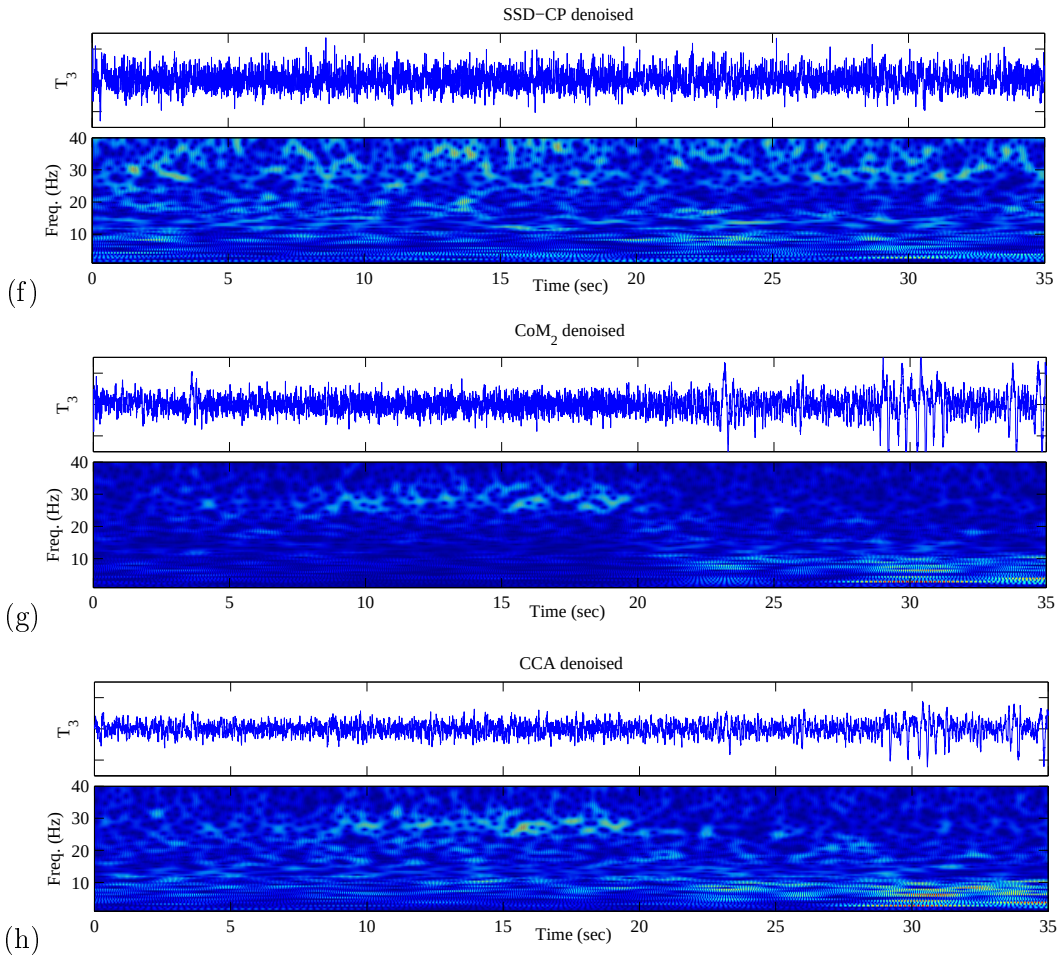


Figure 4.22: (cont.) The time courses and time-frequency representations in the case of the ictal scenario 3 at the channel T3 of: (f) SSD-CP denoised, (g) CoM_2 denoised and (h) CCA denoised EEGs.

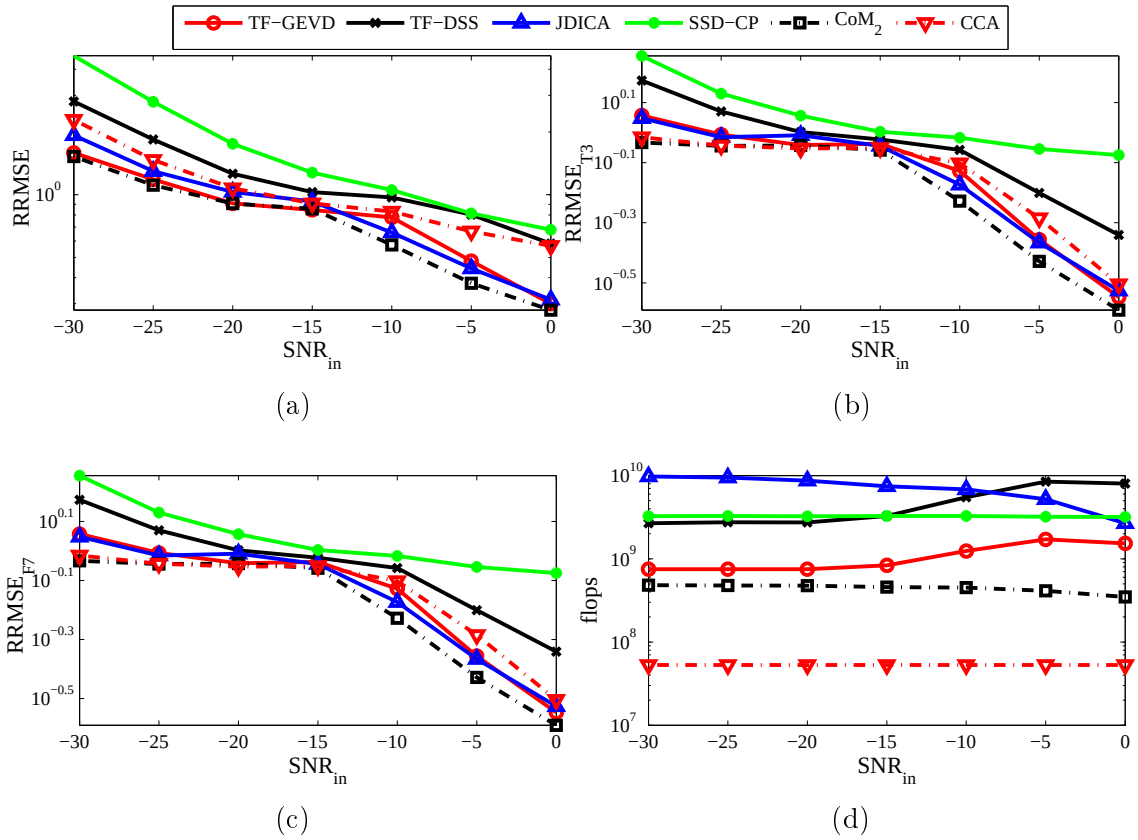


Figure 4.23: Denoising results of the six methods in the case of ictal scenario 3 for different SNR values between -30 and $0dB$: (a) the RRMSE denoising error, (b) the local RRMSE denoising error at channel T3, (c) the local RRMSE denoising error at channel F7 and (d) number of flops.

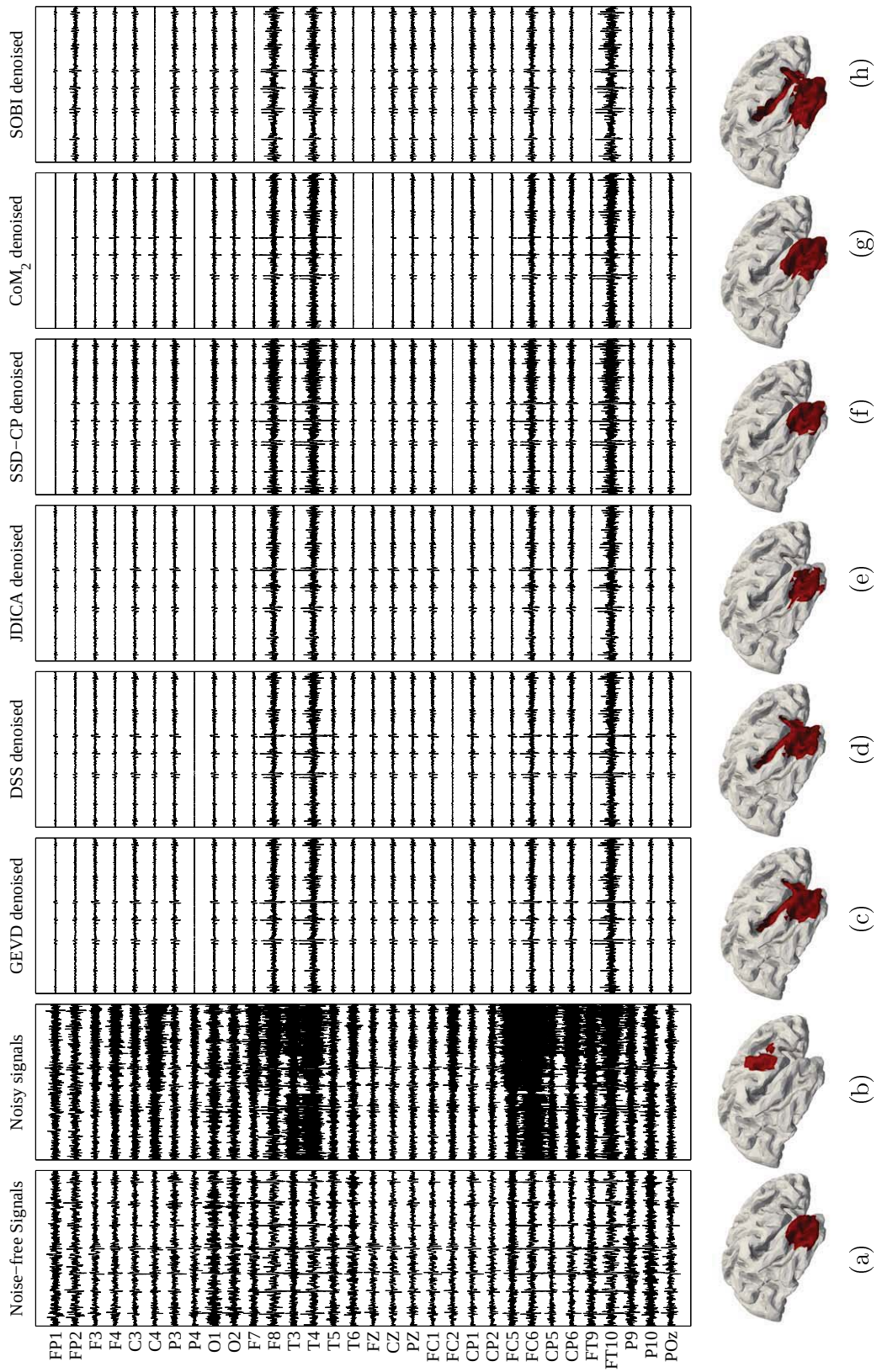


Figure 4.24: Denoising of real interictal spikes data: (a) an epoch containing noise-free interictal spikes, (b) an epoch including spikes hidden in muscle activity, (c-h) EEG denoised by the GEVD, DSS, JDICA, SSD-CP, CoM_2 and CCA methods, respectively. The source localization results obtained from 4-ExSo-MUSIC are depicted at the bottom of each column.

4.3.2 Ictal data

In this section, the denoising algorithms are tested on real ictal data recorded from a 23 years old patient as explained in section 3.2. These data were collected from a long-term video-EEG recording (32 electrodes, 256 Hz, BP [0.3-100] Hz) during which several ictal signals could be recorded. One of these segments is a low noise ictal EEG with very little muscle activity, that can be used as a reference signal. We apply our proposed algorithms on the signal acquired during a second seizure that was highly contaminated with muscle activity. Six channels of both the low noise and the noisy ictal signal are shown in figure 4.25(a) and figure 4.25(b), respectively. As shown in figure 4.25(a), the ictal activity of channel C3 can be segmented to three parts: 1) rhythmic spikes (red box), 2) fast activity (green box) and 3) slower irregular activity (blue box). During the second seizure these activities, except the first rhythmic spikes, are covered with muscular activity as shown in figure 4.25(b).

Figures 4.26(a) and 4.26(b) show the time course and time-frequency representations of the channel C3 for the noise-free and noisy ictal signals. The ictal fast activity is highlighted in the time-frequency representation of the noise-free segment (figure 4.26(a)). This time-frequency signature is almost invisible in the time-frequency representation of the noisy ictal discharge (figure 4.26(b)). Figures 4.26(c-g) show the time course and time-frequency representations for channel C3 of the second ictal signal, after denoising by TF-GEVD, TF-DSS, CCA, CoM_2 and JDICA, respectively. As shown in these figures, the time-frequency signature of the ictal activity appears in the denoised signals. The time course of the denoised signals are cleaner than the noisy data, showing three ictal patterns similar to what had previously been identified during the first noise-free seizure. Qualitatively, it is noteworthy that although all methods provide fairly good denoising, the best result is obtained with TF-DSS, which could retrieved the fast activity with the highest amplitude. CCA and TF-GEVD gave similar results for the fast activity, but with TF-GEVD some of the rhythmic activity has been removed by the denoising procedure. JDICA and CoM_2 can retrieve part of the fast activity but leaves a large amount of muscle activity. Since SSD-CP can not generate acceptable results in this real example, we do not show its results.

4.4 Summary

In this chapter, we studied the performance of the proposed methods. To this end, we first examined the proposed SSD-CP algorithm in some common scenarios in the context of CP decomposition. Then, we studied the performance of the proposed methods to denoise both simulated and real epileptic EEG data with interictal or ictal activity contaminated with muscular activity. The results of this section were presented in four parts: i) simulated interictal data, ii) simulated ictal data, iii) real interictal data and iv) real ictal data. In the next chapter, we will discuss the results given in this chapter and make a conclusion. We then present some ideas for future work.

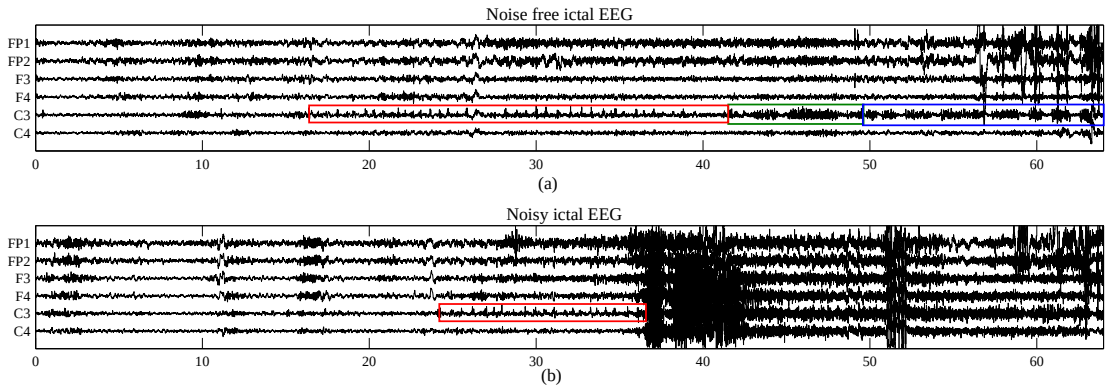


Figure 4.25: A real EEG recorded from an epileptic patient during two separate ictal periods: (a) an almost noise-free ictal EEG. The channel C3 shows ictal activity consists of three parts: 1) rhythmic spikes (red box), 2) fast activity (green box) and 3) slower irregular activity (blue box). (b) another ictal EEG contaminated by muscle activity.

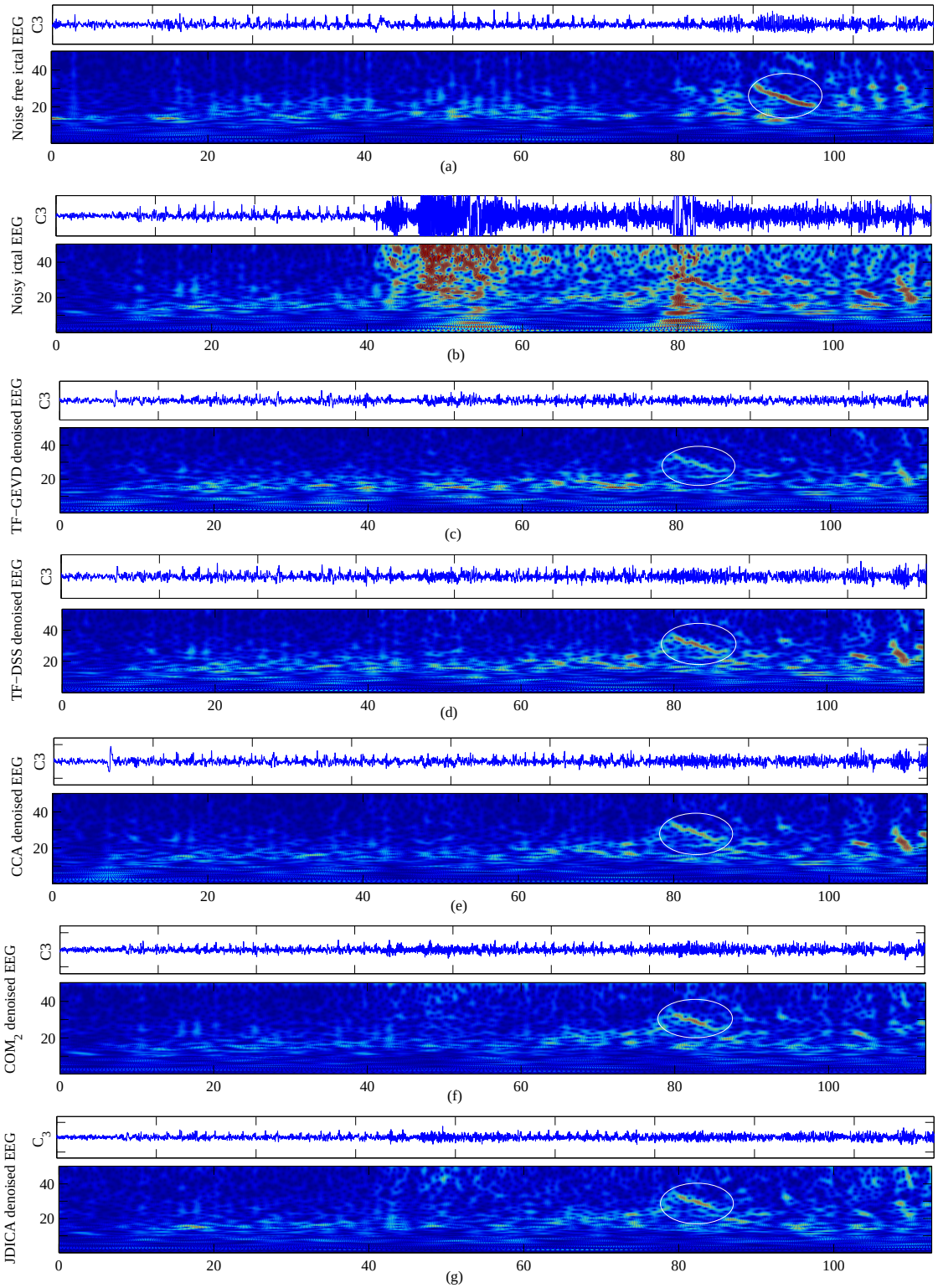


Figure 4.26: The time courses and time-frequency representations at the channel C3 of: (a) the noise-free ictal EEG, (b) the noisy ictal segment, (c) TF-GEVD denoised, (d) TF-DSS denoised, (e) CCA denoised, (f) CoM_2 denoised, (g) JDICA denoised EEGs. Ictal fast activity is shown in the white circle on the time-frequency representation.

Chapter 5

Discussion, Conclusion and Perspectives

In this thesis, we proposed several methods to denoise interictal and ictal EEG signals. These methods include GEVD and DSS methods to denoise interictal data, TF-GEVD and TF-DSS to denoise ictal data and JDICA and SSD-CP to denoise both interictal and ictal signals. We evaluated these methods using both simulated and real epileptic EEG data with interictal or ictal activity contaminated with muscular activity. In this chapter, we first discuss the results given in chapter 4 and make a conclusion. We then present some ideas for future work.

5.1 Discussion and Conclusion

5.1.1 SSD-CP and numerical experiments

In chapter 4, we first examined the proposed SSD-CP algorithm in some common scenarios in the context of CP decomposition. Experimental results showed the efficiency of the proposed method in solving some well-known numerical problems, especially cases of arrays with swamp and bottleneck degeneracies. In contrast with other CP algorithms, the proposed SSD-CP algorithm converged in all these experiments and showed a faster convergence. In another study with real Amino acids fluorescence data, we showed that the proposed SSD-CP method is robust with respect to the overestimation of the rank of the array.

5.1.2 Simulated interictal data

In the second part of our study, we examined the efficiency of the proposed methods for denoising of simulated interictal data with three different scenarios. To this end, we compared the algorithms in terms of RRMSE, local RRMSE, Spike RRMSE and local Spike RRMSE. In terms of RRMSE and local RRMSE, GEVD, DSS, JDICA, CoM_2 and SOBI surpass SSD-CP in denoising interictal data in all three scenarios. For the scenarios 1 and 3, with one source and with two correlated sources, respectively, the GEVD and DSS methods are slightly better than the other methods for high SNRs ($-10dB$ and higher). However, in the case of low SNRs ($-20dB$), JDICA, CoM_2 and

SOBI denoise better than GEVD and DSS. For scenario 2 with two uncorrelated sources, for SNRs equal to and higher than $-10dB$, all methods except SSD-CP produce similar results. But for low SNRs, GEVD and DSS do not work well. Although the results of SSD-CP in terms of RRMSE and local RRMSE are not suitable, it has acceptable results in terms of Spike RRMSE and local Spike RRMSE, especially for SNR values equal to or higher than $-10dB$. For the scenario 1, the Spike RRMSE of SSD-CP is similar to or lower than other methods for SNR values higher than $-20dB$. In this scenario, GEVD, DSS and SSD-CP denoise signals in the intervals around spikes better than the other methods. For scenarios 2 and 3, the Spike RRMSE and local Spike RRMSE of SSD-CP is comparable with other methods. For scenario 3, GEVD and DSS are better than the other algorithms in terms of Spike RRMSE and local Spike RRMSE.

The first noticeable observation from this set of experiments is the difference between the results of SSD-CP in terms of RRMSE and Spike RRMSE. In terms of RRMSE and local RRMSE, the SSD-CP method can not denoise signals as well as the other methods. But in terms of Spike RRMSE and local Spike RRMSE, it has acceptable results. Even in some cases, the Spike RRMSE of SSD-CP is lower than other methods. It shows that SSD-CP can denoise signals in the intervals around spikes. Therefore, the extracted spikes from the SSD-CP denoised signals have true amplitude and shape which is an important factor for the source localization methods.

Another remark that should be considered is the stage where spike detection is used in the denoising method. For the GEVD and DSS methods, the first step is spike detection and the second one is subspace decomposition. But for SSD-CP, JDICA, CoM_2 and SOBI, spike detection is used after source extraction to identify sources of interest. It is clear that for the first group of denoising methods, spike detection is applied on raw data which is more noisy in comparison with the second group. Consequently, for low SNRs ($-15dB$ and lower) spike detection method does not appropriately detect interictal spikes and the performance of GEVD and DSS significantly decreases. We verified this claim by using the manually exacted time support of spikes involved in each epileptic source in the GEVD and DSS methods, and we observed that the performance of these methods considerably increases by using this information.

We also compared the denoising methods in terms of source localization results. The results show that the source localization results are highly correlated with Spike RRMSE and local Spike RRMSE results (and not with RRMSE and local RRMSE ones). In particular, although there exists a significant difference between RRMSE of SSD-CP and that of the others, this difference is smaller in the source localization results, especially for higher SNRs. In some cases, the signals denoised by SSD-CP are even better localized than those denoised by the other methods. Although SSD-CP can not denoise the whole interval of interictal data as well as the other methods, it can better denoise the signals around spikes. In other words, the denoising procedure performed by means of SSD-CP preserves better the topography of brain electrical sources than the other methods. Such a property is crucial for some applications such as the identification of the epileptic region from interictal EEG data.

We also studied the denoising methods based on their numerical complexity. The results illustrate that the GEVD and DSS algorithms have the lowest number of flops.

Among the other methods, JDICA is the most time-consuming. However it provides the best compromise between other deflationary ICA algorithms when the number of channel is low. This comes from the fact that the complexity of JDICA is $O(P^4)$ where P is the number of all independent sources or the dimension of space. We showed in another study performed to denoise pediatric epileptic EEG signals with 12 electrodes, that JDICA needs a fewer number of flops than the other deflationary algorithms.

We conclude from interictal simulation results that GEVD and DSS has the best compromise between the proposed denoising methods and CoM_2 and SOBI, especially for SNRs higher than $-15dB$. For low SNRs ($-15dB$ and $-20dB$), JDICA can be used to give the best performance. SSD-CP can be used for high SNR values ($-10dB$ and higher) especially in the context of interictal source localization.

It should also be noted that all the above-mentioned results were obtained for 32 electrodes and that could be different with more or less sensors.

5.1.3 Simulated ictal data

In the third part of our study, we examined the efficiency of the proposed methods for denoising of simulated ictal data with three different scenarios. The results obtained for ictal data were not similar for all three scenarios. For scenario 1, with one epileptic patch, the TF-GEVD and TF-DSS methods surpass the other methods for SNR values equal to or lower than $-10dB$. For SNR values of $-5dB$ and $0dB$, JDICA give better results than the other algorithms in terms of RRMSE, but local RRMSE at electrode T3 is not as good as that from TF-GEVD and TF-DSS methods. For the second scenario with two uncorrelated sources, the TF-GEVD and TF-DSS methods surpass the other methods for all SNR values. Similar results obtained by TF-GEVD and TF-DSS admit the mathematical results presented in section 2.1.4.4. After TF-GEVD and TF-DSS, the CCA method gives better results than JDICA, CoM_2 and SSD-CP. For the third scenario with two correlated patches, the CoM_2 algorithm surpass the other methods almost for all SNR values. After CoM_2, the TF-GEVD, JDICA and CCA methods give better results than the TF-DSS and SSD-CP algorithms. In all these scenarios, SSD-CP give the worst results almost in all cases.

For ictal scenarios, we observed that SSD-CP does not work well. Note that the CP decomposition is based on the model of data, i.e. for the CP decomposition to be correct, the array to be decomposed should fit the trilinear model in our context. Although the third order array created from ictal data (see section 2.3.4) ideally fit the trilinear model, it diverges from this model when noise (including muscular activity) increases. Consequently, the CP decomposition is not precise in the case of very noisy data and the proposed SSD-CP based denoising method does not work as well as the other methods for low SNR values. However, this does not disprove the quality of the CP decomposition based methods in other contexts. Note that contrarily to our study, in other CP decomposition based denoising or source localization studies, i) the CP decomposition based approaches were compared in the context of source localization of cleaned data and not with methods such as ICA-based approaches [Becker et al., 2012,Phan et al., 2013a,Becker et al., 2014a], or ii) the CP decomposition based methods were tested with simulated non-realistic data which fit the trilinear model [De Vos et al.,

2007], or iii) the CP decomposition based methods were studied with lower level of noise [Deburchgraeve et al., 2009] or iv) different methods were used to produce higher order arrays [Becker et al., 2010, Becker et al., 2012].

In the proposed TF-GEVD and TF-DSS methods, we used the CCA approach to extract the time-frequency information on ictal sources, because CCA has been shown to be the most efficient method to date, to denoise ictal EEG signals contaminated by muscular activity [De Clercq et al., 2006, Hallez et al., 2009, Safieddine, 2012, Karfoul et al., 2015]. However, we observed that in one of the scenarios (the third ictal scenario with correlated sources), CoM_2 and JDICA outperforms CCA. Consequently, we believe that a new framework based on a combination of JDICA (or CoM_2) with TF-GEVD/TF-DSS can be used to denoise ictal signals.

The source localization results were provided only for the first ictal scenario, because for scenarios with two strongly correlated ictal sources, ExSo-MUSIC can not achieve good results even for noise-free signals. The source localization results for the first scenario are in agreement with denoising results, and the signals denoised by TF-GEVD and TF-DSS were localized better than those of other methods especially for low SNRs ($-20dB$ and lower).

In terms of complexity, CCA and CoM_2 have the lowest number of flops and the TF-GEVD and TF-DSS algorithms have also a low number of flops. The SSD-CP and JDICA methods have the highest number of flops. As also explained for interictal data, JDICA can be used when there is a few number of channels.

In summary, for denoising ictal data, TF-GEVD and TF-DSS can be certainly used for single or uncorrelated sources. For scenarios with correlated sources, we may use the combination of JDICA or CoM_2 with the TF-GEVD/TF-DSS methods to achieve good results.

5.1.4 Real interictal data

The results obtained by applying the proposed methods for denoising real interictal data shows that all methods improve the signal and greatly remove the noise. However, comparing the source localization results with that of a noise-free epoch indicates that JDICA, SSD-CP and CoM_2 denoise the provided real interictal data better than GEVD, DSS and SOBI. Since with visual inspection, the time course of all denoised signals are almost similar, the reason behind the difference between source localization results is that the mixing matrix of the epileptic sources is better estimated by some methods and small errors of the estimated mixing matrix lead to high source localization errors. Another remark that should be noticed is the good source localization result of SSD-CP which is in agreement with the source localization results obtained from simulated signals denoised with SSD-CP for high SNRs (see section 5.1.2).

5.1.5 Real ictal data

The results obtained by applying the proposed methods for denoising real ictal data shows that the time course of all denoised signals are cleaner than the noisy data, except for SSD-CP. It is noteworthy that although all methods provide fairly good denoising,

the best result is obtained with TF-DSS, which could retrieve the fast activity with the highest amplitude. CCA and TF-GEVD gave similar results for the fast activity, but with TF-GEVD some of the rhythmic activity has been removed by the denoising procedure. JDICA and CoM_2 can retrieve part of the fast activity but leave a large amount of muscle activity. These results are in agreement with the results obtained by simulated ictal EEGs with one epileptic source.

5.2 Future work

As a future work, we can study the clinical usage of the proposed methods on a large number of cases, both for the interictal and ictal data. We can also study more complicated scenarios with a higher number of epileptic patches with different locations and different correlation states. Moreover, in this thesis, we had to limit the number of channels to 32 for simulations, since the real muscle activity added to the simulated data were obtained from 32 recording channels. Using an advanced EEG recording system with large number of channels will allow us to consider simulated and real data acquired from 256 channels.

As another extension to this work, we can improve the spike detection and spike clustering methods to be able to safely use them in the GEVD and DSS-based methods for denoising interictal data with low SNRs. Moreover, to identify ictal sources by using their time-frequency plot, we can automate this procedure as we did for interictal data. We can also propose sparse denoising methods taking into account the sparsity of the ictal data in a transformed domain. For ictal scenarios with correlated sources, we can also use the combination of JDICA or CoM_2 with TF-GEVD/TF-DSS methods to achieve better results.

Moreover, to obtain better results for SSD-CP, we can test other ways of producing higher order arrays. For example, we can use other time-frequency transforms instead of wavelet. Another idea is to test the use of another way of building the HO array as proposed in [Becker et al., 2012], which consists in computing a spatial short time Fourier transform, since it appeared to give better results provided that we have a sufficient number of sensors (at least 64). We can also test other tensor models, less restrictive than the CP model.

Appendix A

Deriving the DSS framework from the EM algorithm

In this chapter, first the EM algorithm is presented and then the detailed procedure for deriving the DSS framework from the EM algorithm is explained by using the explanations in [Belouchrani and Cardoso, 1995, Bermond and Cardoso, 1999, Sarela and Valpola, 2005].

Based on the N observations of the model $\tilde{\boldsymbol{x}}[n] = \boldsymbol{G}\tilde{\boldsymbol{r}}[n] + \tilde{\boldsymbol{\nu}}[n]$, the problem consists in estimating (i) the mixing matrix $\boldsymbol{G}$ and (ii) some parameters corresponding to the distribution of the sources. The unknown parameters are denoted by $\boldsymbol{\theta}$ as follows:

$$\boldsymbol{\theta} = (\boldsymbol{G}, \eta) \tag{A.1}$$

where η represents some parameters corresponding to the distribution of the source vectors. For convenience we denote:

$$\boldsymbol{R} = [\boldsymbol{r}[1], \boldsymbol{r}[2], \ldots, \boldsymbol{r}[N]], \quad \boldsymbol{R} \in \mathbb{R}^{P' \times N} \tag{A.2}$$

$$\boldsymbol{X} = [\boldsymbol{x}[1], \boldsymbol{x}[2], \ldots, \boldsymbol{x}[N]], \quad \boldsymbol{X} \in \mathbb{R}^{P' \times N} \tag{A.3}$$

The likelihood of the observation matrix $\boldsymbol{X}$ is given by:

$$L(\boldsymbol{X}; \boldsymbol{\theta}) = p(\boldsymbol{X}; \boldsymbol{\theta}) = \int_{\mathbb{R}^{P' \times N}} p(\boldsymbol{X}, \boldsymbol{R}; \boldsymbol{\theta}) d\boldsymbol{R} = \int_{\mathbb{R}^{P' \times N}} p(\boldsymbol{X}|\boldsymbol{R}; \boldsymbol{\theta}) p(\boldsymbol{R}; \boldsymbol{\theta}) d\boldsymbol{R} \tag{A.4}$$

where $p(\boldsymbol{X}, \boldsymbol{R}; \boldsymbol{\theta})$ denotes the joint probability of $\boldsymbol{X}$ and $\boldsymbol{R}$, and $p(\boldsymbol{X}|\boldsymbol{R}; \boldsymbol{\theta})$ is the conditional probability of $\boldsymbol{X}$ given a particular $\boldsymbol{R}$ and $p(\boldsymbol{R}; \boldsymbol{\theta})$ is the a priori probability of $\boldsymbol{R}$. The maximum likelihood estimate $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is given by:

$$\hat{\boldsymbol{\theta}} = arg \max_{\boldsymbol{\theta}} L(\boldsymbol{X}; \boldsymbol{\theta}) \tag{A.5}$$

We can define the log-likelihood of the observations as follows:

$$\mathcal{L}(\boldsymbol{\theta}) = \int_{\mathbb{R}^{P' \times N}} \log p(\boldsymbol{X}, \boldsymbol{R}; \boldsymbol{\theta}) d\boldsymbol{R} = \sum_{n=1}^{N} \int_{\mathbb{R}^{P'}} \log p(\boldsymbol{x}[n], \boldsymbol{r}[n]; \boldsymbol{\theta}) d\boldsymbol{r}[n] \tag{A.6}$$

The last equality is caused from the fact that the sources and therefore the observations are i.i.d. To find the maximum likelihood estimate $\hat{\boldsymbol{\theta}}$, all partial derivatives of $\mathcal{L}(\boldsymbol{\theta})$ should be equal to zero. Therefore we have:

$$\forall j, \quad \frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \theta_j} \equiv 0 \tag{A.7}$$

where θ_j represents the j-th element of $\boldsymbol{\theta}$. The equation (A.7) does not have a closed form solution. So, a gradient-based algorithm is a common method to maximize the log-likelihood (A.6). However, since the cost of computing the gradient of the log-likelihood is very high (in each step, we should run a Monte-Carlo simulation to estimate the probabilities), using a gradient-based method is not advisable. An efficient solution is to use an iterative EM-like algorithm. In this algorithm, in each step, the parameters are re-estimated in such a way that a monotonic increase in the likelihood is guaranteed. Instead of maximizing (A.4), the likelihood is maximized via the maximization of an auxiliary function, called $Q(\boldsymbol{\theta}, \boldsymbol{\theta}')$, related to the Kullback-Leibler information measure, which is a function of the current parameters $\boldsymbol{\theta}'$ and the new parameters $\boldsymbol{\theta}$ as follows:

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}') = \int_{\mathbb{R}^{P' \times N}} p(\boldsymbol{X}, \boldsymbol{R}; \boldsymbol{\theta}') \log p(\boldsymbol{X}, \boldsymbol{R}; \boldsymbol{\theta}) d\boldsymbol{R} \tag{A.8}$$

The key property of the auxiliary function $Q(\boldsymbol{\theta}, \boldsymbol{\theta}')$ is:

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}') \geq Q(\boldsymbol{\theta}', \boldsymbol{\theta}') \Rightarrow \mathcal{L}(\boldsymbol{\theta}) \geq \mathcal{L}(\boldsymbol{\theta}') \tag{A.9}$$

which follows that a sequence $\boldsymbol{\theta}^{(i)}$ of parameters iteratively computed as:

$$\boldsymbol{\theta}^{(i+1)} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(i)}) \tag{A.10}$$

produces a monotonic increase in the likelihood. Under the assumption of independence at different time lags, the partial derivative of the function (A.8) is given by:

$$\begin{aligned} \frac{\partial Q}{\partial \theta_j}(\boldsymbol{\theta}, \boldsymbol{\theta}') &= \sum_{n=1}^{N} \int_{\mathbb{R}^{P'}} p(\boldsymbol{r}[n]|\boldsymbol{x}[n]; \boldsymbol{\theta}') \frac{\partial}{\partial \theta_i} \log p(\boldsymbol{x}[n]|\boldsymbol{r}[n]; \boldsymbol{\theta}) d\boldsymbol{r}[n] \\ &= \sum_{n=1}^{N} \mathrm{E}_{q(\boldsymbol{r}[n])}[\frac{\partial}{\partial \theta_j} \log p(\boldsymbol{x}[n]|\boldsymbol{r}[n]; \boldsymbol{\theta})] \end{aligned} \tag{A.11}$$

where $q(\boldsymbol{r}[n]) = p(\boldsymbol{r}[n]|\boldsymbol{x}[n]; \boldsymbol{\theta}')$ is the posterior distribution of the sources and θ_j corresponds to the elements of $\boldsymbol{\theta}$ related to the mixing matrix $\boldsymbol{G}$ and the noise covariance $\boldsymbol{C_\nu}$. Since the distribution of noise is assumed to be Gaussian ($\boldsymbol{\nu}[n] \sim \mathcal{N}(0, \boldsymbol{C_\nu})$), the posterior distribution of $\boldsymbol{x}[n]$ is Gaussian too:

$$\log p(\boldsymbol{x}[n]|\boldsymbol{r}[n]; \boldsymbol{\theta}) = \frac{-1}{2}(\boldsymbol{x}[n] - \boldsymbol{G}\boldsymbol{r}[n])^{\mathsf{T}} \boldsymbol{C}_{\boldsymbol{\nu}}^{-1}(\boldsymbol{x}[n] - \boldsymbol{G}\boldsymbol{r}[n]) - \frac{1}{2}\log|2\pi \boldsymbol{C_\nu}| \tag{A.12}$$

Therefore, by calculating the partial derivatives of (A.12) with respect to $\boldsymbol{G}$ and $\boldsymbol{C_\nu}$, it is proved that the equations $\frac{\partial Q}{\partial \theta_j}(\boldsymbol{\theta}, \boldsymbol{\theta}') = 0$ (for all θ_js related to the mixing matrix

$\boldsymbol{G}$ and the noise covariance $\boldsymbol{C_\nu}$) admit for the model under consideration a unique closed-form solution as follows:

$$\boldsymbol{G}^{(i+1)} = \overline{\boldsymbol{C}}_{\boldsymbol{xr}}^{(i)} \overline{\boldsymbol{C}}_{\boldsymbol{rr}}^{(i)-1} \tag{A.13}$$

$$\boldsymbol{C}_{\boldsymbol{\nu}}^{(i+1)} = \overline{\boldsymbol{C}}_{\boldsymbol{xx}} - \overline{\boldsymbol{C}}_{\boldsymbol{xr}}^{(i)} \overline{\boldsymbol{C}}_{\boldsymbol{rr}}^{(i)-1} \overline{\boldsymbol{C}}_{\boldsymbol{xr}}^{(i)^{\mathsf{T}}} \tag{A.14}$$

where $\overline{\boldsymbol{C}}_{\boldsymbol{xx}} = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{x}[n]\boldsymbol{x}[n]^{\mathsf{T}}$, and $\overline{\boldsymbol{C}}_{\boldsymbol{xr}}$ and $\overline{\boldsymbol{C}}_{\boldsymbol{rr}}$ are the time average of the non-stationary covariance matrices computed as expectations over $q(\boldsymbol{r})$:

$$\overline{\boldsymbol{C}}_{\boldsymbol{xr}} = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{x}[n] \mathrm{E}_{q(\boldsymbol{r}[n])}\left[\boldsymbol{r}[n]^{\mathsf{T}}\right], \qquad \overline{\boldsymbol{C}}_{\boldsymbol{rr}} = \frac{1}{N}\sum_{n=1}^{N} \mathrm{E}_{q(\boldsymbol{r}[n])}\left[\boldsymbol{r}[n]\boldsymbol{r}[n]^{\mathsf{T}}\right] \tag{A.15}$$

By using (A.15), the EM algorithm to estimate the mixing matrix $\boldsymbol{G} = [\boldsymbol{g}_1, \ldots, \boldsymbol{g}_{P'}]$ has two steps that are computed alternatively. These two steps are represented as follows:

$$E - step\colon \forall n,\ Compute\ \ q(\boldsymbol{r}[n]) = p(\boldsymbol{r}[n]|\boldsymbol{x}[n]; \boldsymbol{\theta}^{(i)}) \tag{A.16}$$

$$M - step\colon Find\ \boldsymbol{G}^{(i+1)} = \arg\max_{\boldsymbol{G}} \sum_{n=1}^{N} \mathrm{E}_{q(\boldsymbol{r}[n])}\left[\log p(\boldsymbol{x}[n]|\boldsymbol{r}[n]; \boldsymbol{\theta})\right] = \overline{\boldsymbol{C}}_{\boldsymbol{xr}}^{(i)} \overline{\boldsymbol{C}}_{\boldsymbol{rr}}^{(i)-1} \tag{A.17}$$

As shown above, in the E-step, the posterior distribution $q(\boldsymbol{r}[n])$ of the sources is calculated based on the known observations and current estimate of the mixing matrix (more accurately the current estimate of $\boldsymbol{\theta}$). In the M-step, the mixing matrix is fitted to the new source estimates.

Since the sources are assumed to be independent, the prior $p(\boldsymbol{r}[n])$ and the likelihood $p(\boldsymbol{x}[n]|\boldsymbol{r}[n]; \boldsymbol{\theta})$ can be factorized [Sarela and Valpola, 2005]. Therefore, the sources are independent in the posterior $q(\boldsymbol{r}[n])$ and it can be proved that the covariance matrix $\overline{\boldsymbol{C}}_{\boldsymbol{rr}}$ is diagonal. Consequently, in the M-step, $\overline{\boldsymbol{C}}_{\boldsymbol{rr}}^{-1}$ reduces to scaling of individual sources. According to the aforementioned information, since in the M-step we need to calculate only $\overline{\boldsymbol{C}}_{\boldsymbol{xr}}$, the E-step can be reduced to calculate an expectation of $\boldsymbol{r}[n]$ over $q(\boldsymbol{r}[n])$. On the other hand, since $q(\boldsymbol{r}[n])$ depends on the observation only through $\boldsymbol{G}^{-1}\boldsymbol{x}[n]$, the expectation $\mathrm{E}_{q(\boldsymbol{r}[n])}[\boldsymbol{r}[n]]$ can be interpreted as a function of $\boldsymbol{G}^{-1}\boldsymbol{x}[n] = \boldsymbol{G}^{\mathsf{T}}\boldsymbol{x}[n]$ denoted by f. So, the EM algorithm can be simplified as follows:

$$\begin{aligned} &E - step : \forall n,\ Compute\ \mathrm{E}_{q(\boldsymbol{r}[n])}[\boldsymbol{r}[n]] = f(\boldsymbol{G}^{(i)^{\mathsf{T}}}\boldsymbol{x}[n]) \\ &M - step : \boldsymbol{G}^{(i+1)} = \overline{\boldsymbol{C}}_{\boldsymbol{xr}}^{(i)} \overline{\boldsymbol{C}}_{\boldsymbol{rr}}^{(i)-1} \end{aligned} \tag{A.18}$$

Generally, in the E-step, the expectation can be computed exactly in some limited cases, e.g. when the source distributions are mixtures of Gaussian. In other cases the expectation can be approximated or can be computed by the Monte-Carlo estimation. In M-step, as mentioned before, $\overline{\boldsymbol{C}}_{\boldsymbol{rr}}^{-1}$ reduces to scaling of individual sources.

Using algorithm (A.18), the algorithm for extracting the sources one-by-one will be as follows. In this algorithm, $r_p^{(i)}[n]$ and $\boldsymbol{w}_p^{(i)}$ represent the p-th source and its corresponding mixing vector in the i-th iteration, respectively. Since $\boldsymbol{G}^{-1} = \boldsymbol{G}^{\mathsf{T}}$, the matrix

containing all estimated mixing vectors, $\boldsymbol{W} = [\boldsymbol{w}_1, ..., \boldsymbol{w}_{P'}] = \boldsymbol{G}$, is the estimated mixing matrix.

Initialize the mixing vector $\boldsymbol{w}_p^{(0)}$

Loop

1) $E - step$:

Computing the expectation of $r_p^{(i)}[n]$ *over* $q(\boldsymbol{r})$:

$$\forall k,\ r_p^{(i)+}[n] = \mathrm{E}_{q(r_p[n])}[r_p[n]] = f(\boldsymbol{w}_p^{(i)^\mathsf{T}} \boldsymbol{x}[n]) \tag{A.19}$$

2) $M - step$:

a) *Calculating the new ML estimate of* $\boldsymbol{w}_p^{(i)}$:

$$\boldsymbol{w}_p^{(i)+} = \sum_{n=1}^{N} \boldsymbol{x}[n] r_p^{(i)+}[n] \tag{A.20}$$

b) *Normalizing the mixing vector* : $\boldsymbol{w}_p^{(i+1)} = \dfrac{\boldsymbol{w}_p^{(i)+}}{||\boldsymbol{w}_p^{(i)+}||}$ (A.21)

Until convergence

For explaining the relationship of the above-mentioned algorithm with (A.18), we can say that the steps (1) and (2-a) can be combined to the following equation:

$$\boldsymbol{w}_p^{(i)+} = \sum_{n=1}^{N} \boldsymbol{x}[n] r_p^{(i)+}[n] = \sum_{n=1}^{N} \boldsymbol{x}[n] \mathrm{E}_{q(\boldsymbol{r}[n])}[r_p^{(i)}[n]] = N(\frac{1}{N} \sum_{n=1}^{N} \boldsymbol{x}[n] \mathrm{E}_{q(\boldsymbol{r}[n])}[r_p^{(i)}[n]]) \tag{A.22}$$

where $\frac{1}{N}\sum_{n=1}^{N} \boldsymbol{x}[n] \mathrm{E}_{q(\boldsymbol{r}[n])}[r_p^{(i)}[n]]$ represents one column of the covariance matrix $\overline{\boldsymbol{C}}_{\boldsymbol{xr}}^{(i)}$. In addition, the normalization in step (2-b) expresses the inversion of the covariance matrix $\overline{\boldsymbol{C}}_{\boldsymbol{rr}}^{(i)}$ in M-step.

Finally, since the noisy estimate of the p-th source in the i-th iteration can be computed by $r_p^{(i)}[n] = \boldsymbol{w}_p^{(i)^\mathsf{T}} \boldsymbol{x}[n]$ which is the mode of likelihood, the E-step may be reinterpreted in two steps and the DSS algorithm is obtained as follows:

1)$E - step$:

a) *Calculating the noisy estimate of one source* : $\forall k,\ r_p^{(i)}[n] = \boldsymbol{w}_p^{(i)^\mathsf{T}} \boldsymbol{x}[n]$ (A.23)

b) *Denoising based on the a priori information* : $\forall k,\ r_p^{(i)+}[n] = f(r_p^{(i)}[n])$ (A.24)

2) $M - step$:

a) *Calculating the new ML estimate of* $\boldsymbol{w}_p^{(i)}$:

$$\boldsymbol{w}_p^{(i)+} = \sum_{n=1}^{N} \boldsymbol{x}[n] r_p^{(i)+}[n] \tag{A.25}$$

b) *Normalizing the mixing vector* : $\boldsymbol{w}_p^{(i+1)} = \dfrac{\boldsymbol{w}_p^{(i)+}}{||\boldsymbol{w}_p^{(i)+}||}$ (A.26)

Until convergence

Appendix B

The wavelet-based spike peak detection algorithm

In this chapter, the wavelet-based spike peak detection algorithm based on the method proposed in [Senhadji et al., 1995, Senhadji et al., 1997] is given in detail. According to the definition of the Continuous Wavelet Transform (CWT), the coefficients of the wavelet transform of the observed signal from m-th channel, denoted by $\{x_m[n]\}$, with the mother wavelet $\psi(t)$ are defined as follows:

$$D_{a,b}^{(m)} = \frac{1}{f_s}\sum_{n=1}^{N} x_m[n]\psi_{a,b}^{*}[n] \tag{B.1}$$

$$\psi_{a,b}(t) \in L^2(\mathbb{R}), \psi_{a,b}[n] = \frac{1}{\sqrt{|a|}}\psi(\frac{n-b}{a}), a \neq 0, b \in \mathbb{R} \tag{B.2}$$

In this algorithm, complex Morlet wavelet is used as follows:

$$\psi(t) = C(1+\cos 2\pi f_0 t)e^{2i\pi\ell f_0 t}, |t| \leq \frac{1}{2f_0}, \ell \notin \{-1,0,1\} \tag{B.3}$$

where ℓ sets up the number of oscillations of the complex part, f_0 is the normalized frequency and C is a normalization coefficient ($\|\psi\| = 1$) [Senhadji et al., 1995].

In the first stage, S_1, by using the wavelet coefficients $D_{a,b}^{(m)}$ in the m-th channel, we calculate the measure $T_n^{(m)}$ for each time sample n and compare it with the decision threshold $\lambda_1^{(m)}$ as follows:

$$T_n^{(m)} = \sum_{i=1}^{I} \alpha_i |D_{a_i,n}^{(m)}|^2 > \lambda_1^{(m)} \tag{B.4}$$

where parameters I and a_is are chosen such that the coefficients $D_{a_i,n}^{(m)}$ approximately cover the bandwidth of the spikes. Consequently, at the output of the stage S_1, the transient signals are enhanced when compared to the background activity without any distinction between interictal spikes and muscular activity [Senhadji et al., 1995]. On the other hand, the experiments show that for a muscular artifact, compared with a spike, the squared modulus increases for high resolution scales [Senhadji et al., 1995]. This remark gives the idea to build a decision parameter $G_n^{(m)}$ which is calculated as the mean gravity center of the abscissa $\frac{1}{a_i}$'s weighted by $|D_{a_i,n}^{(m)}|^2$ on the detection interval

around each sample point. Therefore, in the stage S_2, the measure $G_n^{(m)}$ is calculated in the m-th channel as follows [Senhadji et al., 1995]:

$$G_n^{(m)} = \frac{\sum_{j=n-w}^{n+w} g_j^{(m)}}{2w+1} \quad where \;\; g_j^{(m)} = \frac{\sum_{i=1}^{I} (\frac{1}{a_i}) |D_{a_i,j}^{(m)}|^2}{\sum_{i=1}^{I} |D_{a_i,j}^{(m)}|^2} \tag{B.5}$$

to separate the interictal spikes from muscular activity. Experimentally, this quantity takes distinct values in the presence of an artifact or a spike. By comparing $G_n^{(m)}$ with a threshold $\lambda_2^{(m)}$, the stage S_2 separates the useful signals from artifacts [Senhadji et al., 1995]. In [Senhadji et al., 1995], a practical method to choose the adaptive thresholds $\lambda_1^{(m)}$ and $\lambda_2^{(m)}$ was proposed.

Appendix C

Values of coefficients in the JDICA algorithm

$$d_{40} = C_{p,p,p,p,\boldsymbol{x}}, \quad d_{31} = 4C_{P',p,p,p,\boldsymbol{x}}, \quad d_{22} = 6C_{P',P',p,p,\boldsymbol{x}}$$
$$d_{13} = 4C_{P',P',P',p,\boldsymbol{x}}, \quad d_{04} = C_{P',P',P',P',\boldsymbol{x}}$$
$$d_{30} = 4 \sum_{\substack{i_1 = 1 \\ i_1 \neq p}}^{P'-1} C_{p,p,p,i_1,\boldsymbol{x}} g_{i_1}^{(1-)}, \quad d_{21} = 12 \sum_{\substack{i_1 = 1 \\ i_1 \neq p}}^{P'-1} C_{P',p,p,i_1,\boldsymbol{x}} g_{i_1}^{(1-)}$$
$$d_{12} = 12 \sum_{\substack{i_1 = 1 \\ i_1 \neq p}}^{P'-1} C_{P',P',p,i_1,\boldsymbol{x}} g_{i_1}^{(1-)}, \quad d_{03} = 4 \sum_{\substack{i_1 = 1 \\ i_1 \neq p}}^{P'-1} C_{P',P',P',i_1,\boldsymbol{x}} g_{i_1}^{(1-)}$$
$$d_{20} = 6 \sum_{\substack{i_1 = 1 \\ i_1 \neq p}}^{P'-1} \Big(C_{p,p,i_1,i_1,\boldsymbol{x}} (g_{i_1}^{(1-)})^2 + 2 \sum_{\substack{i_2 = 1 \\ i_2 \neq p}}^{i_1-1} C_{p,p,i_1,i_2,\boldsymbol{x}} g_{i_1}^{(1-)} g_{i_2}^{(1-)} \Big)$$
$$d_{11} = 12 \sum_{\substack{i_1 = 1 \\ i_1 \neq p}}^{P'-1} \Big(C_{P',P',p,i_1,\boldsymbol{x}} (g_{i_1}^{(1-)})^2 + 2 \sum_{\substack{i_2 = 1 \\ i_2 \neq p}}^{i_1-1} C_{P',p,i_1,i_2,\boldsymbol{x}} g_{i_1}^{(1-)} g_{i_2}^{(1-)} \Big)$$
$$d_{02} = 6 \sum_{\substack{i_1 = 1 \\ i_1 \neq p}}^{P'-1} \Big(C_{P',P',i_1,i_1,\boldsymbol{x}} (g_{i_1}^{(1-)})^2 + 2 \sum_{\substack{i_2 = 1 \\ i_2 \neq p}}^{i_1-1} C_{P',P',i_1,i_2,\boldsymbol{x}} g_{i_1}^{(1-)} g_{i_2}^{(1-)} \Big)$$
$$d_{10} = 4 \sum_{\substack{i_1 = 1 \\ i_1 \neq p}}^{P'-1} \Big(C_{P',i_1,i_1,i_1,\boldsymbol{x}} (g_{i_1}^{(1-)})^3 + 3 \sum_{\substack{i_2 = 1 \\ i_2 \neq p \\ i_2 \neq i_1}}^{P'-1} C_{k,i_1,i_1,i_2,\boldsymbol{x}} (g_{i_1}^{(1-)})^2 g_{i_2}^{(1-)}$$
$$+ 6 \sum_{\substack{i_2 = 1 \\ i_2 \neq p}}^{i_1-1} \sum_{\substack{i_3 = 1 \\ i_3 \neq p}}^{i_2-1} C_{p,i_1,i_2,i_3,\boldsymbol{x}} g_{i_1}^{(1-)} g_{i_2}^{(1-)} g_{i_3}^{(1-)} \Big)$$

$$d_{01} = 4 \sum_{\substack{i_1=1\\ i_1\neq p}}^{P'-1} \Big(C_{P',i_1,i_1,i_1,\boldsymbol{x}} (g_{i_1}^{(1-)})^3 + 3 \sum_{\substack{i_2=1\\ i_2\neq p\\ i_2\neq i_1}}^{P'-1} C_{P',i_1,i_1,i_2,\boldsymbol{x}} (g_{i_1}^{(1-)})^2 g_{i_2}^{(1-)}$$

$$+ 6 \sum_{\substack{i_2=1\\ i_2\neq k}}^{i_1-1} \sum_{\substack{i_3=1\\ i_3\neq p}}^{i_2-1} C_{P',i_1,i_2,i_3,\boldsymbol{x}} g_{i_1}^{(1-)} g_{i_2}^{(1-)} g_{i_3}^{(1-)} \Big)$$

$$d_{00} = \sum_{\substack{i_1=1\\ i_1\neq p}}^{P'-1} \sum_{\substack{i_2=1\\ i_2\neq p}}^{P'-1} \sum_{\substack{i_3=1\\ i_3\neq p}}^{P'-1} \sum_{\substack{i_4=1\\ i_4\neq p}}^{P'-1} C_{i_1,i_2,i_3,i_4,\boldsymbol{x}} g_{i_1}^{(1-)} g_{i_2}^{(1-)} g_{i_3}^{(1-)} g_{i_4}^{(1-)} \tag{C.1}$$

where $C_{i,j,k,\ell,\boldsymbol{x}}$ denoted the (i,j,k,ℓ)-th average FO cumulant of the process $\{\tilde{\boldsymbol{x}}[n]\}$.

$$e_0 = d_{11}(g_{P'}^{(1-)})^2 - d_{11}(g_p^{(1-)})^2 + d_{13}(g_{P'}^{(1-)})^4 - d_{31}(g_p^{(1-)})^4 + (2d_{22} - 4d_{04}) g_p^{(1-)} (g_{P'}^{(1-)})^3 + (3d_{31} - 3d_{13})(g_p^{(1-)})^2 (g_{P'}^{(1-)})^2 + (2d_{20} - 2d_{02}) g_p^{(1-)} g_{P'}^{(1-)} + (4d_{40} - 2d_{22})(g_p^{(1-)})^3 g_{P'}^{(1-)}$$

$$e_1 = (2d_{02} - 2d_{20})(g_p^{(1-)})^2 + (2d_{20} - 2d_{02})(g_{P'}^{(1-)})^2 + (2d_{22} - 4d_{04})(g_{P'}^{(1-)})^4 + (2d_{22} - 4d_{40})(g_p^{(1-)})^4 + (6d_{31} - 10d_{13}) g_p^{(1-)} (g_{P'}^{(1-)})^3 + (6d_{13} - 10d_{31})(g_p^{(1-)})^3 g_{P'}^{(1-)} + (12d_{04} - 12d_{22} + 12d_{40})(g_p^{(1-)})^2 (g_{P'}^{(1-)})^2 - 4d_{11} g_p^{(1-)} g_{P'}^{(1-)}$$

$$e_2 = (3d_{31} - 3d_{13})(g_p^{(1-)})^4 + (3d_{31} - 3d_{13})(g_{P'}^{(1-)})^4 + (12d_{04} - 12d_{22} + 12d_{40}) g_p^{(1-)} (g_{P'}^{(1-)})^3 + (12d_{22} - 12d_{04} - 12d_{40})(g_p^{(1-)})^3 g_{P'}^{(1-)} + (18d_{13} - 18d_{31})(g_p^{(1-)})^2 (g_{P'}^{(1-)})^2$$

$$e_3 = (2d_{02} - 2d_{20})(g_p^{(1-)})^2 + (4d_{04} - 2d_{22})(g_p^{(1-)})^4 + (2d_{20} - 2d_{02})(g_{P'}^{(1-)})^2 + (4d_{40} - 2d_{22})(g_{P'}^{(1-)})^4 + (6d_{13} - 10d_{31}) g_p^{(1-)} (g_{P'}^{(1-)})^3 + (6d_{31} - 10d_{13})(g_p^{(1-)})^3 g_{P'}^{(1-)} + (12d_{22} - 12d_{04} - 12d_{40})(g_p^{(1-)})^2 (g_{P'}^{(1-)})^2 - 4d_{11} g_p^{(1-)} g_{P'}^{(1-)}$$

$$e_4 = d_{11}(g_p^{(1-)})^2 + d_{13}(g_p^{(1-)})^4 - d_{11}(g_{P'}^{(1-)})^2 - d_{31}(g_{P'}^{(1-)})^4 + (4d_{04} - 2d_{22})(g_p^{(1-)})^3 g_{P'}^{(1-)} + (2d_{22} - 4d_{40}) g_p^{(1-)} (g_{P'}^{(1-)})^3 + (3d_{31} - 3d_{13})(g_p^{(1-)})^2 (g_{P'}^{(1-)})^2 + (2d_{02} - 2d_{20}) g_p^{(1-)} g_{P'}^{(1-)}$$

$$f_0 = d_{10} g_{P'}^{(1-)} - d_{01} g_p^{(1-)} + d_{12}(g_{P'}^{(1-)})^3 - d_{21}(g_p^{(1-)})^3 + (2d_{21} - 3d_{03}) g_p^{(1-)} (g_{P'}^{(1-)})^2 + (3d_{30} - 2d_{12})(g_p^{(1-)})^2 g_{P'}^{(1-)}$$

$$f_1 = (2d_{12} - 3d_{30})(g_p^{(1-)})^3 + (2d_{21} - 3d_{03})(g_{P'}^{(1-)})^3 + (6d_{03} - 7d_{21})(g_p^{(1-)})^2 g_{P'}^{(1-)} + (6d_{30} - 7d_{12}) g_p^{(1-)} (g_{P'}^{(1-)})^2 - d_{10} g_p^{(1-)} - d_{01} g_{P'}^{(1-)}$$

$$f_2 = d_{10} g_{P'}^{(1-)} - d_{01} g_p^{(1-)} + (2d_{21} - 3d_{03})(g_p^{(1-)})^3 + (3d_{30} - 2d_{12})(g_{P'}^{(1-)})^3 + (6d_{03} - 7d_{21}) g_p^{(1-)} (g_{P'}^{(1-)})^2 + (7d_{12} - 6d_{30})(g_p^{(1-)})^2 g_{P'}^{(1-)}$$

$$f_3 = (2d_{12} - 3d_{30}) g_p^{(1-)} (g_{P'}^{(1-)})^2 - d_{10} g_p^{(1-)} - d_{12}(g_p^{(1-)})^3 + (2d_{21} - 3d_{03})(g_p^{(1-)})^2 g_{P'}^{(1-)} - d_{21}(g_{P'}^{(1-)})^3 - d_{01} g_{P'}^{(1-)}$$

Appendix D

Values of coefficients in the SSD-CP algorithm

To calculate the values of coefficients in the SSD-CP algorithm, first we define the values v_1 to v_{10}, u_1 to u_3 and w_1 to w_3 as a functions of elements of matrices $\boldsymbol{Y}^{(r)}$, $1 \leq r \leq R$ as follows:

$$v_1 = \sum_{r=1}^{R} |Y_{p,p}^{(r)}|^2, \quad v_2 = \sum_{r=1}^{R} Y_{p,p}^{(r)} Y_{p,q}^{(r)*}, \quad v_3 = \sum_{r=1}^{R} Y_{p,p}^{(r)} Y_{q,p}^{(r)*}, \quad v_4 = \sum_{r=1}^{R} Y_{p,p}^{(r)} Y_{q,q}^{(r)*}$$
$$v_5 = \sum_{r=1}^{R} Y_{p,q}^{(r)} Y_{q,q}^{(r)*}, \quad v_6 = \sum_{r=1}^{R} |Y_{p,q}^{(r)}|^2, \quad v_7 = \sum_{r=1}^{R} Y_{p,q}^{(r)} Y_{q,p}^{(r)*}, \quad v_8 = \sum_{r=1}^{R} |Y_{q,p}^{(r)}|^2$$
$$v_9 = \sum_{r=1}^{R} Y_{q,p}^{(r)} Y_{q,q}^{(r)*}, \quad v_{10} = \sum_{r=1}^{R} |Y_{q,q}^{(r)}|^2$$

$$u_1 = \sum_{r=1}^{R} \sum_{k=p+1}^{q-1} |Y_{kp}^{(r)}|^2, \; u_2 = \sum_{r=1}^{R} \sum_{k=p+1}^{q-1} Y_{kp}^{(r)} Y_{kq}^{(r)*}, \; u_3 = \sum_{r=1}^{R} \sum_{k=p+1}^{q-1} |Y_{kq}^{(r)}|^2$$
$$w_1 = \sum_{r=1}^{R} \sum_{k=p+1}^{q-1} |Y_{qk}^{(r)}|^2 \; w_2 = \sum_{r=1}^{R} \sum_{k=p+1}^{q-1} Y_{qk}^{(r)} Y_{pk}^{(r)*}, \; w_3 = \sum_{r=1}^{R} \sum_{k=p+1}^{q-1} |Y_{pk}^{(r)}|^2 \tag{D.1}$$

Then we define b_0–b_4, c_0–c_4, t_0–t_7 and s_0–s_5 as follows:

$$\begin{aligned}
b_0 &= i(v_5^* - v_2 - u_2 + w_2) \\
b_1 &= -i2v_7^* \\
b_2 &= i(v_3^* - v_9 - u_2 + w_2) \\
b_3 &= -2v_1 + 4v_6 + 4\mathrm{Re}(v_4) - 2v_{10} + 2u_3 - 2u_1 + 2w_3 - 2w_1 \\
b_4 &= 2v_1 - 4\mathrm{Re}(v_4) - 4v_8 + 2v_{10} + 2u_3 - 2u_1 + 2w_3 - 2w_1
\end{aligned} \tag{D.2}$$

$$\begin{aligned}
&c_0 = b_1^2, \qquad c_1 = -4b_2^* b_0^* \\
&c_2 = 2|b_1|^2 - 8\mathrm{Re}(b_2.b_0^*), \qquad c_3 = c_1^*, \qquad c_4 = c_0^*
\end{aligned} \tag{D.3}$$

$$
\begin{aligned}
t_0 &= 2ib_2^*b_1^{*2}b_0^{*2} - 2ib_0b_1^{*4} - 2ib_1^{*2}b_0^{*3} + b_3b_1^{*3}b_0^* \\
t_1 &= -4ib_2^{*2}b_0^{*3} + 4ib_2^*b_0^{*4} - 4ib_2b_1^{*2}b_0^{*2} + b_0b_3b_1^{*3} + 2ib_1b_1^{*3}b_0^* + b_4b_1^*b_0^{*3} + 12ib_0b_2^*b_1^{*2}b_0^* \\
&\quad - 3b_3b_2^*b_1^*b_0^{*2} \\
t_2 &= 8ib_2b_0^{*4} + 10ib_0^2b_2^*b_1^{*2} - 16ib_0b_2^{*2}b_0^{*2} + 6ib_0^2b_1^{*2}b_0^* - 6ib_1b_0b_1^{*3} + 4ib_2b_2^*b_0^{*3} + 4ib_0b_2^*b_0^{*3} \\
&\quad - 6ib_1b_1^*b_0^{*3} - 2ib_1b_2^*b_1^*b_0^{*2} - 3b_2b_3b_1^*b_0^{*2} + 3b_0b_4b_1^*b_0^{*2} + 3b_1b_3b_1^{*2}b_0^* - 6b_0b_3b_2^*b_1^*b_0^* \\
t_3 &= 8ib_2^2b_0^{*3} + 4ib_0^3b_1^{*2} - 12ib_0^2b_2^*b_0^{*2} - 20ib_0^2b_2^{*2}b_0^* + 6ib_1^2b_1^{*2}b_0^* + 20ib_2b_0b_0^{*3} + b_1b_4b_0^{*3} \\
&\quad + 3b_1b_0b_3b_1^{*2} + 4ib_2b_0b_2^*b_0^{*2} - 14ib_2b_1b_1^*b_0^{*2} - 6ib_1b_0b_1^*b_0^{*2} - 3b_1b_3b_2^*b_0^{*2} + 3b_0^2b_4b_1^*b_0^* \\
&\quad + 4ib_2b_0^2b_1^{*2} - 3b_0^2b_3b_2^*b_1^* - 6b_2b_0b_3b_1^*b_0^* + 12ib_1b_0b_2^*b_1^*b_0^* \\
t_4 &= t_3^*, \qquad t_5 = t_2^*, \qquad t_6 = t_1^*, \qquad t_7 = t_0^*
\end{aligned}
\tag{D.4}
$$

$$
\begin{aligned}
s_0 &= -2ib_0b_1^{*3} - 2ib_1^*b_0^{*3} + b_3b_1^{*2}b_0^* + 2ib_2^*b_1^*b_0^{*2} \\
s_1 &= b_4b_0^{*3} + b_0b_3b_1^{*2} - 4ib_2b_1^*b_0^{*2} + 2ib_1b_1^{*2}b_0^* - b_3b_2^*b_0^{*2} + 8ib_0b_2^*b_1^*b_0^* \\
s_2 &= -4ib_1b_0^{*3} - 4ib_1b_0b_1^{*2} - b_2b_3b_0^{*2} + 3b_0b_4b_0^{*2} + 6ib_0^2b_2^*b_1^* + 6ib_0^2b_1^*b_0^* - 4ib_2b_0b_1^*b_0^* \\
&\quad - 2b_0b_3b_2^*b_0^* + 2b_1b_3b_1^*b_0^* \\
s_3 &= s_2^*, \qquad s_4 = s_1^*, \qquad s_5 = s_0^*
\end{aligned}
\tag{D.5}
$$

Finally, we can calculate the coefficients A_ℓ, $(0 \le \ell \le 14)$ according to the predefined parameters:

$$
A_\ell = \sum_{k=\max(0,\ell-7)}^{\min(7,\ell)} t_k t_{\ell-k} - \sum_{i=\max(0,\ell-9)}^{\min(5,\ell)} \sum_{j=\max(0,\ell-9,\ell-i-4)}^{\min(5,\ell-i)} s_i s_j c_{\ell-i-j}
$$

Bibliography

[Ami, 1998] (1998). Amino acids fluorescence data, last accessed January, 2015, http://www.models.kvl.dk/amino_acid_fluo.

[Abtahi et al., 2014] Abtahi, F., Seoane, F., Lindecrantz, K., and Löfgren, N. (2014). Elimination of ECG artefacts in foetal EEG using ensemble average subtraction and wavelet denoising methods: A simulation. In *XIII Mediterranean Conference on Medical and Biological Engineering and Computing 2013*, pages 551–554.

[Acar et al., 2007] Acar, E., Aykut-Bingol, C., Bingol, H., Bro, R., and Yener, B. (2007). Multiway analysis of epilepsy tensors. *Bioinformatics*, 23(13):i10–i18.

[Acar et al., 2011] Acar, E., Dunlavy, D., and Kolda, T. (2011). A scalable optimization approach for fitting canonical tensor decompositions. *Journal of Chemometrics*, 25(2):67–86.

[Acir et al., 2005] Acir, N., Oztura, I., Kuntalp, M., Baklan, B., and Guzelis, C. (2005). Automatic detection of epileptiform events in EEG by a three-stage procedure based on artificial neural networks. *IEEE Transactions on Biomedecal Engineering*, 52(1):30–40.

[Ahmadian et al., 2013] Ahmadian, P., Sanei, S., Ascari, L., Gonzalez Villanueva, L., and Umilta, M. A. (2013). Constrained blind source extraction of readiness potentials from EEG. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 21(4):567–575.

[Akhbari et al., 2010] Akhbari, M., Babaie-Zadeh, M., Fatemizadeh, E., and Jutten, C. (2010). An entropy based method for activation detection of functional MRI data using independent component analysis. In *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, pages 2014–2017.

[Akhtar et al., 2012] Akhtar, M. T., Mitsuhashi, W., and James, C. J. (2012). Employing spatially constrained ICA and wavelet denoising, for automatic removal of artifacts from multichannel EEG data. *Signal Processing*, 92(2):401–416.

[Alarcon et al., 1995] Alarcon, G., Binnie, C., Elwes, R., and Polkey, C. (1995). Power spectrum and intracranial EEG patterns at seizure onset in partial epilepsy. *Electroencephalography and Clinical Neurophysiology*, 94(5):326–337.

[Albera et al., 2012] Albera, L., Kachenoura, A., Comon, P., Karfoul, A., Wendling, F., Senhadji, L., and Merlet, I. (2012). ICA-based EEG denoising: a comparative

analysis of fifteen methods. *Special Issue of the Bulletin of the Polish Academy of Sciences*, pages 407–418.

[Ans et al., 1985] Ans, B., Hérault, J., and Jutten, C. (1985). Adaptive neural architectures: detection of primitives. *Proc. of COGNITIVA'85*, pages 593–597.

[Bajaj and Pachori, 2012] Bajaj, V. and Pachori, R. B. (2012). Classification of seizure and nonseizure EEG signals using empirical mode decomposition. *IEEE Transactions on Information Technology in Biomedicine*, 16(6):1135–1142.

[Bancaud et al., 1962] Bancaud, J., Bonis, A., Talairach, J., Schaub, C., Colomb, D., and Szikla, G. (1962). Stereotaxic functional exploration of epilepsies of supplementary areas of mesial surface of hemisphere. In *Electroencephalography and Clinical Neurophysiology*, volume 14, page 788. ELSEVIER SCI Ireland Ltd Customer Relations Manager, Bay 15, Shannon Industrial Estate Co, Clare, Ireland.

[Becker et al., 2014a] Becker, H., Albera, L., Comon, P., Gribonval, R., Wendling, F., Merlet, I., et al. (2014a). A performance study of various brain source imaging approaches. In *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, pages 5910–5914.

[Becker et al., 2014b] Becker, H., Albera, L., Comon, P., Haardt, M., Birot, G., Wendling, F., Gavaret, M., Bénar, C., and Merlet, I. (2014b). EEG extended source localization: tensor-based vs. conventional methods. *NeuroImage*, 96:143–157.

[Becker et al., 2010] Becker, H., Comon, P., Albera, L., Haardt, M., and Merlet, I. (2010). Multiway Space-Time-Wave-Vector analysis for source localization and extraction. In *EUSIPCO 10, XVIII European Signal Processing Conference*, pages 1349–1353, Aalborg, Denmark.

[Becker et al., 2012] Becker, H., Comon, P., Albera, L., Haardt, M., and Merlet, I. (2012). Multi-way Space-Time-Wave-Vector analysis for EEG source separation. *Elsevier Signal Processing*, 92(4):1021–1031.

[Belouchrani et al., 1997] Belouchrani, A., Abed-Meraim, K., Cardoso, J. F., and Moulines, E. (1997). A blind source separation technique using second-order statistics. *IEEE Transactions On Signal Processing*, 45(2):434–444.

[Belouchrani and Cardoso, 1995] Belouchrani, A. and Cardoso, J. (1995). Maximum likelihood source separation by the expectation-maximization technique: deterministic and stochastic implementation. In *International Symposium on Nonlinear Theory and Applications NOLTA95*, pages 49–53, Tokyo, Japan.

[Bermond and Cardoso, 1999] Bermond, O. and Cardoso, J. (1999). Approximate likelihood for noisy mixtures. In *The First International Workshop on Independent Component Analysis and Signal Separation (ICA99)*, pages 325–330, Aussois, France.

[Bezdek et al., 1984] Bezdek, J. C., Ehrlich, R., and Full, W. (1984). FCM: The fuzzy c-means clustering algorithm. *Computers and Geosciences*, 10(2):191–203.

[Blanco-Velasco et al., 2008] Blanco-Velasco, M., Weng, B., and Barner, K. E. (2008). ECG signal denoising and baseline wander correction based on the empirical mode decomposition. *Computers in Biology and Medicine*, 38(1):1–13.

[Bono et al., 2014] Bono, V., Jamal, W., Das, S., and Maharatna, K. (2014). Artifact reduction in multichannel pervasive EEG using hybrid WPT-ICA and WPT-EMD signal decomposition techniques.

[Boubchir and Boashash, 2013] Boubchir, L. and Boashash, B. (2013). Wavelet denoising based on the MAP estimation using the BKF prior with application to images and EEG signals. *IEEE Transactions on Signal Processing*, 61(8):1880–1894.

[Bourien et al., 2007] Bourien, J., Ruel, J., Senhadji, L., and Puel, J. L. (2007). Comparison of three spike detectors dedicated to single unit action potentials of the auditory nerve. In *Proceedings of the 29th Annual International Conference of the IEEE EMBS*, pages 1430–1433, Lyon, France.

[Bro, 1997] Bro, R. (1997). PARAFAC, tutorial and applications. *Chemometrics and Intelligent Laboratory Systems*, 38(2):149–171.

[Bro, 1998] Bro, R. (1998). *Multi-way Analysis in the Food Industry: Models, Algorithms, and Applications*. PhD thesis, University of Amsterdam, Amstrdam, Netherlands.

[Cardoso and Souloumiac, 1993] Cardoso, J.-F. and Souloumiac, A. (1993). Blind beamforming for non-Gaussian signals. In *IEE Proceedings F (Radar and Signal Processing)*, volume 140, pages 362–370.

[Carroll and Chang, 1970] Carroll, J. D. and Chang, J.-J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of Eckart-Young decomposition. *Psychometrika*, 35(3):283–319.

[Castillo et al., 2013] Castillo, E., Morales, D. P., García, A., Martínez-Martí, F., Parrilla, L., and Palma, A. J. (2013). Noise suppression in ECG signals through efficient one-step wavelet processing techniques. *Journal of Applied Mathematics*, 2013.

[Chaozhu et al., 2013] Chaozhu, Z., Siyao, L., and Abdullah, A. K. (2013). A new blind source separation method to remove artifact in EEG signals. In *Third International Conference on Instrumentation, Measurement, Computer, Communication and Control (IMCCC)*, pages 1430–1433.

[Chatrian, 1985] Chatrian, G. (1985). Ten percent electrode system for topographic studies of spontaneous and evoked EEG activity. *The American Journal of EEG Technology*, 25:83–92.

[Chen et al., 2011] Chen, Y., Han, D., and QI, L. (2011). New ALS methods with extrapolating search directions and optimal step size for complex-valued tensor decompositions. *IEEE Transactions on Signal Processing*, 59(12):5888–5898.

[Cherkassky and Kilts, 2001] Cherkassky, V. and Kilts, S. (2001). Myopotential denoising of ECG signals using wavelet thresholding methods. *Neural Networks*, 14(8):1129–1137.

[Comon, 1994] Comon, P. (1994). Independent Component Analysis, a new concept ? *Signal Processing*, 36(3):287–314.

[Comon and Jutten, 2010] Comon, P. and Jutten, C. (2010). *Handbook of blind source separation*. Academic Press.

[Comon et al., 2009] Comon, P., Luciani, X., and Almeida, A. L. F. D. (2009). Tensor decompositions, alternating least squares and other tales. *Journal of Chemometrics*, 23(7-8):393–405.

[Congedo et al., 2008] Congedo, M., Gouy-Pailler, C., and Jutten, C. (2008). On the blind source separation of human electroencephalogram by approximate joint diagonalization of second order statistics. *Clinical Neurophysiology*, 119(12):2677–2686.

[Cosandier-Rimélé et al., 2007] Cosandier-Rimélé, D., Badier, J.-M., Chauvel, P., and Wendling, F. (2007). Modeling and interpretation of scalp-EEG and depth-EEG signals during interictal activity. In *29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 1, pages 4277–4280.

[Cosandier-Rimélé et al., 2012] Cosandier-Rimélé, D., Bartolomei, F., Merlet, I., Chauvel, P., and Wendling, F. (2012). Recording of fast activity at the onset of partial seizures: Depth EEG vs. scalp EEG. *NeuroImage*, 59(4):3474–3487.

[Cosandier-Rimélé et al., 2010] Cosandier-Rimélé, D., Merlet, I., Bartolomei, F., Badier, J., and Wendling, F. (2010). Computational modeling of epileptic activity: from cortical sources to EEG signals. *Journal of Clinical Neurophysiology*, 27(6):465–470.

[Cosandier-Rimélé et al., 2008] Cosandier-Rimélé, D., Merlet, I., Badier, J., Chauvel, P., and Wendling, F. (2008). The neuronal sources of EEG: Modeling of simultaneous scalp and intracerebral recordings in epilepsy. *NeuroImage*, 42(1):135–146.

[Crespel et al., 2006] Crespel, A., Gélisse, P., Bureau, M., and Genton, P. (2006). Atlas d'électroencéphalographie, tome II. les épilepsies: de l'EEG aux syndromes, john libbey eurotext, montrouge.

[Crespo-Garcia et al., 2008] Crespo-Garcia, M., Atienza, M., and Cantero, J. L. (2008). Muscle artifact removal from human sleep EEG by using independent component analysis. *Annals of Biomedical Engineering*, 36(3):467–475.

[De Clercq et al., 2006] De Clercq, W., Vergult, A., Vanrumste, B., Van Paesschen, W., and Van Huffel, S. (2006). Canonical correlation analysis applied to remove muscle artifacts from the electroencephalogram. *IEEE Transactions on Biomedical Engineering*, 53(12):2583–2587.

[De Lathauwer, 2006] De Lathauwer, L. (2006). A link between the canonical decomposition in multilinear algebra and simultaneous matrix diagonalization. *SIAM Journal On Matrix Analysis and Applications*, 28(3):642–666.

[De Lathauwer et al., 2007] De Lathauwer, L., Castaing, J., and Cardoso, J. (2007). Fourth-order cumulant-based blind identification of underdetermined mixtures. *IEEE Transactions on Signal Processing*, 55(6):2965–2973.

[De Lathauwer et al., 2000] De Lathauwer, L., De Moor, B., and Vandewalle, J. (2000). Fetal electrocardiogram extraction by blind source subspace separation. *IEEE Transactions on Biomedical Engineering*, 47(5):567–572.

[De Lathauwer et al., 2000] De Lathauwer, L., De Moor, B., and Vandewalle, J. (2000). A multilinear singular value decomposition. *SIAM Journal on Matrix Analysis and Applications*, 21(4):1253–1278.

[De Lathauwer et al., 2004] De Lathauwer, L., De Moor, B., and Vandewalle, J. (2004). Computation of the canonical decomposition by means of a simultaneous generalized Schur decomposition. *SIAM Journal on Matrix Analysis and Applications*, 26(2):295–327.

[De Vos et al., 2007] De Vos, M., De Lathauwer, L., Vanrumste, B., Van Huffel, S., and Van Paesschen, W. (2007). Canonical decomposition of ictal scalp EEG and accurate source localisation: Principles and simulation study. *Computational Intelligence and Neuroscience*, 2007.

[De Vos et al., 2006] De Vos, M., De Lathauwer, L., Vergult, A., De Clercq, W., Van Paesschen, W., and Van Huffel, S. (2006). Jacobi iterations for spatially constrained independent component analysis. In *Proceeding of the 7th International Conference on Methematics in Signal Processing (IMA 2007), Cirencester, England.*

[Deburchgraeve et al., 2009] Deburchgraeve, W., Cherian, P., De Vos, M., Swarte, R., Blok, J., Visser, G. H., Govaert, P., and Van Huffel, S. (2009). Neonatal seizure localization using PARAFAC decomposition. *Clinical Neurophysiology*, 120(10):1787–1796.

[Delfosse and Loubaton, 1995] Delfosse, N. and Loubaton, P. (1995). Adaptive blind separation of independent sources: a deflation approach. *Signal Processing*, 45(1):59–83.

[Delorme et al., 2007] Delorme, A., Sejnowski, T., and Makeig, S. (2007). Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *NeuroImage*, 34(4):1443–1449.

[Dempster et al., 1977] Dempster, A., Laird, N., and Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological),*, 39(1):1–38.

[Elul, 1972] Elul, R. (1972). The genesis of the EEG. *International Review of Neurobiology*, 15(2):227–272.

[Fatima et al., 2013] Fatima, Z., Quraan, M. A., Kovacevic, N., and McIntosh, A. R. (2013). ICA-based artifact correction improves spatial localization of adaptive spatial filters in MEG. *NeuroImage*, 78:284–294.

[Fitzgibbon et al., 2007] Fitzgibbon, S. P., Powers, D. M., Pope, K. J., and Clark, C. R. (2007). Removal of EEG noise and artifact using blind source separation. *Journal of Clinical Neurophysiology*, 24(3):232–243.

[Fleureau et al., 2011a] Fleureau, J., Kachenoura, A., Albera, L., Nunes, J.-C., and Senhadji, L. (2011a). Multivariate empirical mode decomposition and application to multichannel filtering. *Signal Processing*, 91(12):2783–2792.

[Fleureau et al., 2011b] Fleureau, J., Nunes, J.-C., Kachenoura, A., Albera, L., and Senhadji, L. (2011b). Turning tangent empirical mode decomposition: a framework for mono-and multivariate signals. *IEEE Transactions on Signal Processing*, 59(3):1309–1316.

[Galka et al., 2011] Galka, A., Wong, K. F. K., Ozaki, T., Muhle, H., Stephani, U., and Siniatchkin, M. (2011). Decomposition of neurological multivariate time series by state space modelling. *Bulletin of Mathematical Biology*, 73(2):285–324.

[Ge et al., 2014] Ge, S., Han, M., and Hong, X. (2014). A fully automatic ocular artifact removal from EEG based on fourth-order tensor method. *Biomedical Engineering Letters*, 4(1):55–63.

[Gonçalves et al., 2003] Gonçalves, S. I., de Munck, J. C., Verbunt, J. P., Bijma, F., Heethaar, R. M., and Lopes da Silva, F. (2003). In vivo measurement of the brain and skull resistivities using an EIT-based method and realistic models for the head. *IEEE Transactions on Biomedical Engineering*, 50(6):754–767.

[Goncharova et al., 2003] Goncharova, I., McFarland, D. J., Vaughan, T. M., and Wolpaw, J. R. (2003). EMG contamination of EEG: spectral and topographical characteristics. *Clinical Neurophysiology*, 114(9):1580–1593.

[Gonzalez and Woods, 2002] Gonzalez, R. C. and Woods, R. E. (2002). *Digital image processing*. Prentice Hall, Upper Saddle River, New Jersey 07458, 2nd edition.

[Griffanti et al., 2014] Griffanti, L., Salimi-Khorshidi, G., Beckmann, C. F., Auerbach, E. J., Douaud, G., Sexton, C. E., Zsoldos, E., Ebmeier, K. P., Filippini, N., Mackay, C. E., et al. (2014). ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging. *NeuroImage*, 95:232–247.

[Guccione et al., 2013] Guccione, P., Mascolo, L., Nico, G., Taurisano, P., Blasi, G., Fazio, L., and Bertolino, A. (2013). Functional brain networks and schizophrenia analysis with fMRI by Multiset Canonical Correlation Analysis. In *2nd International Conference on Advances in Biomedical Engineering (ICABME)*, pages 207–210.

[Gunduz et al., 2008] Gunduz, A., Sanchez, J. C., and Principe, J. C. (2008). Electrocorticographic interictal spike removal via denoising source separation for improved

neuroprosthesis control. In *30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS 2008*, pages 5224–5227.

[Haardt and Nossek, 1998] Haardt, M. and Nossek, J. (1998). Simultaneous Schur decomposition of several nonsymmetric matrices to achieve automatic pairing in multidimensional harmonic retrieval problems. *IEEE Transactions on Signal Processing*, 46(1):161–169.

[Hajipour et al., 2012] Hajipour, S., Shamsollahi, M. B., Albera, L., and Merlet, I. (2012). Noise cancelation of epileptic interictal EEG data based on generalized eigenvalue decomposition. In *TSP 12, 2012 IEEE Thirty-Fifth International Conference on Telecommunications and Signal Processing*, Prague, Czech Republic.

[Halder et al., 2007] Halder, S., Bensch, M., Mellinger, J., Bogdan, M., Kübler, A., Birbaumer, N., and Rosenstiel, W. (2007). Online artifact removal for brain-computer interfaces using support vector machines and blind source separation. *Computational Intelligence and Neuroscience*, 2007.

[Hallez et al., 2009] Hallez, H., De Vos, M., Vanrumste, B., Van Hese, P., Assecondi, S., Van Laere, K., Dupont, P., Van Paesschen, W., Van Huffel, S., and Lemahieu, I. (2009). Removing muscle and eye artifacts using blind source separation techniques in ictal EEG source imaging. *Clinical Neurophysiology*, 120(7):1262–1272.

[Hämäläinen et al., 1993] Hämäläinen, M., Hari, R., Ilmoniemi, R. J., Knuutila, J., and Lounasmaa, O. V. (1993). Magnetoencephalography-theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics*, 65(2):413–497.

[Hamalainen and Sarvas, 1989] Hamalainen, M. S. and Sarvas, J. (1989). Realistic conductivity geometry model of the human head for interpretation of neuromagnetic data. *IEEE Transactions on Biomedical Engineering*, 36(2):165–171.

[Hariharan et al., 2006] Hariharan, H., Gribok, A., Abidi, M. A., and Koschan, A. (2006). Image fusion and enhancement via empirical mode decomposition. *Journal of Pattern Recognition Research*, 1(1):16–32.

[Harshman and Lundy, 1994] Harshman, R. A. and Lundy, M. E. (1994). PARAFAC: Parallel factor analysis. *Computational Statistics and Data Analysis*, 18(1):39 – 72.

[Hassan et al., 2011] Hassan, M., Boudaoud, S., Terrien, J., Karlsson, B., and Marque, C. (2011). Combination of canonical correlation analysis and empirical mode decomposition applied to denoising the labor electrohysterogram. *IEEE Transactions on Biomedical Engineering*, 58(9):2441–2447.

[Hesse and James, 2006] Hesse, C. W. and James, C. J. (2006). On semi-blind source separation using spatial constraints with applications in EEG analysis. *IEEE Transactions on Biomedical Engineering*, 53(12):2525–2534.

[Huang and Pan, 2006] Huang, H. and Pan, J. (2006). Speech pitch determination based on Hilbert-Huang transform. *Signal Processing*, 86(4):792–803.

[Huang et al., 1998] Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N.-C., Tung, C. C., and Liu, H. H. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 454(1971):903–995.

[Hyvärinen and Oja, 1997] Hyvärinen, A. and Oja, E. (1997). A fast fixed-point algorithm for independent component analysis. *Neural Computation*, 9(7):1483–1492.

[Ille et al., 2002] Ille, N., Berg, P., and Scherg, M. (2002). Artifact correction of the ongoing EEG using spatial filters based on artifact and brain signal topographies. *Journal of Clinical Neurophysiology: Official Publication of the American Electroencephalographic Society*, 19(2):113.

[Indiradevia et al., 2008] Indiradevia, K., Eliasa, E., Sathidevia, P., Nayakb, S. D., and Radhakrishnanb, K. (2008). A multi-level wavelet approach for automatic detection of epileptic spikes in the electroencephalogram. *Computers in Biology and Medicine*, 38(7):805–816.

[Iriarte et al., 2006] Iriarte, J., Urrestarazu, E., Artieda, J., Valencia, M., LeVan, P., Viteri, C., and Alegre, M. (2006). Independent component analysis in the study of focal seizures. *Journal of Clinical Neurophysiology*, 23(6):551–558.

[James and Gibson, 2003] James, C. J. and Gibson, O. J. (2003). Temporally constrained ICA: an application to artifact rejection in electromagnetic brain signal analysis. *IEEE Transactions on Biomedical Engineering*, 50(9):1108–1116.

[Jansen and Rit, 1995] Jansen, B. H. and Rit, V. G. (1995). Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. *Biological Cybernetics*, 73(4):357–366.

[Jansen et al., 1993] Jansen, B. H., Zouridakis, G., and Brandt, M. E. (1993). A neurophysiologically-based mathematical model of flash visual evoked potentials. *Biological Cybernetics*, 68(3):275–283.

[Jasper, 1958] Jasper, H. H. (1958). The ten twenty electrode system of the international federation. *Electroencephalography and Clinical Neurophysiology*, 10:371–375.

[Ji et al., 2011] Ji, Z., Sugi, T., Goto, S., Wang, X., Ikeda, A., Nagamine, T., Shibasaki, H., and Nakamura, M. (2011). An automatic spike detection system based on elimination of false positives using the large-area context in the scalp EEG. *IEEE Transactions on Biomedical Engineering*, 58(9):2478–2488.

[Jiang and Sidiropoulos, 2004] Jiang, T. and Sidiropoulos, N. (2004). Kruskal's permutation lemma and the identification of CANDECOMP/PARAFAC and bilinear models with constant modulus constraints. *IEEE Transactions on Signal Processing*, 52(9):2625–2636.

[Joyce et al., 2004] Joyce, C. A., Gorodnitsky, I. F., and Kutas, M. (2004). Automatic removal of eye movement and blink artifacts from EEG data using blind component separation. *Psychophysiology*, 41(2):313–325.

[Jung et al., 2000] Jung, T.-P., Makeig, S., Humphries, C., Lee, T.-W., Mckeown, M. J., Iragui, V., and Sejnowski, T. J. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, 37(02):163–178.

[Kachenoura et al., 2008] Kachenoura, A., Albera, L., Senhadji, L., and Comon, P. (2008). ICA: a potential tool for BCI systems. *IEEE Signal Processing Magazine*, 25(1):57–68.

[Kachenoura et al., 2011] Kachenoura, A., Safieddine, D., Albera, L., Birot, G., Wendling, F., Senhadji, L., Merlet, I., et al. (2011). Blind source separation methods applied to muscle artefacts removing from epileptic EEG recording: A comparative study. In *RITS 2011 (Colloque National Recherche en Imagerie et Technologies pour la Santé)*.

[Kang and Zhizeng, 2012] Kang, D. and Zhizeng, L. (2012). A method of denoising multi-channel EEG signals fast based on PCA and DEBSS algorithm. In *International Conference on Computer Science and Electronics Engineering (ICCSEE)*, volume 3, pages 322–326.

[Karfoul et al., 2011] Karfoul, A., Albera, L., and Lathauwer, L. D. (2011). Iterative methods for the canonical decomposition of multi-way arrays: Application to blind underdetermined mixture identification. *Signal Processing*, 91(8):1789–1802.

[Karfoul et al., 2015] Karfoul, A., Kachenoura, A., Albera, L., Safieddine, D., Pasnicu, A., Wendling, F., Senhadji, L., Merlet, I., et al. (2015). Muscle artifact removal in ictal scalp-EEG based on blind source separation. pages 485–488.

[Kawaguchi et al., 2013] Kawaguchi, H., Kume, T., and Kobayashi, T. (2013). A background EEG removal method combining PCA with multivariate empirical mode decomposition for event-related potential measurements. *IEEJ Transactions on Electrical and Electronic Engineering*, 8(S1):S53–S60.

[Khatwani and Tiwari, 2013] Khatwani, P. and Tiwari, A. (2013). A survey on different noise removal techniques of EEG signals. *International Journal of Advanced Research in Computer and Communication*, 2(2).

[Kiers, 1998] Kiers, H. (1998). A three-step algorithm for CANDECOMP/PARAFAC analysis of large data sets with multicollinearity. *Chemometrics*, 12(3):155–171.

[Koldovsky et al., 2012] Koldovsky, Z., Phan, A. H., Tichavsky, P., and Cichocki, A. (2012). A treatment of EEG data by underdetermined blind source separation for motor imagery classification. In *Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*, pages 1484–1488.

[Kruskal, 1977] Kruskal, J. B. (1977). Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear Algebra and Applications*, 18(2):95–138.

[Lai and Tsai, 2010] Lai, C.-C. and Tsai, C.-C. (2010). Digital image watermarking using discrete wavelet transform and singular value decomposition. *IEEE Transactions on Instrumentation and Measurement*, 59(11):3060–3063.

[Latif et al., 2006] Latif, M. A., Sanei, S., Chambers, J., and Shoker, L. (2006). Localization of abnormal EEG sources using blind source separation partially constrained by the locations of known sources. *IEEE Signal Processing Letters*, 13(3):117–120.

[Lee and Lee, 2013] Lee, K. J. and Lee, B. (2013). Removing ECG artifacts from the EMG: A comparison between combining empirical-mode decomposition and independent component analysis and other filtering methods. In *13th International Conference on Control, Automation and Systems (ICCAS)*, pages 181–184.

[Lee et al., 1999] Lee, T.-W., Girolami, M., and Sejnowski, T. J. (1999). Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. *Neural computation*, 11(2):417–441.

[LeVan et al., 2006] LeVan, P., Urrestarazu, E., and Gotman, J. (2006). A system for automatic artifact removal in ictal scalp EEG based on independent component analysis and bayesian classification. *Clinical Neurophysiology*, 117(4):912–927.

[Liang et al., 2005a] Liang, H., Bressler, S. L., Buffalo, E. A., Desimone, R., and Fries, P. (2005a). Empirical mode decomposition of field potentials from macaque V4 in visual spatial attention. *Biological Cybernetics*, 92(6):380–392.

[Liang et al., 2005b] Liang, H., Lin, Q.-H., and Chen, J. (2005b). Application of the empirical mode decomposition to the analysis of esophageal manometric data in gastroesophageal reflux disease. *IEEE Transactions on Biomedical Engineering*, 52(10):1692–1701.

[Liang et al., 2000] Liang, H., Lin, Z., and McCallum, R. (2000). Artifact reduction in electrogastrogram based on empirical mode decomposition method. *Medical and Biological Engineering and Computing*, 38(1):35–41.

[Lin et al., 2013] Lin, Y.-P., Wang, Y., and Jung, T.-P. (2013). A mobile SSVEP-based brain-computer interface for freely moving humans: The robustness of canonical correlation analysis to motion artifacts. In *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 1350–1353.

[Looney et al., 2008] Looney, D., Li, L., Rutkowski, T. M., Mandic, D. P., and Cichocki, A. (2008). Ocular artifacts removal from EEG using EMD. In *Advances in Cognitive Neurodynamics ICCN 2007*, pages 831–835.

[Luciani and Albera, 2010] Luciani, X. and Albera, L. (2010). Joint eigenvalue decomposition using polar matrix factorization. In *Latent Variable Analysis and Signal*

Separation LVA-ICA 2010, volume 6365 of *Lecture Notes in Computer Science*, pages 555–562, Saint Malo France. Springer.

[Luciani and Albera, 2011] Luciani, X. and Albera, L. (2011). Semi-algebraic canonical decomposition of multi-way arrays and joint eigenvalue decomposition. In *ICASSP 11, 2011 IEEE International Conference on Acoustics Speech and Signal Processing*, pages 4104–4107, Prague, Czech Republic.

[Luciani and Albera, 2014] Luciani, X. and Albera, L. (2014). Canonical Polyadic Decomposition based on joint eigenvalue decomposition. *Chemometrics and Intelligent Laboratory Systems*, 132:152–167.

[Makeig et al., 2000] Makeig, S., Bell, T., Lee, T., Jung, T., Enghoff, S., et al. (2000). EEGLAB: ICA toolbox for psychophysiological research. *WWW Site, Swartz Center for Computational Neuroscience, Institute of Neural Computation, University of San Diego California.*

[Mammone et al., 2012] Mammone, N., La Foresta, F., and Morabito, F. C. (2012). Automatic artifact rejection from multichannel scalp EEG by wavelet ICA. *IEEE Sensors Journal*, 12(3):533–542.

[Mamun et al., 2013] Mamun, M., Al-Kadi, M., and Marufuzzaman, M. (2013). Effectiveness of Wavelet denoising on electroencephalogram signals. *Journal of Applied Research and Technology*, 11(1):156–160.

[Mangin et al., 1995] Mangin, J.-F., Frouin, V., Bloch, I., Régis, J., and López-Krahe, J. (1995). From 3D magnetic resonance images to structural representations of the cortex topography using topology preserving deformations. *Journal of Mathematical Imaging and Vision*, 5(4):297–318.

[Mateo et al., 2013] Mateo, J., Torres, A., Soria, C., and Santos, J. (2013). A method for removing noise from continuous brain signal recordings. *Computers and Electrical Engineering*, 39(5):1561–1570.

[Meijs et al., 1989] Meijs, J. W., Weier, O. W., Peters, M. J., and van Oosterom, A. (1989). On the numerical accuracy of the boundary element method (EEG application). *IEEE Transactions on Biomedical Engineering*, 36(10):1038–1049.

[Melia et al., 2014] Melia, U., Clariá, F., Vallverdú, M., and Caminal, P. (2014). Filtering and thresholding the analytic signal envelope in order to improve peak and spike noise reduction in EEG signals. *Medical Engineering and Physics*, 36(4):547–553.

[Mert and Akan, 2014] Mert, A. and Akan, A. (2014). EEG denoising based on empirical mode decomposition and mutual information. In *XIII Mediterranean Conference on Medical and Biological Engineering and Computing 2013*, pages 631–634.

[Mi, 2014] Mi, J.-X. (2014). A novel algorithm for Independent Component Analysis with Reference and methods for its applications. *PloS one*, 9(5):e93984.

[Mijovic et al., 2010a] Mijovic, B., De Vos, M., Gligorijevic, I., Taelman, J., and Van Huffel, S. (2010a). Source separation from single-channel recordings by combining empirical-mode decomposition and independent component analysis. *IEEE Transactions on Biomedical Engineering*, 57(9):2188–2196.

[Mijovic et al., 2010b] Mijovic, B., De Vos, M., Gligorijevic, I., and Van Huffel, S. (2010b). Combining EMD with ICA for extracting independent sources from single channel and two-channel data. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 5387–5390.

[Miwakeichi et al., 2004] Miwakeichi, F., Martınez-Montes, E., Valdés-Sosa, P. A., Nishiyama, N., Mizuhara, H., and Yamaguchi, Y. (2004). Decomposing EEG data into space–time–frequency components using parallel factor analysis. *NeuroImage*, 22(3):1035–1045.

[Molla et al., 2012] Molla, M. K. I., Tanaka, T., and Rutkowski, T. M. (2012). Multivariate EMD based approach to EOG artifacts separation from EEG. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 653–656.

[Molla et al., 2010] Molla, M. K. I., Tanaka, T., Rutkowski, T. M., and Cichocki, A. (2010). Separation of EOG artifacts from EEG signals using bivariate EMD. In *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, pages 562–565.

[Morlet, 1983] Morlet, J. (1983). Sampling theory and wave propagation. In *Issues in Acoustic Signal-Image Processing and Recognition*, pages 233–261. Springer.

[Mørup et al., 2006] Mørup, M., Hansen, L. K., Herrmann, C. S., Parnas, J., and Arnfred, S. M. (2006). Parallel factor analysis as an exploratory tool for wavelet transformed event-related EEG. *NeuroImage*, 29(3):938–947.

[Mosher and Leahy, 1999] Mosher, J. C. and Leahy, R. M. (1999). Source localization using recursively applied and projected (RAP) MUSIC. *IEEE Transactions on Signal Processing*, 47(2):332–340.

[Mourad and Niazy, 2013] Mourad, N. and Niazy, R. K. (2013). Automatic correction of eye blink artifact in single channel EEG recording using EMD and OMP. In *Proceedings of the 21st European Signal Processing Conference (EUSIPCO)*, pages 1–5.

[Naik et al., 2013] Naik, G. R., Guo, Y., and Nguyen, H. (2013). A new approach to improve the quality of biosensor signals using fast independent component analysis: Feasibility study using EMG recordings. In *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 1927–1929.

[Nam et al., 2002] Nam, H., Yim, T.-G., Han, S. K., Oh, J.-B., and Lee, S. K. (2002). Independent component analysis of ictal EEG in medial temporal lobe epilepsy. *Epilepsia*, 43(2):160–164.

[Navarro et al., 2012] Navarro, X., Porée, F., and Carrault, G. (2012). ECG removal in preterm EEG combining empirical mode decomposition and adaptive filtering. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 661–664.

[Niknazar et al., 2014] Niknazar, M., Becker, H., Rivet, B., Jutten, C., and Comon, P. (2014). Blind source separation of underdetermined mixtures of event-related sources. *Signal Processing*, 101:52–64.

[Nion and Lathauwer, 2008] Nion, D. and Lathauwer, L. D. (2008). An enhanced line search scheme for complex-valued tensor decompositions. Application in DS-CDMA. *Signal Processing*, 88(3):749–755.

[Oikonomou et al., 2007a] Oikonomou, V., Tzallas, A., and Fotiadis, D. (2007a). A Kalman filter based methodology for EEG spike enhancement. *Computer Methods and Programs in Biomedicine*, 85(2):101–108.

[Oikonomou et al., 2007b] Oikonomou, V. P., Tzallas, A. T., and Fotiadis, D. I. (2007b). A kalman filter based methodology for EEG spike enhancement. *Computer Methods and Programs in Biomedicine*, 85(2):101–108.

[Oostenveld and Praamstra, 2001] Oostenveld, R. and Praamstra, P. (2001). The five percent electrode system for high-resolution EEG and ERP measurements. *Clinical Neurophysiology*, 112(4):713–719.

[Paatero, 2000] Paatero, P. (2000). Construction and analysis of degenerate PARAFAC models. *Journal of Chemometrics*, 14(3):284–299.

[Parra and Sajda, 2004] Parra, L. and Sajda, P. (2004). Blind Source Separation via Generalized Eigenvalue Decomposition. *Journal of Machine Learning Research*, 4(7-8):1261–1269.

[Patel et al., 2014] Patel, A. X., Kundu, P., Rubinov, M., Jones, P. S., Vértes, P. E., Ersche, K. D., Suckling, J., and Bullmore, E. T. (2014). A wavelet method for modeling and despiking motion artifacts from resting-state fMRI time series. *NeuroImage*, 95:287–304.

[Pfurtscheller and Fischer, 1978] Pfurtscheller, G. and Fischer, G. (1978). A new approach to spike detection using a combination of inverse and matched filter techniques. *Electroencephalography and Clinical Neurophysiology*, 44(2):243–247.

[Phan et al., 2013a] Phan, A.-H., Tichavsky, P., and Cichocki, A. (2013a). CANDECOMP/PARAFAC decomposition of high-order tensors through tensor reshaping. *IEEE Transactions on Signal Processing*, 61(17-20):4847–4860.

[Phan et al., 2013b] Phan, A.-H., Tichavsky, P., and Cichocki, A. (2013b). Fast alternating LS algorithms for high order CANDECOMP/PARAFAC tensor factorizations. *IEEE Transactions on Signal Processing*, 61(17-20):4834–4846.

[Phan et al., 2013c] Phan, A.-H., Tichavsky, P., and Cichocki, A. (2013c). Low complexity damped Gauss–Newton algorithms for CANDECOMP/PARAFAC. *SIAM Journal on Matrix Analysis and Applications*, 34(1):126–147.

[Phlypo et al., 2010a] Phlypo, R., Zarzoso, V., and Lemahieu, I. (2010a). Atrial activity estimation from atrial fibrillation ECGs by blind source extraction based on a conditional maximum likelihood approach. *Medical and Biological Engineering and Computing*, 48(5):483–488.

[Phlypo et al., 2010b] Phlypo, R., Zarzoso, V., and Lemahieu, I. (2010b). Source extraction by maximizing the variance in the conditional distribution tails. *Signal Processing, IEEE Transactions on*, 58(1):305–316.

[Qi, 2012] Qi, W. (2012). EOG artifacts removal in EEG measurements for affective interaction with brain computer interface. In *Eighth International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP)*, pages 471–475.

[Raghavendra and Dutt, 2011] Raghavendra, B. and Dutt, D. N. (2011). Wavelet enhanced CCA for minimization of ocular and muscle artifacts in EEG. *World Academy of Science, Engineering and Technology*, 57(6):1027–1032.

[Rajih et al., 2008] Rajih, M., Comon, P., and Harshman, R. (2008). Enhanced Line Search: A novel method to accelerate PARAFAC. *SIAM Journal in Matrix Analysis and Applications*, 30(3):1128–1147.

[Roemer and Haardt, 2008] Roemer, F. and Haardt, M. (2008). A closed-form solution for multilinear PARAFAC decompositions. In *SAM 08, Fifth IEEE Sensor Array and Multichannel Signal Processing Workshop*, pages 487–491.

[Roemer and Haardt, 2013] Roemer, F. and Haardt, M. (2013). A semi-algebraic framework for approximate CP decompositions via simultaneous matrix diagonalizations (SECSI). *Signal Processing*, 93(9):2722–2738.

[Romero et al., 2009] Romero, S., Mananas, M., and Barbanoj, M. J. (2009). Ocular reduction in EEG signals based on adaptive filtering, regression and blind source separation. *Annals of Biomedical Engineering*, 37(1):176–191.

[Romo Vázquez et al., 2012] Romo Vázquez, R., Vélez-Pérez, H., Ranta, R., Louis Dorr, V., Maquin, D., and Maillard, L. (2012). Blind source separation, wavelet denoising and discriminant analysis for EEG artefacts and noise cancelling. *Biomedical Signal Processing and Control*, 7(4):389–400.

[Rosenfeld and Johnston, 1973] Rosenfeld, A. and Johnston, E. (1973). Angle detection on digital curves. *IEEE Transactions on Computers*, C-22(9):875–878.

[Rui et al., 2013] Rui, T., Zhang, Q., Zhou, Y., and Xing, J. (2013). Object tracking using particle filter in the wavelet subspace. *Neurocomputing*, 119:125–130.

[Rutkowski et al., 2009] Rutkowski, T. M., Cichocki, A., Tanaka, T., Ralescu, A. L., and Mandic, D. P. (2009). Clustering of spectral patterns based on EMD components of EEG channels with applications to neurophysiological signals separation. In *Advances in Neuro-Information Processing*, pages 453–460. Springer.

[Rutkowski et al., 2010] Rutkowski, T. M., Mandic, D. P., Cichocki, A., and Przybyszewski, A. W. (2010). EMD approach to multichannel EEG data - the amplitude and phase components clustering analysis. *Journal of Circuits, Systems, and Computers*, 19(01):215–229.

[Safieddine, 2012] Safieddine, D. (2012). *Apport des méthodes statistiques et déterministes à la réduction d'artefacts et de bruits dans les signaux EEG épileptiques.* PhD thesis, Université de Rennes 1.

[Safieddine et al., 2012] Safieddine, D., Kachenoura, A., Albera, L., Birot, G., Karfoul, A., Pasnicu, A., Biraben, A., Wendling, F., Senhadji, L., and Merlet, I. (2012). Removal of muscle artifact from EEG data: comparison between stochastic (ICA and CCA) and deterministic (EMD and wavelet-based) approaches. *EURASIP Journal on Advances in Signal Processing*, (1):1–15.

[Safieddine et al., 2011] Safieddine, D., Kachenoura, A., Albera, L., Birot, G., Wendling, F., Senhadji, L., and Merlet, I. (2011). ICA versus CCA pour le débruitage de signaux épileptiques intercritiques: une étude comparative de performances basée sur la localisation de la zone épileptogène. *IRBM*, 32(5):298–301.

[Salis et al., 2013] Salis, C. I., Malissovas, A. E., Bizopoulos, P. A., Tzallas, A. T., Angelidis, P. A., and Tsalikakis, D. G. (2013). Denoising simulated EEG signals: A comparative study of EMD, wavelet transform and kalman filter. In *IEEE 13th International Conference on Bioinformatics and Bioengineering (BIBE)*, pages 1–4.

[Samadi et al., 2013] Samadi, S., Amini, L., Cosandier-Rimélé, D., Soltanian-Zadeh, H., and Jutten, C. (2013). Reference-based source separation method for identification of brain regions involved in a reference state from intracerebral EEG. *IEEE Transactions on Biomedical Engineering*, 60(7):1983–1992.

[Samadi and Shamsollahi, 2014] Samadi, S. and Shamsollahi, M. B. (2014). ECG noise reduction using empirical mode decomposition based on combination of instantaneous half period and soft-thresholding. In *Middle East Conference on Biomedical Engineering (MECBME)*, pages 244–248.

[Sameni and Gouy-Pailler, 2014] Sameni, R. and Gouy-Pailler, C. (2014). An iterative subspace denoising algorithm for removing electroencephalogram ocular artifacts. *Journal of Neuroscience Methods*, 225:97–105.

[Sameni et al., 2010] Sameni, R., Jutten, C., and Shamsollahi, M. B. (2010). A deflation procedure for subspace decomposition. *Signal Processing, IEEE Transactions on*, 58(4):2363–2374.

[Sanchez and Kowalski, 1986] Sanchez, E. and Kowalski, B. R. (1986). Generalized rank annihilation factor analysis. *Analytical Chemistry*, 58(2):496–499.

[Sanchez and Kowalski, 1990] Sanchez, E. and Kowalski, B. R. (1990). Tensorial resolution: a direct trilinear decomposition. *Journal of Chemometrics*, 4(1):29–45.

[Santillan-Guzman et al., 2013] Santillan-Guzman, A., Fischer, M., Heute, U., and Schmidt, G. (2013). Real-time Empirical Mode Decomposition for EEG signal enhancement. In *Proceedings of the 21st European Signal Processing Conference (EUSIPCO)*, pages 1–5.

[Sarela and Valpola, 2005] Sarela, J. and Valpola, H. (2005). Denoising Source Separation. *Journal of Machine Learning Research*, 6:233–272.

[Scolaro et al., 2012] Scolaro, G. R., de Azevedo, F. M., Boos, C. F., and Walz, R. (2012). Wavelet filter to attenuate the background activity and high frequencies in EEG signals applied in the automatic identification of epileptiform events. *INTECH*.

[Senhadji et al., 1997] Senhadji, L., Bellanger, J.-J., Carrault, G., et al. (1997). EEG spike detectors based on different decompositions: a comparative study. *Time Frequency and Wavelets in Biomedical Signal Processing*, pages 407–421.

[Senhadji et al., 1995] Senhadji, L., Dillenseger, J. L., Wendling, F., Rocha, C., and Kinie, A. (1995). Wavelet analysis of EEG for three-dimensional mapping of epileptic events. *Annals of Biomedical Engineering*, 23(5):543–552.

[Senhadji and Wendling, 2002] Senhadji, L. and Wendling, F. (2002). Epileptic transient detection: wavelets and time-frequency approaches. *Neurophysiol Clin*, 32(3):175–192.

[Shamsollahi et al., 1996] Shamsollahi, M., Senhadji, L., Bouquin-Jeannes, L., et al. (1996). Detection and localization of complex SEEG patterns in epileptic seizures using time-frequency analysis. In *Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis*, pages 105–108.

[Shapoori et al., 2013] Shapoori, S., Wang, W., and Sanei, S. (2013). A constrained approach for extraction of pre-ictal discharges from scalp EEG. In *IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, pages 1–5.

[Shen et al., 2013] Shen, K., Yu, K., Bandla, A., Sun, Y., Thakor, N., and Li, X. (2013). Multiple time-lag canonical correlation analysis for removing muscular artifacts in EEG. In *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 6792–6795.

[Sidiropoulos and Bro, 2000] Sidiropoulos, N. and Bro, R. (2000). On the uniqueness of multilinear decomposition of n-way arrays. *Journal of Chemometrics*, 14(3):229–239.

[Soomro et al., 2013] Soomro, M. H., Badruddin, N., Yusoff, M. Z., and Malik, A. S. (2013). A method for automatic removal of eye blink artifacts from EEG based on EMD-ICA. In *IEEE 9th International Colloquium on Signal Processing and its Applications (CSPA)*, pages 129–134.

[Sørensen and De Lathauwer, 2010] Sørensen, M. and De Lathauwer, L. (2010). New simultaneous generalized Schur decomposition methods for the computation of the canonical polyadic decomposition. In *ASILOMAR 10, 2010 Conference on Signals, Systems and Computers*, pages 13–17, Monterey, USA.

[Srinivasan and Ravichandran, 2013] Srinivasan, M. and Ravichandran, N. (2013). A new technique for face recognition using 2D-Gabor Wavelet Transform with 2D-Hidden Markov Model approach. In *International Conference on Signal Processing Image Processing and Pattern Recognition (ICSIPR)*, pages 151–156.

[Stegman and Sidiropoulos, 2007] Stegman, A. and Sidiropoulos, N. (2007). On kruskal's uniqueness condition for the Candecomp/Parafac decomposition. *Linear Algebra and its Applications*, 420(2-3):540–552.

[Subasi and Kiymik, 2010] Subasi, A. and Kiymik, M. K. (2010). Muscle fatigue detection in EMG using time–frequency methods, ICA and neural networks. *Journal of Medical Systems*, 34(4):777–785.

[Subramaniyam et al., 2013] Subramaniyam, N. P., Peltola, J., Tanskanen, J. M., Wendel-Mitoraj, K., Hyttinen, J., and Malmivuo, J. (2013). Source localization based on ictal electroencephalographic recordings. In *International Symposium on Biomedical Engineering and Medical Physics*, pages 26–30.

[Tang et al., 2005] Tang, A. C., Sutherland, M. T., and McKinney, C. J. (2005). Validation of SOBI components from high-density EEG. *NeuroImage*, 25(2):539–553.

[Tichavsky et al., 2013] Tichavsky, P., Phan, A. H., and Cichocki, A. (2013). A further improvement of a fast damped GAUSS-NEWTON algorithm for CANDECOMP/PARAFAC tensor decomposition. In *Proceeding of IEEE Int. Conf. Acoustics, Speech, Signal Processing, ICASSP*, pages 5964–5968.

[Tomasi and Bro, 2006] Tomasi, G. and Bro, R. (2006). A comparison of algorithms for fitting the PARAFAC model. *Computational Statistics and Data Analysis*, 50(7):1700–1734.

[Tong et al., 1990] Tong, L., Soon, V., Huang, Y., and Liu, R. (1990). AMUSE: a new blind identification algorithm. In *IEEE International Symposium on Circuits and Systems*, pages 1784–1787.

[Urrestarazu et al., 2004] Urrestarazu, E., Iriarte, J., Alegre, M., Valencia, M., Viteri, C., and Artieda, J. (2004). Independent component analysis removing artifacts in ictal recordings. *Epilepsia*, 45(9):1071–1078.

[Vergult et al., 2007] Vergult, A., De Clercq, W., Palmini, A., Vanrumste, B., Dupont, P., Van Huffel, S., and Van Paesschen, W. (2007). Improving the interpretation of ictal scalp EEG: BSS–CCA algorithm for muscle artifact removal. *Epilepsia*, 48(5):950–958.

[Vigário, 1997] Vigário, R. N. (1997). Extraction of ocular artefacts from EEG using independent component analysis. *Electroencephalography and Clinical Neurophysiology*, 103(3):395–404.

[Villringer and Chance, 1997] Villringer, A. and Chance, B. (1997). Non-invasive optical spectroscopy and imaging of human brain function. *Trends Neuroscience*, 20(10):435–442.

[von Tscharner et al., 2011] von Tscharner, V., Eskofier, B., and Federolf, P. (2011). Removal of the electrocardiogram signal from surface EMG recordings using nonlinearly scaled wavelets. *Journal of Electromyography and Kinesiology*, 21(4):683–688.

[Vos et al., 2007] Vos, M. D., Vergult, A., Lathauwer, L. D., Clercq, W. D., Huffel, S. V., Dupont, P., Palmini, A., and Paesschen, W. V. (2007). Canonical decomposition of ictal EEG reliably detects the seizure onset zone. *NeuroImage*, 37(3):844–854.

[Wallstrom et al., 2004] Wallstrom, G. L., Kass, R. E., Miller, A., Cohn, J. F., and Fox, N. A. (2004). Automatic correction of ocular artifacts in the EEG: a comparison of regression-based and component-based methods. *International Journal of Psychophysiology*, 53(2):105–119.

[Wang et al., 2013] Wang, T., Lin, L., Zhang, A., Peng, X., and Zhan, C. A. (2013). EMD-based EEG signal enhancement for auditory evoked potential recovery under high stimulus-rate paradigm. *Biomedical Signal Processing and Control*, 8(6):858–868.

[Wang et al., 2004] Wang, Y., Sutherland, M. T., Sanfratello, L. L., and Tang, A. C. (2004). Single-trial classification of ERPS using second-order blind identification (SOBI). In *Proceedings of 2004 International Conference on Machine Learning and Cybernetics*, volume 7, pages 4246–4251.

[Ward, 1963] Ward, J. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301):236–244.

[Wendling et al., 2001] Wendling, F., Bartolomei, F., Bellanger, J., and Chauvel, P. (2001). Interpretation of interdependencies in epileptic signals using a macroscopic physiological model of the EEG. *Clinical Neurophysiology*, 112(7):1201–1218.

[Wendling et al., 2002] Wendling, F., Bartolomei, F., Bellanger, J., and Chauvel, P. (2002). Epileptic fast activity can be explained by a model of impaired GABAergic dendritic inhibition. *European Journal of Neuroscience*, 15(9):1499–1508.

[Wendling et al., 2003] Wendling, F., Bartolomei, F., Bellanger, J.-J., Bourien, J., and Chauvel, P. (2003). Epileptic fast intracerebral EEG activity: evidence for spatial decorrelation at seizure onset. *Brain*, 126(6):1449–1459.

[Wendling et al., 2000] Wendling, F., Bellanger, J.-J., Bartolomei, F., and Chauvel, P. (2000). Relevance of nonlinear lumped-parameter models in the analysis of depth-EEG epileptic signals. *Biological Cybernetics*, 83(4):367–378.

[Wendling et al., 2005] Wendling, F., Hernandez, A., Bellanger, J., Chauvel, P., and Bartolomei, F. (2005). Interictal to ictal transition in human temporal lobe epilepsy: insights from a computational model of intracerebral EEG. *Journal of Clinical Neurophysiology*, 22(5):343.

[Wilson and Emerson, 2002] Wilson, S. B. and Emerson, R. (2002). Spike detection: a review and comparison of algorithms. *Clinical Neurophysiology*, 113(12):1873–1881.

[Zanow and Peters, 1995] Zanow, F. and Peters, M. (1995). Individually shaped volume conductor models of the head in EEG source localisation. *Medical and Biological Engineering and Computing*, 33(4):582–588.

[Zarzoso and Comon, 2010] Zarzoso, V. and Comon, P. (2010). Robust independent component analysis by iterative maximization of the kurtosis contrast with algebraic optimal step size. *IEEE Transactions on Neural Networks*, 21(2):248–261.

[Zarzoso et al., 1997] Zarzoso, V., Nandi, A., and Bacharakis, E. (1997). Maternal and foetal ecg separation using blind source separation methods. *IMA Journal of Mathematics Applied in Medicine and Biology*, 14(3):207–225.

[Zhang and Zhou, 2013] Zhang, X. and Zhou, P. (2013). Filtering of surface EMG using ensemble empirical mode decomposition. *Medical Engineering and Physics*, 35(4):537–542.

[Zima et al., 2012] Zima, M., Tichavskỳ, P., Paul, K., and Krajča, V. (2012). Robust removal of short-duration artifacts in long neonatal EEG recordings using wavelet-enhanced ICA and adaptive combining of tentative reconstructions. *Physiological Measurement*, 33(8):N39.

[Zou et al., 2012] Zou, Y., Hart, J., and Jafari, R. (2012). Automatic EEG artifact removal based on ica and hierarchical clustering. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 649–652.

[Zumer et al., 2007] Zumer, J. M., Attias, H. T., Sekihara, K., and Nagarajan, S. S. (2007). A probabilistic algorithm integrating source localization and noise suppression for MEG and EEG data. *NeuroImage*, 37(1):102–115.