

KATHOLIEKE UNIVERSITEIT
LEUVEN

Arenberg Doctoral School of Science, Engineering & Technology
Faculty of Engineering
Department of Electrical Engineering

DECOMPOSITION METHODS WITH APPLICATIONS IN NEUROSCIENCE

Maarten De Vos

Dissertation presented in partial
fulfillment of the requirements for
the degree of Doctor
in electrical engineering

December 2009

DECOMPOSITION METHODS WITH APPLICATIONS IN NEUROSCIENCE

Maarten De Vos

Jury:
Prof. dr. ir. Y. Willems, president
Prof. dr. ir. S. Van Huffel, promotor
Prof. dr. ir. L. De Lathauwer, promotor
Prof. dr. W. Van Paesschen, promotor
Prof. dr. ir. D. Vandermeulen
Prof. dr. P. Dupont
Prof. dr. ir. M. Moonen
Prof. dr. J.-C. Pesquet (C.N.R.S)
Prof. dr. I. Lemahieu (UGent)

Dissertation presented in partial fulfillment of the requirements for the degree of Doctor in electrical engineering

December 2009

© Katholieke Universiteit Leuven – Faculty of Electrical Engineering
Kasteelpark Arenberg 10, B-3001 Leuven (Belgium)

Alle rechten voorbehouden. Niets uit deze uitgave mag worden vermenigvuldigd en/of openbaar gemaakt worden door middel van druk, fotocopie, microfilm, elektronisch of op welke andere wijze ook zonder voorafgaande schriftelijke toestemming van de uitgever.

All rights reserved. No part of the publication may be reproduced in any form by print, photoprint, microfilm or any other means without written permission from the publisher.

Legal depot number D/2009/7515/140
ISBN number 978-94-6018-158-0

Voorwoord

Een doctoraat schrijf je misschien wel grotendeels alleen, maar zonder de constante bijdragen van anderen, breng je het nooit tot een goed einde. Het is daarom mijn genoegen alle mensen te bedanken die mij geholpen en gesteund hebben. Op welke manier dan ook.

Vooreerst een woord van dank aan mijn legioen van promotoren. Waren ze zo talrijk omdat ik veel steun nodig had? Of wilden ze samen sterk staan om mijn persoonlijke, onmogelijke ideeën te bekampen? In elk geval, met de globale leiding van prof. Sabine, het wiskundig inzicht van prof. Lieven, en de medische kennis van prof. Wim kan geen enkel doctoraat mislukken. Sabine, bedankt voor alle steun bij mijn wilde plannen, en de vrijheid om steeds mijn dromen achterna te kunnen jagen. Wim, bedankt voor alle bemoediging en hulp tijdens onze epileptische avonturen. En dan ... Lieven; ik ben je veel dank verschuldigd. Ons eerste echte gesprek was toen we toevallig dezelfde trein namen. Vanaf dan veranderde veel. Er gingen door jou deuren open in de tensorwereld, je stond klaar om ideeën voor algoritmes te spuien en mijn papers (te :)) grondig te herschrijven.

Met dankbaarheid wil ik ook alle financiële instellingen bedanken die mijn onderzoek hebben ondersteund. Dit werk was voornamelijk gesponsord door een beurs van het Instituut voor Wetenschap en Technologie (IWT) Vlaanderen.

In de wereld van epilepsie kwam ik ook geregeld in contact met de lokale SPECT-goeroe, prof. Patrick, wiens gesprekken mij inspireerden, en prof. Guido, die altijd weer klaar stond om nieuwe data aan te leveren, data te interpreteren, of gewoon de rosé koel wou zetten.

Prof. D. Vandermeulen, bedankt om raadgever te zijn tijdens dit doctoraat en prof. M. Moonen, bedankt om de jury te vervoegen. Ik dank van harte prof. I. Lemahieu, voor de samenwerking met MEDISIP Gent. Un grand merci a prof. J.-C. Pesquet. C'est un grand honeur que vous vouliez être membre de mon jury.

Mijn bureau was zonder meer een serre van productiviteit en collegialiteit. Katrien, bedankt voor mijn plantje levend te houden in droge tijden, en voor nog zoveel

meer. Wouter, het was me altijd een genoegen om samen met jou de auto, de trein of het vliegtuig te nemen. Ik bedank ook alle vroegere en huidige mensen van Biomed en ESAT voor warme gesprekken of praktische hulp.

Tijdens mijn doctoraat kwam ik via verschillende projecten in contact met collega's van andere universiteiten. Steeds discussieerde ik met genoegen met de mensen uit San Fransisco, Rotterdam, Gent, Marseille, Helsinki, ... Deze gesprekken hebben mijn werk positief beïnvloed.

Tot slot is er "de familie" die bedankt wordt voor hun inspanningen naast het werk: voor de hotel- en limo service, voor hun aanmoedigingen,.... Vele violisten en muzikanten droegen bij tot mijn ontspanning, ook zij verdienen een eervolle vermelding.

And there is Monica, who always cheered me up or surprised me with a little poem. You earn a fluffy and passionate hug. You always gave me the energy to carry on, the hope when i most needed it and the laugh that kept me sane.

Abstract

The brain is the most important signal processing unit in the human body. It is responsible for receiving, processing and storing information. One of the possibilities to study brain functioning is by placing electrodes on the scalp and recording the synchronous neuronal activity of the brain.
Such a recording measures a combination of active processes in the whole brain. Unfortunately, it is also contaminated by artifacts. By extracting the artifacts and removing them, cleaned recordings can be investigated. Furthermore, it is easier to look at specific brain activities, like an epileptic seizure, than at a combination. In this thesis, we present different mathematical techniques that can be used to extract individual contributing sources from the measured signals for this purpose. We focused on Canonical Correlation Analysis (CCA), Independent Component Analysis (ICA) and Canonical/ Parallel Factor Analysis (CPA).
We show that the properties of the sources, extracted with CCA are appropriate to extract muscle artifacts from the brain recordings. We validated this in a study on speech production. We illustrate that ICA algorithms can be used to remove eye artifacts. An important topic in epilepsy research is also accurate localisation of the epileptogenic focus. Based on the brain signals recorded during an epileptic seizure, the localisation of this focus can be derived. However, artifacts often obscure this seizure. In a first step, we removed eye and muscle artifacts to improve the localisation of this focus. In a second step, we developed a method based on CPA that directly extracts the epileptic source. The localisation of this source provides then information on the focus.
We also developed some new algorithms. We show how to incorporate spatial constraints into ICA. These constraints are related to how much the sources are present in the different recording electrodes. We show that incorporating such prior knowledge improves the accuracy of the estimation of the sources.
A last chapter deals with a new algorithm that combines two decomposition methods: ICA and CPA. We discuss that combining the constraints underlying both decomposition methods is useful in practical applications and show that in several situations the decomposition of the new method outperforms both ordinary ICA and CPA.

De hersenen zijn de belangrijkste plaats voor informatieverwerking in het menselijk lichaam. Ze zijn verantwoordelijk voor het ontvangen, verwerken en opslaan van informatie. Een van de mogelijkheden om de werking van de hersenen te bestuderen is door middel van het plaatsen van electroden op de hoofdhuid. Hiermee kan men synchrone neuronale activiteit van de hersenen opmeten.
Zulke meting registreert een combinatie van actieve processen in de volledige hersenen. Jammer genoeg is de meting ook bevuild door artefacten. Door de artefacten softwarematig te verwijderen, kunnen propere hersensignalen onderzocht worden. Meer nog, het is gemakklijker om specifieke hersenactiviteit, zoals een epileptische aanval, te bestuderen dan naar een combinatie te kijken. In dit doctoraat presenteren we verschillende wiskundige technieken die gebruikt kunnen worden om individuele bronnen te schatten uit opgemeten signalen. De klemtoon ligt bij Canonieke Correlatie Analyse (CCA), Onafhankelijke Componenten Analyse (ICA) en de Canonieke Ontbinding (CPA).
We tonen dat de eigenschappen van de bronsignalen, geëxtraheerd met CCA geschikt zijn om spiersignalen te extraheren uit hersensignalen. We hebben dit gevalideerd in een studie naar spraakproductie. We tonen dat ICA-algoritmes kunnen gebruikt worden om oogartefacten te verwijderen. Een belangrijk onderwerp in onderzoek rond epilepsie is het bekomen van een nauwkeurige localisatie van de epileptische focus. Op basis van de hersensignalen, opgemeten tijdens een epileptische aanval kan de localisatie van de focus afgeleid worden. Soms is de aanval echter niet zichtbaar door de aanwezigheid van artefacten. In een eerste stap hebben we oog- en spierartefacten verwijderd om de localisatie van de focus te verbeteren. In een tweede stap hebben we een methode ontwikkeld op basis van CPA die de epileptische bron extraheert. De localisatie van deze bron geeft dan informatie over de focus.
We hebben ook nieuwe algoritmes ontwikkeld. We tonen hoe spatiële beperkingen kunnen opgelegd worden in ICA. Deze beperkingen hebben betrekking tot hoeveel de bronsignalen aanwezig zijn in de verschillende electrodes. We tonen dat het in rekening brengen van zulke a priori kennis de accuraatheid van de schatting van de bronnen verbetert.
Een laatste hoofdstuk heeft betrekking tot een nieuw algoritme dat twee decompositiemethodes combineert: ICA en CPA. We leggen uit dat de combinatie zinvol is in praktische toepassingen en tonen dat de nieuwe methode in verschillende situaties beter is dan standaard ICA en CPA.

Nomenclature

List of symbols

$a, b, \ldots$	scalars
$\mathbf{a}, \mathbf{b}, \ldots$	vectors
$\mathbf{A}, \mathbf{B}, \ldots$	matrices
$\mathcal{A}, \mathcal{B}, \ldots$	tensors
$a_{ijk}, (\mathcal{A})_{ijk}$	element of $\mathcal{A}$ at position (i, j, k)
$\mathcal{A}_{i=\alpha}$	subtensor of $\mathcal{A}$ obtained by fixing the index i
$\mathbf{A}_{(n)}$	mode-n matrix representation of a tensor $\mathcal{A}$
$\mathbf{I}$	identity matrix
$I, I_1, I_2, I_3, N, R, \ldots$	special scalars and upper bounds
i	$\sqrt{-1}$

Sets of numbers

$\mathbb{R}$	the set of real numbers
$\mathbb{C}$	the set of complex numbers

Basic operations

$\mathbf{A}^T$	transpose of the matrix $\mathbf{A}$
$\mathbf{A}^{-1}$	inverse
$\mathbf{A}^\dagger$	pseudoinverse
$\operatorname{diag}(\mathbf{v})$	square diagonal matrix with vector $\mathbf{v}$ as diagonal
$\otimes$	Kronecker product $\mathbf{A} \otimes \mathbf{B} := [a_{ij}\mathbf{B}]$
$\odot$	Khatri-Rao product $\mathbf{A} \odot \mathbf{B} = [\mathbf{a}_1 \otimes \mathbf{b}_1 \ldots \mathbf{a}_R \otimes \mathbf{b}_R]$
$*$	element-wise (Hadamard) product $\mathbf{A} * \mathbf{B} := [a_{ij} b_{ij}]$

$[\mathbf{a}\,\mathbf{b}]$	matrix with columns $\mathbf{a}$ and $\mathbf{b}$
$\lvert a\rvert$	absolute value of a
$\langle \cdot, \cdot \rangle$	inner/scalar product
$\lVert \mathbf{A} \rVert$	Frobenius norm, $\sqrt{\text{trace}(\mathbf{A}^T\mathbf{A})}$
$\text{trace}(\mathbf{A})$	trace of the matrix $\mathbf{A}$, $\sum a_{ii}$
$\text{rank}(\cdot)$	rank
$\text{rank}_n(\mathcal{A})$	mode-n rank of $\mathcal{A}$
Σ	sum
$\prod$	product
$\mathbf{a} \circ \mathbf{b}$	outer product of vectors $\mathbf{a}$ and $\mathbf{b}$
$\mathcal{A} \times_n \mathbf{M}$	mode-n product of a tensor $\mathcal{A}$ and a matrix $\mathbf{M}$
$\in$	belongs to

Abbreviations

AQSES	Accurate Quantification of Short-echo Time Magnetic Resonance Spectroscopic Signals
BCM	Block Component Model
BER	Bit Error Ratio
BOLD	Blood Oxygen Level Dependent
BSS	Blind Source Separation
Candecomp	Canonical Decomposition
CCA	Canonical Correlation Analysis
CPA	Candecomp/Parafac Analysis
CSF	Cerebrospinal Fluid
ECoG	Electrocorticography
EDS	Exponentially Damped Sinusoidal
EEG	Electroencephalogram
EMD	Empirical Mode Decomposition
EMG	Electromyogram
EMU	Epilepsy Monitoring Unit
EVD	EigenValue Decomposition
ERP	Event Related Potential
EOG	Electro-oculogram
fMRI	functional Magnetic Resonance Imaging
GLM	General Linear Model
HOS	Higher-Order Statistics
HOSVD	Higher-Order Singular Value Decomposition
hrf	hemodynamic response function
ICA	Independent Component Analysis
JADE	Joint Approximate Diagonalization of Eigenmatrices
LS-SVM	Least Squares - Support Vector Machine
MEG	Magnetoencephalogram
MR	Magnetic Resonance
MRI	Magnetic Resonance Imaging
MRS	Magnetic Resonance Spectroscopy
MRSI	Magnetic Resonance Spectroscopic Imaging
MUSIC	Multiple Signal Classification
Parafac	Parallel Factor Analysis
PCA	Principal Component Analysis
RAP-MUSIC	Recursively Applied and Projected MUSIC
RRE	Relative Residual Energy
SISCOM	Subtraction ictal SPECT co-registered to MRI
SPM	Statistical Parameter Mapping
SOBI	Second Order Blind Identification
SPECT	Single Photon Emission Computerized Tomography
SNR	Signal to Noise Ratio
SVD	Singular Value Decomposition

Contents

Contents ix

1 Overview 1

1.1 Introduction 1

1.2 Contributions in the thesis 3

1.3 Chapter-by-chapter overview 4

2 Background: Brain monitoring and epilepsy 7

2.1 Monitoring brain activity 7

2.1.1 The electrical activity of synchronously firing neurons can be measured with the EEG 7

2.1.2 Application areas of EEG 19

2.1.3 Increased blood flow can be measured with fMRI 21

2.1.4 Combining EEG and fMRI 22

2.2 Brain dysfunction in epilepsy 24

2.2.1 Epileptic activity in the EEG 24

2.2.2 Presurgical evaluation 24

2.3 Conclusion 27

3 Background: Mathematical decomposition techniques 29

3.1 Introduction 29

3.2 Matrix rank and the singular value decomposition (SVD) 33
3.3 Rank in multilinear algebra 35
3.4 One-dimensional signal decomposition techniques 36
3.5 Two-dimensional matrix decomposition techniques 38
3.5.1 Principal Component Analysis (PCA) 38
3.5.2 Independent Component Analysis (ICA) 38
3.5.3 Canonical Correlation Analysis (CCA) 45
3.5.4 Other matrix decompositions 48
3.5.5 A remark on the evaluation of decompositions 48
3.6 Higher-order tensor decomposition techniques 49
3.6.1 Canonical/Parallel factor Analysis (CPA) 49
3.6.2 The Tucker3 model 53
3.6.3 The Block Component Model (BCM) 54
3.6.4 Other tensor decompositions 55
3.7 Conclusion 56

4 CCA **57**

4.1 CCA for tumor segmentation 57
4.1.1 Introduction 57
4.1.2 Applying CCA to spectra 59
4.1.3 A two-step CCA-based method 61
4.1.4 Applying CCA to features 62
4.1.5 Material 64
4.1.6 Simulation study 65
4.1.7 *In vivo* data 70
4.1.8 Summary on segmentation of MRS data. 72
4.2 BSS-CCA for removal of muscle artifacts from epileptic seizure recordings 75

4.2.1 Method . . . 75
4.2.2 Illustration . . . 75
4.2.3 Discussion . . . 76
4.3 BSS-CCA for the removal of short bursts of EMG activity during speech production . . . 78
4.3.1 Introduction . . . 78
4.3.2 Materials and Methods . . . 81
4.3.3 Results . . . 83
4.3.4 Discussion . . . 86
4.4 Conclusion . . . 89

5 (Spatially constrained) ICA 95

5.1 Removal of eye artifacts with BSS . . . 95
5.1.1 Introduction . . . 95
5.1.2 Illustration . . . 98
5.1.3 Discussion . . . 100
5.2 Imposing prior knowledge to ICA . . . 101
5.3 Imposing spatial constraints . . . 103
5.3.1 The mixing vector of one source is exactly known. . . . 103
5.3.2 The mixing vectors of more sources are exactly known. . . . 106
5.3.3 The mixing vector of one source is known with a limited accuracy. . . . 107
5.3.4 Weighted constraint . . . 107
5.4 Simulation study . . . 112
5.5 Illustration of scICA for eye artifact removal . . . 113
5.6 Discussion on ICA with fixed constraint . . . 116
5.7 Conclusion . . . 118

6 Removing artifacts prior to ictal source localisation 119

6.1 Introduction . . . 119
6.2 Materials . . . 121
6.3 Estimating the sources by RAP-MUSIC . . . 124
6.4 Evaluation of EEG dipole source localisation . . . 125
6.5 Results and discussion . . . 127
6.6 Conclusion . . . 129

7 Extracting ictal activity with CPA 135
7.1 Introduction . . . 135
7.2 Method . . . 137
7.3 Simulation study . . . 141
7.3.1 Simulation setup . . . 141
7.3.2 Results of simulation study . . . 146
7.4 Clinical validation on ictal EEGs . . . 151
7.4.1 Materials . . . 151
7.4.2 Statistics . . . 152
7.4.3 Results of validation study . . . 152
7.5 Discussion . . . 154
7.6 Conclusion . . . 159

8 A combination of Parallel Factor and Independent Component Analysis 161
8.1 Introduction . . . 161
8.2 Combination of ICA and CPA: ICA-CPA. . . . 164
8.2.1 Tensor pICA . . . 164
8.2.2 Imposing independence constraints in CPA: ICA-CPA . . . 165
8.3 Numerical experiments . . . 167
8.4 Discussion . . . 170
8.5 Conclusion . . . 172

9 Conclusions and open problems **173**

9.1 General conclusions of the thesis 173

9.1.1 Applications 173

9.1.2 Mathematical algorithms 174

9.2 Future work and open problems 175

Bibliography **181**

Publication list **213**

Curriculum vitae **219**

Chapter 1

Overview

1.1 Introduction

Most PhD theseses have two subjects. First, there is the nice story the man in the street understands. Of course you never tell anybody that only one percent of your work is truly related to stories. Second, there is the more advanced stuff you really did research about and you published about. Fortunately, the man in the street does not want to know about it. This chapter will make you familiar with the story. I will describe the global idea behind "Decomposition methods with applications in neuroscience". The next chapter will give some more biomedical background. The rest of the book will mainly focus on more advanced mathematical signal processing techniques.

Neuroscience is the field devoted to the scientific study of the nervous system. More in particular we focus on the brain. Brain (dys)function can be monitored with several techniques. The most important non-invasive techniques are ElectroEncephaloGram (EEG) and Functional Magnetic Resonance Imaging (fMRI). The EEG measures directly the electrical brain activity as a function of time. FMRI measures increased blood flow as an indirect respons to increased brain activity. Both techniques give a global picture of what is going on in the whole brain and, unfortunately, also pick up artifacts. Most of the time however, one is more interested in specific brain activities, *e.g.*, an epileptic signal, a specific wave arising in a particular brain region, ..., instead of looking at the combination of brain sources. The goal of this thesis is to develop and improve mathematical techniques that extract reliably the individual contributing sources from the measured signals.

Just to introduce the topic, Fig. 1.1 shows 10 seconds of an EEG registration in

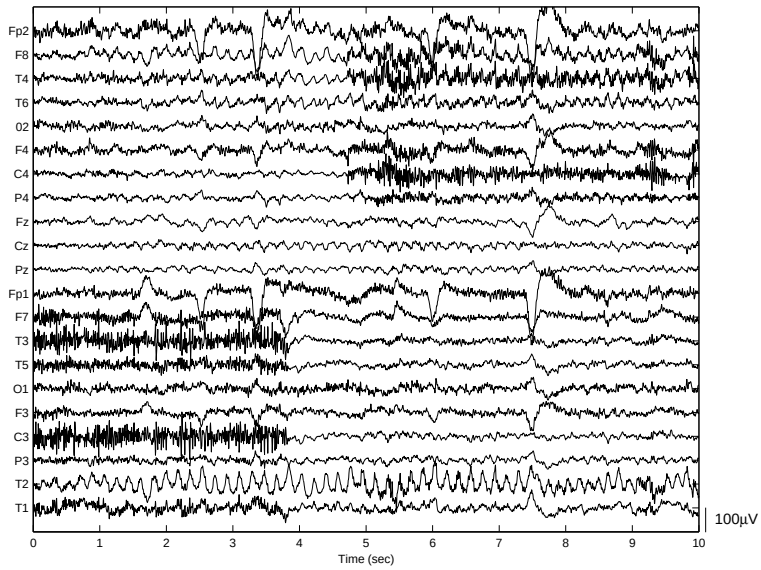


Figure 1.1: *10 sec of a 21 channel, scalp EEG recording from a long-term EMU. The epileptic seizure is mainly visible on the $T2$, $F8$ and $T4$ electrodes. The seizure EEG is contaminated with eye blinks and muscle artifact.*

an Epilepsy Monitoring Unit (EMU) during an epileptic seizure. The 21 different channels are labeled according to their well-defined position on the head. The seizure activity is the rhythmical wave mainly present on the $T2$, $F8$ and $T4$ electrodes. Furthermore, on the channels above the eyes, Fp2 and Fp1, sharp deflections occur around seconds 2.5, 3.2, 6 and 7.5. These deflections are ocular artifacts. The high-frequency noise, present on different channels, but mainly at T4, C4, T3 and C3, is related to muscle activity. These artifacts often obscure EEG recordings.

The goal of mathematical decomposition methods is now to decompose such an EEG measurement in an accurate way into the different contributing signal components. In this example, the contributing signal components are the epileptic signal, ocular and muscular artifacts and some background brain activity. A good decomposition allows us to remove the interfering artifacts. This enables neurologists to look more accurately at the measured brain activity. Ideally, one brain signal of interest can be extracted and its time course and localization can be investigated.

The concept of Blind Source Separation (BSS) can be used to estimate the contributing sources from the measured mixture. There are different conditions to be fulfilled in order to reliably decompose mixed signals. Very different BSS

solutions can be obtained when different assumptions about the underlying sources are made. Potential and limitations of several BSS methods will be discussed throughout this thesis. We focus mainly on Canonical Correlation Analysis (CCA), Independent Component Analysis (ICA) and CPA (Canonical Decomposition (Candecomp)/ Parallel Factor Analysis (Parafac)). The example shown in Fig. 1.1 will be used throughout the rest of this PhD thesis to illustrate the explained mathematical concepts.

1.2 Contributions in the thesis

The personal contributions in this thesis can be organised in two parts. Several problems were solved by using or slightly modifying existing techniques (Applications). Besides, we also developed some new methods (Algorithms).

Applications

1. We developed a CCA based method to segment and classify brain tumors.
2. We developed a BSS-CCA based method to remove short bursts of muscle activity from EEG recordings during speech production.
3. We removed artifacts from epileptic seizure recordings as preprocessing technique for improved dipole localisation.
4. We developed a CPA based method for accurate seizure localisation and investigated potential and limitations.

Algorithms

5. We developed an ICA algorithm with exactly known mixing vector(s).
6. We developed an ICA algorithm for situations when one mixing vector is approximately known.
7. We developed an algorithm to combine ICA and CPA.

1.3 Chapter-by-chapter overview

Chapter 2

The next chapter will provide relevant background information for the applications we worked on in this thesis. In order to be able to develop signal processing techniques for analyzing biomedical data, it is important to know how these data are generated and how they should be interpreted. In a first part, we will discuss non-invasive brain monitoring techniques. In a second part, we will give more detail about the neurological disorder epilepsy as this was an important application throughout this thesis.

Chapter 3

Chapter 3 will review the mathematics necessary for a good understanding of the next chapters. Different data decomposition techniques will be discussed and differences between them will be highlighted.

Chapter 4

In this chapter, we illustrate a first decomposition method, Canonical Correlation Analysis (CCA). CCA is originally a method for comparing different datasets. This will be illustrated with the application of brain tumor segmentation. Tumor segmentation is not a topic in neuroscience. However, it is an interesting biomedical application to illustrate CCA. We will also provide some background on the used data in this chapter. Second, we illustrate the use of CCA as a blind decomposition method able to remove muscle artifacts from continuous epilepsy recordings. Third, we show that it can be used to remove short bursts of muscle activity superimposed on speech production recordings. This can become a breakthrough, as, up to now, research on speech production is not possible because of the artifacts.

Chapter 5

Independent Component Analysis (ICA) is a decomposition method able to blindly extract independent sources from a linear mixture. It is a known technique to remove ocular artifacts from EEG recordings. Furthermore, we develop a new algorithm to incorporate prior knowledge about a mixing vector during the extraction of the independent sources. This algorithm, called scICA, outperforms fully blind ICA in accuracy.

Chapter 6

Brain recordings in epileptic patients are sometimes severely obscured by different types of artifacts. Decomposition methods presented in chapter 4 and 5 enable us to separate the brain recordings into individually contributing signals and removing different kinds of artifacts. We show in this chapter the results of a pilot study. We illustrate that localization of ictal brain activity benefits from removing artifacts.

Chapter 7

Instead of first removing artifacts from brain recordings and afterwards localizing the remaining EEG component, the optimal way would be to extract immediately the activity of interest, the ictal activity. In this chapter, we develop a new CPA method that is able to reliably extract the epileptic signal.

Chapter 8

This chapter deals with combining two decomposition models: ICA and CPA. We impose independence constraints in the multidimensional CPA method and show with simulations when this new ICA-CPA method outperforms ordinary ICA and CPA.

Chapter 9

The last chapter summarizes the important findings of this work, and gives several suggestions for further research.

Chapter 2

Background: Brain monitoring and epilepsy

This chapter introduces some medical background. First, brain functioning and measurement tools are discussed. Afterwards, more details about epilepsy monitoring are given.

2.1 Monitoring brain activity

2.1.1 The electrical activity of synchronously firing neurons can be measured with the EEG

Anatomy of the brain

The brain can be divided into two symmetric hemispheres. The cortex of each hemisphere is subdivided into four lobes, *i.e.*, the frontal, parietal, occipital and temporal lobe, as illustrated in Fig. 2.1. These lobes contain many distinct areas that control different functions. In order to localise brain activity in a more precise way, Brodmann divided the brain in smaller regions according to the structure of brain cells. This finer classification results in the Brodmann area's, shown in Fig. 2.2.

Fig. 2.3 shows a horizontal slice through the brain. In the figure, white and grey parts can be distinguished, corresponding to the so-called white and grey matter. The grey matter can be found at the outermost surface of the brain, the cortex, and in some deeper structures. The larger inner part of the brain

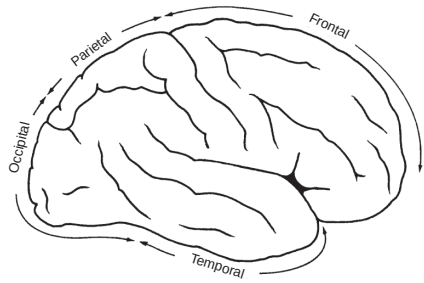


Figure 2.1: *Each hemisphere of the brain is divided in four different lobes. From [184].*

contains the white matter. In order to understand this difference in color, one has to know that the brain consists of about 10^{10} nerve cells or neurons. These neurons can have diverse sizes and shapes, but they all possess a grey cell body, a white myelinated axon, grey dendrites and synaptic terminals, as illustrated in Fig. 2.4. The cell body contains the cell nucleus, in which all vital proteins are synthesized. Neurons correspond with each other via synaptic terminals by releasing chemicals (neurotransmitters). Dendrites extend from the cell body and share with it the function of receiving information from synaptic connections with neighbouring neurons. An axon is a slender tube branching with other cells to form synaptic terminals. It can range in length from a few hundred micrometers to more than a meter.

Electrical brain activity

Neurons, like other cells, are electrically polarized so that their interior is negatively charged with respect to the outside of the cell. This potential difference, called the resting potential, is due to an unequal distribution of mainly Na^+, K^+ and Cl^- ions across the cell membrane. Typical values for this resting potential are around -70 mV [308, 325, 48].

Cells communicate with each other by releasing chemicals at the synaptic terminals, from presynaptic to postsynaptic regions that disturb this resting potential. The postsynaptic potentials alter the neuronal membrane potential by several millivolts. This change in potential last over about 10 ms. Both depolarisation or hyperpolarisation is possible. As every neuron has many synapses connecting to different neurons, the actual potential over a cell membrane is given by the spatial and/or temporal summation of the postsynaptic potentials. When the neuronal cell is depolarized beyond a critical level or threshold, an action potential is generated that proliferates along the axon. When such an action

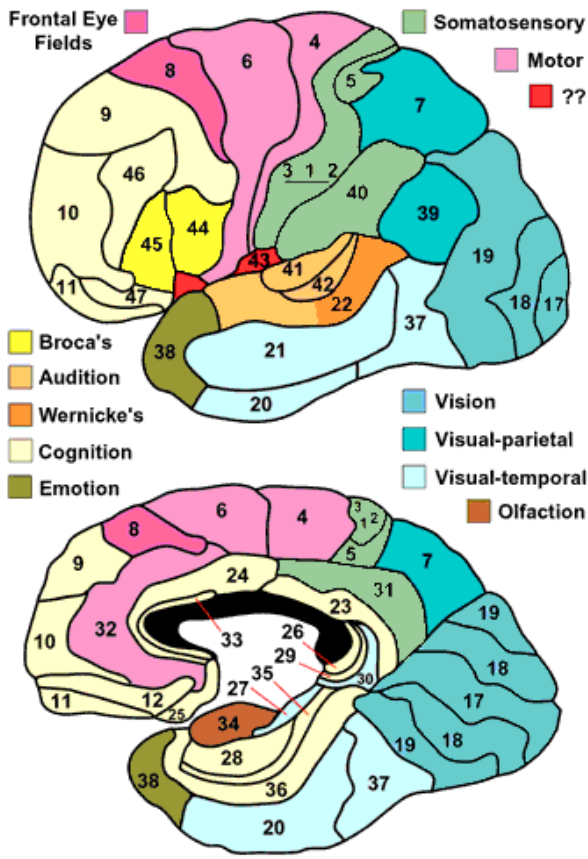


Figure 2.2: *The different Brodmann areas in the brain reveal different functional regions in the brain. From [160].*

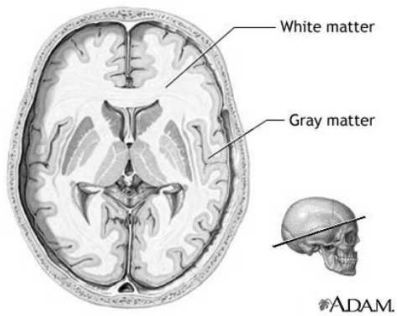


Figure 2.3: *A horizontal slice through the brain. The grey matter, consisting of cell bodies and dendrites, can be found at the outmost surface of the brain (the cortex) and in some deeper structures. The inner part of the brain contains the axons of the nerve cells and is called white matter. From [279].*

potential arrives at a synapse, it is able to release on its turn neurotransmitters to communicate with other neurons. If the potential is excitatory, the chance that an action potential will be triggered, will increase in the postsynaptic cell. When it is inhibitory, the generation of action potentials will be suppressed. This very complex network of chemical and electrical signals controls in a detailed way normal brain function.

Electroencephalography (EEG)

The EEG measures the electrical communication between neurons as a function of time. A change in potential can only be detected when many neurons synchronously de- or hyperpolarize. This happens mainly in vertical oriented large pyramidal cells of the cortex because these neurons are aligned and amplify

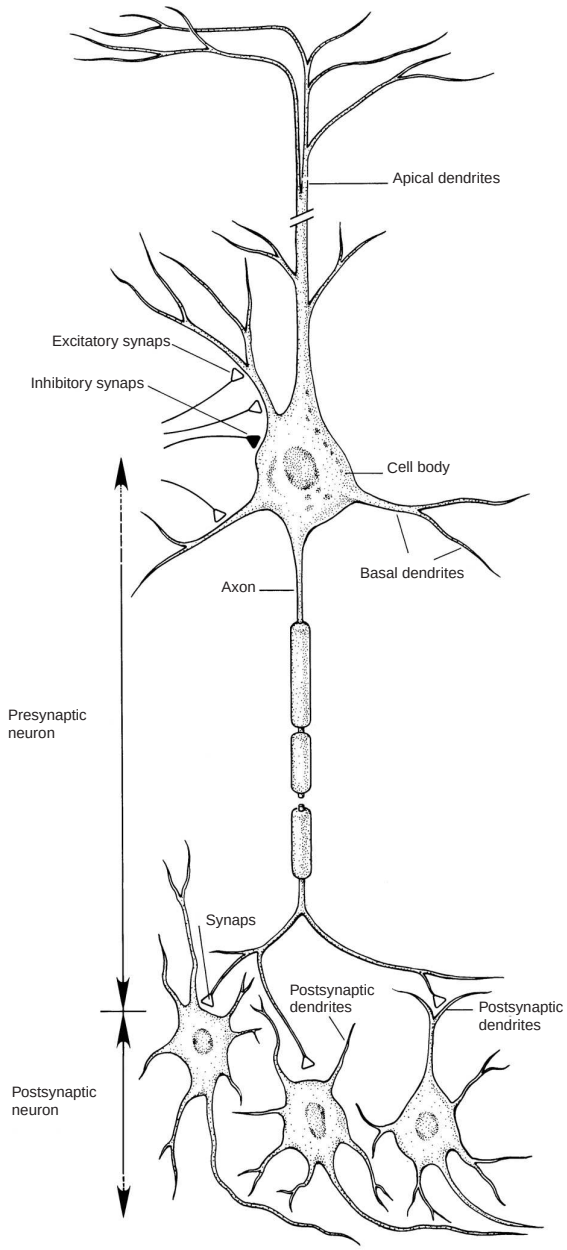


Figure 2.4: *The basic parts of a neuron are the cell body, dendrites, the axon and synaptic terminals. From [48].*

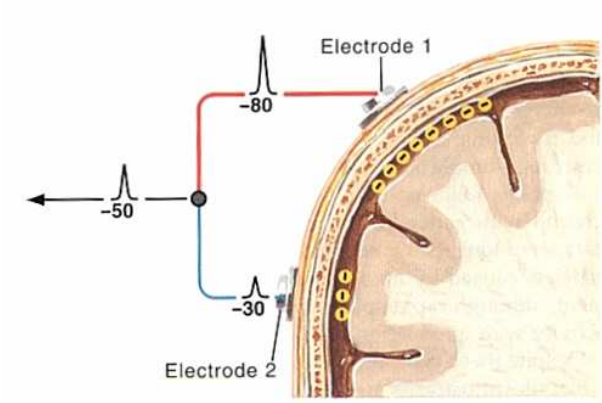


Figure 2.5: *The sum of electrical brain potentials is large enough to be measured by electrodes placed on the human scalp.*

each other's extracellular field [115]. The currents generated by these neurons summate in the extracellular space and are attenuated through the meningeal coverings, spinal fluid, skull and scalp, but can still be detected. When cortical neurons are simultaneously active, the sum of these potentials will be between 10 to 150 μV on the human scalp. The signals, measured between two electrodes constitute the EEG (Fig. 2.5). Hans Berger discovered this in Jena in 1924 [19]. An electroencephalogram measures the potential differences between electrodes as a function of time. The electrical activity is transformed into wave patterns that, in the past, typically were traced onto rolling sheets of paper. In order to be able to localise the measured brain activity, electrodes are placed at well-defined positions on the head. The most common positioning of electrodes on the head is given by the international 10-20 system [276]. This electrode placement provides a uniform coverage of the entire scalp, presented in Fig. 2.6. In an EMU, a total of 21 electrodes are used: besides the 19 electrodes of the 10-20 system also two additional electrodes $T1$ and $T2$ on the temporal regions are used. Each electrode is given a label as illustrated in Fig. 2.7. The names of these electrodes refer to the different brain lobes.

As the EEG measures the time course of potential differences between electrodes, a reference for this potential difference has to be chosen. The most used reference systems are bipolar reference, the Laplacian derivation, the average reference, and the linked-ears reference [116]. With an average reference, all potentials are displayed with respect to the average value of all electrodes. In linked-ears reference, the EEG is displayed with respect to the average of the potentials at the ear lobes (which should be ideally zero). Both are good for visualizing widespread coherent waveforms. These waveforms occur with similar amplitude and phase. In bipolar and Laplacian montage, the shown EEG signals are obtained

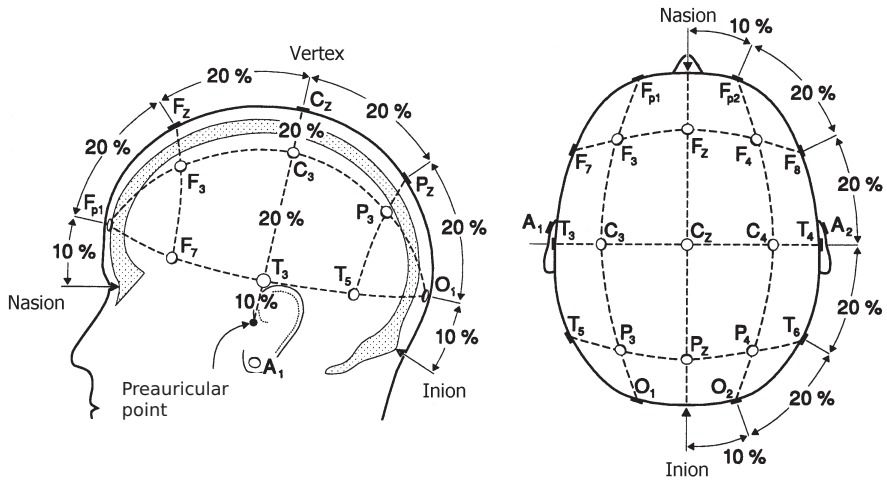


Figure 2.6: *The standard electrode placement according to the 10-20 international system: side view and top view of the head. It uses the distances between bony landmarks (the nasion, inion and preauricular points) of the head to generate a system of lines that run across the head and intersect at intervals of 10 or 20% of their total length. Electrodes are placed at the intersections. From [239].*

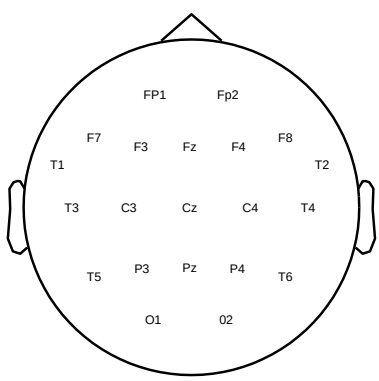


Figure 2.7: *Names and places of electrodes used in the EMU.*

by subtracting neighboring electrode signals. These are best suited for viewing highly localised activity over a limited area of the scalp because the widespread waveforms are filtered out.

Potential and limitations of the EEG

The main advantage of the EEG is that it is a quite cheap and easy way to directly measure brain function. It has an excellent temporal resolution, determined by the sampling frequency. Recent research focuses also on high-frequency oscillations (fast ripples up to 500Hz) [26, 339] using higher sampling frequencies. However, there is still no full consensus about the origin of these oscillations. In clinical applications, the EEG is mostly sampled at 256 Hz. This is sufficient, as the EEG is commonly subdivided in 4 frequency f bands as follows:

- Delta: $0.5 \leq f < 4$Hz;
- Theta: $4 \leq f < 8$Hz;
- Alpha: $8 \leq f < 13$Hz; and
- Beta: $13 \leq f < 30$Hz.

Fig. 2.8 illustrates traces of EEG signals with the rhythms listed above. EEG rhythms are associated with various physiological and mental processes [113, 114]. Mainly changes in a person's state of arousal − from being awake and alert to deeply asleep − are associated with changes in brain waves:

- Beta waves show that a person is actively attending to events and, therefore, is aware of what is happening around him.
- Alpha waves show that a person is awake but is not actively processing visual information. Alpha waves generally are observed when the person is relaxed and inattentive (such as in daydreaming).
- Theta waves are observed when a person has fallen into a light sleep (a stage of sleep from which he or she can be awakened easily).
- Delta waves are observed when a person has entered deep sleep (a stage of sleep from which he or she cannot be awakened easily).

When working with the EEG, it is also important to keep in mind the limitations of the technique. First of all, scalp EEG records only a portion of the underlying brain electrical activity. The mechanisms by which the cortical activity is altered at the

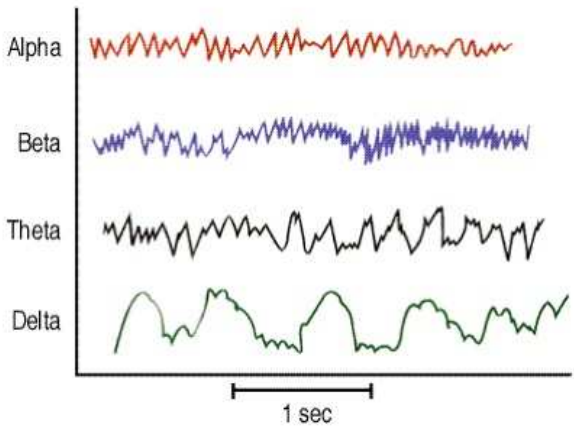


Figure 2.8: *Different brain traces, from bottom to top: examples of (1) delta rhythm (0.5-4Hz); (2) theta rhythm (4-8Hz); (3) beta rhythm (13-30Hz); (4) alpha rhythm (8-13Hz).*

scalp level are not fully understood. From a practical point of view, however, some general principles can be recognized [115, 300]. At a very basic level, it is known that scalp electrodes will be less sensitive to discharges that originate from deep brain structures [124]. Besides, it appears that the coverings of the brain act as 'a 'low-pass" filter [90], so rapidly changing potentials of interest can be averaged out by the different brain layers. The principles of volume conduction provide further explanation for the fact that sometimes the scalp recording electrode may have such a geometric orientation to the generator, often represented by an electrical dipole, that the electrical signal can not be recorded [124]. Dipole orientations that are "aimed" towards the electrode are better recorded.

Second, due to the low spatial sampling, it is not possible to extract precise localising information of interesting brain activity. Based on visual inspection one can lateralize activity or maybe define brain lobes of interest. Research also focused on developing mathematical models for electrical sources and volume conductor models in order to perform quantitative localisation. However, the 3D localisation problem is higly underdetermined and ill-posed. Ill-posed means that the solution can drastically change when the input data are only very slightly modified. This means that the problem is very susceptible to noise. Underdetermined means that many solutions are possible for one recording, *e.g.*, a very focal deep source can generate the same potential distribution as a more wide-spread superficial source. In order to obtain a unique solution, extra constraints about the source are introduced. Several different approaches to solve this source localisation or inverse problem exist based on different assumptions [11, 247]. One assumption is that

the surface potentials are generated by a dense set of dipolar sources distributed on the cortical surface. The most popular method from this "distributed source" family is Loreta [285]. In a second approach, which is more common, a limited number of "equivalent dipoles" are assumed to generate the measured potential distribution [309]. The electrical field generated by a group of pyramidal cells can be approximated sufficiently by a single dipole [67].

The third and maybe the hardest problem in EEG is that scalp electrodes record all potentials and cannot discriminate between brain sources and other electrical signals. All potentials not generated by the brain are called artifacts. One of the goals in signal processing is to remove these artifacts as they hamper the interpretation of the cerebral activity. In order to have a clear picture of the artifactual signal, several types will be discussed. Artifacts may be divided into two categories depending on their origin: physiological artifacts or non-physiological artifacts [110]. Besides, in many applications, also alpha brain activity can be considered as an artifact as it obscures normal brain activity [345]. Non-physiological artifacts are caused mainly by:

- external electrical interference from power lines or electrical equipment and
- electrical malfunctioning of the recording system itself, arising from recording electrodes, electrode positioning, cables or amplifiers. In EMUs it will be mainly caused by disconnected electrodes.

Artifacts are called physiological when they are caused by other parts in the body than the brain. We can distinguish artifacts caused by

- movements of the head, body or scalp, *e.g.*, during tremor,
- moving electrical potentials within the body, *e.g.*, heart muscle contraction,
- changes in skin resistance *e.g.*, due to sweating,
- active muscle cells, *e.g.*, during frowning, chewing, ...,
- movements of the eye ball, *e.g.*, during eye movements and blinking.

As chapter 4 and 5 are dedicated to signal processing techniques that can be used to remove muscle and eye artifacts respectively, we will illustrate these artifacts in more detail. These are also by far the most common artifacts in scalp EEG recordings.

Ocular and muscular artifacts

Different ocular artifacts can be distinguished: eye blinks, vertical and horizontal eye movements [110, 169].

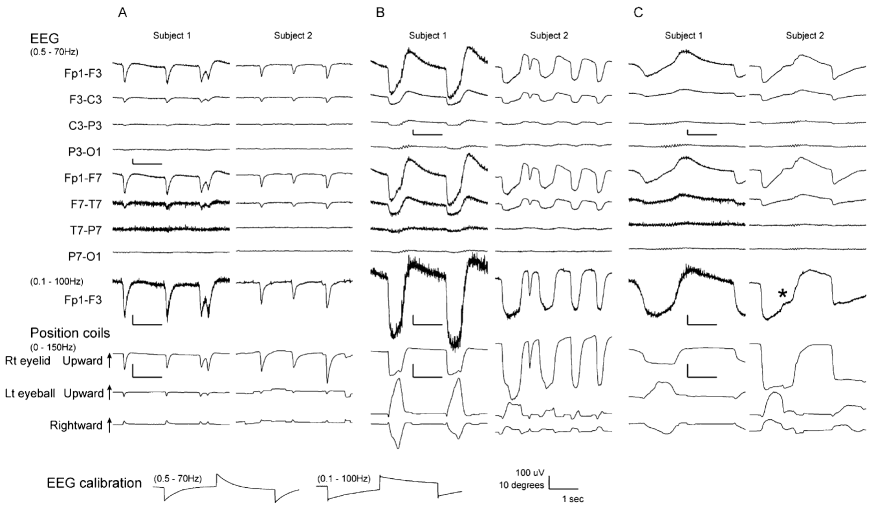


Figure 2.9: *Voluntary eyelid movements. (A) Natural, spontaneous blinks. (B) Forced blinks. (C) Slow, self-paced, eye opening and closure. Conventions are similar to Fig. 1. The asterisk in panel C corresponds to the small negative deflection that occurred after eyelid closure, and was consistent with the time course of the downward eyeball rotation. From [169].*

"In a typical bipolar montage, eye-blinks are characterized by positive deflection in the most anterior electrodes with a rapid amplitude fall-off at more posteriorly located electrodes. Traditionally, these artifacts are explained by an upward rotation of the eyeball during the lid closure, with the eyeball representing a dipole with corneal positivity and retinal negativity" [187, 298]. Their contribution is largest on the frontal $Fp1$ and $Fp2$ electrodes and decreases rapidly with distance from the eyes. Fig. 2.9 illustrates different types of blinks.
Movements of the eyes also cause potential changes, see Fig. 2.10 for different types.
Eye blinks and movements may be monitored by electrodes placed near the eyes. Such a recording is called the ElectroOculoGram (EOG). It allows easy identification of the ocular artifacts. However, in long term recording, often no EOG signals are recorded.
Eye artifacts are annoying in different applications (see below) because (a) they may obscure the EEG, *e.g.*, for detection of the seizure onset zone, (b) they make contaminated Evoked Potentials (EP's) trials useless, (c) they are a source of false positives in automatic seizure detectors. For the removal of eye artifacts, several methods have already been proposed, but a perfect method does not yet exist.

The EEG is frequently contaminated by electrophysiological potentials associated

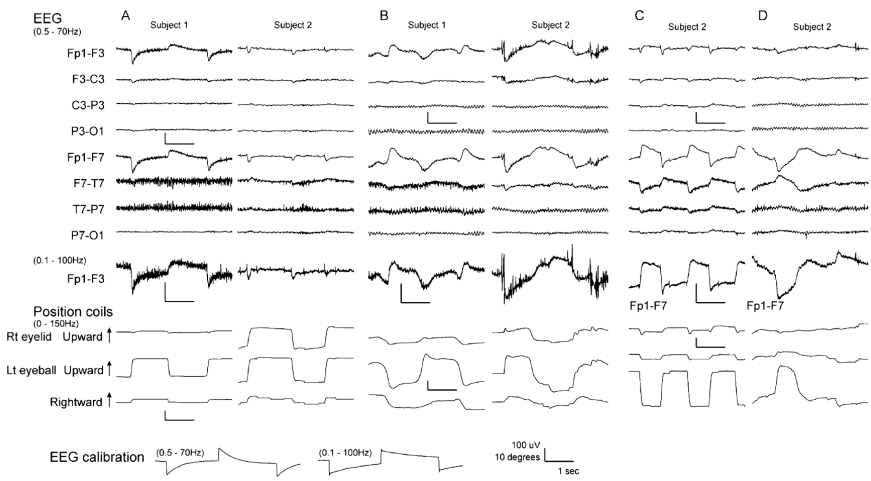


Figure 2.10: *Movements of the left eyeball and right lid recorded simultaneously with scalp EEG during eye saccades. Top 8 channels: EEG in a conventional bipolar montage and common filter setting (0.570 Hz). Ninth channel: the same EEG with the wide low-frequency cutoff filter (0.1 Hz). Note that the conventional filter setting enhances the rapid phase of the EEG waveform. Bottom 3 channels: lid and eyeball movements recorded by the magnetic search coil technique (0150 Hz). Upward deflection indicates upward and rightward rotation of the coil, in the corresponding vertical and horizontal movements. Scale: 100 mV vertically, 1 s horizontally. (A) Vertical saccades between the lower and upper fields of gaze. (B) Vertical voluntary saccades beneath closed eyes. (C) Horizontal saccades between the right and left fields of gaze. (D) Horizontal voluntary saccades beneath closed eyes. From [169].*

with muscle contraction due to speaking, biting, chewing and frowning. These muscle artifacts, also known as ElectroMyoGram (EMG) artifacts, obscure the EEG and complicate the interpretation of the EEG or even fully cover brain activity. It is known that EMG artifacts have a broad frequency distribution from 0 to over 200Hz, which overlaps with the frequency distribution of cerebral activity [128]. Muscle artifacts from scalp and face muscles occur mainly in frontal and temporal regions. Neck muscle activity will be recorded by occipital electrodes. Sometimes it can be reduced by asking the patient to relax or open the mouth slightly. However, during speech production tasks, or during epileptic seizures, it is inevitable and has to be removed by signal processing techniques. Fig. 2.11 shows an EEG recording contaminated with muscle artifact.

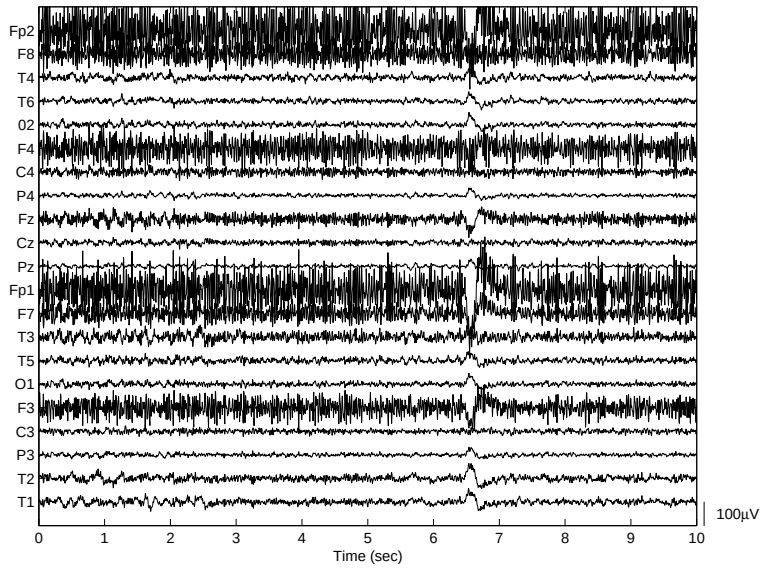


Figure 2.11: *EEG recording contaminated with muscle artifact, strongly visible on the frontal electrodes. An average reference montage was used.*

2.1.2 Application areas of EEG

Scalp EEG is a widely-used tool to study human brain functioning in general. In clinical practice, it is also used for diagnosis and follow-up of neurological disorders or diseases. Different valuable application areas are: neuropharmacology, evoked potential studies, analysis of sleep disorders and long-term monitoring in epilepsy patients and neonati [269]. Neuropharmacology is the field in which the effect of medication on brain function is investigated [356]. An example is the validation of medication for increasing attention in a person. As brain waves related to attention consist mainly of waves with frequencies in the beta range, one can investigate the power in the beta band before, during and after medication to monitor whether medication improves mental concentrations.
EEG is also used in order to study human brain functioning and deviations from normal functioning. In such studies, stimuli are presented to the person, and brain responses on these stimuli are measured. This explains the name "evoked" or "event-related" potentials (ERP). The actual "tasks" depend on the study. They range from showing visual stimuli that only have to be detected to advanced memory tasks. As individual brain responses on stimuli are much smaller than background EEG activity, the same stimulus is typically presented several times. Trials are averaged out and the resulting time-locked EEG waves are enhanced with respect to the background. Such studies are used to detect

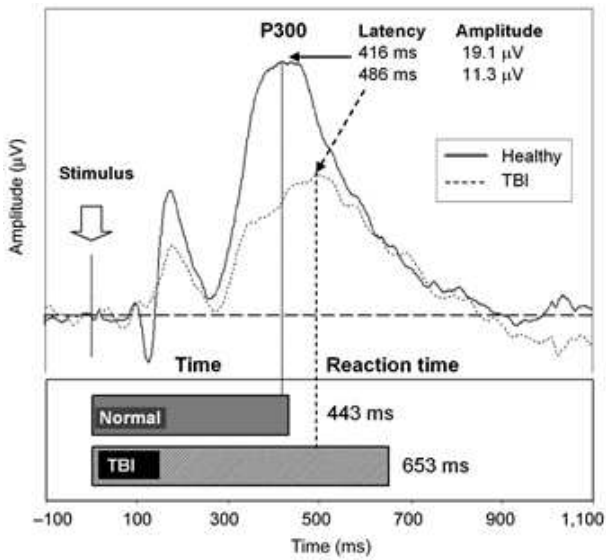


Figure 2.12: *An evoked potential, averaged out over many trials, is the brain response on a certain stimulus and can reveal differences between normal subjects and e.g., Traumatic Brain Injury (TBI) patients. From [224].*

and follow up brain functioning in many neurological diseases, *e.g.*, ADHD, MS [21, 45, 46, 103, 228, 277].
In contrast to the above-mentioned applications, in which the experiments are limited in time, sometimes long-term EEG recordings are performed. Sleep studies in sleep laboratories aim at evaluating patients with sleep disorders by monitoring all possible signals related to sleep: oxygen-saturation, eye movements, EEG, ... Based on the integration of all information provided by the different modalities, sleep pathologies can be detected. The EEG in normal subjects during sleep shows four different sleep stages. Abnormalities in these patterns are directly related to the pathologies. Also important is the detection of sleep apneas [268, 29].
A relatively new field is EEG monitoring of neonates with neurologic disorders. It is an important field as seizures occur in 1 up to 3.5 per 1000 births [353]. As many neonatal seizures are subclinical, *i.e.*, without clinical symptoms, it is extremely important to monitor them with a full EEG. Accurate seizure detection methods are here a very important tool to assist doctors in reviewing the EEG [86].
Finally, the EEG has an important clinical use in the diagnosis of epilepsy. As the main research track in this thesis is built on improved analysis of EEG in epilepsy, section 2.2 will give a more elaborated introduction to this field.

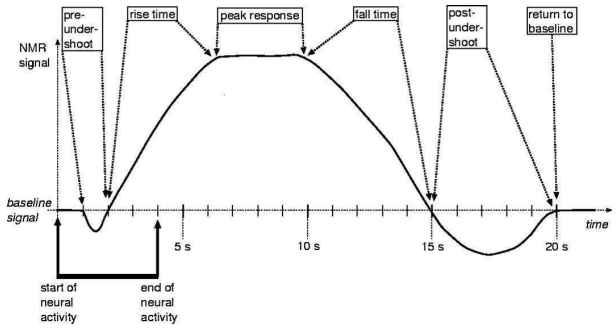


Figure 2.13: *Hemodynamical response function (hrf).*

2.1.3 Increased blood flow can be measured with fMRI

Neuroimaging tools for non-invasive monitoring of the working brain open new possibilities to study brain functionality and to further advance our understanding of the brain. Besides the EEG, also functional Magnetic Resonance Imaging (fMRI) is of extreme importance due to its capacity to locate brain activity with high spatial precision. This section explains the working principle of fMRI without going into detail about basic principles of MR imaging. For more details, we refer to available literature, *e.g.*, [288].

The oxygen demand is higher in the part of the brain that is active. To satisfy this demand, an increased flow of blood is regionally supplied to the population of active neurons. The mechanisms underlying this very local regulation of blood flow are not yet fully understood. Important is however that an excess of oxygen is supplied to the active neurons, leading to an increased concentration of oxygenated blood in the capillaries surrounding the active brain area. These spatial variations in oxygen concentration can be visualised by an MR scanner in an fMRI scan. The physical basis for fMRI imaging is the difference in magnetic properties between oxygenated and deoxygenated blood.

It is important to keep in mind that fMRI —in contrast to EEG— does not directly measure neuronal activity, but relies on the neurovascular coupling to encode information about brain functioning. The big advantage of fMRI is its high spatial resolution. This allows us to study brain functionality with millimeter-precision.

Detection of fMRI

The result of increased blood oxygenation is represented by the BOLD (blood oxygen level dependent) hemodynamic respons function (hrf). This hrf is

visualised in Fig. 2.13. The signal-to-noise ratio (SNR) of fMRI data is very low and fMRI data have most often been analysed by means of statistical methods. The standard analysis of fMRI data is based on building a general linear model (GLM) of the fMRI signal [122]. This GLM is constructed by identifying time points of interest and convolving the signal with the hrf. The signal at each voxel is statistically compared with the constructed model. A statistical map (t or F statistic) is then constructed where the value at each voxel reflects the resemblance between the model and the data and, therefore, the plausibility that this region is activated in reponse to the activity of interest. Thresholding this statistical map reveals then activated brain areas.
Recently, also techniques that are more data driven have been proposed for the analysis of fMRI data. The advantage of data driven techniques is that they are able to reveal unexpected activations, and that no time instants of interest have to be defined [30].

Application area of fMRI

fMRI is part of clinical practice to localise evoked brain activity for a wide range of applications. This is similar to EPs in the EEG. Data-driven methods can also be used to visualize resting state networks, see [59] and references herein.

2.1.4 Combining EEG and fMRI

As mentioned before, EEG has a very good temporal resolution, while fMRI has a very good spatial resolution. Simultaneous measurements would benefit from the complementarity of these techniques because both electrical and vascular activity can be measured at exactly the same time instance. In the past years, the inherent problems of measuring simultaneously electrical and magnetic signals have been solved by introducing MR-compatible electrodes. This opened a new exciting world in neuroscience as we are now able to measure brain activity with high temporal and high spatial resolution. For reviews we refer to [149] and references therein. Of course, in life one gets nothing for free. The price to pay is that new artifacts are introduced in the EEG: the pulse and scanner artifact. The scanner artifact is related to gradient switching of the scanner, and can be quite well removed with template matching algorithms [301]. The pulse artifact is supposed to be mainly caused by the blood pulsation in the arteries. New methods still appear for removing the pulse artifact, indicating that this problem has not yet fully been solved [4, 84, 346]. All the artifact removal techniques are decomposing the measurement data and removing the artifact related components.

Application area of EEG/fMRI

EEG-fMRI is a new technique to study haemodynamic cerebral changes related to electrical brain events. An example of an unprovocable event is epileptic activity. Such epileptic discharges occur at unknown time instants. By detecting them in the EEG, corresponding fMRI activations can be visualised [131, 23]. Data-driven techniques are especially useful when no or not enough epileptic spikes can be detected [302].
The second application field is monitoring the brain during evoked studies. In many evoked studies, the brain is supposed to react in the same way when the same stimulus is presented. In such cases, EEG and fMRI can be acquired and analysed in separated sessions. However, during some tasks (*e.g.*, gambling, attention studies, ...) and in neurological diseases like ADHD, the brain cannot be supposed to respond in an identical way at different time instants. Then it is worth to acquire EEG-fMRI data for analyzing single trial responses [85, 254, 101].

2.2 Brain dysfunction in epilepsy

Epilepsy is a chronical neurological disorder characterized by abnormal unprovoked synchronous neuronal discharges in the brain [22]. This is a rather vague definition. Epilepsy can be further subdivided in many types based on cause, observable manifestations (*e.g.*, disturbance of mental, motor, sensory or autonomic activity), brain location, or as a part of discrete, identifiable medical syndromes [12]. It has to be mentioned that in some cases the cause of epilepsy is even unknown.

2.2.1 Epileptic activity in the EEG

The EEG of patients with epilepsy can be divided in two parts. The clinical manifestation of epilepsy is the epileptic seizure or the ictal period. Between seizures, the EEG is called interictal activity. Both periods have typical waveforms [111, 112, 266]. Ictal activity consists of continuous rhythmical waves or repeated sharp spikes. These can gradually change in shape, frequency, amplitude or spatial distribution as the discharge progresses. In the literature, ictal discharges are said to begin with low amplitude rhythmical (β) activity that increases in amplitude and decreases in frequency. In reality, the (β) activity is not always visible. The ictal activity can stay very focal during the seizure or propagate over the brain. As an example, Fig. 2.15 shows a 5 sec scalp EEG recording from a long-term EMU. The epileptic activity is present in the left frontal region of the brain, at the $Fp1$, $F7$ and $F3$ electrodes.

Interictal activity consists of spikes or sharp waves. These may combine with slow waves to form a spike-and-wave complex. Fig. 2.14 shows a 4 s EEG recording containing several epileptic spikes in a bipolar montage. This activity has a high amplitude for the EEG signals $T6 - O4$ and $P4 - O2$. Interictal discharges are usually intermittent with little or no variation of shape. A huge number of spikes indicates a very bad state of the brain. Diagnosis of epilepsy is often based on the presence of spikes.
The exact shape of epileptiform activity has a large variety between patients. However, most adult patients have only one or a few interictal and ictal patterns.

2.2.2 Presurgical evaluation

Epilepsy is a common disease with an important medical and socioeconomic impact [152], affecting around 35.000 persons in Flanders. Around 30% have epileptic seizures that cannot be controlled by anti-epileptic drugs. This is called refractory

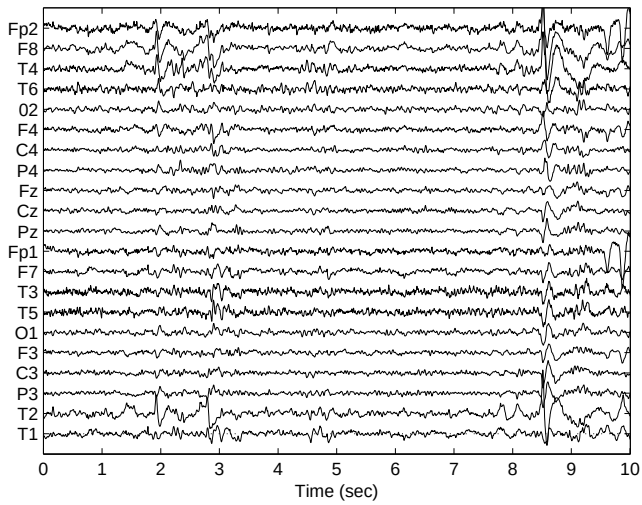


Figure 2.14: *Epoch of 4 s interictal EEG containing several epileptic spikes. A bipolar montage was used. This activity has a high amplitude for the EEG signals* $T6 - O2$ *and* $P4 - O2$.

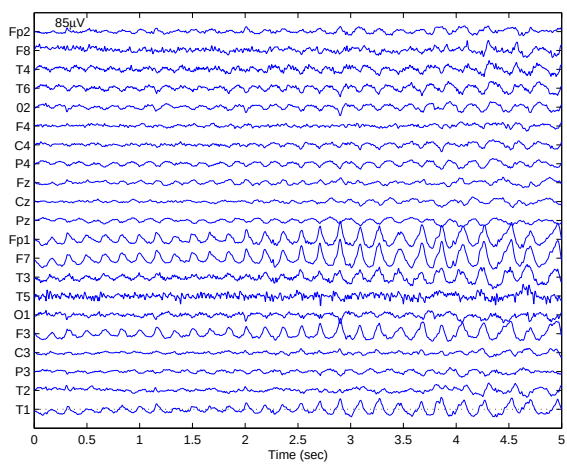


Figure 2.15: *Five sec of a scalp EEG recording from a long-term EMU. Clean ictal activity can be seen on the* $Fp1$, $F7$ *and* $F3$ *electrodes.*

epilepsy. Up to 3% of patients with epilepsy are candidates for epilepsy surgery in which neurosurgeons try to perform accurate brain surgery to cure the epilepsy. A majority of patients with refractory partial epilepsy who are considered good candidates after a presurgical evaluation are rendered seizure free after surgery. These patients have a significantly improved quality of life (relationships, studies, driving, employment) and improvements in cognitive functions, and have less psychiatric morbidity [333].
Successful surgery relates directly to the confidence with which the "epileptogenic zone", theoretically defined as the minimum volume of brain tissue whose resection is both necessary and sufficient for seizure ablation, can be identified [304]. This epileptogenic zone is a theoretical concept and a single, non-invasive gold standard technique for defining the epileptogenic zone is still lacking. In clinical practice, it is estimated based on the irritative and the seizure onset zone. The irritative zone is the cerebral generator of interictal epileptic activity recorded on scalp EEG. It is closely related to this epileptogenic zone but remains anatomically difficult to define using EEG alone. The seizure onset zone is the region in which the seizures actually originate.

New imaging techniques and breakthroughs in the mapping of the human brain are enabling to pinpoint the foci of the seizures more accurately. Nowadays, a variety of diagnostic methods are used to estimate the epileptogenic zone area during a presurgical evaluation. In our epilepsy monitoring unit (EMU), this localisation is based on the combination of behavioral ictal symptoms and signs on video recordings of seizures, Magnetic Resonance Imaging (MRI)-visible epileptic lesions, regions of hyperperfusion on ictal single photon emission computerized tomography (SPECT), and interictal and ictal epileptic activity on EEG recordings.

Surgical candidates are monitored continuously during about five days with Video-EEG monitoring. The anti-epileptic treatment is decreased or stopped to record several habitual seizures during this week. A structural MRI scan can reveal abnormalities causing the epilepsy [118]. This is especially useful in cortical dysplasia [106].

Ictal SPECT can visualize the ictal onset zone after injecting ^{99m}Tc ethyl cysteinate dimer (^{99m}Tc-ECD) as a blood flow tracer. As soon as possible after the start of an epileptic seizure, the tracer should be injected. It rapidly crosses the bloodbrain barrier. Around 85% is taken up by the brain during the first pass. Cerebral uptake is complete within 2 minutes and less than 5% is redistributed later. An ictal SPECT scans the brain and shows cerebral perfusion. Mostly SPECT data is visualized by comparing the ictal with interictal SPECT data. Methodologically, after computer-aided voxel-based coregistration, this is called subtraction ictal SPECT co-registered to MRI (SISCOM) [278].

Additional techniques that are sometimes used are magnetoencephalography

(MEG) and intracranial recordings [323, 304, 321, 41, 40]. MEG is an imaging technique used to measure the magnetic fields produced by electrical activity in the brain via extremely sensitive devices such as superconducting quantum interference devices (SQUIDs). The advantage over EEG is that magnetic signals are less distorted by passing through the skull and scalp, and a higher signal-to-noise ratio can be obtained. The disadvantage is that MEG electrodes are only perceptive for tangentially oriented neurons of a current source in a spherical volume conductor.

Intracranial recordings or electrocorticography (ECoG) is the practice of using electrodes placed directly on the exposed surface of the brain to record electrical activity from the cerebral cortex. ECoG may be performed either in the operating room during surgery (intraoperative ECoG) to localise brain functions or outside surgery (extraoperative ECoG). As the surgical incision into the skull requires to implant the electrode grid, ECoG is an invasive procedure. Some people call ECoG the gold standard for defining epileptogenic zones in clinical practice as it can provide precise localisation and extent of the lesion. These recordings are also not contaminated by artifacts [148]. However, one should keep in mind that grid electrodes can only by implanted in a small area in the brain. There is always the risk of not covering the suspicious region. Furthermore, invasive examinations carry a risk of complications such as bleedings and infections in 2% of the patients [300].

To conclude, results of the different analyses are coregistered and visualised on a multimodal imaging platform. When the findings of the different examinations suspect the same area to cause epilepsy and provided that there is no overlap with eloquent cortex, surgery can be considered.

2.3 Conclusion

This chapter mentioned the principles of measurement techniques in neuroscience and it has given some background in epilepsy. The EEG measures the synchronised electrical activity of the brain. In the EEG, the epileptic seizure can be recognised as a rhythmical wave starting in a specific brain region, the epileptogenic zone.

Chapter 3

Background: Mathematical decomposition techniques

This chapter introduces the mathematics, necessary for a good understanding of the thesis. The introduction provides some basic definitions. Afterwards, the definition of rank in linear and multi-linear algebra is discussed. Finally, an overview of important decomposition techniques is given.

3.1 Introduction

Stacking data

A signal in one dimension can be represented by a vector $\mathbf{s} \in \mathbb{R}^{1\times J}$. Two-dimensional data can be stored in a matrix $\mathbf{X} \in \mathbb{R}^{I\times J}$. When recorded data are intrinsically more than two-dimensional, the data is stored in a higher-order tensor $\mathcal{A} \in \mathbb{R}^{I_1\times I_2\times\ldots\times I_n}$. We first define our notation. Italic capitals are used to denote index upper bounds (e.g., $i = 1, 2, \ldots, I$). Scalars are denoted by lower-case letters (a, b, ...), vectors are written in bold-face lower-case ($\mathbf{a}$, $\mathbf{b}$, ...) or capitals (A, B ...), matrices correspond to bold-face capitals ($\mathbf{A}$, $\mathbf{B}$, ...) and tensors are written as calligraphic letters ($\mathcal{A}$, $\mathcal{B}$, ...). The notation will be consistently used for lower-order parts of a given structure. Hence, the entry with row index i and column index j in a matrix $\mathbf{A}$, i.e. $(\mathbf{A})_{ij}$, is symbolised by a_{ij}.

For tensors, a mode-n vector of a tensor $\mathcal{A} \in \mathbb{R}^{I_1\times I_2\times I_3}$ is the I_n-dimensional vector obtained from $\mathcal{A}$ by varying the index i_n and keeping the other indices fixed [181]. Alternatively, the mode-n vectors of a third-order tensor can be called

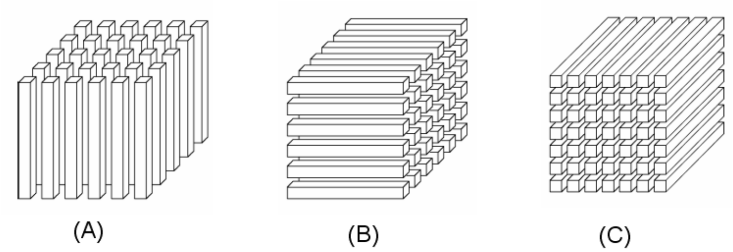


Figure 3.1: *(A) Columns, (B) rows and (C) tubes in a third-order tensor.*

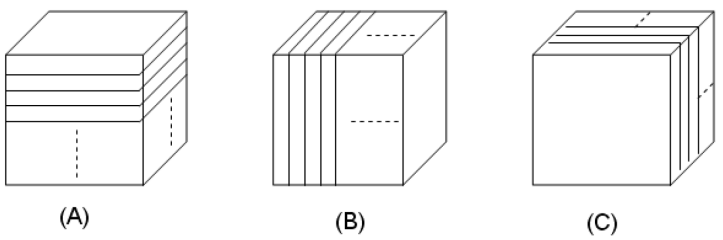


Figure 3.2: *(A) Horizontal, (B) vertical and (C) frontal slices.*

columns, rows and tubes as illustrated in Fig. 3.1. By fixing only 1 mode, a matrix or a tensor slice is obtained, as shown in Fig. 3.2. By aligning these matrix slices, one can unfold tensors and obtain a matrix [338]. This is visualised in Fig. 3.3. We define unfolding in formulas as:

Definition 3.1 *Assume an Nth-order tensor* $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$. *The matrix unfolding* $\mathbf{A}_{(n)} \in \mathbb{R}^{I_n \times (I_{n+1} I_{n+2} \ldots I_N I_1 I_2 \ldots I_{n-1})}$ *contains the element* $a_{i_1 i_2 \ldots i_N}$ *at the position with row number* i_n *and column number equal to* $(i_{n+1} - 1) I_{n+2} I_{n+3} \ldots I_N I_1 I_2 \ldots I_{n-1} +$
$(i_{n+2} - 1) I_{n+3} I_{n+4} \ldots I_N I_1 I_2 \ldots I_{n-1} + \ldots + (i_N - 1) I_1 I_2 \ldots I_{n-1} + (i_1 - 1) I_2 I_3 \ldots I_{n-1} +$
$(i_2 - 1) I_3 I_4 \ldots I_{n-1} + \ldots + i_{n-1}$.

For a third order tensor this definition becomes:

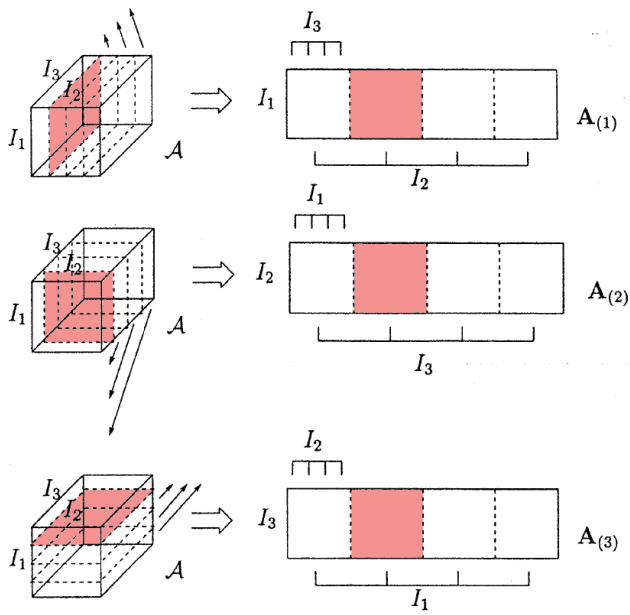


Figure 3.3: *Unfolding of the* $(I_1 \times I_2 \times I_3)$*-tensor* $\mathcal{A}$ *to the* $(I_1 \times I_2 I_3)$*-matrix* $\mathbf{A}_{(1)}$*, the* $(I_2 \times I_3 I_1)$*-matrix* $\mathbf{A}_{(2)}$ *and the* $(I_3 \times I_1 I_2)$*-matrix* $\mathbf{A}_{(3)}$*. From [210].*

Definition 3.2 *The standard matrix representation* $\boldsymbol{X}_{(1)} \in \mathbb{R}^{I_1 \times I_2 \cdot I_3}$*,* $\boldsymbol{X}_{(2)} \in \mathbb{R}^{I_2 \times I_1 \cdot I_3}$ *and* $\boldsymbol{X}_{(3)} \in \mathbb{R}^{I_3 \times I_1 \cdot I_2}$ *of a tensor* $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ *are defined by*

$$
\begin{aligned}
(\boldsymbol{X}_{(1)})_{i_1,(i_2-1)I_3+i_3} &= (\mathcal{X})_{i_1 i_2 i_3}, \\
(\boldsymbol{X}_{(2)})_{i_2,(i_3-1)I_1+i_1} &= (\mathcal{X})_{i_1 i_2 i_3}, \\
(\boldsymbol{X}_{(3)})_{i_3,(i_1-1)I_2+i_2} &= (\mathcal{X})_{i_1 i_2 i_3},
\end{aligned}
$$

for all values of the indices [181].

The standard vector representation $(\mathcal{X})_{I_1 I_2 I_3} = \mathbf{x} \in \mathbb{R}^{I_1 I_2 I_3 \times 1}$ of $\mathcal{X}$ can be defined as

$$
\mathbf{x}_{(i_1-1)I_2 I_3+(i_2-1)I_3+i_3} = (\mathcal{X})_{i_1 i_2 i_3},
$$

for all values of the indices. Note that in these definitions indices to the right vary more rapidly than indices to the left.

For the vectorisation of a matrix $\mathbf{X} = [\mathbf{x}_1\ \mathbf{x}_2\ \ldots]$ we keep the following convention: $\text{vec}(\mathbf{X}) = [\mathbf{x}_1^T\ \mathbf{x}_2^T\ \ldots]^T$. The $(N \times N)$ identity matrix is represented by $\mathbf{I}_{N\times N}$. The $(N \times M)$ zero matrix is denoted by $\mathbf{0}_{N\times M}$. $\mathbf{1}_J$ represents a column vector of all ones of length J. The zero tensor is denoted by $\mathcal{O}$.

Basic definitions

To avoid confusion, we define clearly some important operators and concepts. The symbol $\otimes$ denotes the Kronecker product,

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \ldots \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \ldots \\ \vdots & \vdots & \end{pmatrix},$$

and $\odot$ represents the Khatri-Rao or column-wise Kronecker product [296]:

$$\mathbf{A} \odot \mathbf{B} = (\mathbf{a}_1 \otimes \mathbf{b}_1 \ldots \mathbf{a}_R \otimes \mathbf{b}_R).$$

The superscripts $\cdot^T$ and $\cdot^\dagger$ denote the transpose and Moore-Penrose pseudo-inverse, respectively. The operator $\text{diag}(\cdot)$ stacks its scalar arguments in a square diagonal matrix.

In this paper we denote the Tucker mode-n product $\times_n$ in the same way as in [53, 55, 210, 211, 215]; in the literature the symbol $\bullet_n$ is also used.

Definition 3.3 *Consider* $\mathcal{X} \in \mathbb{R}^{I_1\times I_2\times I_3}$ *and* $\mathbf{A} \in \mathbb{R}^{J_1\times I_1}$, $\mathbf{B} \in \mathbb{R}^{J_2\times I_2}$, $\mathbf{C} \in \mathbb{R}^{J_3\times I_3}$. *Then the* Tucker mode-1 product $\mathcal{X} \times_1 \mathbf{A}$, mode-2 product $\mathcal{X} \times_2 \mathbf{B}$ and mode-3 product $\mathcal{X} \times_3 \mathbf{C}$ *are defined by*

$$\begin{aligned}
(\mathcal{X} \times_1 \mathbf{A})_{j_1 i_2 i_3} &= \sum_{i_1=1}^{I_1} x_{i_1 i_2 i_3} a_{j_1 i_1}, \qquad \forall j_1, i_2, i_3, \\
(\mathcal{X} \times_2 \mathbf{B})_{i_1 j_2 i_3} &= \sum_{i_2=1}^{I_2} x_{i_1 i_2 i_3} b_{j_2 i_2}, \qquad \forall i_1, j_2, i_3, \\
(\mathcal{X} \times_3 \mathbf{C})_{i_1 i_2 j_3} &= \sum_{i_3=1}^{I_3} x_{i_1 i_2 i_3} c_{j_3 i_3}, \qquad \forall i_1, i_2, j_3,
\end{aligned}$$

respectively [338].

Definition 3.4 *The* Frobenius norm *of a tensor* $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$ *is defined as*

$$\|\mathcal{X}\| = \left(\sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} |x_{ijk}|^2 \right)^{\frac{1}{2}}.$$

Definition 3.5 *The outer product* $\mathcal{A} \circ \mathcal{B}$ *of a tensor* $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_P}$ *and a tensor* $\mathcal{B} \in \mathbb{R}^{J_1 \times J_2 \times \ldots \times J_Q}$ *is the tensor defined by*

$$(\mathcal{A} \circ \mathcal{B})_{i_1 i_2 \ldots i_P j_1 j_2 \ldots j_Q} = a_{i_1 i_2 \ldots i_P} b_{j_1 j_2 \ldots j_Q},$$

for all values of the indices.

For example, the outer product $\mathcal{X}$ of three vectors $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ is defined by $x_{ijk} = a_i b_j c_k$, for all values of the indices.

Definition 3.6 *The* Kruskal rank *or* k-rank *of a matrix* $\mathbf{A}$*, denoted by* $\text{rank}_k(\mathbf{A})$ *or* $k_{\mathbf{A}}$*, is the maximal number* r *such that any set of* r *columns of* $\mathbf{A}$ *is linearly independent.*

3.2 Matrix rank and the singular value decomposition (SVD)

Two-dimensional data can be stored in a matrix $\mathbf{X} \in \mathbb{R}^{I \times J}$.

Definition 3.7 *The number of linearly independent rows in a matrix is called the row rank of the matrix, and the number of linearly independent columns is called the column rank.*

These ranks can also be seen as the dimension of respectively the row space and the column space of the matrix. One of the basic properties in linear algebra is that for every matrix the row rank is equal to the column rank and this is called *the* rank of the matrix. The rank of a matrix $\mathbf{A}$ will be denoted by $\text{rank}(\mathbf{A})$ or $r_{\mathbf{A}}$. As an example; the outer product of two non-zero vectors $\mathbf{a}$ and $\mathbf{b}$ equals a matrix of rank 1, because all rows and columns are multiples of each other.

A basic concept in linear algebra is the Singular Value Decomposition (SVD).

Definition 3.8 *Consider a matrix* $\boldsymbol{X} \in \mathbb{R}^{I \times J}$ *with rank* R*, the SVD is defined as*

$$\boldsymbol{U} \cdot \Sigma \cdot \boldsymbol{V}^T \tag{3.1}$$

with $\boldsymbol{U} \in \mathbb{R}^{I \times I}$ and $\boldsymbol{V} \in \mathbb{R}^{J \times J}$ orthogonal matrices and Σ a non-negative diagonal matrix. The elements on the diagonal of Σ are called the singular values σ_{ii}. These singular values are sorted in non-increasing order.

The SVD also enables us to write $\mathbf{X}$ as a sum of rank-1 terms:

$$\mathbf{X} = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \ldots + \sigma_R \mathbf{u}_R V_R^T \tag{3.2}$$

with $\mathbf{u}_j$ and $\mathbf{v}_j$ the jth column of $\mathbf{U}$ and $\mathbf{V}$ respectively. We can also write the individual elements of $\mathbf{X}$ as:

$$x_{ij} = \sum_{r=1}^{R} u_{ir} \sigma_{rr} v_{jr}. \tag{3.3}$$

Theorem 3.1 *The rank of a matrix $\boldsymbol{X}$ can (also) be defined as the minimal number of rank-1 terms whose sum equals $\boldsymbol{X}$.*

This SVD decomposition, restricted to R components, is visualised in Fig. 3.4.

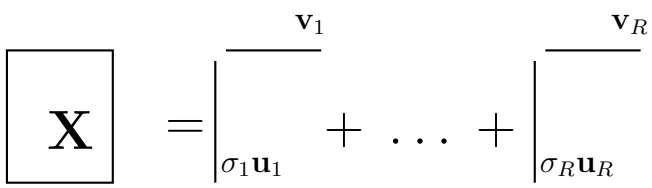


Figure 3.4: *Singular Value Decomposition of $\boldsymbol{X}$ restricted to R components.*

The SVD is important because of the famous Theorem of Eckart-Young. If we define the truncated SVD as the SVD restricted to the first ρ components, we have:

Theorem 3.2 *Let the rank-ρ matrix $\boldsymbol{Y}$ with $\rho \leqslant R$ be given by the truncated SVD of $\boldsymbol{X}$ (of rank R), then $\boldsymbol{Y}$ is the best rank-ρ approximation of $\boldsymbol{X}$ in the least-squares sense, i.e., $||\boldsymbol{X} - \boldsymbol{Y}||^2$ is minimal.*

This theorem tells us that the SVD provides all the best rank-ρ approximations of $\mathbf{X}, \rho = 1, \ldots, R$.

3.3 Rank in multilinear algebra

We consider a third-order tensor $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$. There are several ways to generalise the rank of a matrix.

Definition 3.9 *The mode-n rank of $\mathcal{X}$ is the rank of the subspace spanned by the mode-n vectors, n=1,2,3.*

This rank definition is analogous to that of the column and row rank of matrices.

Definition 3.10 *We say that a third-order tensor is* rank-(L, M, N) *if its mode-1 rank, mode-2 rank and mode-3 rank are equal to L, M and N, respectively.*

Definition 3.11 The *rank of a tensor $\mathcal{X}$ is defined as the smallest number of rank-*1 *tensors whose sum equals $\mathcal{X}$.*

Again, a rank-1 tensor equals the outer product of three non-zero vectors [199]. There are major differences between the rank of matrices and tensors. The different mode-n ranks can be different in the higher-order case. The rank of a tensor can be larger than L, M or N.

In contrast with linear algebra, where the best rank-ρ approximation of a matrix is well-defined, there is no higher order variant of Theorem 3.2 in multilinear algebra. A counterexample was given in [315], inspired by [190]. Let $\mathbf{a}_i, \mathbf{b}_i \in \mathbb{R}^{1 \times J}, i = 1, 2, 3$. Let

$$\mathcal{X} = \mathbf{a}_1 \circ \mathbf{a}_2 \circ \mathbf{b}_3 + \mathbf{a}_1 \circ \mathbf{b}_2 \circ \mathbf{a}_3 + \mathbf{b}_1 \circ \mathbf{a}_2 \circ \mathbf{a}_3 \tag{3.4}$$

and for $n \in \mathbb{N}$,

$$\mathcal{X}_n = \mathbf{a}_1 \circ \mathbf{a}_2 \circ (\mathbf{b}_3 - n\mathbf{a}_3) + (\mathbf{a}_1 + \frac{1}{n}\mathbf{b}_1) \circ (\mathbf{a}_2 + \frac{1}{n}\mathbf{b}_2) \circ n(\mathbf{a}_3) \tag{3.5}$$

For this tensor, the rank of $\mathcal{X}$ is different from three if $\mathbf{a}_i, \mathbf{b}_i$ are linearly independent. However, the rank of $\mathcal{X}_n$ equals or is larger than two and

$$\lim_n \mathcal{X}_{n \to \infty} = \mathcal{X}. \tag{3.6}$$

3.4 One-dimensional signal decomposition techniques

A signal in one dimension can be represented by a vector $\mathbf{s} \in \mathbb{R}^{1\times J}$. If this signal still contains different signal components, one could exploit the available time and frequency information contained in the signal. We review some concepts, as they can also be used as building blocks for more complex decomposition techniques.
The discrete Fourier transform (DFT) is a specific kind of Fourier transform [92]. It transforms one function in the time domain into its counterpart in the frequency domain. It decomposes a signal as a sum of different sinusoidal components.
Discrete wavelets [62, 341] decompose a signal into signals with frequencies in a limited frequency band instead of having individual frequencies. The signal is separated successively into slow and fast components using a pair of finite impulse response filters. The slow component in each stage is again submitted to two similar filters in the next stage so that every stage extracts slower components than the previous stage. The discrete wavelet can be useful to visualize slower and faster components, but it is very improbable that contributing parts of the signal will be nicely separated by the decomposition.
Matching pursuit [238] is a numerical technique that estimates the "best matching" projections of data onto an over-complete dictionary $\mathcal{D}$. The basic idea is to represent a signal $\mathbf{f}(t)$ from Hilbert space $\mathcal{H}$ as a weighted sum of functions g_{γ_n} (called atoms) taken from $\mathcal{D}$:

$$\mathbf{f}(t) = \sum_{n=0}^{\infty} a_n g_{\gamma_n}. \tag{3.7}$$

This technique decomposes on a very local scale the signal and has already been applied to individual channels of the EEG, see *e.g.*, [97].
A new adaptive decomposition technique was introduced, the empirical mode decomposition (EMD) [162]. It represents a signal, that can be non-stationary of have non-linear dynamics, as a sum of simpler components with amplitude- and frequency- modulated parameters. The authors of [162] claims that this technique is able to reveal components, overlapping in both time and frequency, that cannot be separated by other filtering techniques. However, there is no mathematical theory for this method.

Another way to extract contributing signal components from a one-dimensional vector is by re-arranging this vector into a more-dimensional structure in order to estimate the underlying components. In [343], the components are assumed to be exponentially damped sinusoids. Since exponentially damped sinusoids are a set of basis functions for real continuous functions, they can be used to model any arbitrary signal sufficiently close, provided that the model order is high enough. In the parametric Exponentially Damped Sinusoidal (EDS) model, the signal is modeled as a finite sum of K discrete-time exponentially damped complex

sinusoids:

$$\begin{aligned} x(n) &= \sum_{k=1}^{K} a_k \exp(j\phi_k) \exp\left((-d_k + j2\pi f_k) n\Delta t\right) + e(n), \\ n &= 0, 1, \ldots, N-1; \end{aligned} \tag{3.8}$$

where $j = \sqrt{-1}$, $n\Delta t$ is the time lapse between the origin and the sample $x(n)$, Δt is the sampling interval, N is equal to the number of data points and $e(n)$ is white gaussian noise. The parameters of this model are the amplitudes a_k, the phases ϕ_k, the damping factors d_k and the frequencies f_k. The model function (3.8) can be rewritten in terms of complex amplitudes $c_k = a_k \exp(j\phi_k)$ and signal poles $z_k = \exp\left((-d_k + j2\pi f_k)\Delta t\right)$ as follows

$$\begin{aligned} x(n) &= \sum_{k=1}^{K} c_k z_k^n + e(n), \\ n &= 0, 1, \ldots, N-1. \end{aligned} \tag{3.9}$$

The EDS model can be used to approximate all real-valued signals since any real-valued sinusoid can be expressed as a superposition of two conjugated complex exponentials. For more information and extensions, we refer to [270, 150, 221, 258, 257].

A last concept worth mentioning, is the so-called single channel ICA [64]. This technique aims at extracting several independent sources from one single measurement. However, more research is still needed to validate the proposed algorithm(s) or to develop new algorithms.

3.5 Two-dimensional matrix decomposition techniques

We will discuss three decompositions: Principal Component Analysis (PCA), Independent Component Analysis (ICA) and Canonical Correlation Analysis (CCA). We conclude with a note on the quality of a decomposition.

3.5.1 Principal Component Analysis (PCA)

PCA [287, 99] of a matrix $\mathbf{X} \in \mathbb{R}^{I \times J}$ with rank R computes orthogonal matrices $\mathbf{A} \in \mathbb{R}^{I \times R}$ and $\mathbf{B} \in \mathbb{R}^{J \times R}$ as

$$\mathbf{X} = \mathbf{A}\mathbf{B}^T \text{ for which } \|\mathbf{X} - \mathbf{A}\mathbf{B}^T\|^2 \text{ is minimal.} \tag{3.10}$$

The columns of $\mathbf{A} \in \mathbb{R}^{I \times R}$ are called the factors and the columns of $\mathbf{B} \in \mathbb{R}^{J \times R}$ the loadings. The factors are linear combinations of the measurements and are uncorrelated. Alternatively, we can write:

$$x_{ij} = \sum_{r=1}^{R} a_{ir} b_{jr}. \tag{3.11}$$

It has been derived that a PCA solution can be obtained as the truncated SVD of $\mathbf{X}$. However, a PCA solution is not unique:

$$\mathbf{A} \cdot \mathbf{B}^T = (\mathbf{A}\mathbf{Q}) \cdot (\mathbf{Q}^T\mathbf{B}^T) \text{ for } \mathbf{Q} \text{ with } \mathbf{Q}\mathbf{Q}^T = \mathbf{Q}^T\mathbf{Q} = \mathbf{I}_R \tag{3.12}$$

gives an equivalent solution of the minimisation problem (3.10). So every orthogonal rotation $\mathbf{Q}$ gives a different basis for the factor space spanned by the columns of $\mathbf{A}$. If the rotation $\mathbf{Q}$ yields a simple structure in the loadings matrix $(\mathbf{Q}^T\mathbf{B}^T)$, then the factors are easier to interpret. PCA and SVD are closely related: they both provide the decomposition in orthogonal components. The PCA decomposition can also be visualised by Fig. 3.4. The picture tells us that with PCA a matrix can be decomposed into rank-1 signals that are orthogonal both in column and in row space.

3.5.2 Independent Component Analysis (ICA)

We start with some important definitions. Then, we introduce the ICA problem and we will explain how it can be solved. Afterwards, we discuss some specific algorithms. We conclude this section by giving more details on Jacobi rotations for simultaneous diagonalisation, because it is often used in ICA algorithms.

Some definitions.

The basic higher-order statistics (HOS) are higher-order moments and cumulants. For random vectors these quantities are higher-order tensors. We have the following definitions.

Definition 3.12 *The Nth order* moment *of a real stochastic vector* $\mathbf{s}$ *is defined by*

$$\mathcal{M}_{\mathbf{s}}^{(N)} \stackrel{\text{def}}{=} \mathrm{E}\{\mathbf{s} \circ \mathbf{s} \circ \ldots \circ \mathbf{s}\}, \tag{3.13}$$

in which the expectation over the Nth order outer product is performed in a component-wise manner.

The first-order moment is the mean of the stochastic vector. The second-order moment is the correlation matrix [272].

Definition 3.13 *For a real zero-mean stochastic vector* $\mathbf{s}$ *the* cumulants *up to order 4 are given by:*

$$(C_{\mathbf{s}})_i = \mathrm{Cum}(s_i) \stackrel{\text{def}}{=} \mathrm{E}\{s_i\}, \tag{3.14}$$

$$(\mathbf{C}_{\mathbf{s}})_{i_1 i_2} = \mathrm{Cum}(s_{i_1}, s_{i_2}) \stackrel{\text{def}}{=} \mathrm{E}\{s_{i_1} s_{i_2}\}, \tag{3.15}$$

$$(\mathcal{C}_{\mathbf{s}}^{(3)})_{i_1 i_2 i_3} = \mathrm{Cum}(s_{i_1}, s_{i_2}, s_{i_3}) \stackrel{\text{def}}{=} \mathrm{E}\{s_{i_1} s_{i_2} s_{i_3}\}, \tag{3.16}$$

$$\begin{aligned}(\mathcal{C}_{\mathbf{s}}^{(4)})_{i_1 i_2 i_3 i_4} = \mathrm{Cum}(s_{i_1}, s_{i_2}, s_{i_3}, s_{i_4}) \stackrel{\text{def}}{=} & \mathrm{E}\{s_{i_1} s_{i_2} s_{i_3} s_{i_4}\} \\ & - \mathrm{E}\{s_{i_1} s_{i_2}\}\mathrm{E}\{s_{i_3} s_{i_4}\} \\ & - \mathrm{E}\{s_{i_1} s_{i_3}\}\mathrm{E}\{s_{i_2} s_{i_4}\} \\ & - \mathrm{E}\{s_{i_1} s_{i_4}\}\mathrm{E}\{s_{i_2} s_{i_3}\}. \end{aligned} \tag{3.17}$$

For every component s_i of $\mathbf{s}$ that has a non-zero mean, s_i has to be replaced in these formulas, except Eq. (3.14), by $s_i - \mathrm{E}\{s_i\}$.

The first-order cumulant is also equal to the mean of the stochastic vector. The second-order cumulant is the covariance matrix. The definition for arbitrary cumulant orders can be found in [209, 271, 272]. Cumulants are said to be marginal if

ICA Problem.

Assume the basic linear statistical model

$$\mathbf{Y} = \mathbf{M} \cdot \mathbf{X} \quad (+\mathbf{N}) \tag{3.18}$$

where $\mathbf{Y} \in \mathbb{R}^{I \times J}$ is called the observation matrix and $\mathbf{X} \in \mathbb{R}^{P \times J}$ the source matrix. $\mathbf{N} \in \mathbb{R}^{I \times J}$ is additive noise, which will not be explicitly modeled. $\mathbf{M} \in \mathbb{R}^{I \times P}$ is the mixing matrix. I is the number of sensors, P is the number of sources, and J the number of samples in each observation.
The goal is to estimate the mixing matrix $\mathbf{M}$, and/or the source matrix $\mathbf{X}$, given only the observation matrix $\mathbf{Y}$. In this work, we restrict our discussion to the so-called overdetermined case, i.e., $P \leqslant I$. For strategies to solve the underdetermined ICA problem, where there are more sources than sensors, we refer to [206, 205] and references herein. We do not consider convolutive models either. For this, we refer to [42].
We note that ICA is subject to two basic indeterminacies. First, it is impossible to determine the norm of the mixing vectors, the columns of $\mathbf{M}$ in Eq. (3.18), since a rescaling of these vectors can be compensated by the inverse scaling of the sources. Similarly, the ordering of the source signals cannot be identified.

ICA Solution.

In order to solve the problem, the hypothesis that the sources are mutually statistically independent is often made [35]. With this hypothesis, sources with gaussian distributions cannot be estimated [60]. Besides the assumption of statistical independence, another common assumption is that the sources are temporally colored (see below). In this way, also gaussian distributions can be estimated.
Also other models can be used in order to estimate underlying sources, *e.g.*, assuming discrete sources [133] or exploiting the fact that sources are localised in the time-frequency domain [364].

In order to solve the problem of estimating statistically independent sources, *contrast functions* have been introduced [95]. The global maximum of these functions is only reached when the sources are fully separated. So, in order to solve the source separation problem, one has to define a contrast function and estimate its global optimum.

As the extraction of sources is based on the the statistical independence of these sources, it is natural to minimize a criterion that measures the dependence of the

source signals. Such a criterion can be the mutual information [50, 289], defined for R sources $\mathbf{y}_r, r = 1, \ldots, R$ as:

$$I(\mathbf{y}_1, \ldots, \mathbf{y}_R) = \sum_{r=1}^{R} H(\mathbf{y}_r) - H(\mathbf{y}_1, \ldots, \mathbf{y}_R) \tag{3.19}$$

with H the (marginal or joint) entropy. A possible contrast function can thus be $-I(\mathbf{y}_1, \ldots, \mathbf{y}_R)$.

Afterwards, it was realised that minimizing the mutual information leads to maximizing the marginal cumulants of order 3 and 4 [51]. This was further generalised [255]: the maximisation of arbitrary higher-order marginal cumulants can be interpreted as the maximisation of a contrast function.

These higher order cumulants are higher order tensors. Maximisation of the marginal cumulants corresponds to eliminating the cross cumulants, or, said in an other way, diagonalizing the tensor. Due to multilinear properties, a tensor can in general not be exactly diagonalised. This explains why there is a variety of algebraic techniques "to diagonalize the tensor as much as possible". A theoretical overview of ICA methods is given in [54] and an overview of some practical algorithms is given in [164, 209].

Pre-whitening.

From an algebraic point of view, saying that source signals in the rows of $\mathbf{X}$ are uncorrelated means that the covariance matrix of $\mathbf{X}$, $\mathbf{C}_\mathbf{X}$, is a diagonal matrix. Saying that the signals in $\mathbf{X}$ are independent also implies that all the higher-order cumulants of $\mathbf{X}$, $\mathcal{C}_\mathbf{X}^{(N)}$, are diagonal. With the linear model from equation (3.18), we can now write:

$$\mathbf{C}_\mathbf{Y} = \mathbf{C}_\mathbf{X} \times_1 \mathbf{M} \times_2 \mathbf{M} + \mathbf{C}_\mathbf{N} \tag{3.20}$$

$$\mathcal{C}_\mathbf{Y}^{(N)} = \mathcal{C}_\mathbf{X}^{(N)} \times_1 \mathbf{M} \times_2 \mathbf{M} \ldots \times_N \mathbf{M} + \mathcal{C}_\mathbf{N}^{(N)} \tag{3.21}$$

in which $\mathbf{C}_\mathbf{X}$ and $\mathcal{C}_\mathbf{X}^{(N)}$ are diagonal and $\mathcal{C}_\mathbf{N}^{(N)}$ drops if the noise is gaussian. Different kinds of algebraic ICA algorithms can be distinguished, depending on how equations (3.20) and (3.21) are combined. One approach is to solve the ICA problem using only higher-order statistics. For algebraic techniques, we refer to [32, 206, 208]. Equations (3.20) and (3.21) can also be combined in a more or less balanced way [361, 215]. However, most algorithms start with a so-called

pre-whitening step. In this pre-whitening step, as much information as possible is extracted from (3.20). In this stage, the mixing matrix can be identified up to an orthogonal rotation matrix. Afterwards, (3.21) has to be solved in some approximative manner.

In the pre-whitening step, the observation matrix $\mathbf{Y}$ is transformed into an other matrix of which the corresponding stochastic variables have unit covariance. One starts with the computation of the covariance matrix $\mathbf{C_Y}$ of the signal. We assume for convenience of notation that the signals are zero-mean. We have:

$$\mathbf{C_Y} = \frac{1}{J}\mathbf{Y} \cdot \mathbf{Y}^T \tag{3.22}$$

$$= \frac{1}{J}(\mathbf{M} \cdot \mathbf{X})(\mathbf{M} \cdot \mathbf{X})^T \tag{3.23}$$

$$= \mathbf{M} \cdot \mathbf{C_X} \cdot \mathbf{M}^T \tag{3.24}$$

$$= (\mathbf{US}) \cdot (\mathbf{US})^T. \tag{3.25}$$

In the last step, we assume without loss of generality that the source signals have unit variance, and we replace $\mathbf{M}$ by its singular value decomposition $\mathbf{USV}^T$. A possible whitening matrix is then given by the pseudo-inverse of $\mathbf{US}$. We denote the pre-whitened observation matrix as

$$\mathbf{Y}_w = (\mathbf{US})^\dagger \cdot \mathbf{Y} \in \mathbb{R}^{P \times J}. \tag{3.26}$$

The pre-whitened mixing matrix is denoted by

$$\mathbf{M}_w = (\mathbf{US})^\dagger \cdot \mathbf{M} = \mathbf{V}^T. \tag{3.27}$$

In many applications of ICA (e.g., EEG, fMRI, ...), the number of sensors is larger than the number of sources. In the pre-whitening step, the dimensionality of the ICA problem is reduced by projecting $\mathbf{Y}$ onto the signal subspace. From the equations, it can be seen that the mixing matrix $\mathbf{M} = \mathbf{USV}^T$ can only be estimated from (3.25) up to an orthogonal matrix $\mathbf{V}$.

Although many ICA algorithms start with a pre-whitening step, this step has the disadvantage that the calculations are directly affected by additive Gaussian noise. Recall that (3.21) is blind to additive Gaussian noise. Errors introduced in the pre-whitening step cannot be completely compensated in the higher-order step [34, 213].

Joint Approximate Diagonalisation of Eigenmatrices (JADE).

The assumption of mutually statistically independent sources means algebraically that all the higher-order cumulants of $\mathbf{X}$ are diagonal tensors. From (3.21) it can be seen that the mixing matrix can be estimated as the matrix that diagonalizes the higher-order cumulant. This mixing matrix will be an orthogonal matrix when the observation matrix is first pre-whitened ($\mathbf{M}_w = \mathbf{V}^T$). A computationally efficient algorithm, called Joint Approximate Diagonalisation of Eigenmatrices (JADE) was introduced in [37].

When the cumulant slices of $\mathcal{C}^{(4)}_{\mathbf{Y}_w}$ are rearranged into a matrix $\mathbf{C}^{(4)}_{\mathbf{Y}_w} \in \mathbb{R}^{P^2 \times P^2}$, it was shown that the mixing matrix can be estimated as the orthogonal matrix that approximately *simultaneously* diagonalizes the matrices that correspond to a basis of the column span of this matrix. The best choice for this set of matrices are the P dominant eigenmatrices of $\mathbf{C}^{(4)}_{\mathbf{Y}_w}$, which we will denote by $\mathbf{A}_p \in \mathbb{R}^{P \times P}, p = 1, \ldots, P$. So in order to estimate the mixing matrix, the P dominant eigenmatrices $\mathbf{A}_p$ of $\mathbf{C}^{(4)}_{\mathbf{Y}_w}$ have to be jointly (approximately) diagonalised. For the orthogonal simultaneous matrix diagonalisation, one can resort to the Jacobi technique developed in [38, 37].

Second Order Blind Identification (SOBI).

Another assumption that can be made to estimate $\mathbf{M}$ in (3.18) is that the sources are mutually uncorrelated but autocorrelated individually. In this case, an algorithm based on second order statistics can be used to estimate the underlying sources, called Second Order Blind Identification (SOBI) [17, 18]. For other methods exploiting temporal characteristics, see [88] and references therein.

Under our working hypothesis, not only the covariance matrix of the sources has to be diagonal, but also all the covariance matrices at different time lags τ. The mixing matrix after pre-whitening $\mathbf{M}_w$ is then the matrix that jointly diagonalizes all the covariance matrices:

$$\mathbf{C}_{\mathbf{Y}_w}(\tau) = \frac{1}{J}\mathbf{Y}_w(t)\mathbf{Y}_w(t+\tau)^T \tag{3.28}$$

$$= \mathbf{M}_w \cdot \mathbf{C}_{\mathbf{X}}(\tau) \cdot \mathbf{M}_w^T \qquad \forall\tau \tag{3.29}$$

After pre-whitening of the observation matrix, the problem reduces to an orthogonal approximate simultaneous diagonalisation of matrices $\mathbf{C}_{\mathbf{Y}_w}(\tau)$. These matrices can also be referred to as $\mathbf{A}_p, p = 1, \ldots, P$ (cf. JADE).

Some other algorithms.

Other algorithms that are popular in neuroscience are FastICA [163] and Infomax [16]. Both algorithms are pre-whitening based. FastICA estimates the sources by maximising the fourth-order marginal cumulants in an approximate way. In this way, the distributions of the sources will be as non-gaussian as possible. Infomax estimates the unmixing matrix with a neural network, minimizing the mutual information of the estimated sources. Although these algorithms are popular, from an theoretical point of view, they are not optimal, see *e.g.*, [363]. Furthermore, a recent article [63] illustrated that both algorithms exploit more the sparseness than the independence of the sources in the application of extracting brain sources from fMRI.

Jacobi angles for simultaneous diagonalisation

A Jacobi rotation matrix is equal to the identity matrix except for the entries at positions i and j. It is of the form

$$\mathbf{J}(i,j,c,s) = \begin{pmatrix} 1 & \dots & 0 & \dots & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & \dots & c & \dots & s & \dots & 0 \\ \vdots & & \vdots & \ddots & \vdots & & \vdots \\ 0 & \dots & -s & \dots & c & \dots & 0 \\ \vdots & & \vdots & & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \dots & 0 & \dots & 1 \end{pmatrix} \tag{3.30}$$

with $c = cos(\theta), s = sin(\theta)$. In [37], it is shown that the computation of the optimal Jacobi angles for approximate simultaneous diagonalisation reduces to an eigenvalue problem. We repeat that this technique can be used for the orthogonal diagonalisation of both cumulant slices and covariance matrices. We recall that we denote these matrices by $\mathbf{A}_p, p = 1, \dots, P$. The goal is to maximize the objective function

$$f = \sum_{p=1}^{P} ||diag(\mathbf{J}^T \mathbf{A}_p \mathbf{J})||^2 \tag{3.31}$$

where $diag(\cdot)$ is the diagonal of the matrix and $\mathbf{J}$ a Jacobi rotation. As can be seen from (3.30), $\mathbf{J}$ only affects the elements in rows and columns i and j. The

corresponding contrast function is given by

$$f = \sum_{p=1}^{P} \tilde{a}_{iip}^2 + \tilde{a}_{jjp}^2 \tag{3.32}$$

where
$$\tilde{a}_{iip} = \begin{pmatrix} c & -s \end{pmatrix} \begin{pmatrix} a_{iip} & a_{ijp} \\ a_{jip} & a_{jjp} \end{pmatrix} \begin{pmatrix} c \\ -s \end{pmatrix} \tag{3.33}$$

and
$$\tilde{a}_{jjp} = \begin{pmatrix} s & c \end{pmatrix} \begin{pmatrix} a_{iip} & a_{ijp} \\ a_{jip} & a_{jjp} \end{pmatrix} \begin{pmatrix} s \\ c \end{pmatrix}, \tag{3.34}$$

with $(A_p)_{ij} = a_{ijp}$ denoting the entry at row i and column j of $\mathbf{A}_p$, $1 \leqslant i < j \leqslant P$. It was shown that this contrast can also be written as:

$$f = \mathbf{w}^T \cdot \mathbf{G} \cdot \mathbf{w} \tag{3.35}$$

where (3.36)

$$\mathbf{w} = \begin{pmatrix} \cos(2\theta) & \sin(2\theta) \end{pmatrix}^T, \tag{3.37}$$

$$\mathbf{G} = \sum_p \mathbf{z}_p \mathbf{z}_p^T \tag{3.38}$$

and
$$\mathbf{z}_p = \frac{1}{2} \begin{pmatrix} a_{iip} - a_{jjp} \\ a_{ijp} + a_{jip} \end{pmatrix} \tag{3.39}$$

Maximisation of (3.35) is an eigenvalue problem. Namely, the rotation angle θ for the optimal Jacobi rotation follows from the dominant eigenvector $\mathbf{w}$ of $\mathbf{G}$. In a Jacobi iteration, one iterates over different choices of i and j [125].

Remark: This procedure can be generalised to the case of complex data. The overall approach remains the same, but the optimal Jacobi rotation follows from the dominant eigenvector of a real symmetric (3x3) matrix instead of a real symmetric (2x2) matrix [37].

3.5.3 Canonical Correlation Analysis (CCA)

As CCA was originally introduced as a method to compare two datasets, we review the method in its original setting. Afterwards, we discuss its use as a BSS method.

CCA to assess the similarity between two datasets

We describe the main idea of CCA [158]. For more details, we refer to [25, 120]. Ordinary correlation analysis quantifies the relation between (realizations of) two variables $\mathbf{a}(t)$ and $\mathbf{b}(t)$ by means of a correlation coefficient ρ:

$$\rho = \frac{Cov[\mathbf{a}, \mathbf{b}]}{\sqrt{V[\mathbf{a}]V[\mathbf{b}]}} \tag{3.40}$$

in which Cov and V indicate respectively the covariance and variance. CCA is a multivariate extension of ordinary correlation analysis. Consider matrices containing realisations of two multivariate zero-mean vectors: $\mathbf{A}=[\mathbf{a}_1(t), \ldots, \mathbf{a}_m(t)]^T$ and $\mathbf{B}=[\mathbf{b}_1(t), \ldots, \mathbf{b}_n(t)]^T$, $t = 1, \ldots, N$. Two new vectors, $\tilde{\mathbf{a}}$ and $\tilde{\mathbf{b}}$, are generated as a linear combination of the components in $\mathbf{A}$ and $\mathbf{B}$:

$$\begin{aligned} \tilde{\mathbf{a}} &= w_{\mathbf{a}_1}\mathbf{a}_1 + \ldots + w_{\mathbf{a}_m}\mathbf{a}_m = \mathbf{w}_\mathbf{a}^T\mathbf{A}, \\ \tilde{\mathbf{b}} &= w_{\mathbf{b}_1}\mathbf{b}_1 + \ldots + w_{\mathbf{b}_m}\mathbf{b}_m = \mathbf{w}_\mathbf{b}^T\mathbf{B}, \end{aligned} \tag{3.41}$$

CCA computes the coefficient vectors $\mathbf{w_a}$ and $\mathbf{w_b}$ that maximize the correlation between $\tilde{\mathbf{a}}$ and $\tilde{\mathbf{b}}$. The coefficients are called regression weights and $\tilde{\mathbf{a}}$ and $\tilde{\mathbf{b}}$ are called canonical variates. The resulting correlation coefficient is called the canonical correlation coefficient.
It can be shown that finding these regression weights corresponds to solving an eigenvalue problem.
By inserting (3.41) into the definition of the correlation coefficient (3.40), and assuming that $\mathbf{A}$ and $\mathbf{B}$ are zero-mean, we obtain:

$$\rho = \frac{\mathbf{w}_\mathbf{a}^T\mathbf{C_{ab}}\mathbf{w_b}}{\sqrt{(\mathbf{w}_\mathbf{a}^T\mathbf{C_{aa}}\mathbf{w_a})(\mathbf{w}_\mathbf{b}^T\mathbf{C_{bb}}\mathbf{w_b})}}, \tag{3.42}$$

with $\mathbf{C_{aa}}$ and $\mathbf{C_{bb}}$ the auto-covariance matrices of $\mathbf{A}$ and $\mathbf{B}$ respectively and $\mathbf{C_{ab}}$ the cross-covariance matrix of $\mathbf{A}$ and $\mathbf{B}$. ρ is a function of $\mathbf{w_a}$ and $\mathbf{w_b}$. In order to maximise the correlation coefficient, we impose the partial derivatives with respect to $\mathbf{w_a}$ and $\mathbf{w_b}$ to be zero. This results in the following set of equations:

$$\begin{aligned} \mathbf{C}_\mathbf{aa}^{-1}\mathbf{C_{ab}}\mathbf{C}_\mathbf{bb}^{-1}\mathbf{C_{ba}}\mathbf{w}_{\mathbf{a}_i} &= \rho^2\mathbf{w}_{\mathbf{a}_i}, \\ \mathbf{C}_\mathbf{bb}^{-1}\mathbf{C_{ba}}\mathbf{C}_\mathbf{aa}^{-1}\mathbf{C_{ab}}\mathbf{w}_{\mathbf{b}_i} &= \rho^2\mathbf{w}_{b_i}. \end{aligned} \tag{3.43}$$

These are eigenvalue decompositions. The matrices $\mathbf{C}_\mathbf{aa}^{-1}\mathbf{C_{ab}}\mathbf{C}_\mathbf{bb}^{-1}\mathbf{C_{ba}}$ and $\mathbf{C}_\mathbf{bb}^{-1}\mathbf{C_{ba}}\mathbf{C}_\mathbf{aa}^{-1}\mathbf{C_{ab}}$ have the same eigenvalues. The vectors $\mathbf{w_a}$ and $\mathbf{w_b}$ we are looking for, are the eigenvectors corresponding to the largest eigenvalue. This

eigenvalue is the square of the maximal correlation between the canonical variates.

Concerning implementation, a reliable algorithm is described in [125]. We will also discuss in the next chapter how non-negativity constraints can be incorporated.

CCA as a BSS technique

In chapter 4, we will show that CCA can also be used to decompose one dataset into contributing sources.

As in eq. (3.18), let us define $\mathbf{X}$ as the matrix of source signals and $\mathbf{Y}$ as the observed mixture. From $\mathbf{Y}$, we construct two new matrices $\mathbf{A}$ and $\mathbf{B}$, defined as $\mathbf{A} = \mathbf{Y}(:, 1 : T-1)$ and $\mathbf{B} = \mathbf{Y}(:, 2 : T)$. Note that we use matlab notation (:) to define the range of columns/rows involved. $\mathbf{A}$ is thus the observation matrix and $\mathbf{B}$ is the observation matrix shifted by one time sample. CCA will look for the linear combination of the mixed signals ($\mathbf{Y}(:, 1 : T-1)$) that correlates best with the linear combination of the mixed signals shifted by one sample ($\mathbf{Y}(:, 2 : T)$) in order to extract the source signals $\mathbf{X}$. Mathematically, CCA is computed by solving the matrix equations (). We focus now on how the source signals are extracted. Looking for a common basis between a dataset ($\mathbf{A}$) and its shifted variant ($\mathbf{B}$) means explaining variance common to both.

The first extracted source will be the linear combination of the mixed signals with maximal autocorrelation at time lag 1. This source explains most of the cross variance between $\mathbf{A}$ and $\mathbf{B}$. This first extracted source is called the canonical correlation component. The regression weights defining the unmixing, will be the first row of the demixing matrix $\mathbf{W}$, see eq. (3.41). The data is then projected on the orthogonal complement of the canonical correlation component (here we exploit the mutual orthogonality of the different sources), and a similar procedure is used to find the second source signal. The second source signal will be the linear combination with maximal autocorrelation, this time under the constraint that it is uncorrelated with the first component. The coefficients defining this unmixing will define the second row of $\mathbf{W}$. This procedure of finding autocorrelated source signals is repeated until the data are fully decomposed. In contrast to ICA, the source signals extracted by CCA have a fixed order (based on their autocorrelation).

BSS-CCA decomposes the observation matrix $\mathbf{Y}$ into sources:

$$\mathbf{W} \cdot \mathbf{Y} = \mathbf{X}. \tag{3.44}$$

These sources are sorted in order of decreasing autocorrelation (highly autocorrelated sources ranked first, weakly autocorrelated ones ranked last). So, also with CCA, a mixture can be decomposed into its contributing sources. Here the main

hypothesis is that the sources are mutually uncorrelated and have an individual different autocorrelation.

3.5.4 Other matrix decompositions

Many other matrix decomposition techniques can be computed, *e.g.*, factor models [185], generalised SVD [253, 5] non-negative matrix factorisation [220, 47], sparse coding [109], CUR decomposition [235], ... These techniques can in certain applications be used to extract sources of relevance. Non-negative constraints are relevant and have already been applied in neuroscience [47]. We refer to the literature for more details.

3.5.5 A remark on the evaluation of decompositions

There are many fields where matrix decomposition techniques are used: speech, telecommunication, biomedical applications, ... Throughout this PhD, we will focus on decompositions of brain recordings.
In all these applications, we will be confronted with the question: "How well is our measurement decomposed?" or "How well are the sources estimated?". It has already been mentioned that the relevance of the different assumptions made about the sources will be reflected in the quality of the decomposition.
Furthermore, although there is an intuitive link between the estimation of the mixing matrix (separation index) and the estimation of the sources (identification index), it is shown in [213] that a small difference in the separation index may correspond to a large variation in the identification index.

3.6 Higher-order tensor decomposition techniques

3.6.1 Canonical/Parallel factor Analysis (CPA)

Model.

The trilinear Canonical Decomposition (Candecomp) [39], also known as Parallel Factor Analysis (Parafac) [141, 198, 320], or referred to as CPA (**C**andecomp/**P**arafac Analysis) is a generalisation of the SVD to higher order tensors [154, 153, 44].
It is defined for a three-way array $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$ as:

$$x_{ijk} = \sum_{r=1}^{R} a_{ir} b_{jr} c_{kr} + e_{ijk} \tag{3.45}$$

where R is the number of components used in CPA and e_{ijk} are the residuals containing the unexplained variation. A pictorial representation of CPA is given in Fig. 3.5. CPA is a trilinear model: fixing the parameters in two modes, x_{ijk} is expressed as a linear function of the remaining parameters. Although the definition can be seen as a generalisation of the SVD, its properties cannot. The vectors in the different modes are not necessarily orthogonal, the different components are not ordered and computing CPA with different number of rank-one terms will in general give different sets of rank-one terms.

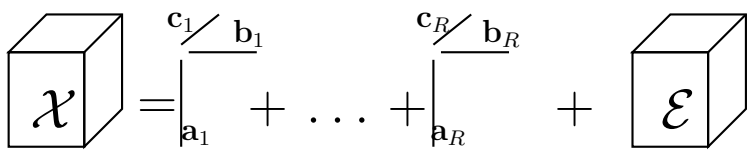


Figure 3.5: *CPA with R components.*

CPA can be expressed using the Khatri-Rao product [320].
Stack the elements of $\mathcal{X}$ in a matrix $\mathbf{X} \in \mathbb{R}^{IJ \times K}$ as

$$\mathbf{X}_{(i-1)J+j,k} = x_{ijk} \tag{3.46}$$

Collect the elements a_{ir} in $\mathbf{A} \in \mathbb{R}^{I \times R}$; b_{jr} in $\mathbf{B} \in \mathbb{R}^{J \times R}$ and c_{kr} in $\mathbf{C} \in \mathbb{R}^{K \times R}$. Then

$$\mathbf{X} = (\mathbf{A} \odot \mathbf{B}) \cdot \mathbf{C}^T. \tag{3.47}$$

Uniqueness.

CPA has the advantage that it is unique under mild conditions [199, 318, 329]. When

$$\text{rank}_k(\mathbf{A}) + \text{rank}_k(\mathbf{B}) + \text{rank}_k(\mathbf{C}) \geqslant 2R + 2 \tag{3.48}$$

is satisfied, CPA is unique up to permutation and scaling. If one of the dimensions is larger than R, a less restrictive condition was derived in [202]. Although this condition guarantees uniquess, it can in real-world applications still be useful to impose extra constraints to increase the robustness of the estimation.

Computation of CPA.

Originally, an Alternating Least-Squares (ALS) algorithm was proposed in order to minimise the least squares cost function for the computation of CPA [320]:

$$\min_{\mathbf{A},\mathbf{B},\mathbf{C}} ||\mathbf{X}^T - \mathbf{C}(\mathbf{A} \odot \mathbf{B})^T||^2. \tag{3.49}$$

In order to fit the model using alternating least squares, one should compute the update in one mode, keeping the two other modes fixed. Due to the symmetry of the model in the different modes, the updates for all modes are essentially identical with the role of the different modes shifted. Given $\mathbf{A}$ and $\mathbf{B}$, $\mathbf{C}$ can be estimated from the classical linear least squares procedure:

$$\min_{\mathbf{C}} ||\mathbf{X}^T - \mathbf{C}\mathbf{Z}^T||^2 \tag{3.50}$$

where $\mathbf{Z}$ equals $\mathbf{A} \odot \mathbf{B}$. The updates of the two other modes can be done in very similar ways. The ALS algorithm computes CPA by updating in an alternating way $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$, while keeping the matrices in the two other modes fixed.
Afterwards, it was also shown that CPA can in theory be computed from an eigenvalue decomposition (EVD) [222, 307] under certain assumptions among which the most restrictive is that $R \leqslant \min\{I, J\}$. This results in a faster computation. However, when the model is only approximately valid, this will only form the initialisation of the ALS-procedure.
In [212], it is shown that the computation of CPA, based on a simultaneous EVD is actually more robust than a single EVD. This again implied the rank condition $R \leqslant \min\{I, J\}$. We review the computational scheme here. First, $\mathbf{A}$ and $\mathbf{B}$ are estimated and $\mathbf{C}$ is then computed from equation (3.50). Any vector in the range of matrix $\mathbf{X}$, can be represented by an $I \times J$ matrix that can be decomposed as:

$$\mathbf{V} = \mathbf{A} \cdot \mathbf{D} \cdot \mathbf{B}^T \tag{3.51}$$

with $\mathbf{D}$ diagonal. If the range is spanned by R matrices $\mathbf{V}_1, \mathbf{V}_2, \ldots, \mathbf{V}_R$, the computation of the canonical decomposition can be obtained by the simultaneous decomposition of the set $\{\mathbf{V}_r\}_{(1 \leqslant r \leqslant R)}$.

$$\mathbf{V}_1 = \mathbf{A} \cdot \mathbf{D}_1 \cdot \mathbf{B}^T \tag{3.52}$$

$$\mathbf{V}_2 = \mathbf{A} \cdot \mathbf{D}_2 \cdot \mathbf{B}^T \tag{3.53}$$

$$\vdots$$

$$\mathbf{V}_R = \mathbf{A} \cdot \mathbf{D}_R \cdot \mathbf{B}^T \tag{3.54}$$

The best choice for these matrices in order to span the full range of this mapping consists of the R dominant left singular matrices of the matrix in eq.(3.46) [211]. These $\{\mathbf{V}_r\}$ can thus be obtained from the singular value decomposition of $\mathbf{X}$. In order to deal with these equations in a numerically proper way, the problem can be formulated in terms of orthogonal unknowns [349, 212]. Introducing a QR-factorisation $\mathbf{A} = \mathbf{Q}^T\mathbf{R}$ and an RQ-factorisation $\mathbf{B}^T = \tilde{\mathbf{R}}\mathbf{Z}^T$, leads to a simultaneous generalised Schur decomposition:

$$\mathbf{Q}\mathbf{V}_1\mathbf{Z} = \mathbf{R} \cdot \mathbf{D}_1 \cdot \tilde{\mathbf{R}} \tag{3.55}$$

$$\mathbf{Q}\mathbf{V}_2\mathbf{Z} = \mathbf{R} \cdot \mathbf{D}_2 \cdot \tilde{\mathbf{R}} \tag{3.56}$$

$$\vdots$$

$$\mathbf{Q}\mathbf{V}_K\mathbf{Z} = \mathbf{R} \cdot \mathbf{D}_K \cdot \tilde{\mathbf{R}}. \tag{3.57}$$

This simultaneous generalised Schur decomposition can be computed by an extended QZ-iteration [349] or by a Jacobi-iteration [212]. From this decomposition $\mathbf{A}$ and $\mathbf{B}$ can be derived.

Recently, in [202] it has been shown that the canonical components can also be obtained from a simultaneous matrix diagonalisation with a much less severe restriction on R.

Consider again the equations (3.46) and (3.47). $\mathbf{X}$ can also be factorised in the form:

$$\mathbf{X} = \mathbf{E} \cdot \mathbf{F}^T, \tag{3.58}$$

with $\mathbf{E} \in \mathbb{R}^{IJ \times R}$ and $\mathbf{F} \in \mathbb{R}^{K \times R}$ full column rank. This implies the following equality:

$$(\mathbf{A} \odot \mathbf{B}) = \mathbf{E} \cdot \mathbf{W} \tag{3.59}$$

for some nonsingular $\mathbf{W} \in \mathbb{R}^{R \times R}$. $\mathbf{A}$ and $\mathbf{B}$ can be found if only we know $\mathbf{W}$. In order to find $\mathbf{W}$, each column of $\mathbf{E}$ has to be stacked in a tensor $\mathcal{E} \in \mathbb{R}^{I \times J \times R}$ as follows:

$$(\mathcal{E})_{ijr} = (\mathbf{E})_{(i-1)J+j,r} \qquad \forall i, j, r \tag{3.60}$$

In [202], it is proven that $\mathbf{W}$ can then be computed as the matrix that simultaneously diagonalizes a set of symmetric matrices $\{\mathbf{M}_r\}_{(1 \leqslant r \leqslant R)}$. These matrices $\{\mathbf{M}_r\}$ are obtained as the R linearly independent solutions of the kernel

$$\sum_{s,t=1}^{R} m_{st} \mathcal{P}_{st} = \mathcal{O} \qquad \forall s, t \tag{3.61}$$

under the constraint $m_{st} = m_{ts}$ and with $\mathcal{P}_{st} \in \mathbb{R}^{I \times I \times J \times J}, 1 \leqslant r, s \leqslant R$ defined by:

$$(\mathcal{P}_{st})_{ijkl} = e_{ikr}e_{jls} + e_{iks}e_{jlr} - e_{ilr}e_{jks} - e_{ils}e_{jkr} \; \forall i, j, k, l \tag{3.62}$$

Once $\mathbf{W}$ is known, $\mathbf{A}$ and $\mathbf{B}$ can be computed from (3.59) and $\mathbf{C} = \mathbf{F} \cdot \mathbf{W}^{-T}$.

Line search was proposed to speed up the convergence of the original ALS algorithm. In [274, 295] it is shown how the optimal step size can be computed in the real and complex case respectively. This method is called ALS with enhanced line search (ELS).
Other computational schemes have also been proposed, see *e.g.*, [283, 335, 354, 263].

Degeneracy

Sometimes, the CPA algorithm converges slower and slower and the CPA solution displays the following strange pattern:

- $\mathbf{a}_i \approx \pm\mathbf{a}_j, \mathbf{b}_i \approx \pm\mathbf{b}_j$ and $\mathbf{c}_i \approx \pm\mathbf{c}_j$
- $\cos(\mathbf{a}_i\mathbf{a}_j) \cdot \cos(\mathbf{b}_i\mathbf{b}_j) \cdot \cos(\mathbf{c}_i\mathbf{c}_j)$ tends towards -1
- elements of $\mathbf{c}_i$ and $\mathbf{c}_j$ become arbitrarily large (if the first two modes are normalised)

This is called two-factor degeneracy. Degenerate solutions occur when the best rank-R approximation of $\mathcal{X}$ does not exist. In these cases, the CPA objective function has no minimum, only an infimum. For more information, we refer to [146, 200, 328].

Applications of CPA.

Since 1970, CPA has slowly found its way in various disciplines such as Chemometrics and food technology [320], exploratory data analysis [198], wireless communications and array processing [313, 312], blind source separation [52, 205, 107], web link analysis [10]. Also neuroscience provided multi-way datasets and neuroscientists quickly discovered CPA for analysing them [252, 108, 104].

3.6.2 The Tucker3 model

Focusing on different properties of the SVD, another higher-order generalisation can be defined. In [210] the following tensor decomposition, named the Higher-Order Singular Value Decomposition (HOSVD) was discussed. It is a normalised version of the Tucker3 model [337, 338], with all-orthogonality and ordering constraints.

Theorem 3.3 (Nth-Order Singular Value Decomposition) *Every $(I_1 \times I_2 \times \ldots \times I_N)$-tensor $\mathcal{X}$ can be written as the product*

$$\mathcal{X} = \mathcal{S} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \ldots \times_N \mathbf{U}^{(N)}, \tag{3.63}$$

in which:

- $\mathbf{U}^{(n)} = \left[U_1^{(n)} U_2^{(n)} \ldots U_{I_n}^{(n)}\right]$ *is a unitary $(I_n \times I_n)$-matrix,*
- $\mathcal{S}$ *is a $(I_1 \times I_2 \times \ldots \times I_N)$-tensor of which the subtensors $\mathcal{S}_{i_n=\alpha}$, obtained by fixing the nth index to α, have the properties of:*
  - *all-orthogonality: two subtensors $\mathcal{S}_{i_n=\alpha}$ and $\mathcal{S}_{i_n=\beta}$ are orthogonal for all possible values of n, α and β subject to $\alpha \neq \beta$:*

$$\langle \mathcal{S}_{i_n=\alpha}, \mathcal{S}_{i_n=\beta} \rangle = 0 \quad \text{when} \quad \alpha \neq \beta, \tag{3.64}$$

  - *ordering:*

$$\|\mathcal{S}_{i_n=1}\| \geqslant \|\mathcal{S}_{i_n=2}\| \geqslant \ldots \geqslant \|\mathcal{S}_{i_n=I_n}\| \geqslant 0 \tag{3.65}$$

  for all possible values of n.

The Frobenius-norms $\|\mathcal{S}_{i_n=i}\|$, symbolised by $\sigma_i^{(n)}$, are n-mode singular values of $\mathcal{X}$ and the vector $U_i^{(n)}$ is an ith n-mode singular vector. The decomposition is visualised for third-order tensors in Fig. 3.6.

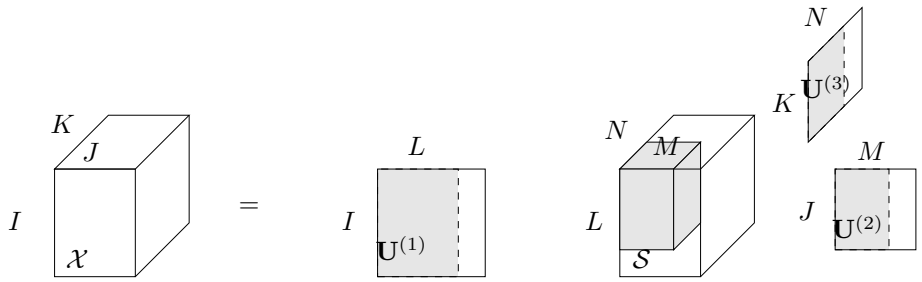


Figure 3.6: *Visualisation of the HOSVD / Tucker decomposition. From [210].*

For a tensor $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$, Theorem 3.3 says that it is always possible to find orthogonal transformations, defined by $\mathbf{U}^{(1)}$,$\mathbf{U}^{(2)}$ and $\mathbf{U}^{(3)}$, such that $\mathcal{S} = \mathcal{X} \times_1 \mathbf{U}^{(1)^T} \times_2 \mathbf{U}^{(2)^T} \times_3 \mathbf{U}^{(3)^T}$ is all-orthogonal and ordered. With a matrix being orthogonal with respect to another matrix when the sum of the products of the corresponding entries equals zero, all-orthogonality implies that the different horizontal, vertical and frontal slices of $\mathcal{S}$ are mutually orthogonal. The ordering constraint imposes that the Frobenius-norm of these slices does not increase when the index in that mode is increased.

Application of Tucker

Applications of Tucker can be found in web-mining [331], computer vision [355], psychometrics [183], ... There are few applications of the Tucker model in signal separation, because the components are not that straightforward to interpret but perhaps the most important application is the compression of the data before computing CPA [182, 167, 168].

3.6.3 The Block Component Model (BCM)

Recently, the Block Component Model has been introduced [203, 204, 214, 207, 273]. It consists of different families of decompositions. We only define two classes here because they are particularly interesting for potential applications. We also visualize them.

Definition 3.14 *A* decomposition of a tensor $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$ in a sum of rank-$(L_r, L_r, 1)$ terms, $1 \leqslant r \leqslant R$, *is a decomposition of* $\mathcal{X}$ *of the form*

$$\mathcal{X} = \sum_{r=1}^{R} \mathbf{E}_r \circ \mathbf{c}_r, \tag{3.66}$$

in which the $(I \times J)$ *matrix* $\mathbf{E}_r$ *is rank-*L_r*,* $1 \leqslant r \leqslant R$*.*

If we factorize $\mathbf{E}_r$ as $\mathbf{A}_r \cdot \mathbf{D}_r \cdot \mathbf{B}_r^T$, in which the $(I \times L_r)$ matrix $\mathbf{A}_r$ and the $(J \times L_r)$ matrix $\mathbf{B}_r$ are rank-L_r, $r = 1, \ldots, R$, then we can write (3.66) as

$$\mathcal{X} = \sum_{r=1}^{R} (\mathbf{A}_r \cdot \mathbf{D}_r \cdot \mathbf{B}_r^T) \circ \mathbf{c}_r. \tag{3.67}$$

This is visualised in Fig. 3.7. One can see that this decomposition does not impose a strong rank-1 structure as in CP, but that it allows to model the tensor as a sum of outer products of a low-rank matrix and a vector.

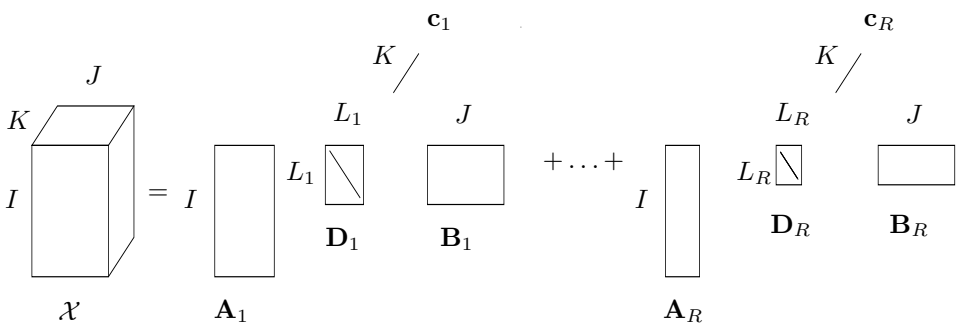


Figure 3.7: *Visualisation of the decomposition of a tensor in a sum of rank-*$(L_r, L_r, 1)$ *terms,* $1 \leqslant r \leqslant R$*. From [204].*

Definition 3.15 *A* decomposition of a tensor $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$ in a sum of rank-(L, M, N) terms *is a decomposition of* $\mathcal{X}$ *of the form*

$$\mathcal{X} = \sum_{r=1}^{R} \mathcal{D}_r \bullet_1 \mathbf{A}_r \bullet_2 \mathbf{B}_r \bullet_3 \mathbf{C}_r, \tag{3.68}$$

in which $\mathcal{D}_r \in \mathbb{R}^{L \times M \times N}$ *are full rank-*(L, M, N)*, and in which* $\mathbf{A}_r \in \mathbb{R}^{I \times L}$*,* $\mathbf{B}_r \in \mathbb{R}^{J \times M}$ *and* $\mathbf{C}_r \in \mathbb{R}^{K \times N}$ *are nonsingular,* $1 \leqslant r \leqslant R$*.*

This is visualised in Fig. 3.8. It can be seen that CPA is a specific instance of this decomposition with $L = M = N = 1$, and Tucker the other extreme with $R = 1$.

3.6.4 Other tensor decompositions

Similar to the matrix case, one can impose non-negativity constraints [259, 47], ... Other decompositions include Parafac2 [142], Indscal [39], S-Parafac [145], CANDELINC [172], DEDICOM [143].

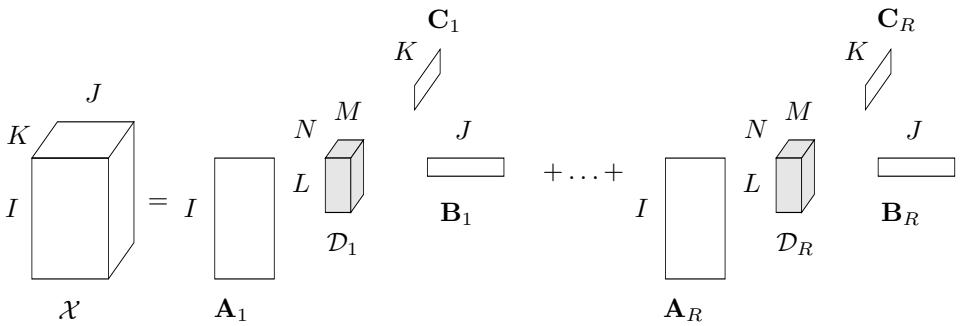


Figure 3.8: *Visualisation of the decomposition of a tensor in a sum of rank-(L, M, N) terms. From [204].*

3.7 Conclusion

In this chapter, we introduced the basic mathematical concepts that will be used in this thesis. Furthermore, we defined the decomposition methods (ICA, CCA, CPA) that will be further explored and applied in this thesis.

Chapter 4

Canonical Correlation Analysis in nosologic brain imaging and muscle artifact removal

We introduced the mathematics of CCA in section 3.5.3 and explained that it is originally a technique to compare datasets. We also mentioned that CCA can be used as a BSS technique for decomposing a dataset into source signals. In this chapter, we illustrate first the use of CCA for comparing different datasets with the application of brain tumor segmentation and classification. As the dataset consists of Magnetic Resonance Spectroscopy (MRS) data, we also provide some background on this topic. Second, we illustrate the use of CCA as a BSS method that can be used for muscle artifact removal from continuous epilepsy recordings. Third, we explore its use as a technique for removing short bursts of muscle activity superposed on speech production recordings.

4.1 CCA for tumor segmentation

4.1.1 Introduction

MRS is increasingly being used in the diagnosis of brain tumors, to choose the most appropriate therapy and to perform therapy follow-up. With single voxel proton MRS, tissue concentrations of metabolites can be determined in a well-defined voxel in a non-invasive way. This can reduce the need for a biopsy, which is not without risks for the patient [20]. However, analysis of MR spectra requires a lot

of expertise from the radiologist. To reduce the need of human interaction, several classifiers have been proposed that try to distinguish tumors in an automatic way, *e.g.*, [332, 293, 226, 96, 91, 360]. Although single voxel proton MRS is a relatively fast method to characterise the metabolic signature of the tumor, radiologists are also interested in its shape, size and (metabolic) heterogeneity. This information can be obtained with multivoxel proton MRS Imaging (MRSI), in combination with high-quality anatomical MR images [265]. During an MRSI examination, metabolic information is measured in a grid of voxels rather than in one voxel. This allows radiologists to extract histopathological information from each voxel as well as information about the heterogeneity of the tumor. However, due to the large amount of data that is acquired and needs to be processed during the investigation of a patient, there is an even stronger need for automatic 'classification' strategies than for single voxel MRS. More correctly, one should talk about 'segmentation' strategies as the goal is to segment an MRS image into regions of different tissues. We will explain how the proposed segmentation method is also able to label the detected segments simultaneously. A variety of methods is available in the literature, and in particular pattern recognition methods are commonly used to segment MRSI data [293, 317]. The result of a segmentation is a nosologic image [100]. Such an image summarizes the available information by coloring each pixel according to the determined histopathological class. In this way, nosologic images can be routinely and easily used by radiologists for tumor diagnosis.

In this study, we investigate the accuracy and robustness of a Canonical Correlation Analysis (CCA) based method to segment short echo-time brain MRSI data. The main advantage of CCA is the possibility to exploit both spatial and spectral information during the analysis of the spectrum in a voxel. Since it can be expected that the chance to have a normal voxel is much higher when the surrounding voxels are normal as well, including spatial information can improve robustness. As much spectral information as possible is also incorporated by generating a representative model spectrum for each tissue class that allows for realistic variations. CCA has been shown to be a fast and robust method for automated segmentation of long echo-time prostate MRSI data [219]. In [217], the same CCA-based method is used to segment long echo-time brain MRSI data of patients with one known tumor type, glioblastoma.

We extend this study in various ways.
First, we focus on short echo-time spectra of the brain. We use short echo-time spectra, because they are more relevant than long echo-time spectra for tumor classification and grading due to the presence of signals of more metabolites, such as myo-inositol, glutamate and glutamine [43, 13, 105, 236, 360]. However, interpretation of such spectra is complicated by the presence of a broad baseline, originating from macromolecules and lipids [311, 156, 15]. The lipid signals may be relevant as indicator of tumor grade [8, 134], but the shape of the macromolecular

baseline is in general unpredictable, especially in pathological cases.
Second, for brain tumor diagnosis, it is not sufficient to segment an MRSI image in normal and tumor tissue. Several types and grades of brain tumors may exist in different patients. For choosing the most appropriate therapy, it is very important to distinguish between different grades of tumor. In order to deal with different tumor types, we introduce a two-step CCA approach. We will explain how the first step is a tumor typing step, and how in the second step a segmentation is obtained.
Third, it is shown how to combine multimodal information and the influence of adding MRI information to the MRSI data is investigated. Four MR images with different contrasts (see section 4.1.5) were acquired for every patient. This extra information has been shown to improve the classification performance in other pattern recognition studies [100, 317]. To integrate MRSI and MRI information in a CCA-based method, a dimension reduction has to be performed on the MR spectra. This can be done by the extraction of characteristic features [316]. Most other classifiers or pattern recognition methods also use features as input variables, because they cannot deal with high-dimensional input spaces. As the CCA-method can be applied either to a MR spectrum or to a set of features, it is very interesting to compare the results obtained with different CCA-approaches.

CCA was introduced in chapter 3. We explain how CCA is applied for nosologic imaging of the brain (section 4.1.2). In section 4.1.6, we describe how we constructed simulations. Further, the accuracy of the method is investigated on baseline-corrected MR spectra. We also compare these results with the results of the CCA-method on features, without and with MRI information. We show results on *in vivo* MRSI in section 4.1.7.

4.1.2 Applying CCA to spectra

During an MRSI acquisition, MR spectra are measured in a grid of voxels. Based on the information contained in these MR spectra, the goal is to segment the image in different regions. All the voxels in one region belong to the same tissue type. Each voxel can be assigned individually to a certain tissue type by using a sophisticated single voxel classifier like LS-SVM [91] or simply by computing the correlation or distance between the spectrum in the considered voxel and a known tissue type model [100]. However, with this approach, the spatial information provided by multivoxel MRS is not used. When all the surrounding voxels contain healthy tissue, the considered voxel also has a high probability to contain healthy tissue. CCA exploits this spatial information by comparing the MR spectrum in the voxel under investigation and in some surrounding voxels with a model spectrum of a certain tissue. The matrices $\mathbf{A}$ and $\mathbf{B}$ in section 3.5.3 are defined as follows. $\mathbf{A}$ contains the spectrum of the voxel under investigation and some

spectra of the surrounding voxels. Several spatial models can be used. In this study, we used the spatial 'star' model, defined as:

$$\mathbf{A} = \begin{pmatrix} \mathbf{a}_5 \\ \frac{\mathbf{a}_5+\mathbf{a}_2}{2} \\ \frac{\mathbf{a}_5+\mathbf{a}_4}{2} \\ \frac{\mathbf{a}_5+\mathbf{a}_6}{2} \\ \frac{\mathbf{a}_5+\mathbf{a}_8}{2} \end{pmatrix}$$

A schematic representation of the different voxels is given in Fig. 4.1.

B contains the model spectrum and realistic variation between different spectra. In this way, CCA also integrates some spectral information in the spectral model. The first component $\mathbf{b}_1$ of this spectral model for each tissue type is constructed as the mean of a database containing *in vivo* spectra, characteristic of that tissue type. To model natural inter-spectral and inter-patient variability, a variational component $\mathbf{b}_2$ is computed as follows. First the computed mean is subtracted from all *in vivo* spectra in the database. Second, PCA is performed on this mean-subtracted database. The first principal component contains most of the variations, and is used as $\mathbf{b}_2$.

To obtain a segmentation with CCA, canonical correlation coefficients are computed with the chosen spatial model and the spectral models of different tissue types. A visual representation of the analysis of one voxel (a_5) with a certain tissue type is given in Fig. 4.1.
The computed canonical correlation coefficients indicate how likely it is that the considered voxel belongs to the different tissue types. In the case of three tissue types, these three canonical correlation coefficients can be displayed in the three-dimensional coordinate system (first canonical correlation coefficient, second canonical correlation coefficient and third canonical correlation coefficient). The distribution of the voxels in this coordinate system is displayed for a simulated MRS image with normal, mixed and tumor tissue in Fig. 4.2. Three clouds of points can be distinguished, corresponding to the three present tissue types. The voxels indicated by circles, have a high correlation with normal tissue; the voxels indicated with stars, have a high correlation with tumor tissue and the voxels indicated with diamonds lie between both clouds. A clustering method [250] can be applied to these canonical correlation coefficients. However, in order to avoid well-known problems of clustering algorithms (*e.g.*, initialisation, number of clusters, ...), the highest canonical correlation coefficient is used to assign a voxel to a certain tissue type.

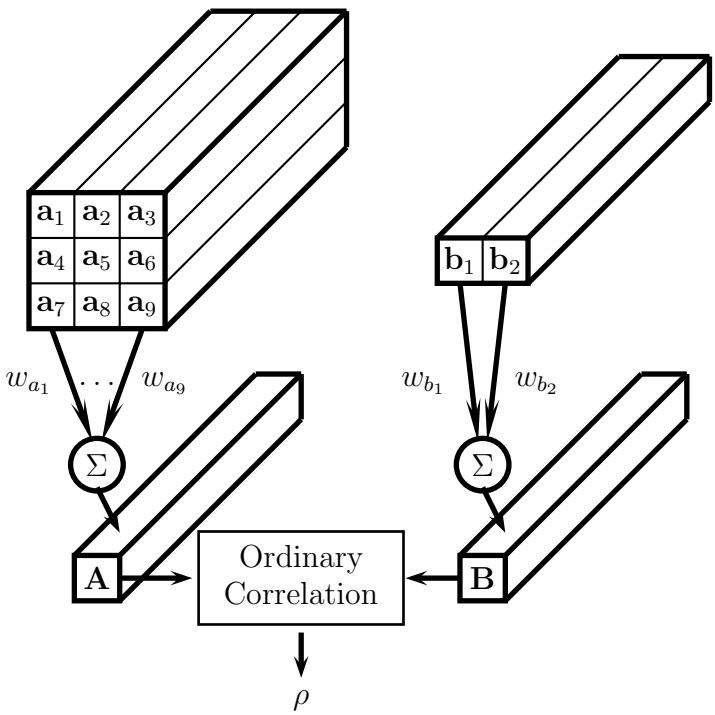


Figure 4.1: *Schematical representation of CCA for the analysis of MRSI. We want to assign voxel $\boldsymbol{a}_5$ to a tissue type. The other a_i are surrounding voxels that will also be used in the analysis of voxel $\boldsymbol{a}_5$. $\boldsymbol{b}_1$ is the model spectrum of a certain tissue, and $\boldsymbol{b}_2$ takes into account realistic variations. CCA computes the regression weights $\boldsymbol{w}_{a_i}$ and w_{b_i} in order to maximize the correlation between $\boldsymbol{A}$ and $\boldsymbol{B}$*

4.1.3 A two-step CCA-based method

As there exist several grades of brain tumor malignancy, a model spectrum has to be constructed for each tumor type and for each corresponding mixed tissue since it can be expected that only one tumor type is present in a patient. This mixed tissue is composed of fractions of the model spectrum for healthy and tumor tissue.
CCA is able to discriminate between different tumor types, but we noticed that voxels with mixed tissue are not always correctly detected. To avoid this problem a two-step method is proposed. The first step is a tumor typing step and the

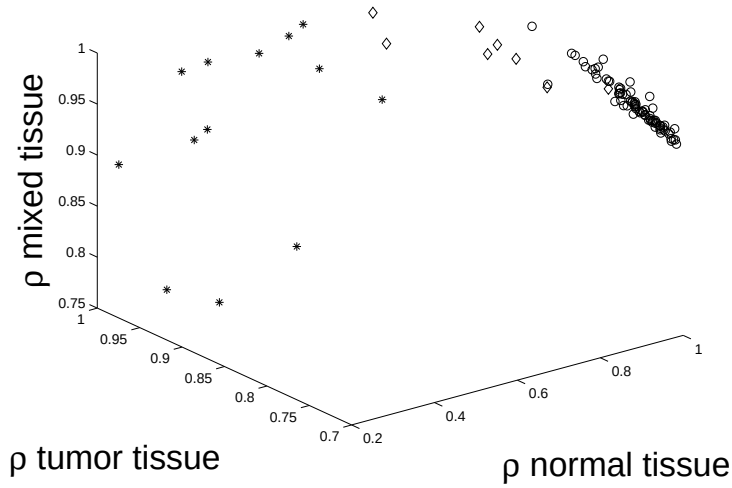


Figure 4.2: *All voxels in an MRSI grid are displayed in the coordinate system of the computed canonical correlation coefficients for different tissue types. Three clouds (circles, diamonds and stars) corresponding to the three present tissue types are visible.*

second the segmentation step. In the first step only the model spectra of normal tissue, cerebro-spinal fluid (CSF) and the different tumor spectra are used. As pathologists type the tumor with respect to the most malignant tissue, this step is performed to detect the most malignant tissue present. In the second step, CCA is applied with the model spectrum from CSF, normal tissue, the detected tumor tissue and corresponding mixed tissue. In this step, the heterogeneity of the tumor is shown.

4.1.4 Applying CCA to features

When we want to integrate the available MRSI and MRI information in the CCA-based method, harmonisation of both input spaces is necessary. This multimodal integration requires two steps, which are described.
At most four images with different contrasts are acquired during a study. These four extra variables per voxel will not influence the result, when they are added to the full spectra of 1024 variables. The spectra have to be summarized into representative features first. This dimension reduction can be done by applying PCA or quantification of the spectra. The latter approach is most meaningful, since the quantified amplitudes are a direct estimate of the concentrations of

the most important metabolites. Several quantification methods for short echo time spectra can be used. Amares is conceived for long echo-time MRS data quantification [348] and requires too much user interaction in the quantification of short echo-time MRS data [248]. At this moment there exist three easy-to-use quantitation tools for short echo-time MRS data: the commercially available LCModel [294], the freely available QUEST [297] and the open-source package AQSES [292]. For the present study AQSES was used, which fits an *in vivo* spectrum in the complex time domain as a linear combination of *in vitro* metabolite spectra. To avoid overfitting, only the 8 most important metabolites (Myo-inositol, Choline, Creatine, Glutamine, Glutamate, NAA, Alanine and Lactate) were quantified, together with the lipid peaks at 0.9 and 1.3 ppm. These metabolite levels formed the first 10 values in the set of features.
In addition to feature reduction, also the spatial resolution of both imaging modalities has to be matched. This can be done in several ways: lowering the resolution of the MR image to MRSI voxel size, keeping the MRI resolution and interpolating between MRSI voxels, or a combination of both. A detailed comparison of the best way to integrate both modalities is beyond the scope of this study. We decided to lower the resolution of the MR image to MRSI voxel size.
The images with 4 contrasts were aligned with respect to the spectroscopic image and the MR pixel intensities were averaged out over a spectroscopic voxel. These MRI variables were scaled to be in the same range as the quantified metabolite concentrations. These 4 image variables formed the last 4 values in the set of features.

Although there is no conceptual difference between the application of the CCA method to spectra or to features, there is a difference in implementation. The aim of the spatial model of CCA is to reduce the noise by 'averaging out' some voxels. In order to obtain a realistic solution, all the regression weights of $\mathbf{A}$ should have the same sign. When the CCA-method is applied to spectra, the mathematically optimal solution satisfies most of the time this constraint. However, in the feature case, this constraint is violated with the implementation given in [219], and has to be imposed explicitly. In [61, 280, 120], it is proven that an elegant constrained CCA solution exists in the special case where the regression weights are imposed to be non-negative. Such a constrained CCA solution equals an *unconstrained* CCA solution when one or several variables have been excluded. Formally, consider the sequence $\mathbb{N}_k = \{1, 2, \ldots, k\}$ where k is a positive integer and all its possible subsequences $\mathbb{W}_k = \{\mathbf{c} | \mathbf{c} \subseteq \mathbb{N}_k\}$. When $\mathbf{a}$ is a m-dimensional vector and $\mathbf{c} \in \mathbb{W}_m$, ${}_{\mathbf{c}}\mathbf{a}$ denotes the vector obtained when only the components listed in $\mathbf{c}$ are kept. Denoting the canonical correlation coefficient between $\mathbf{a}$ and $\mathbf{b}$ by $\rho(\mathbf{a}, \mathbf{b})$, the solution ρ of the constrained CCA problem is found in the following set:

$$\bigcup_{\mathbf{c} \in \mathbb{W}_m} \{\rho({}_{\mathbf{c}}\mathbf{a}, \mathbf{b})\}. \tag{4.1}$$

In order to find the constrained canonical correlation coefficient, one must solve the ordinary CCA problem for all possible subvectors ${}_{\mathbf{a}}\mathbf{x}$ with $\mathbf{y}$, and pick the largest correlation for which the regression weights fulfill the non-negativity constraint.

We summarise the feature-based method in Algorithm 1.

Algorithm 1 *Two-step CCA-based segmentation*

1: Input: grid of spectra (MRSI) and set of model spectra.
2: Tumor typing step, for every voxel:

1. Compute constrained CCA with spectral model of normal tissue
2. Compute constrained CCA with spectral model of Cerebrospinal Fluid (CSF)
3. Compute constrained CCA with spectral model of first tumor type
4. ...
5. Compute constrained CCA with spectral model of last tumor type

and type the voxel according to the highest canonical correlation coefficient. The highest tumor grade of all voxels is detected and used in the segmentation step.
3: Segmentation step, for every voxel:

1. Compute constrained CCA with spectral model of corresponding mixed tissue
2. Compare this canonical coefficient with the coefficients, computed in the tumor typing step of normal tissue, CSF and detected tumor tissue

and type the voxel according to the highest canonical correlation coefficient.
4: Output: MRS image in which all the voxels are assigned to normal tissue, CSF, tumor tissue or mixed tissue.

Notice that with this algorithm, we exclude the possibility of segmenting heterogeneous tumors. When one is interested in the heterogeneity of the tumor, one should not perform the segmentation step.

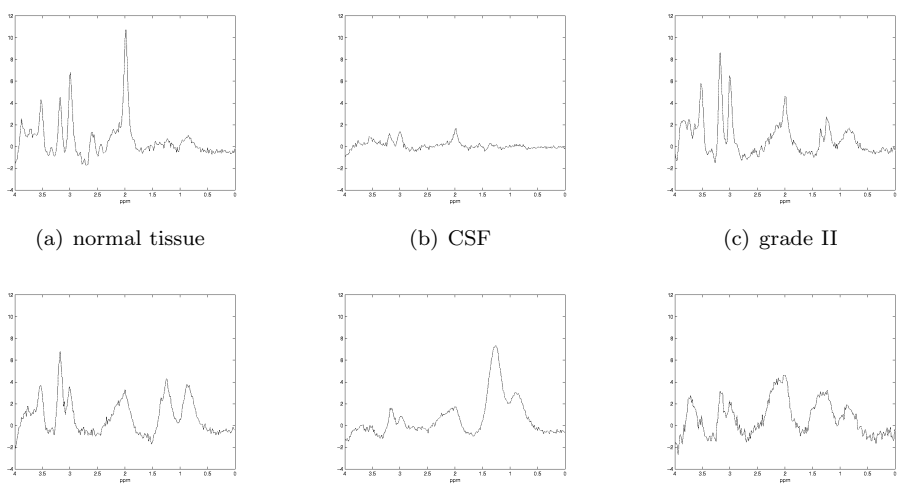
4.1.5 Material

In this study, data from 25 patients with a brain tumor and from 4 volunteers are used. These data were obtained within the EU project INTERPRET [159], and acquisition has been approved by the ethical committee of the University Medical Center Nijmegen, and from all patients written informed consent was obtained. Each patient's tumor was diagnosed by consensus on a histopathological study. The diagnosis was done according to the rules of the World Health Organization (WHO). 9 patients had a glial brain tumor of grade II (consisting of astrocytic, oligoastrocytic and oligodendroglial tumors), 5 of grade III (consisting of astrocytic, oligoastrocytic and oligodendroglial tumors) and 7 of grade IV (glioblastoma) and 3 patients were diagnosed with a meningioma tumor. The MR data were acquired with a 1.5 T Siemens Vision whole body scanner, using

a CP-head coil. For every patient first 4 MR images were acquired: T_1 weighted (TE/TR=15/644ms), T_2 weighted (TE/TR=16/3100ms), proton density weighted (TE/TR=98/3100ms) and a Gadolinium enhanced T_1 image (15 ml 0.5 M Gd-DTPA). Also both water suppressed and unsuppressed proton MR Spectroscopic Images were acquired. This MRSI data set was acquired using a 2D STEAM pulse sequence with the STEAM box positioned in a transversal plane through the brain showing the largest tumor in the Gd-image. MRSI parameters were: $16 \times 16 \times 1024$ samples, TR/TE/TM=2000 or 2500/20/30 ms, slice thickness = 12.5 or 15 mm, FOV = 200 mm, spectral width = 1000 Hz and NS=2. Eddy current correction was performed [188], followed by water removal with HLSVD-PRO [218], and finally a simple baseline correction was done by filtering the time signal with a highpass filter with a filterwidth of 5 ms, followed by the subtraction of the residual from the original time domain signal. Zero order phase was already corrected during eddy current correction. Manually, the first order phase of the average spectrum of a patient was used to reset the first order phase of all spectra of that patient. Finally, all spectra were normalised with respect to the water signal [336].

4.1.6 Simulation study

Construction of simulated MRS images



(a) normal tissue (b) CSF (c) grade II

(d) grade III (e) glioblastoma (f) meningioma

Figure 4.3: *The first component ($\boldsymbol{y}_1$) of the spectral model for the different tissue types, present in this study.*

From the *in vivo* data set, a large database of representative spectra for the different tumor types, normal tissue and CSF was created for simulating purposes. This database was divided in two parts. From each tissue type, 25 spectra were used to construct the spectral model, shown in Fig. 4.3. As can be seen, the spectra show that the different tissue types are characterised by different contributions of the metabolites of interest. The other spectra were used to simulate 10x10 MRS images for each tissue type as shown in Fig. 4.4. The mixed spectra for all tumor types in these MRS images were artificially generated by averaging out a normal and a tumor spectrum.

Three types of simulated MRS images were created in order to evaluate the performance of CCA. In the first experiment MR spectra were used as input, the second experiment was carried out with a set of 10 quantified metabolite concentrations and the third experiment used a set of 14 features (10 metabolite concentration variables and 4 image variables) as input. For the experiments, the database was randomly split 5 times in order to obtain different model spectra and simulated MRS images. As there were four different tumor types in the database (grade II, grade III, glioblastoma and meningioma), the CCA-method was evaluated on 20 images. For comparison, the result of ordinary correlation was computed.
For experiment 2, the whole database was quantified. Using the same partitioning of the database as in the spectra case, the 20 corresponding grids of sets of metabolite levels were created.
For experiment 3, the image variables were added to the quantified metabolite levels.

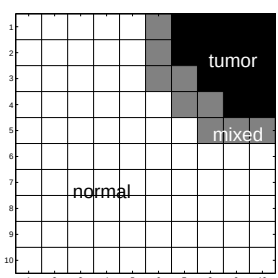


Figure 4.4: *Artificially created MRS image, partitioned into normal, mixed and tumor tissue. This partition was used for all the simulated MRS images.*

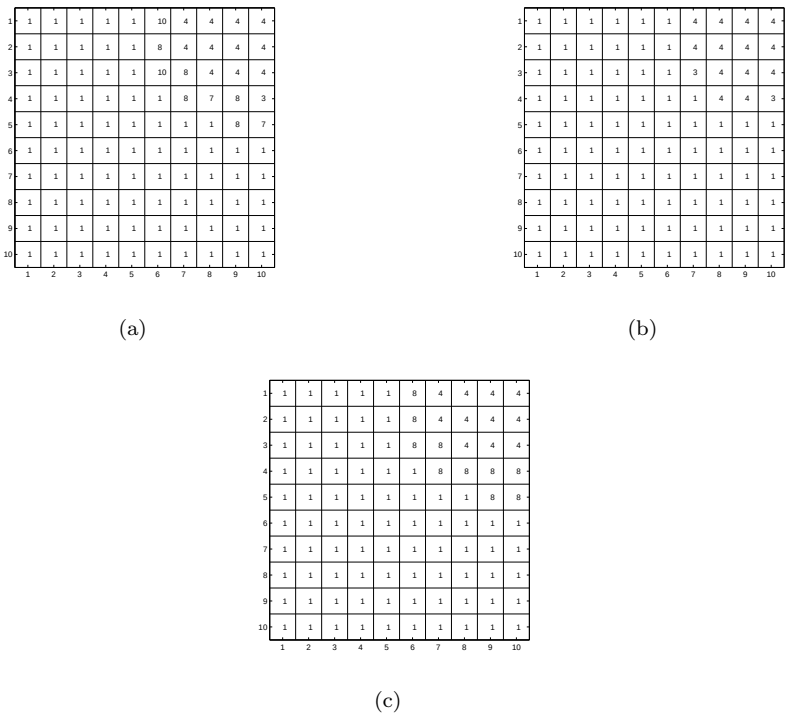
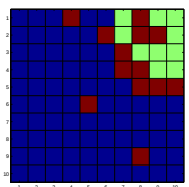


Figure 4.5: *The result of CCA on spectra using the spatial 'star' model applied to a simulated MRSI example containing a grade III tumor, (a) when all the constructed spectral models are used, (b) step one for tumor labeling: only spectral models of normal and tumor tissue are used, (c) step two for tissue segmentation: only spectral models of normal tissue, the detected tumor and corresponding mixed tissue are used. Explanation of numbers: 1 normal; 2 CSF; 3 grade II; 4 grade III; 5 glioblastoma; 6 meningioma; 7 mixed grade II; 8 mixed grade III; 9 mixed glioblastoma; 10 mixed meningioma.*

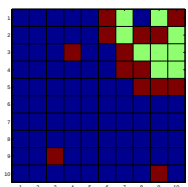
Results

Fig. 4.5 demonstrates the necessity of the two-step method. The left figure shows the result on a simulated MRSI example of grade III when all the constructed spectral models are used. A large region with normal tissue (1) and a region with tumor tissue of grade III (4) are detected. In the transition region, voxels are assigned to grade II (3), mixed tissue of grade II (7), mixed tissue of grade III (8) and mixed tissue of meningioma (10). We simulated the image with only grade III spectra and spectra of mixed grade III. A major part of the voxels are assigned incorrectly. In Fig. 4.5b, the first step of the proposed two-step method is applied: only spectral models of normal and pure tumor are used. A region with normal

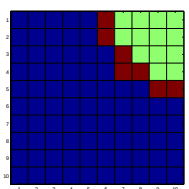
tissue (1) and a region with grade III tumor (4) are segmented. In addition, two voxels are assigned to grade II (3). As clinicians type the tumor with respect to the most malignant tissue present, the detected tumor type is convincingly grade III. In our second step (Fig. 4.5c), the MRS image is analysed with spectral models of normal tissue (1), tumor (4) and mixed (8) tissue of grade III. Three homogeneous regions are segmented. Compared to Fig. 4.5a, this result is certainly more correct and easier to interpret.



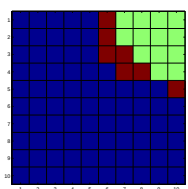
(a) correlation - spectra



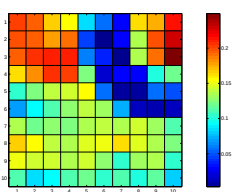
(b) correlation - features



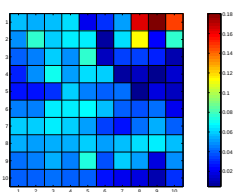
(c) CCA - spectra



(d) CCA - features



(e) reliability of (c)



(f) reliability of (d)

Figure 4.6: *The result of (a) ordinary correlation on spectra, (b) ordinary correlation on features, (c) CCA on spectra, (d) CCA on features. (e) and (f) show the smallest difference between two canonical correlation coefficients when CCA is computed using spectra and features respectively.*

For all the simulated MRS images, CCA found in the first step the correct tumor type. The simulation may be biased by the fact that the simulated MRSI are constructed with a database containing only carefully selected spectra. In the rest of this section, we only discuss the segmentation accuracy.

Before discussing more general results, it is useful to have a closer look at the segmentation result for one simulated MRS image. Fig. 4.6 compares the result of ordinary correlation (a (spectra) and b (14 features)) with the result of CCA (c (spectra) and d (14 features)) using the spatial 'star' model on a grade II tumor. The red, green and blue regions represent respectively the detected tumor, mixed and normal tissue. The example clearly shows the advantage of using spatial information: ordinary correlation does not segment homogeneous regions, but assigns a lot of isolated voxels to the wrong tissue type. Especially the distinction between tumor tissue and mixed tissue is not made correctly. The CCA-based method correctly detects the tumor region, surrounded by a band of mixed tissue. This results in a segmentation accuracy of 98% for CCA. It is observed that most of the misclassified voxels are corner voxels. This is logical, because corner voxels are surrounded by fewer voxels of the same tissue type. On Fig. 4.6, we cannot see differences between the CCA-method applied to spectra and features. Both perform in this case similarly. It is however interesting to note that misclassified voxels in the spectra case not necessarily correspond to misclassified voxels in the feature case.
A question that can arise is how reliable the result is. CCA gives a canonical correlation coefficient that is a measure of the similarity between the voxel under investigation and a model spectrum. This canonical correlation coefficient is for all voxels in the spectra case between 0.91 and 0.98, and in the features case between 0.95 and 0.999. This is much higher than the correlation coefficients of ordinary correlation (0.7-0.8). However no direct measure of reliability can be derived from the canonical correlation coefficient itself. The canonical correlation coefficient will rather reflect the amount of noise/artefacts in the data. A measure of reliability that can be introduced is the smallest difference between two canonical correlation coefficients, shown in Fig. 4.6e-f. When two canonical correlation coefficients are very close to each other, the probability of misclassification is much higher than in the case when one canonical correlation coefficient is higher than the other one. Fig. 4.6e nicely reflects this idea: in the region of the mixed tissue, the difference is smaller than in the other regions, because the correlation with all three tissue types is high. However in the feature case, Fig. 4.6f does not add any further information. This means that a better measure for reliability has to be proposed.

	spectra	metabolite levels	metabolite levels + image vars
CCA	97.3	95	98.2
ordinary correlation	91	93.7	95.5

Table 4.1: *The segmentation accuracy of the segmented image by CCA methods cross-correlation. The displayed value is the mean value of the results over 20 artificially generated MRS images with different tumor types. Values in %.*

To compare the different CCA approaches in a more general way, the segmentation

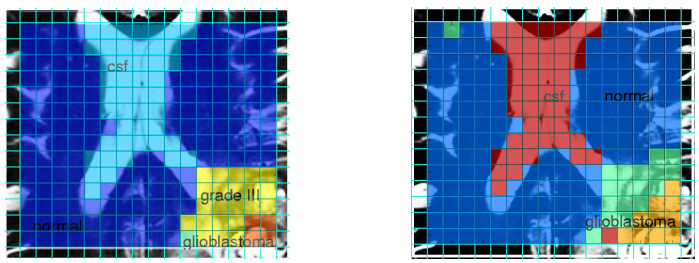
accuracy of the 20 segmented images is obtained by simply counting the number of correctly classified voxels. The average value is given in Table 4.1. The result of ordinary correlation is added as reference value. The columns of the table correspond to three kinds of input vectors: the MR spectra, the quantified metabolite levels or the features.
The lower performance of segmentation accuracy of the ordinary correlation analysis compared to CCA using a 'star' model is confirmed in Table 1. When the different input strategies are compared, it is seen that quantifying the spectra before applying CCA reduces the segmentation accuracy. This approach should never be used, as the quantification step can introduce errors. However, quantifying the spectra makes sense when image information is added. The values displayed in the third column are higher than the values in the second column, which shows that adding image information really improves the segmentation result. The accuracy with feature sets of 14 variables is slightly higher than the accuracy of the spectra case. However, the differences are very small. Keeping in mind that we only work with 20 carefully constructed simulated examples, we can state that the spectra and feature approaches perform similarly in terms of accuracy.

4.1.7 *In vivo* data

Also several *in vivo* MRS images were analysed. The CCA-results for four patients are discussed below. We focus in this part on differences between the two valuable approaches, namely CCA using spectra and CCA using feature sets of 14 variables. All the results are obtained with a spatial 'star' model. Care was taken that the model spectrum was constructed with 25 spectra of other patients. So independence is guaranteed. The colors of the segments in the figures 4.7 to 4.10 have a specific meaning. As the tissue types differ in the tumor typing and segmentation steps, it is not possible to use in both steps the same color legend. After the tumor typing step, dark blue is used for normal tissue, light blue for CSF, green for grade II, yellow for grade III, orange for glioblastoma and red for meningioma. After the segmentation step, blue represent normal tissue, green mixed tissue, orange the detected tumor tissue and red is used for CSF.
Fig. 4.7 shows the result of CCA using features for a patient with a glioblastoma tumor. In the first step several glioblastoma voxels are detected, but also several voxels at the boundary of the glioblastoma region are assigned to grade III. This could be expected since tumors are known to be very heterogeneous. The border between normal glial tissue and the center of the tumor is often lower in grade. The diagnosis of the method is correctly glioblastoma. In the second step most 'grade III voxels' in Fig. 4.7a are assigned to mixed tissue in Fig. 4.7b as the model spectra of grade III and mixed glioblastoma are very similar. It can also be seen that the segmented regions with CSF correspond nicely to the ventricles in the MR image.

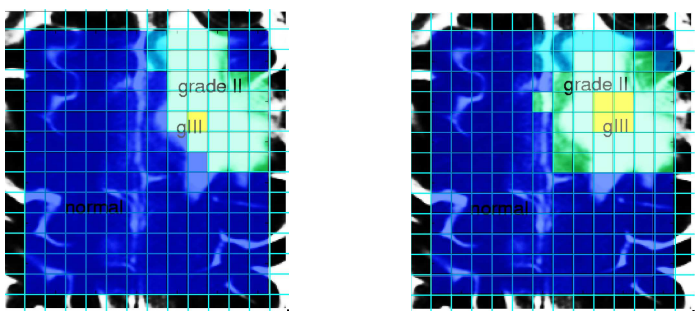
In Fig. 4.8 the results for spectra and features are compared after the tumor typing step on a tumor of grade III. Although the detected tumor region is large, the CCA method using spectra (Fig. 4.8a) only detects one voxel of grade III in the first step. This tumor is probably very heterogeneous and only a very small part of the tumor will really be of grade III. For this patient, the diagnosis based on the most malignant tumor present, will still be correct (grade III) but this example illustrates the limitations of methods that improve robustness by including spatial information. When the tumor region is too small, the tumor will be hidden by the surrounding voxels. Compared to the result obtained with features (Fig. 4.8b) we can clearly state that the result of CCA using features is better, as there is more convincingly a grade III region detected (like in [317]). In the second step, both approaches gave similar results.
Fig. 4.9 compares the segmentation results of both approaches on a meningioma tumor. The first step is not shown as a region with meningioma is convincingly detected by both CCA approaches. In the second step, the present meningioma tumor is nicely segmented in both figures. But it is also seen that the CCA method using spectra segments three lines as a (homogeneous) region of mixed tissue. These lines contain normal spectra. However, the spectra in the lower part are of minor quality compared to the spectra in the upper part of the grid.
Finally, Fig. 4.10 compares for the same patient the results of both approaches after step 1 and step 2. For this patient no histopathological agreement was reached. The pathologists who validated the tumor disagreed between grade II and grade III. Both CCA using spectra and CCA using features classify the large area as grade III. The result of ordinary correlation is not shown. Ordinary correlation assigned two voxels to glioblastoma tissue (grade IV). These are clearly misclassifications which will lead to the wrong diagnosis. These misclassifications are avoided by the use of spatial information. The most interesting part of this example is however the right ventricle. CCA using spectra assigns these CSF-voxels incorrectly to grade II (step 1) or mixed tissue (step 2). When features are used as input, the two ventricles are nicely segmented and correctly classified. This is certainly due to the incorporation of the image variables because in the MR images the ventricles are clearly visible. It can also be noted that in this figure and in Fig. 4.8, the segmentation boundaries differ between both approaches. However, it cannot be said which segmentation is the best, because no gold standard is available. A comparison between the obtained segmentation and the high-resolution MR images shows that the boundary of mixed voxels obtained with the method makes sense.
In [219], an extensive study was carried out in order to demonstrate the computational efficiency of the CCA-based segmentation. A segmentation based on the original MR spectra took a few seconds, the segmentation based on features was even faster due to the limited number of features. However, quantitation of the spectra takes 10 to 20 minutes with AQSES. Taking this into account, the overall segmentation on the original MR spectra is the fastest segmentation. Finally, we have to remark that the use of this accurate method has one drawback. A reliable

and representative model spectrum has to be constructed for the present tissue types, in order to obtain a meaningful result. To construct all the spectral models, a database of carefully validated spectra has to be generated. This takes time, especially for rare tumor types. Constructing a spectral model is even 'impossible' when there are 'unexpected' tissue types in a new MRSI dataset. In this study, 'unexpected' tissue types are *e.g.*, edema, necrosis, ... At the moment of this study, a large database of all possibly expected tissue types (adults and children) was starting to be constructed during the European eTumour project [161]. With this database, optimal spectral models can be generated.



(a) tumor typing step (b) segmentation step

Figure 4.7: *The result of CCA using features (a) after the tumor typing step and (b) after the segmentation step.*

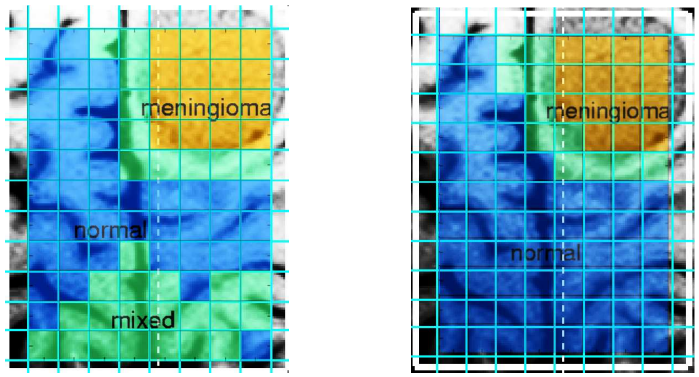


(a) CCA on spectra (b) CCA on features

Figure 4.8: *The result of CCA after the tumor typing step (a) using spectra and (b) using features.*

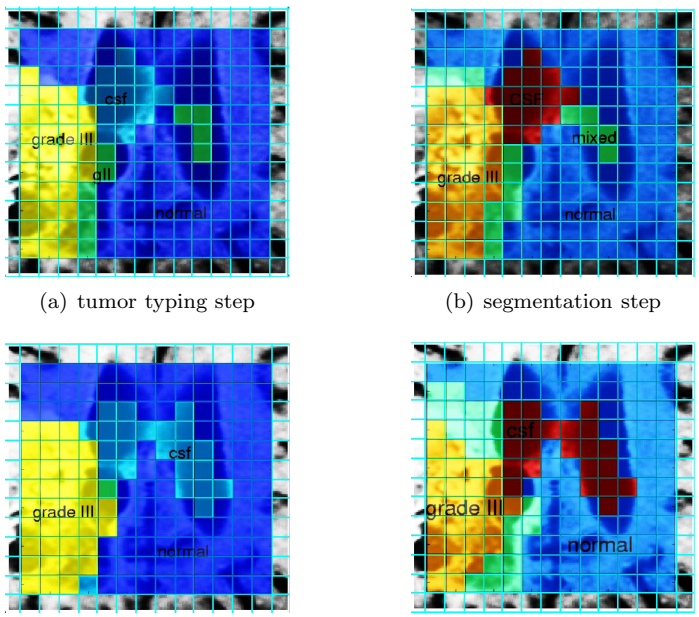
4.1.8 Summary on segmentation of MRS data.

CCA is a technique that quantifies the relationship between two multivariate vectors. In this section, we wanted to illustrate the CCA-based method. It was



(a) CCA on spectra (b) CCA on features

Figure 4.9: *The result of CCA after the tumor typing step (a) applied to spectra and (b) applied to features for a patient with a meningioma tumor*



(a) tumor typing step (b) segmentation step

(c) tumor typing step (d) segmentation step

Figure 4.10: *The result of CCA using spectra (a) after the tumor typing step and (b) after the segmentation step and the result of CCA using features in (c) and (d)*

introduced in the MRS world to translate short echo-time brain MRS images into easy-to-understand nosologic images. In order to distinguish in a reliable way

between different brain tumor types, a two-step method is presented. The first step is a tumor typing step and in the second step the heterogeneity of the detected tumor is shown. In addition, we introduce a CCA-based method on feature sets. These feature sets integrate information of MRSI and MRI. We also show that the CCA-based method on such feature sets outperforms the CCA-based method on the whole spectra in *in vivo* cases. These results were published in [77]. The results look promising, although no golden standard was available at the time of analysis. Furthermore, only 4 patients could be analysed. More recently, research on brain tumor recognition explored the use of multi-class LS-SVMs [231] or incorporated segmentation techniques of the original MRI, after which the segments can be classified [232]. Unfortunately, LS-SVM based classifications were never compared to our CCA-based method.

4.2 BSS-CCA for removal of muscle artifacts from epileptic seizure recordings

4.2.1 Method

As described in section 2.1.1, the EEG is often contaminated by muscle artifacts, certainly during epileptic seizures. In this section, we illustrate a recently proposed method for muscle artifact removal in EEG, based on Canonical Correlation Analysis (CCA) as a Blind Source Separation technique (BSS): BSS-CCA. In [49], the method was shown to remove successfully the muscle artifact without altering the recorded underlying brain activity. BSS-CCA had already been used for the cleaning of EEG in [189]. In [49] the method was shown to outperform a low pass filter and an ICA-based technique for muscle artifact removal in accuracy.

CCA can be used as a BSS technique when certain assumptions are valid: (1) the mixing should be linear and instantaneous, (2) the sources have to be mutually uncorrelated, (3) the sources have to have a different autocorrelation structure, and (4) the number of sources equals at most the number of sensors. The third assumption is crucial to obtain a good decomposition. In the case of EEG recordings, assumptions (1) and (2) are satisfied, because volume conduction is assumed to be linear and without appreciable delay in the scalp sensor measurement [290, 227]. The ideal sources correspond to brain activity and muscle artifacts. These sources are uncorrelated. It was demonstrated in [49] that brain sources have a much higher autocorrelation than muscle sources. This means that assumption (3) is satisfied, and a good decomposition can be obtained. Furthermore, because of the ordering of the CCA sources according to decreasing autocorrelation, [49] proposed a semi-automatic method, where the least autocorrelated sources were removed one by one until enough muscle activity was removed. After the removal of a muscular source, the EEG is reconstructed without this source. This can be done by projecting the other sources back onto the scalp. The method is semi-automatic, because a clinician has to visually decide which threshold to select in order to separate brain sources from muscle sources. It has to be mentioned that the effective number of sources contributing to the scalp EEG is not known, so assumption (4) cannot be proved to be valid.

4.2.2 Illustration

We illustrate the BSS-CCA decomposition for the epoch of multichannel EEG containing ictal EEG and muscle artifacts shown in Fig. 1.1, repeated in 4.11 (a).

The (normalized) sources as obtained by BSS-CCA are depicted in Fig. 4.12 (a) and their corresponding autocorrelation is shown in Fig. 4.12 (b).

We can observe that the muscle activity is well separated from the rhythmical ictal activity (mainly present in source 3, but also visible in source 2 and 4), and that the muscle activity is present in the lowest autocorrelated CCA sources. Fig. 4.11 (b) shows the reconstructed EEG after excluding sources (9-21). It can be seen that muscle activity has been removed and that the sharp ictal theta wave is not altered.

4.2.3 Discussion

Although we limited ourselves to only one illustration of the BSS-CCA method, the example is representative. The method was formerly validated to remove muscle artifacts more accurately than low pass filters. The frequency spectrum of the muscle artifacts overlaps with that of the brain activity signal. This implies that low pass filters are not capable of removing the whole artifact without altering the underlying brain signal.

ICA has also been proposed to clean ictal EEG from the EMG contamination [357, 340]. Although ICA might sometimes be able to remove parts of tonic, continuous EMG activity in epileptic context, it has been shown to be far from optimal [49]. This is probably due to the fact that the underlying sources are not really statistically independent, and can not be separated with *e.g.*, JADE. Furthermore, the selection of muscle sources after the ICA decomposition is not straightforward, as the sources are not ordered. Recently in [233], ICA has been used to remove muscle artifacts from sleep recordings. It is hard to judge if ICA was appropriate for their goal, as the authors did not show cleaned EEG data. The comparison of decomposition methods will be discussed in the next chapter.

BSS-CCA has been validated on a large database of ictal seizures [350]. The authors showed that using BSS-CCA as a preprocessing step on ictal EEGs improved the localisation and recognition of ictal activity. Moreover, the short computational time makes the proposed method applicable in clinical practice. Also seizure detection was shown to benefit from this muscle removal tool [351].

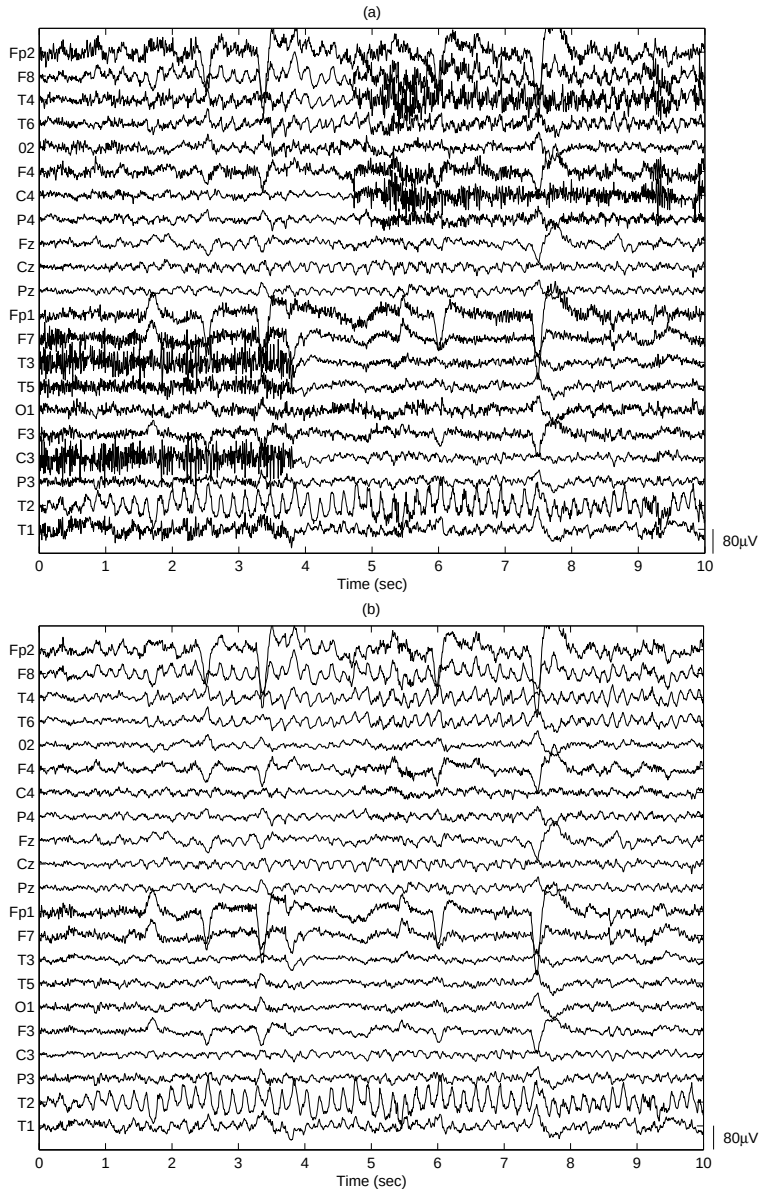
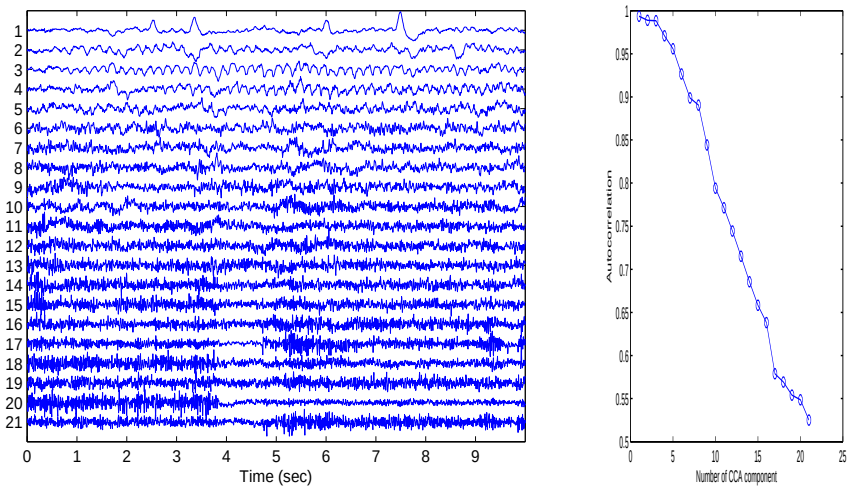


Figure 4.11: *(a) Original EEG epoch; (b) EEG reconstruction after excluding CCA sources with lowest autocorrelation. One can observe that the muscle artifact was removed, while the sharp ictal theta activity in the right temporal lobe was not altered.*



(a) (b)

Figure 4.12: *(a) Sources estimated by BSS-CCA on the ictal EEG recording shown in Fig. 4.11 (a). (b) The autocorrelation of the CCA components.*

4.3 BSS-CCA for the removal of short bursts of EMG activity during speech production

4.3.1 Introduction

Psychologists and neuroscientists have made extensive use of brain activity measures to construct models of language processing [330]. In most of these studies, human participants are requested to understand or decode language (*e.g.*, reading or listening to utterances). By contrast, the research on how the brain produces spoken language is considerably under-developed (for a review see [166]). This is partly due to the artifacts induced by overt speaking on the signals measured by imagery techniques. For example, articulatory movements tend to reduce the signal to noise ratio in functional magnetic resonance imaging. Magneto- and Electro-encephalography suffer from the contamination of the brain activity (evoked fields or potentials) by the Electromyographic (EMG) activity of the face muscles involved in overt speaking.

Previous research has shown how these artifacts may render the data uninterpretable. For example, [245] reported larger slow potentials in an EEG experiment during both the production of polysyllabic words, and in analogous non-speech

articulatory movements (*e.g.*, puffing). The critical finding was that these slow potentials were left-lateralized during word production but not in the control condition. This effect was interpreted as a reflection of linguistic processing tied to left inferior frontal cortical activity. However, shortly after publication, this interpretation was questioned based on methodological concerns [256]. Recordings of EMG activity of various articulators suggested that the reported lateralization effect may not have occurred during the preparation of the verbal response, but instead during articulation itself. Consequently, it may be a mere consequence of EMG activity. [28] investigated this issue in close detail (see also references therein) and also compared speech to non-speech articulatory responses, while they measured both EEG scalp activity as well as the EMG of several articulators. Results showed strong correlations between the activity recorded by electrodes placed over the articulators and most of the activity picked by the electrodes placed over the scalp (with the notable exception of the vertex scalp electrode). Further analyses were conducted in which the muscle–related activity was included as a co-variate factor. They did not reveal any lateralization of speech related cortical potentials. Besides questioning previous findings, the general conclusion was that pre-vocalization potentials are severely confounded by muscle artifacts.
In light of these difficulties, scholars investigating speech production have resorted to a variety of strategies. The most radical one has been to investigate cognitive stages of speech production without speech being overtly produced during experimental trials (speech was produced during interleaved filler trials). In an influential article [344], participants were asked to press buttons in response to visually presented pictures; critically, the responses were guided by linguistic properties of the pictures' names. For example, participants pressed one of two buttons depending on the first phoneme of the picture's name. In this way, participants had to process linguistic information without having to overtly speak. A somewhat similar strategy was adopted by [173]. In their experimental paradigm, participants named pictures while their EEG activity was recorded. However, they did not name the pictures directly after their presentation but when a visual cue was presented to them. In the critical experimental trials, participants heard a distractor word just after the presentation of the picture. A semantic manipulation was performed on this word. The EEG activity of interest was recorded between this distractor word and the visual cue. In this way, the data were free of EMG activity. The downside is that they presumably reflected a combination of brain activities elicited by speaking and listening.

These experimental protocols have provided information about the neural basis of linguistic processing. However, what is being investigated in these studies may not reflect processes of natural speech production. This is because the complex instructions that participants are requested to follow presumably induce considerable non-speech brain activity (*e.g.*, attention focused on decision making in button press or on withholding verbal responses). Another strategy has been

to elicit speech in a spontaneous manner (*e.g.*, immediate picture naming) and thus obtain heavily muscle-contaminated EEG recording. Scientists following this approach try to remove the muscle artifact with low-pass filters. For example, [244] or [123] used 10 or 12 Hz low pass filters. This approach is not without problems, however. The frequency spectrum of muscle artifacts has been shown to largely overlap with that of brain signals of theoretical interest [128]. [119] showed that EMG activity is present both in the alpha (8-13 Hz) and in the beta bands (13-20Hz). [128] reported that facial EMG has a broad frequency distribution, from almost DC to more than 200Hz, even with weak muscle contraction. Moreover, heavy low pass filtering prevents investigations of EEG activity present above 10 Hz (*e.g.*, beta band), and is known to significantly reduce the amplitude of phasic activities. Visually evoked potentials, for example, often last around 100 ms and will therefore be clearly suppressed after 10 Hz low-pass filtering [230].

We explore now the use of BSS-CCA for the removal of these EMG bursts. We will discuss why this method can be appropriate for this purpose.

Application of BSS for muscle burst removal

We again consider the model

$$\mathbf{Y} = \mathbf{M} \cdot \mathbf{X} \tag{4.2}$$

and aim at estimating the underlying, generating sources from the observed mixture. This is carried out by introducing the unmixing matrix $\mathbf{W} \in \mathbb{R}^{P \times I}$ such that the estimated sources

$$\tilde{\mathbf{X}} = \mathbf{W} \cdot \mathbf{Y}, \tag{4.3}$$

approximate the unknown physiological source signals in $\mathbf{X}$, up to a scaling factor. Ideally, $\mathbf{W}$ is the inverse of the unknown mixing matrix $\mathbf{M}$, up to scaling and permutation. We will now show why BSS-CCA can be used. In the results section, we will show that ICA is not able to remove EMG contamination from speech recordings.

We already mentioned that in [49], a BSS-CCA method was introduced for separating brain and muscle sources in epilepsy. The estimated sources were sorted in order of decreasing autocorrelation (highly autocorrelated sources ranked first, weakly autocorrelated sources ranked last). After BSS-CCA, the sources with the highest autocorrelation should correspond to EEG, while the sources with the lowest autocorrelation should correspond to EMG. With this method, the authors could reveal the ictal activity otherwise completely masked in the background tonic EMG activity.
In speech production research, EMG contamination from the muscles involved

in articulating words is not a tonic, continuous activity, but instead appears as short EMG bursts, that are localised in time around the window of interest. A second difference with [49] is that in epilepsy, a semi-automatic method was used. Neurologists visually removed muscle components one by one until ictal activity became visible. In the current application, we select the EMG sources automatically based on spectral properties (see below).

The goal of this study was to investigate whether the BSS-CCA method could be used in practice to automatically distinguish cortical and EMG signals in electrophysiological recordings obtained during spoken language production. We did so on a dataset recorded using the picture naming task (a popular method for eliciting speech in psycho- and neuro-linguistic experiments).

4.3.2 Materials and Methods

Participants and task

12 right-handed native french-speakers (3 females) with normal or corrected to normal vision, participated in the experiment (mean age: 24.5). They all gave their informed consent. Participants were presented with line drawings pictures representing objects that they had to name as fast and accurately as possible [3]. Forty-five line drawings (11×11 cm) of common objects were used as stimuli. Each drawing was randomly repeated twice during the experiment. Vocal responses were recorded with a voice-key (Eprime 2.0 Professional, Pittsburgh, PA: Psychology Software Tools).

Electrophysiological recordings

The EEG was recorded by means of 64 Ag/AgCl pre-amplified electrodes (BIOSEMI, Amsterdam) (10-20 system positions). The sampling rate was 512 Hz (anti-aliasing filters: DC to 104 Hz, 3 db/octave). The passive reference was obtained by offline averaging the signal recorded over the left and right mastoïds. The vertical electro-oculogram (EOG) was recorded by means of two electrodes (same type as EEG) just above and below the left eye, respectively, and the horizontal EOG was recorded with two electrodes positioned over the two outer canthi.

Electrophysiological processing

After acquisition, the EEG data was filtered (high pass = 0.16 Hz). Eye movement artifacts were corrected using the statistical method of [132]. The continuous EEG

was epoched off-line, time-locked to stimulus presentation, starting from -0.2 s up to 2 s after stimulus onset. BSS-CCA decomposition was computed for each epoch separately, obtaining a $\mathbf{W}$ matrix and source signals for each trial (of 2.2 seconds). This was motivated by the fact that different muscular patterns are associated with different words used in this experiment [241], and hence that the decomposition should be word-dependent. The sources obtained with BSS-CCA are ordered according to their autocorrelation. Because the autocorrelation of a source is an abstract value, another criterion has to be used to select the sources that are considered to be EMG. This criterion can be based on the Power Spectral Density (PSD), because EMG sources can be assumed to also have more high-frequency content than brain sources. Components were considered to be EMG activity if their average power in the EMG band (approximated by 15-30 Hz) was at least one seventh of the average power in the EEG band (approximated by 0-15 Hz). These values were empirically determined based on visual interpretation of the recordings. We will show that the chosen values (15 and 7) are not critical at all for the proposed method. In fact, our criterion removed the least autocorrelated sources. This confirms that decomposing EEG for separating brain and muscle sources with BSS-CCA is appropriate. The contributions of the source signals identified as EMG were removed from the surface EEG by setting to zero the corresponding columns in the $\mathbf{M}$ matrix. This $\mathbf{M}$ is estimated as the inverse of $\mathbf{W}$. The new mixing matrix $\mathbf{M_{clean}}$ is then used to reconstruct the denoised EEG signal matrix $\mathbf{D_{clean}}$:

$$\mathbf{D_{clean}} = \mathbf{M_{clean}} \cdot \mathbf{X}. \tag{4.4}$$

$\mathbf{D_{clean}}$ will contain sources not identified as muscle activity. After removing muscle artifacts, all other artifacts (*e.g.*, related to eye movements, recording problems, persistent muscle activity) were rejected after a trial-by-trial visual inspection of monopolar recordings and the clean EEG segments were re-epoched, centered on the verbal response onset.

For comparison, instead of removing EMG sources with BSS-CCA, we applied a low-pass filter (10 Hz, 24db/oct), and we also applied an ICA algorithm based on higher-order statistics (RobustICA [362]).

Statistical evaluation

In order to quantify the removal of EMG in a more objective way, we computed the amplitude differences for all the subjects between the first negative peak and the following positive peak on one electrode (Fp2). We computed this for three datasets: the raw data, the data cleaned with the low-pass filter, and the data cleaned with BSS-CCA. We compared the obtained values with a standard t-test.

4.3.3 Results

The average reaction time (RT) across subjects was 651 ± 72 ms and the average error rate, defined as the rate of wrong namings during the task, was 1,31% ± 0,96 %.
We will illustrate the effect of BSS-CCA on one representative subject with considerable EMG contamination. Although we only show results for this subject, very similar figures were obtained for the other subjects. After trial-by-trial visual inspection, 67 % of the trials were left for further analysis.

Impact of BSS-CCA on speech-related EEG activity denoising

[49] has already demonstrated the validity of BSS-CCA to remove EMG artifacts by means of simulations. Evaluating the validity of the method on real data is a much more difficult matter, since, by definition, the unaffected signal is unknown. We will first present several analyses comparing directly the raw to the cleaned data.

Fig.4.13 presents an example of five central channels (FPz, AFz, Fz, FCz and Cz). The top panel presents the raw data, the middle one plots the identified EMG components, and the bottom panel presents the EEG data after EMG removal. It is to be noted that the removed components clearly represent EMG burst around reaction time, and no low frequency modulation. In contrast, the cleaned traces present slow modulations, typical of EEG activity with, apparently, no visible EMG activity.

The previous figure shows the cleaning on only one epoch (but several electrodes). We now visualize for one electrode the impact of denoising on all the trials with the ERP-image technique [179]. All the trials are represented as parallel color lines, where color codes are used for the polarity and the strength of the activity. The trials are sorted as a function of RT, represented as the black S-Shaped line (trials for shortest RT are at the bottom, trials with the longest RT are at the top). Fig.4.14 presents the ERP-images of the original epochs, and of the epochs after cleaning with BSS-CCA for channels Fp2 and T7. We can observe a reduction of the high frequency noise (visible as a pixelisation of the plot), especially around the response time.

Since EMG contamination is mainly located in high frequencies (see previous figures), we can evaluate the impact of BSS-CCA cleaning more directly on the power spectrum of the recorded activities. Fig. 4.15 shows the averaged spectral content of all epochs on selected channels represented topographically in the original data (black), after EMG removal with BSS-CCA (blue) and after low-

pass filtering (red). While the unfiltered EEG contains a lot of high frequency content, the low-pass filter and the BSS-CCA method dramatically reduces high frequency and reveals the expected $1/f^{\alpha}$ shape in the spectrum. Note that the actual value of α differs between the two methods. It can not be told from this figure which spectra are the best. However, the spectrum after BSS-CCA is closer to the original than the spectrum after filtering. We can also see that, prior to applying BSS-CCA, there is more high frequency content on the frontal and temporal compared to the central, parietal and occipital recording sites. After muscle removal, however, an $1/f^{\alpha}$ shape of the spectrum is visible on all electrodes.

To compare BSS-CCA and filtering more in detail, we plot the averaged ERP on two channels (Fp2 and P3) in Fig. 4.16. In both figures can be seen that the ERP cleaned with BSS-CCA resembles much more the ERP of the raw data than after filtering. Furthermore, the amplitude difference between the first negative peak and the following positive peak is clearly larger after BSS-CCA than after filtering.

In the next figure, we also compare the removal with BSS-CCA and the removal with ICA based on higher-order statistics. ICA based on higher-order statistics was previously proposed to clean EEG from muscle artifacts [340]. The main disadvantage that was mentioned for using ICA is the selection of sources related to EMG. To overcome this problem, we used exactly the same criterion as we proposed for BSS-CCA. The figure clearly illustrates that the problem is not the selection of EMG components, but the decomposition itself. Cleaning with ICA removes less EMG and follows less the raw ERP.

In Fig. 4.18, we show that the actual thresholds for selecting components is not relevant. We compared three different parameter settings on the averaged ERP (border frequency 15, and removing an EMG source when there is more than 1/7 of the power above this frequency; 16 and 1/10; 13 and 1/5). The figure shows that the actual threshold does not influence the final result.

Finally, we evaluated where the removed components were stemming from by plotting the topography of the removed muscle components (Fig. 4.19). To do so, one needs to provide a representative topography. Plotting the topography of an individual removed component is not appropriate since it might not be representative. Instead, we normalise the estimated source signals to unit variance, and weighted the mixing vectors by the appropriate variance. The average of all mixing vectors is then a representative estimate of the topography of the EMG sources. Indeed, consider again the estimation

$$\hat{\mathbf{X}} = \mathbf{W} \cdot \mathbf{Y} \tag{4.5}$$

for the decomposition

$$\mathbf{Y} = \mathbf{M} \cdot \mathbf{X}. \tag{4.6}$$

After removal of the EMG sources, we can construct two datasets: $\mathbf{Y_{clean}}$, containing the cleaned EEG and $\mathbf{Y_{EMG}}$ containing muscle artifacts:

$$\mathbf{Y_{clean}} = \mathbf{M_{clean}} \cdot \mathbf{X} \text{ and } \mathbf{Y_{EMG}} = \mathbf{M_{EMG}} \cdot \mathbf{X} \tag{4.7}$$

If the sources are ordered by decreasing auto-correlation, $\mathbf{M_{clean}}$ and $\mathbf{M_{EMG}}$ are defined as:

$$\mathbf{M_{clean}} = \mathbf{M} \cdot \begin{pmatrix} 1 & 0 & 0 & \dots & \\ 0 & 1 & 0 & \dots & \\ \vdots & \vdots & \ddots & & \\ & & & 0 & \\ & & & & \ddots \end{pmatrix} \text{ and } \mathbf{M_{EMG}} = \mathbf{M} \cdot \begin{pmatrix} 0 & 0 & \dots & & \\ \vdots & \ddots & & & \\ & & 1 & & \\ & & & 1 & \\ & & & & \ddots \end{pmatrix}. \tag{4.8}$$

Only the columns from the first sources will be retained in $\mathbf{M_{clean}}$ and only the last columns will be retained in $\mathbf{M_{EMG}}$. The coefficients in $\mathbf{M_{EMG}}$ give an estimate of how much each EMG source contributes on the observed signal. These coefficients are constant within trial and they do not reflect the exact EMG contribution at any time instant, as the time courses of the components are ignored while they can have all possible values. A global representation of EMG topography can be obtained nonetheless by averaging all the EMG mixing vectors because an actual mixing topography will be given by a weighted sum of these mixing vectors. A similar reasoning can be applied to the effect of filtering via Fourier transform, and it has already been derived in the literature that Singular Value Decomposition (SVD) can be interpreted as a filter [140]. We average this estimate over trials and plot it on a 3D head. Most activation is frontal as expected, as many muscles involved in speech production are facial. Also some EMG contributions of the neck muscles are observed.

In order to validate the applicability of the BSS-CCA method on all the subjects, we compared the amplitude differences between the raw data, the data cleaned with BSS-CCA and the data cleaned with filtering. The averaged result are shown in Table 4.2. The amplitude difference was significantly greater after BSS-CCA than after heavy low-pass filtering ($t(11) = 5,54, p < 0,001$). There was no significant difference between the amplitude difference measured on the raw data compared to after BSS-CCA ($t(11) = 1,51, p = 0,10$).

4.3.4 Discussion

Psycholinguists have long avoided electrophysiological (EEG and/or MEG) investigations of spoken language because of the, justified, fear of the artifacts induced by facial EMG. Here we propose a solution to this problem based on a BSS technique that exploits the difference in autocorrelation between brain and muscle signals in order to separate them. This method was originally developed, and validated, for long term epilepsy recordings. In the present study, we adapted and automated it to remove short bursts of myographic activity related to speech production. Validating artifact removal techniques is a challenging task since neither the EMG artifact nor the EEG signal related to speech production are well known. We thus investigated in detail the effects of the proposed method based on CCA in various ways.

BSS-CCA successfully reduced the EMG artifacts in the EEG signal recorded during speech. The efficiency of BSS-CCA is clearly visible on a trial by trial basis (Figures 4.13 to 4.14) but also on the power spectra of the grand averages (Fig. 4.15). Before muscle artifact removal, the power spectra showed lots of high frequency activity. After cleaning, the power spectra had a clear $1/f^{\alpha}$ shape on all electrodes. This is exactly the shape of the EEG power spectrum in [128] in the situation where subjects were asked not to contract facial muscles. This shape model is not a very strong one, and can also be obtained with filtering, although the actual α value in filtering will be higher, and the spectra after filtering will less follow the original spectrum.

Based on our results, we can certainly claim that the removal of EMG contamination with BSS-CCA outperformed the removal with filtering (Fig. 4.16) and with ICA (Fig. 4.17). That filtering at 10 Hz removes also brain activity is quite well-known, and again confirmed by the figure.
Also the use of higher-order ICA seems to be problematic for speech-induced EMG. ICA on higher order statistics requires lots of samples in order to estimate accurately the mutually statistical independence. As every trial is limited in time, there are not enough samples for ICA to reliably decompose the EEG into brain and muscle activity when one wants to compute ICA on a word by word basis. No strong explanations can be given why ICA does not separate the brain and EMG sources. It is probably due to the fact that these sources are not statistically independent. The superiority of BSS-CCA over ICA is reflected in our figure: there is more muscle artifact left after ICA than after BSS-CCA while the shape of the original ERP is less followed.

In order to further validate the removal, we studied the part of the signal that was removed by the algorithm. The topography (Fig. 4.19) of the rejected data indicates that most of the muscle activity was frontal. Also in the raw data, we observed that the impact of facial muscle contraction was related

Table 4.2: *The average amplitude difference between the first negative peak and the following positive peak between the 12 subjects. It can be seen that filtering largely reduces this amplitude difference.*

Raw data	BSS-CCA	filtering
8,36 ($\sigma = 3,25$)	7,96 ($\sigma = 3,05$)	6,11 ($\sigma = 2,54$)

to the location of the recording electrode. These topographic differences are confirmed by the physical position of the articulation muscles, i.e. facial, and thus reflected on frontal electrodes. The locations where most activity was found on these topographies also corresponds to where the power spectra of the raw data contained most high frequencies.

The last validation concerned the peak-to-peak values. Based on the statistical test performed on all the subjects, no significant difference could be seen between the values on raw data and the values after BSS-CCA.

Altogether, although the validation stays a difficult problem, our results tend us to believe that we successfully removed the most prominent EMG artifacts present in the EEG signal without significantly altering the actual brain signal. However, although the global picture is convincing, the question remains if a removed source contained purely EMG and thus whether or not BSS-CCA removes some brain activity along with the EMG artifacts in EEG. We have started to study event-related potentials elicited by speech production that had not been described previously, see also [299]. The future may further clarify possible limitations in the next experiments we will perform. Also other neuroscientific research groups can use this algorithm (freely available at www.neurology-kuleuven.be/index.php?id=210) for cleaning their event-related potentials and start thorough speech research. Covering activated muscles with more electrodes can maybe further improve the separation.

The BSS-CCA algorithm clearly outperforms the existing filtering method used to overcome the problem of EMG artifacts elicited by articulation in the EEG signal without avoiding direct overt speech in the experimental tasks. This method therefore enables neuroscientific investigations of spoken language production beyond current practice.

The results of this method in epilepsy and language processing experiments are very promising. They also open perspectives towards new applications, both with continuous and event-related EEG. One could think of addressing the problem of EMG in neuromuscular paralysis [358] or EMG during ongoing mental tasks.

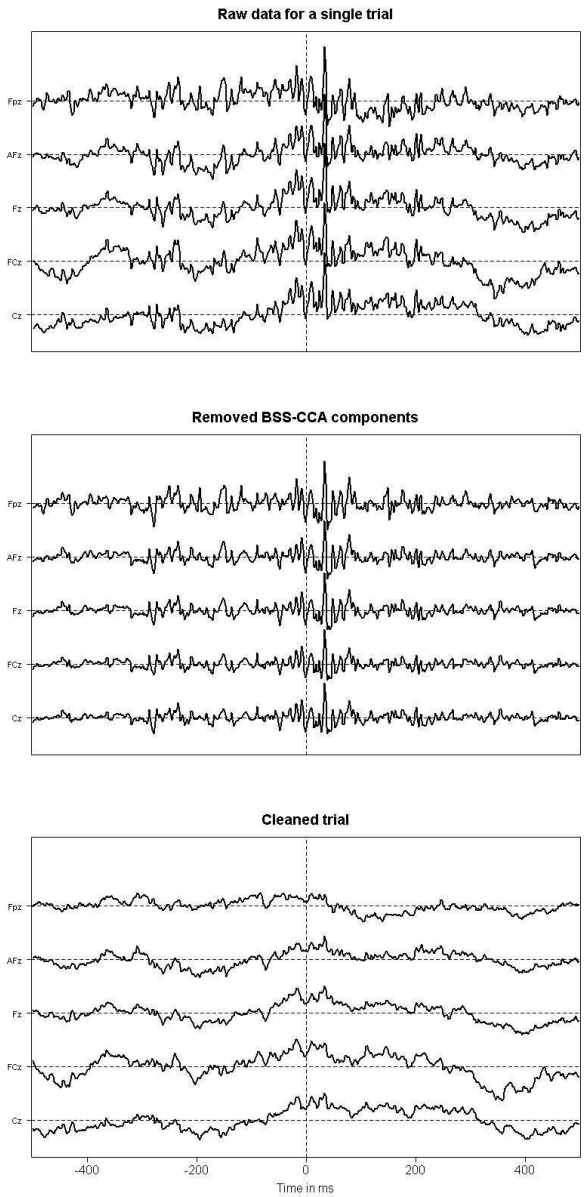


Figure 4.13: *Single trial of EEG data on 5 channels around voice onset (0 ms). We show the original evoked potential, the components that BSS-CCA removed and the evoked potential after cleaning with BSS-CCA. The removed components correspond to high frequency activity, and the cleaned trial still contains the low-frequency fluctuation.*

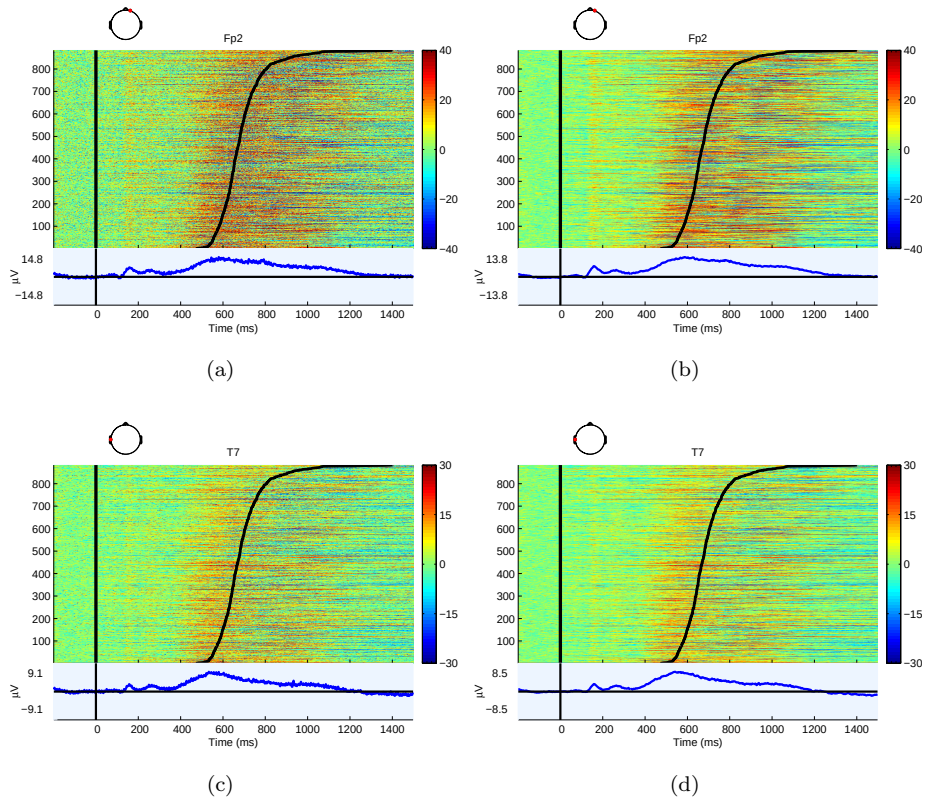


Figure 4.14: *Raster plot of the single trials at channel Fp2 and T7 before (a,c) and after (b,d) muscle removal, ordered according to reaction time. The black line represents reaction time and the blue line the ERP. After BSS-CCA, less high-frequency activity, supposed to be related to EMG contamination, is visible.*

4.4 Conclusion

CCA is a method to assess the similarity between two datasets. We illustrated this with the segmentation of MRS Images.
Furthermore, CCA can be used as a BSS method. BSS-CCA was shown to be a useful semi-automatic method to remove muscle artifacts from continuous EEG recordings in epilepsy.
We proposed an automatic version to clean muscle artifacts from epoched speech production recordings. We faced the inherent problem of validation, as nor the artifact, neither the evoked potential is known. However, our results are promising and certainly outperform low-pass filtering. The method opens perspectives for

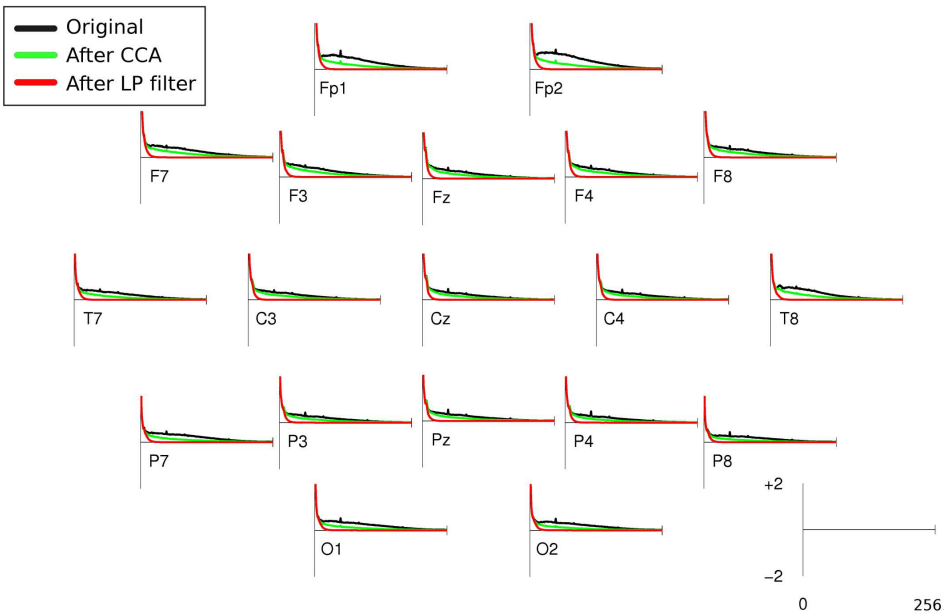


Figure 4.15: *Averaged Fourier transform over the trials of selected channels, equally distributed over the head. Black is the raw spectrum, green represents the spectrum after BSS-CCA and the spectra after filtering is shown in red. The green and red spectra have a* $\frac{1}{f^{\alpha}}$ *shape. However, the green spectrum follows the original spectrum more closely.*

research on brain activity involving speech production.

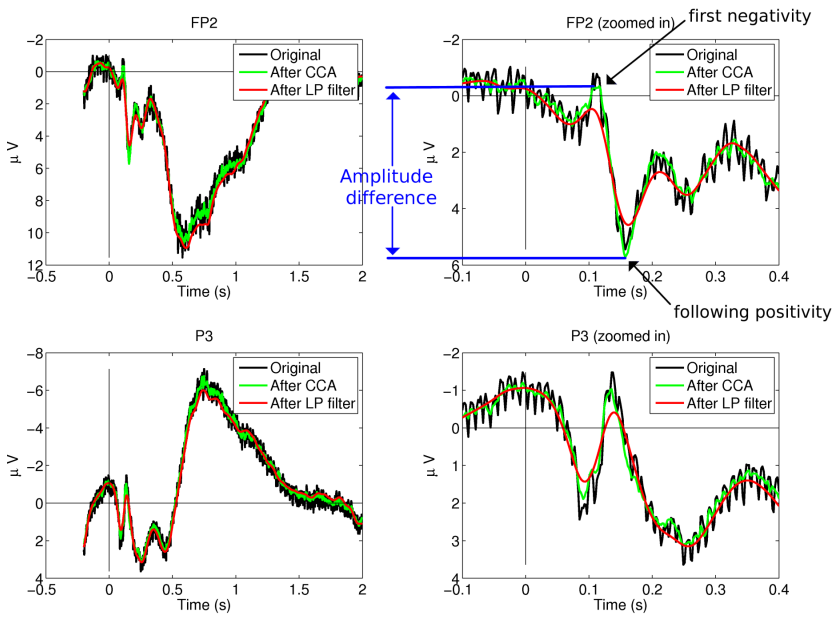


Figure 4.16: *Averaged ERP on two channels. We can see that BSS-CCA follows much more closely the ERP from the raw data than filtering.*

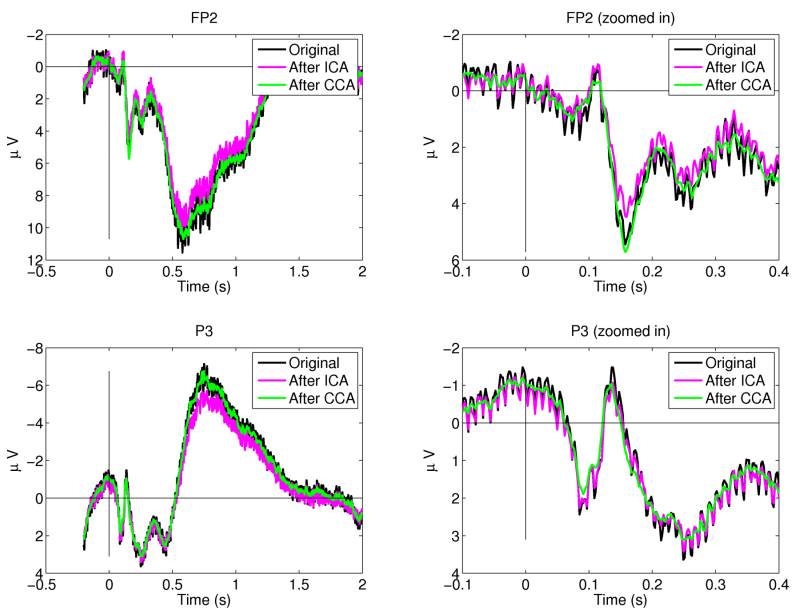


Figure 4.17: *Averaged ERP on two channels. We can see that BSS-CCA outperforms ICA.*

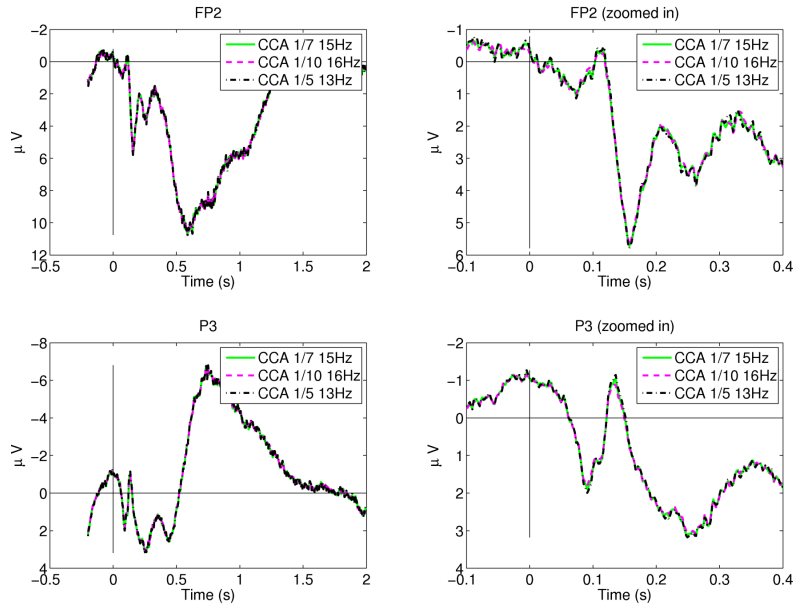
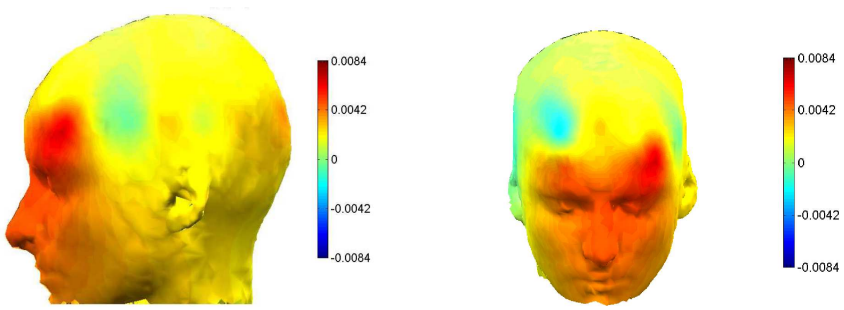


Figure 4.18: *Averaged ERP on two channels. We can see that the parameters for source selection do not influence the final ERP much.*



(a) (b)

Figure 4.19: *Averaged topography of the removed muscle components. Mainly frontal activity, corresponding to cheek muscles, can be observed. Note that the colors outside the EEG electrodes are not very reliable, due to extrapolation.*

Chapter 5

Eye artifact removal and spatially constrained ICA

Independent Component Analysis (ICA) is a Blind Source Separation (BSS) method that blindly decomposes a linear mixture in statistically independent sources. In this chapter, we illustrate the decomposition of EEG by means of ICA and show its use to remove eye artifacts. We also discuss in a more global way decomposition methods.
Eye artifacts are caused by the rotation of the electrically charged eye ball. Because the position of the eye ball is fixed with respect to the electrodes, we have prior knowledge of this artifact: we know one mixing vector. In the second part of this chapter, we derive several algorithms to include such prior knowledge in the computation of ICA, which we will denote by spatially constrained ICA (scICA). We validate the improvement in accuracy of this constrained decomposition with respect to standard ICA. We conclude the chapter by illustrating the use of scICA for the removal of eye artifacts.

5.1 Removal of eye artifacts with BSS

5.1.1 Introduction

The EEG is a non-invasive tool that offers millisecond temporal resolution for the study of neural mechanisms underlying mental activity. As described in section 2.1.1, it is often contaminated by eye artifacts. Eye blinks also produce electrical potentials creating significant electro-oculographic (EOG) artifacts in the EEG.

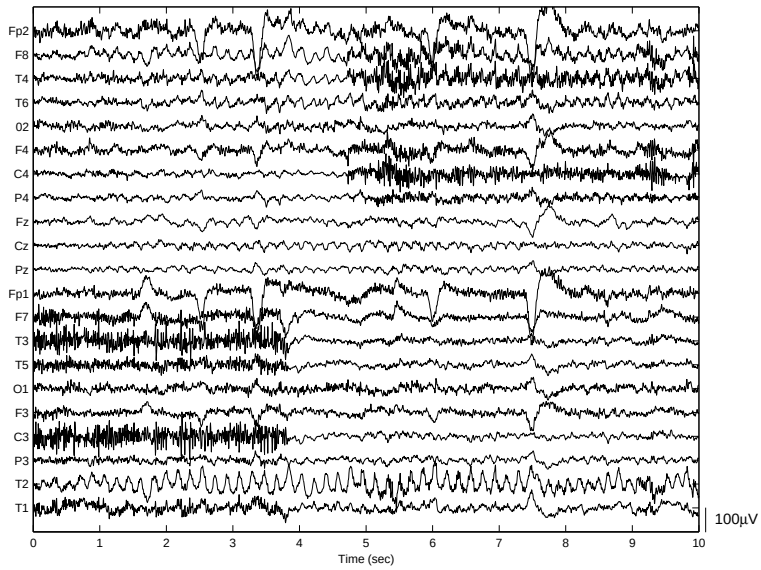
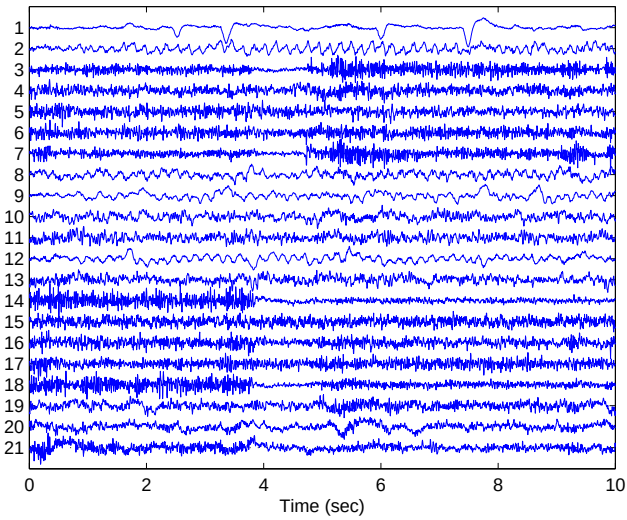


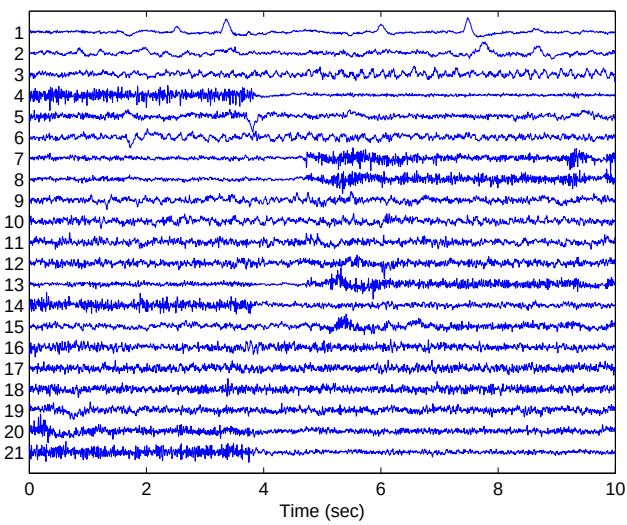
Figure 5.1: *10 sec of an EEG recording from a long-term EMU. The eye artifact is mainly visible on the $Fp1$ and $Fp2$ electrodes.*

EOG artifacts have to be removed for several reasons. First, they can obscure the EEG at the time of seizure onset. This can be problematic in the setting of a preoperative evaluation of patients with refractory epilepsy, since the ictal recordings are crucial for the localisation of the epileptogenic zone. Also in an Event Related Potential (ERP) study, trials contaminated with eye artifacts are not useful. Furthermore, these artifacts are a source of false positives in automatic epileptic seizure detection systems.

Several EOG correction methods have been proposed based on different concepts. Traditionally, EOG artifacts have been corrected using regression-based techniques [57]. These techniques require the EOG signal. Later, BSS methods have been introduced that blindly separated brain signals from artifacts in EEG recordings [237, 178]. These BSS methods are nowadays (semi-)automated and outperform in most comparisons the traditional regression-based methods [352, 303]. These methods do not require the EOG reference signal and can be used in an EMU. In EMU's, the EOG is not recorded simultaneously with the EEG since the EOG electrodes are cumbersome for the patient who is monitored for typically one week.



(a)



(b)

Figure 5.2: *Figure (a) and (b) show the sources estimated with respectively SOBI and JADE. The first source contains the eye blink.*

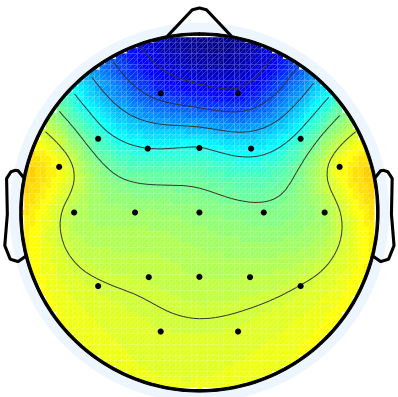
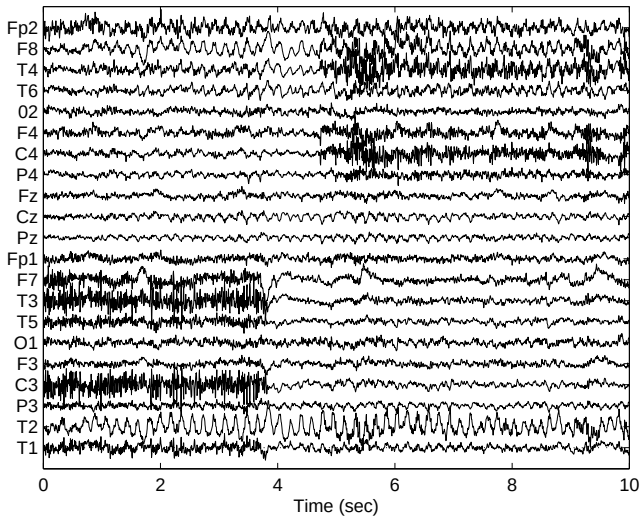


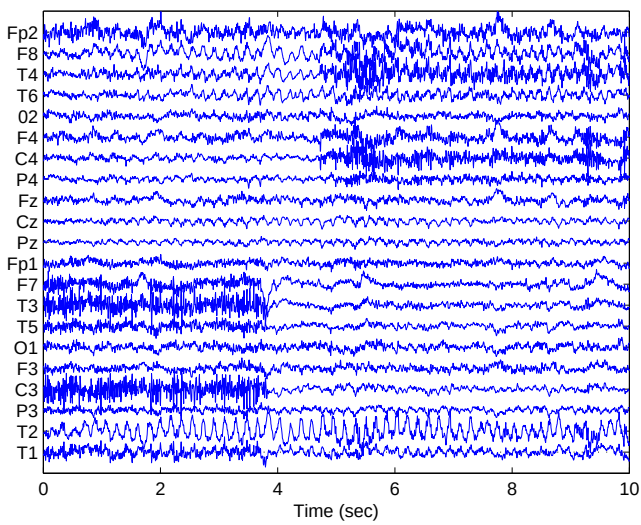
Figure 5.3: *The topography of an eye blink. It can be easily recognised: highest values are in frontal regions, and the topography is very symmetric.*

5.1.2 Illustration

In the EEG of Fig. 1.1, repeated in Fig. 5.1, also eye artifacts were present. How well can they be removed with BSS algorithms? To answer this question, we computed the contributing sources with SOBI and JADE (see section 3.5.2). These sources are visualised in Fig. 5.2. We can clearly recognise the eye-related source as the first estimated source in both decompositions, because this source has clear peaks at the time-instants where an eye-blink occurred. Also the mixing topography (visualised in Fig. 5.3 for the first source of SOBI) nicely corresponds to the expected topography of an eye artifact. It is symmetric over the hemispheres, and has maximal amplitudes in the frontal regions. When we reconstruct the EEG without the artifact, we obtain the EEGs as shown in Fig. 5.4. We can see that the recording, cleaned with SOBI, does not contain residual eye signal on Fp1, while the recording, cleaned with JADE, still does (see *e.g.*, channel Fp2). The EEG epoch contained 2500 datapoints. This is enough to reliably estimate the higher-order cumulants in JADE [177]. The result that SOBI outperforms JADE is in agreement with the literature. It is generally stated that methods exploiting the temporal coloration of the source signals outperform methods based on HOS for the estimation of the eye artifact [135, 303]. Transients, *e.g.*, epileptic spikes, will be well separated with ICA algorithms exploiting HOS [194].



(a)



(b)

Figure 5.4: *Figure (a) and (b) show the EEG without eye blink, by removing the first source estimated by respectively SOBI and JADE.*

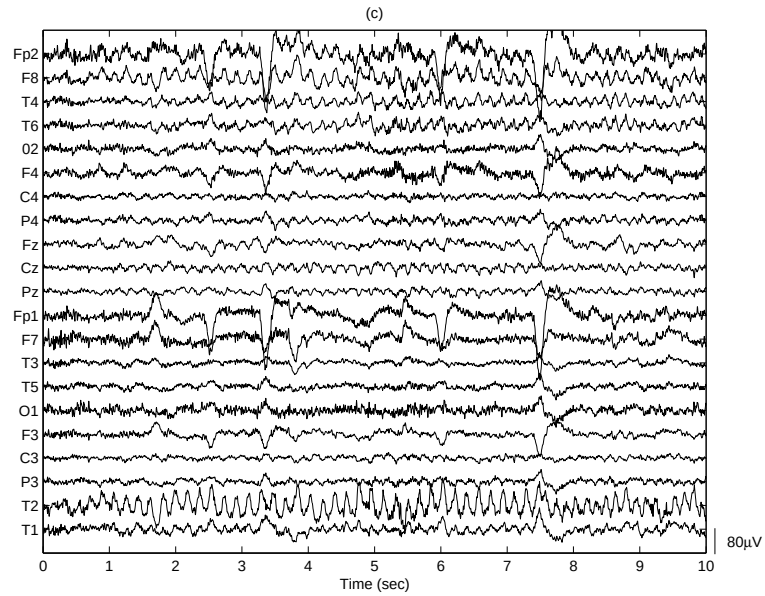


Figure 5.5: *EEG reconstruction based on JADE components without muscle components.*

Although some people claim that it is easy to remove muscle artifacts with ICA algorithms based on HOS, *e.g.*, [56], a more detailed look at the sources estimated with JADE reveals that it is not straightforward. First, on top of the rhythmical sources (3, 6, 9 and 10), there is still muscle activity present. As we mentioned in the previous chapter, BSS-CCA largely outperforms ICA for the separation of muscle artifacts, both in epileptic recordings and in speech-related ERPs. Furthermore, it is hard to determine in ICA which are the muscle related components. By reconstructing the EEG without the (visually determined) muscle-related sources number 4, 7, 8, 13, 14, 16–19, 20 and 21, we obtain the EEG visualised in Fig. 5.5. There is clearly some muscle activity left.

5.1.3 Discussion

The validation of an artifact removal tool is a very important but very difficult task. Often, neither the artifact, nor the underlying brain activity is known. Of course one has to try to validate a new technique carefully. The hypotheses that are made, should match the properties of the sources and the obtained sources should be examined in detail. But at the same time, one should keep in mind the limitations of validation. Although some studies aim at quantifying for example ocular or muscle removal tools, they are limited to synthetic data. When a linear

mixture of two independent artificial signals is simulated, it can be separated with BSS. However, this does not guarantee that the BSS method will also reliably separate *in vivo* EEG data. We illustrated our experience with BSS by means of just one representative example. This example reflects the experience gained through the years by dealing with all kinds of different *in vivo* EEGs. In the former chapter, BSS-CCA successfully removed muscle artifacts from the EEG recording. In this chapter, SOBI successfully removed eye artifacts. With *succesfully removed*, we mean that no artifact was left and no visual difference could be observed in the rest of the signal.

Recently, it has also been questioned whether brain activity can really be separated into independent brain sources, as all the neurons are connected with each other [225]. In practice, the estimated sources will be as independent as possible, and BSS is a useful tool to remove artifacts and investigate individual brain processes. Certainly the eye artifacts are independent from brain activity, and can in most cases be well separated. For an exception, see section 5.5.
Nowadays, brain neurophysiology is very interested in brain networks [240]. Such networks can perhaps be modeled with Independent Subspace Analysis (ISA) [65, 33], identifying groups of dependent sources. However, more research has to be done to investigate this path.

5.2 Imposing prior knowledge to ICA

Standard ICA has two limitations, which form the starting point in this section. First, the extracted components are not ordered. When someone is only looking for the presence of one component, all the extracted components have to be visually inspected. Usually, complex features have to be computed in order to recognize the signal of interest. Second, in some applications, there is prior knowledge about the source signals being sought. ICA methods are black box techniques that do not exploit this available prior knowledge.

In order to improve the performance, several attempts have been made to incorporate the available signal characteristics and develop "semi-blind" source separation techniques. If one has information on the statistical distributions, one could incorporate prior knowledge in a Bayesian framework, see e.g. [322, 191, 234]. This has the disadvantage of having to specify all priors for the distribution of the model parameters. Furthermore, no closed form equations can be derived. If the source distributions are not fully known, but it is known that some components have large second- or higher-order statistics, a so-called "robust BSS approach" has been introduced in [305]. (This should not be confused with "robust ICA" introduced in [362].) When only kurtosis signs are known, a BSS algorithm is proposed in [363]. When information about the shape of one or more source

signals is assumed, this prior knowledge can be used as a constraint in the ICA optimisation algorithm, leading to the introduction of temporally constrained ICA [170]. Following this idea, in e.g. [229, 175, 31] other optimisation algorithms are proposed to extract sources resembling a reference signal.

Besides trying to use information about the source signals, a second way to constrain ICA is to impose constraints on the columns of the mixing matrix. This is useful when the spatial distribution of one (or more) source(s) is (approximately) known. This is called spatially constrained ICA (scICA). In [151], a first algorithm was proposed. At every iteration step, the BSS criterion was optimised, and afterwards the spatial constraint was enforced. [151] also contains a thorough discussion on the difference between hard and soft constraints and on the trade-off between imposing prior knowledge constraints and statistical independence constraints. Indeed, in some situations inaccurate prior knowledge can conflict with assumptions of statistical independence. In [216], it is mentioned (without giving details) that spatial constraints can be incorporated with Lagrange multipliers to solve an unconstrained problem. In [165], another spatially constrained algorithm is mentioned.

This type of ICA with prior knowledge is potentially very important for applications. For instance, in wireless communications, the known mixing vector could correspond to a transmitter of which the position is known not to change from one transmitted block to the other. Below, we present a biomedical application.

If a mixing vector is known, it may be tempting to work by deflation. That is, one could project the data on the known vector and continue with the residu. This approach is incorrect. It implies that the remaining mixing vectors lie in the orthogonal complement of the known mixing vector, which is not necessarily true. If the mixing matrix is orthogonal, *i.e.*, after a pre-whitening, then the greedy approach may work. However, pre-whitening in the case where a mixing vector is known, is still an open problem, a fortiori when the problem is strictly overdetermined. In the next sections, we explain how this problem can be solved. We propose a new algorithm to incorporate prior knowledge into the popular ICA algorithms JADE [37] and SOBI [18]. The algorithm assumes that there is information about one or more mixing vectors. We first show how the prior knowledge should be taken into account in the pre-whitening step. Next, we show how the prior knowledge should be taken into account in the separation of the pre-whitened data. We derive detailed closed-form formulas based on numerical algebra. We also evaluate with a simulation study the improvement in performance that can be obtained by incorporating prior knowledge.

The rest of this chapter is organized as follows. The necessary background on ICA algorithms, in particular JADE and SOBI algorithms, was already given in chapter 3. In section 5.3, we develop our ideas about scICA. Two variants are discussed. A first algorithm computes the ICA decomposition when a mixing vector is exactly known (section 5.3.1). Then, we explain how to generalize the algorithm for the case where more mixing vectors are known. We present in section 5.3.4 a second algorithm, with a weighted constraint on one mixing vector. The next section validates the methods with fixed constraints in a simulation study (section 5.4). We illustrate this technique by means of a real-life application, namely accurate eye artifact removal from EEG data (section 5.5). We conclude the chapter with a discussion.

5.3 Imposing spatial constraints

Remark: In this section, we make an exception on our notation. Vectors will be noted by bold capitals ($A, B, \ldots$).

There are two ways to incorporate prior knowledge about a mixing vector during ICA computation. Either we know a mixing vector exactly or we only know it approximately. In the former case, the extracted mixing matrix should have the given mixing vector as one of its columns. This will be discussed in section 5.3.1. In the latter case, we can weigh the degree of confidence in the given mixing vector against the independence constraint. The use of prior knowledge can then help the ICA algorithm to converge towards the desired solution. This will be discussed in section 5.3.4.

5.3.1 The mixing vector of one source is exactly known.

We start from the model in section 3.18:

$$\mathbf{Y} = \mathbf{M} \cdot \mathbf{X}. \tag{5.1}$$

However, the formulation of the rest of this chapter is in terms of complex numbers. We assume that the observations are I-dimensional, that there are P sources and that we know exactly one mixing vector K. We first explain how we make sure that the mixing vector is fully contained in the signal subspace. This step is necessary for reducing the dimensionality of the problem. Then we conduct a full pre-whitening.In the next step, we estimate the full mixing matrix .

Dimensionality reduction

We have to determine the subspace of dimension P that includes K and explains as much of the observed data as possible. Formally, we have to find a matrix $\tilde{\mathbf{U}} \in \mathbb{C}^{I \times P}$ with mutually orthonormal columns that maximises $||\tilde{\mathbf{U}}^H \mathbf{Y}||$ under the constraint that $K \in span(\tilde{\mathbf{U}})$. This can be done as follows:

1. Normalise K to unit length. The result will be the first column of $\tilde{\mathbf{U}}$, denoted by $\tilde{U}_1$. In this way, $K \in span(\tilde{\mathbf{U}})$.
2. Project the observation matrix $\mathbf{Y}$ on the orthogonal complement of $\tilde{U}_1$ in order to obtain $\mathbf{Y}_2$. We have $\mathbf{Y}_2 = (\mathbf{I} - \tilde{U}_1 \tilde{U}_1^H) \cdot \mathbf{Y}$.
3. Define $\tilde{U}_2$ to $\tilde{U}_P$ as the $P-1$ dominant left singular vectors of $\mathbf{Y}_2$. These vectors allow us to maximise $||\tilde{\mathbf{U}}^H \mathbf{Y}||$.

In order to reduce the dimensionality, we now project the data matrix on the subspace that we have formed:

$$\mathbf{Y}_s = \tilde{\mathbf{U}}^H \cdot \mathbf{Y}, \tag{5.2}$$

where $\mathbf{Y}_s \in \mathbb{C}^{P \times J}$. (The subscript refers to "signal subspace".) The mixing matrix in the P-dimensional space is denoted by $\mathbf{M}_s$. We have

$$\mathbf{M} = \tilde{\mathbf{U}} \cdot \mathbf{M}_s. \tag{5.3}$$

Pre-whitening

After the dimensionality reduction step, we know the column space of the mixing matrix. This is the column space of $\tilde{\mathbf{U}}$. After projection on this subspace, we solve the square $(P \times P)$ ICA problem. We start with a pre-whitening:

$$\mathbf{C}_{\mathbf{Y}_s} = \mathbf{Y}_s \cdot \mathbf{Y}_s^H \tag{5.4}$$

$$= \mathbf{M}_s \cdot \mathbf{C}_{\mathbf{X}} \cdot \mathbf{M}_s^H \tag{5.5}$$

$$= (\mathbf{QS}) \cdot (\mathbf{QS})^H, \tag{5.6}$$

in which the last equation shows the Eigen Value Decomposition (EVD) of $\mathbf{C}_{\mathbf{Y}_s}$. The singular value decomposition of the mixing matrix after dimensionality reduction is given by:

$$\mathbf{M}_s = \mathbf{QSV}^H, \tag{5.7}$$

in which the orthogonal matrix $\mathbf{V}$ is still unknown. $\mathbf{Q}$ and $\mathbf{S}$ are obtained from the EVD (5.6). The overall pre-whitening matrix $\mathbf{M}_w \in \mathbb{C}^{I\times P}$ that takes into account the spatial prior knowledge, is given by

$$\mathbf{M}_w = \tilde{\mathbf{U}} \cdot \mathbf{QS}. \tag{5.8}$$

The pre-whitened data are denoted by

$$\mathbf{Y}_w = \mathbf{M}_w^{\dagger} \cdot \mathbf{Y} = \mathbf{S}^{\dagger}\mathbf{Q}^H\tilde{\mathbf{U}}^H \cdot \mathbf{Y} \in \mathbb{C}^{P\times J}. \tag{5.9}$$

The pre-whitened observations are related to the sources as follows:

$$\mathbf{Y}_w = \mathbf{V}^H \cdot \mathbf{X}. \tag{5.10}$$

Determination of V using prior knowledge

The orthogonal matrix $\mathbf{V} \in \mathbb{C}^{P\times P}$ can be determined as follows in order to estimate the full mixing matrix:

1. We know the mixing vector in the original I-dimensional space, K, and its normalised variant $\tilde{U}_1$. Consequently, we know the first column of $\mathbf{M}_s$ after dimensionality reduction, which we will denote as $M_{s,1}$. From equation (5.7) we then also know the first column of $\mathbf{V}$, denoted as V_1. The procedure is as follows. From equations (3.18) and (5.2), we have that $M_{s,1} = \tilde{\mathbf{U}}^H\tilde{U}_1 = E_1$, in which E_1 is the first canonical unit vector with 1 as its first entry and 0 as the other entries. From equation (5.7), we have that

   $$M_{s,1} = \mathbf{QS} \cdot V_1 \tag{5.11}$$

   in which V_1 is the first column of $\mathbf{V}^H$. Consequently,

   $$V_1 = \mathbf{S}^{\dagger}\mathbf{Q}^H \cdot M_{s,1}. \tag{5.12}$$

   V_1 is subsequently normalised to unit length, which we can do because of the ICA scaling ambiguity.

2. The matrix $\mathbf{V}$ is orthogonal. By definition, its last $(P-1)$ columns $V_2, \ldots, V_P$ are orthogonal to V_1. Let $\mathbf{V}_1^{\perp} \in \mathbb{C}^{P\times(P-1)}$ denote an arbitrary column-wise orthonormal matrix of which the columns form an orthonormal basis for the orthogonal complement of V_1. Then we have $[V_2 \ldots V_P] = \mathbf{V}_1^{\perp} \cdot \mathbf{W}$, in which $\mathbf{W} \in \mathbb{C}^{(P-1)\times(P-1)}$ is an orthogonal matrix. Our goal is the estimation of $\mathbf{W}$. Let $\tilde{\mathbf{X}} \in \mathbb{C}^{(P-1)\times J}$ be the matrix obtained from $\mathbf{X}$ after dropping the first row, i.e., $\tilde{\mathbf{X}}$ contains all the sources except from the one of which the

mixing vector is known. Projection of $\mathbf{Y}_w$ on the orthogonal complement of V_1 now yields:

$$\tilde{\mathbf{Y}}_w = (\mathbf{V}_1^{\perp})^H \mathbf{Y}_w = \mathbf{W} \cdot \tilde{\mathbf{X}}. \tag{5.13}$$

Equation (5.13) is a classical ICA problem after pre-whitening, without constraints on the mixing matrix due to prior knowledge. This problem may be solved by means of, e.g., standard JADE or SOBI.

Summarizing our derivation, our estimate of the mixing matrix $\mathbf{M}$ is given by

$$\hat{\mathbf{M}} = \tilde{\mathbf{U}} \cdot \mathbf{QS} \cdot [V_1 \, \mathbf{V}_1^{\perp}] \cdot \begin{bmatrix} 1 & 0 \\ 0 & \mathbf{W} \end{bmatrix}. \tag{5.14}$$

5.3.2 The mixing vectors of more sources are exactly known.

We now assume that L vectors are known.

Dimensionality reduction

The first step in section 5.3.1 is replaced by the computation of an orthonormal basis of the span of the L known vectors. In the second step, we project on the orthogonal complement of this basis. In the third step, we extract $P - L$ singular vectors. This yields a column-wise orthonormal $\tilde{\mathbf{U}} \in \mathbb{C}^{I \times P}$ which we use for dimensionality reduction as in (5.3).

Pre-whitening

The pre-whitening can proceed as in section 5.3.1. Equation (5.7) remains valid.

Determination of V using prior knowledge

Like in equation (5.12), we have that

$$V_l = \mathbf{S}^{\dagger} \mathbf{Q}^H M_{s,l} \qquad l = 1, \ldots, L. \tag{5.15}$$

If these vectors are not orthogonal, then we have a conflict between the spatial prior knowledge and the assumption of statistical independence. We distinguish between two cases.

A. We can assume that the spatial prior knowledge is reliable and hence, that the vectors V_l have been correctly estimated. Since we want to stay as closely as possible to the situation where $\mathbf{V}$ is orthogonal, we will determine $V_{L+1}, \ldots, V_P$ such that they are mutually orthonormal and orthogonal to $\text{span}([V_1 \ldots V_L])$. Let $\mathbf{V}_P^{\perp} \in \mathbb{C}^{P \times (P-L)}$ denote an arbitrary column-wise orthonormal matrix of which the columns form an orthonormal basis for the orthogonal complement of $[V_1 \ldots V_L]$. Then we can proceed as in section 5.3.1 and reduce the problem to an unconstrained $((P-L) \times (P-L))$ ICA problem.

B. We can assume that the statistical independence constraint is more reliable than the prior knowledge. However, we try to stay as closely as possible to the known vectors. Consequently, we replace the vectors $V_l, l = 1, \ldots, L$, by column-wise orthonormal vectors $Z_l, l = 1, \ldots, L$, that are as close as possible in the least squares sense, up to an irrelevant scaling. We stack the vectors V_l in $\mathbf{V}_{1:L} \in \mathbb{C}^{P \times L}$. Formally, we try to estimate the column-wise orthonormal matrix $\mathbf{Z} \in \mathbb{C}^{P \times L}$ and the diagonal matrix $\mathbf{D}$ that minimise

$$||\mathbf{V}_{1:L} \cdot \mathbf{D} - \mathbf{Z}||^2.$$

This can be done as follows [314]:

1. Normalise the columns of $\mathbf{V}_{1:L}$ to unit length.
2. Iterate until convergence:
   a) The computation of $\mathbf{Z}$, given $\mathbf{D}$, is an orthogonal Procrustes problem [125]. This means that, if the SVD of $\mathbf{V}_{1:L} \cdot \mathbf{D}$ is given by $\mathbf{ACB}^H$, then the optimal $\mathbf{Z}$ is given by $\mathbf{AB}^H$.
   b) The computation of $\mathbf{D}$, given $\mathbf{Z}$, is a standard linear least squares problem. We have: $\mathbf{D} = (diag(\mathbf{Z}^H \cdot \mathbf{V}_{1:L}))^{-1}$.

Then we can proceed as in section 5.3.1 and reduce the problem to an unconstrained $((P-L) \times (P-L))$ ICA problem.

The algorithm is summarized in Algorithm 2.

5.3.3 The mixing vector of one source is known with a limited accuracy.

5.3.4 Weighted constraint

We derive a second algorithm that can be used when one disposes of an estimate of a mixing vector K, that has however limited accuracy. In this section, we limit ourselves to real-valued data.

Algorithm 2 scICA, Case: mixture $\mathbf{Y}$ containing P sources of which L mixing vectors M_l are known.

1: Compute an orthonormal basis of the known mixing vectors and stack it in $\tilde{\mathbf{U}}_{1:L}$.
2: Project the data on the orthogonal complement of this basis. $\mathbf{Y}_2 = (\mathbf{I} - \tilde{\mathbf{U}}_{1:L}\tilde{\mathbf{U}}_{1:L}^H) \cdot \mathbf{Y}$.
3: Compute the $P - L$ dominant left singular vectors of $\mathbf{Y}_2$ and stack them in $\tilde{\mathbf{U}}_{(L+1):P}$.
4: Project the data on $\tilde{\mathbf{U}} = [\tilde{\mathbf{U}}_{1:L}\ \tilde{\mathbf{U}}_{(L+1):P}] : \mathbf{Y}_s = \tilde{\mathbf{U}}^H \cdot \mathbf{Y}$.
5: Whiten the data: $\mathbf{Y}_s = \mathbf{QSV}^H \Rightarrow \mathbf{Y}_w = \mathbf{S}^\dagger\mathbf{Q}^H \cdot \mathbf{Y}_s$.
6: Whiten the known vectors: $V_l = \mathbf{S}^\dagger\mathbf{Q}^H\tilde{\mathbf{U}}^H \cdot M_l$, $l = 1, \ldots, L$.
7: Normalise V_l to unit length, $l = 1, \ldots, L$.
8: If the vectors V_l should be orthonormal, then iterate until convergence:

a) $\mathbf{V}_{1:L} \cdot \mathbf{D} = \mathbf{ACB}^H \Rightarrow \mathbf{Z} = \mathbf{AB}^H$.

b) $\mathbf{D} = (diag(\mathbf{Z}^H \cdot \mathbf{V}_{1:L}))^{-1}$.

9: Compute the orthogonal complement $\mathbf{V}_{1:L}^\perp$ of $\mathbf{V}_{1:L}$ and project $\mathbf{Y}_w$ on it: $\tilde{\mathbf{Y}}_w = (\mathbf{V}_{1:L}^\perp)^H\mathbf{Y}_w$.
10: Solve a classical $((P - L) \times (P - L))$ ICA problem: $\tilde{\mathbf{Y}}_w = \mathbf{W} \cdot \tilde{\mathbf{X}}$.
11: The mixing matrix can be estimated as:
$\hat{\mathbf{M}} = \tilde{\mathbf{U}} \cdot \mathbf{QS} \cdot [V_{1:L}\ \mathbf{V}_{1:L}^\perp] \cdot \begin{bmatrix} \mathbf{I}_{L\times L} & 0 \\ 0 & \mathbf{W} \end{bmatrix}$.
12: The sources can be estimated as $\hat{\mathbf{X}} = (\hat{\mathbf{M}})^\dagger \cdot \mathbf{Y}$.

Pre-whitening step and dimensionality reduction

Similarly to §5.3.1, we have to define a signal subspace of dimension P. This time, we are only up to a certain level confident that the subspace fully includes K. In the definition of $\tilde{\mathbf{U}}$, we have to find a trade-off between maximizing the norm of the projection of the data on $\tilde{\mathbf{U}}$ and the constraint that K is contained in span($\tilde{\mathbf{U}}$). The following procedure fulfils our objectives:

1. Compute the covariance matrix of the observation matrix:

$$\begin{aligned}\mathbf{C}_2^Y &= \frac{1}{J}\mathbf{Y} \cdot \mathbf{Y}^T \\ &= \mathbf{M} \cdot \mathbf{C}_2^X \cdot \mathbf{M}^T \\ &= (\mathbf{US}) \cdot (\mathbf{US})^T.\end{aligned}$$

2. Keep the $J-1$ eigenvectors $[U_1, \ldots, U_{J-1}]$ corresponding to the $J-1$ largest eigenvalues. We call the resulting matrix $\tilde{\mathbf{U}}_{1..J-1} \in \mathbb{R}^{1\times(J-1)}$

3. Compute the component of K orthogonal to this subspace as follows: $\tilde{K^o} = (\mathbf{I} - \tilde{\mathbf{U}}_{1:J-1}\tilde{\mathbf{U}}_{1:J-1}^T)K$. We normalize this vector to unit length in order to obtain K^o.

4. The Jth component $\tilde{U}_J$ is then defined by a weighted average of K^o and U_J: $\tilde{U}_J = K^o \cos\alpha_1 + U_J \sin\alpha_1$. The weighting factor takes into account

the confidence in the accuracy of K on one hand and the confidence in the Jth eigenvector, on the other hand. A small value of the angle α_1 expresses great confidence in the vector K. A value of α_1 close to $\frac{\pi}{2}$ expresses great confidence in the dominant eigenspace of $\mathbf{C}_2^Y$. Since both K^o and U_J are orthogonal to $span(\tilde{\mathbf{U}}_{1..J-1})$, $\tilde{U}_J$ is also orthogonal to $span(\tilde{\mathbf{U}}_{1..J-1})$.

5. We construct the column-wise orthonormal matrix $\tilde{\mathbf{U}} = \left[U_1, \ldots, U_{J-1}, \frac{\tilde{U}}{||\tilde{U}_J||}\right] \in \mathbb{R}^{I \times J}$.

We prove that the matrix $\tilde{\mathbf{U}}$ diagonalises $\mathbf{C}_2^Y$. We know that

$$\mathbf{U}^T \mathbf{C}_2^Y \mathbf{U} = \mathbf{S}^2. \tag{5.16}$$

Since $\tilde{U}_J$ is in the orthogonal complement of $U_1, \ldots, U_{J-1}$, it is a linear combination of $U_J, \ldots, U_I$. Hence, we have:

$$\tilde{\mathbf{U}} = \mathbf{U} \begin{bmatrix} 1 & 0 & & \\ 0 & 1 & 0 & \\ & & \ddots & \\ & & 0 & c_1 \\ & & 0 & c_2 \\ & & 0 & \vdots \end{bmatrix}. \tag{5.17}$$

Substitution of (5.17) in (5.16) yields:

$$\tilde{\mathbf{U}}^T \mathbf{C}_2^Y \tilde{\mathbf{U}} = \begin{bmatrix} 1 & 0 & & & & \\ 0 & 1 & 0 & & & \\ & & \ddots & & & \\ & & 0 & c_1 & c_2 & \ldots \end{bmatrix} \cdot \mathbf{S}^2 \cdot \begin{bmatrix} 1 & 0 & & \\ 0 & 1 & 0 & \\ & & \ddots & \\ & & 0 & c_1 \\ & & 0 & c_2 \\ & & 0 & \vdots \end{bmatrix} \tag{5.18}$$

$$= \tilde{\mathbf{S}}^2. \tag{5.19}$$

in which the latter matrix is diagonal. The overall pre-whitening matrix $\mathbf{M}_w \in \mathbb{R}^{I \times P}$ that takes into account the spatial prior knowledge, is given by $\mathbf{M}_w = \tilde{\mathbf{U}}\tilde{\mathbf{S}}$. The prewhitened data are denoted by

$$\mathbf{Y}_w = (\mathbf{M}_w)^\dagger \mathbf{Y} = \tilde{\mathbf{S}}^\dagger \tilde{\mathbf{U}}^T \mathbf{Y} \in \mathbb{R}^{P \times J}. \tag{5.20}$$

The pre-whitened data are related to the sources a follows:

$$\mathbf{Y}_w = \mathbf{V}^T \mathbf{X} \tag{5.21}$$

in which $\mathbf{V} \in \mathbb{R}^{P \times P}$ is orthogonal.

Determination of V using prior knowledge

Recall that JADE and SOBI rely on a Jacobi iteration. In the ICA decomposition with a weighted spatial constraint, we will still compute $\mathbf{U}$ by means of a Jacobi iteration. Now, we derive how the angle θ of the Jacobi matrix $\mathbf{J}$ can be computed such that the spatial constraint is taken into account. We impose that the overall first mixing vector is similar to K. The first vector of the pre-whitened mixing matrix should be similar to the vector K_w, defined by

$$K_w = \mathbf{M}_w^{\dagger} \cdot K. \tag{5.22}$$

We normalise this K_w. For a Jacobi iteration acting on sources and mixing vectors i and j with $i \neq 1$, the spatial constraint does not play a role. The problem is identical to the JADE/SOBI core problem and equation (3.32) defines the optimal rotation. When $i = 1$, we add a term to the contrast function that expresses that the estimated pre-whitened first mixing vector should become close to K_w.

Formally, we replace (3.32) by

$$f = \sum_{p}^{P} \tilde{a}_{iip}^2 + \tilde{a}_{jjp}^2 + \alpha_2 (K_w^T (\widetilde{M}_{new})_1)^2 \tag{5.23}$$

with $\alpha_2 > 0$ a weighting factor and $(\widetilde{M}_{new})_1$ the updated version of the first pre-whitened mixing vector. Vector $(\widetilde{M}_{new})_1$ is the first column of the product of the current and all the previous Jacobi matrices. A large value of α_2 expresses a great confidence in the value of K_w.

Consider an iteration step in which a Jacobi matrix is computed that works on i=1 and j. Let the current estimate of the pre-whitened mixing matrix be denoted by $\tilde{\mathbf{M}}_{old}$ and the updated estimate by $\tilde{\mathbf{M}}_{new}$. $\tilde{\mathbf{M}}_{new}$ and $\tilde{\mathbf{M}}_{old}$ are related by:

$$\tilde{\mathbf{M}}_{new} = \tilde{\mathbf{M}}_{old} \cdot \mathbf{J}^T. \tag{5.24}$$

The first column of the updated mixing matrix $\tilde{\mathbf{M}}_{new}$ can be written as:

$$(\widetilde{M}_{new})_1 = s \cdot (\widetilde{M}_{old})_j + c \cdot (\widetilde{M}_{old})_1. \tag{5.25}$$

We examine how the second term in (5.23) depends on the Jacobi angle θ.

We have:

$$\alpha_2 (K_s^T (\widetilde{M}_{new})_1)^2 = \alpha_2 \left[K_s^T (s \cdot (\widetilde{M}_{old})_j + c \cdot (\widetilde{M}_{old})_1) \right]^2 \tag{5.26}$$

$$= \alpha_2 \left[s^2 (K_s^T (\widetilde{M}_{old})_j)^2 + 2cs (K_s^T (\widetilde{M}_{old})_j K_s^T (\widetilde{M}_{old})_1) + c^2 (K_s^T (\widetilde{M}_{old})_1)^2 \right] \tag{5.27}$$

Let us define

$$f_1 = \alpha_2((K_s^T(\widetilde{M}_{old})_j)^2,$$

$$f_2 = \alpha_2(K_s^T(\widetilde{M}_{old})_1)^2$$

and
$$f_3 = 2\alpha_2(K_s^T(\widetilde{M}_{old})_j K_s^T(\widetilde{M}_{old})_1), \tag{5.28}$$

Then (5.27) can be written as:

$$\alpha_2(K_s^T(\widetilde{M}_{new})_1)^2 = f_1\cos^2(\theta) + f_2\sin^2(\theta) + f_3\cos(\theta)\sin(\theta). \tag{5.29}$$

With the substitutions

$$\cos(2\theta) = \cos^2(\theta) - \sin^2(\theta), \tag{5.30}$$

$$\sin(2\theta) = 2\cos(\theta)\sin(\theta), \tag{5.31}$$

and
$$1 = \cos^2(\theta) + \sin^2(\theta), \tag{5.32}$$

expression (5.23) can be rewritten as a homogeneous polynomial in $\cos(\theta)$ and $\sin(\theta)$:

$$\begin{aligned} f = {} & g_{11}(\cos^4(\theta) + \sin^4(\theta) - 2\cos^2(\theta)\sin^2(\theta)) \\ & +(2g_{12} + 2g_{21})(\sin(\theta)\cos^3(\theta) - \sin^3(\theta)\cos(\theta)) \\ & +4g_{22}\sin^2(\theta)cos^2(\theta) + f_1(\cos^4(\theta) + \cos^2(\theta)\sin^2(\theta)) \\ & +f_2(\sin^4(\theta) + \cos^2(\theta)\sin^2(\theta)) \\ & +f_3(\cos^3(\theta)\sin(\theta) + \cos(\theta)\sin^3(\theta)), \end{aligned}$$

in which g_{ij} is the (i,j)th entry of $\mathbf{G}$ in (3.38). In order to compute the θ that maximizes this contrast function, we set its derivative with respect to θ to zero:

$$\begin{aligned} & \cos^4(\theta)(2g_{21} + 2g_{12} + f_3) + \cos^3(\theta)\sin(\theta)(8g_{22} - 8g_{11} - 2f_1 + 2f_2) \\ & + \cos^2(\theta)\sin^2(\theta)(-12g_{12} - 12g_{21}) + \cos(\theta)\sin^3(\theta)(8g_{11} + 2f_2 - 2f_1 - 8g_{22}) \\ & + \sin^4(\theta)(-f_3 + 2g_{12} + 2g_{21}) = 0. \end{aligned} \tag{5.33}$$

Dividing this homogeneous equation by $\cos^4(\theta)$ yields a polynomial of degree 4 in $\tan(\theta)$. Selecting the real root that maximises (5.23) yields the optimal Jacobi rotation angle θ.

5.4 Simulation study

In this section, we show the performance of the proposed algorithm with fixed constraint(s) for three synthetic datasets and compare them with standard ICA (JADE).
The independent sources are drawn from a binary distribution (1 or -1), with an equal probability of both values. These sources form the rows of $\mathbf{X}$. We consider 600 samples of each source. Observation matrices $\tilde{\mathbf{Y}} \in \mathbb{R}^{6\times 600}$ are generated in the following way:

$$\tilde{\mathbf{Y}} = \frac{\mathbf{Y}}{||\mathbf{Y}||_F} + \sigma_N \frac{\mathbf{N}}{||\mathbf{N}||_F}, \tag{5.34}$$

in which $\mathbf{Y} = \mathbf{M} \cdot \mathbf{X}$ exactly satisfies a linear mixture model with $P = 4$ independent sources and in which $\mathbf{N}$ represents zero-mean gaussian noise.

We conducted Monte Carlo simulations consisting of 250 runs. We evaluated the performance of the different algorithms by means of the Bit Error Rate (BER). The rows of the estimated sources $\tilde{\mathbf{X}}$ were optimally ordered and scaled.

We considered three datasets. The entries from the mixing matrix $\mathbf{M}$ are drawn from a zero-mean unit-variance gaussian distribution. White gaussian noise was added. In the first dataset (Fig. 5.6), we gave all sources equal contributions in the mixture and we compared standard ("blind") JADE to scICA with 1, 3 and 4 mixing vectors perfectly known. The figure shows that having perfect knowledge of one mixing vector improves the estimation of the sources compared to standard JADE. The knowledge of adding other *fixed* mixing vectors (case A in section 5.3.2) does not improve estimation further. The figure also illustrates that having perfect knowledge of the full mixing matrix, does not imply the best source estimation. This ambiguity between accurate mixing matrix estimation and accurate source separation was already noticed in [213]. However, by having knowledge about more mixing vectors but giving more importance to the statistical independence (case B in section 5.3.2), even a better source estimation can be obtained.

In the second simulation, we generated mixtures in the same way, but we gave the first source a 10 times weaker contribution. This situation corresponds to a mixing matrix with a higher condition number. The prior knowledge concerned the weak source. From Fig. 5.7, it can be seen that knowing the mixing vector improves the estimation of the sources a lot in this case. At high noise levels, some sources are not well estimated by standard ICA. By using prior knowledge, these sources can be estimated much more accurately.

In the third simulation (Fig. 5.8), we consider mixtures $\in \mathbb{R}^{8\times 600}$ containing 5 sources. Colored noise was generated by multiplying white noise with a matrix containing elements drawn from a gaussian distribution. This noise was added to the mixture. The figure illustrates that knowing 1 source improves the estimation

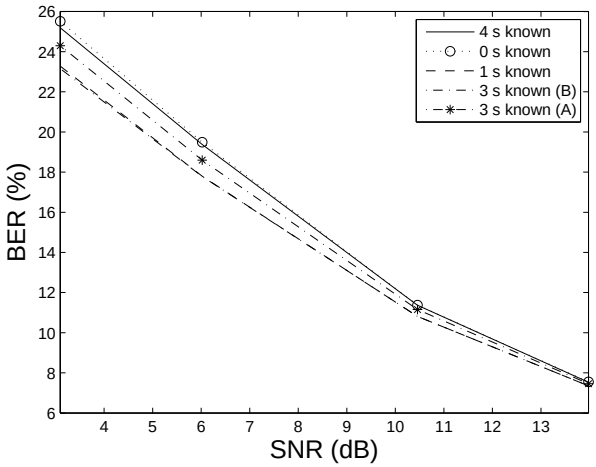


Figure 5.6: *The mean value of BER as a function of the SNR (dB) for unconstrained JADE (dotted), when the mixing vector of 1 source is known (dashed), when the mixing vector of 3 sources are known (dash-dot)(A and B refer to option A and B), and when the mixing matrix is fully known (solid). All sources have the same variance.*

compared to blind JADE. Knowing more mixing vectors exactly can further improve the estimation, even with fixing the known mixing vectors (case A in section 5.3.2). This is partially due to a better estimation of the signal subspace.

These simulations show that the accuracy can certainly be increased by incorporating spatial prior knowledge. Furthermore, as we mentioned in the introduction, the components obtained from standard ICA are not ordered and have to be inspected afterwards. ICA with prior knowledge puts the signal(s) of interest in the first component(s). So postprocessing is not needed anymore, at least when the components of interest correspond to the components for which prior knowledge is available.

5.5 Illustration of scICA for eye artifact removal

We now show an application of scICA in neuroscience. We illustrate its use by means of the removal of eye blink artifacts from EEG recordings. The spatial distribution of eye blinks can be supposed known from other ICA decompositions. We concatenated five EEG epochs of rest EEG contaminated with blinks from the

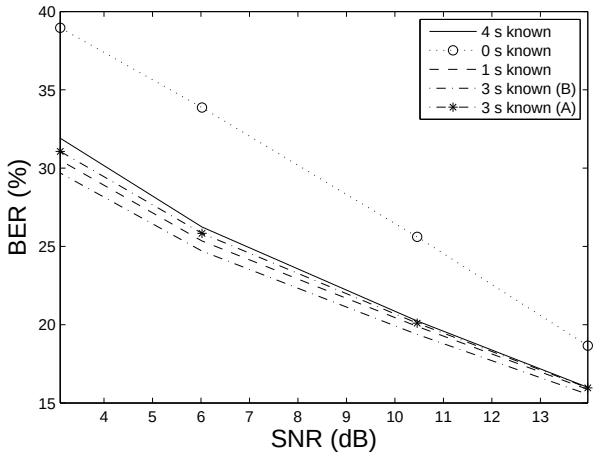


Figure 5.7: *The mean value of BER as a function of the SNR (dB) for unconstrained JADE (dotted), when the mixing vector of 1 source is known (dashed), when the mixing vector of 3 sources are known (dash-dot)(A and B refer to option A and B), and when the mixing matrix is fully known (solid). The one source with known mixing vector has a 10 times smaller variance than the sources with unknown mixing vector.*

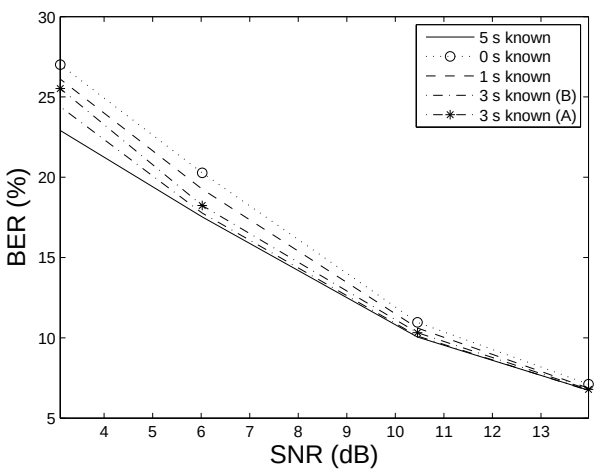


Figure 5.8: *The mean value of BER as a function of the SNR (dB) for unconstrained JADE (dotted), when the mixing vector of 1 source is known (dashed), when the mixing vector of 3 sources is known (dash-dot) (A and B refer to option A and B), and when the mixing matrix is fully known (solid). The noise was colored.*

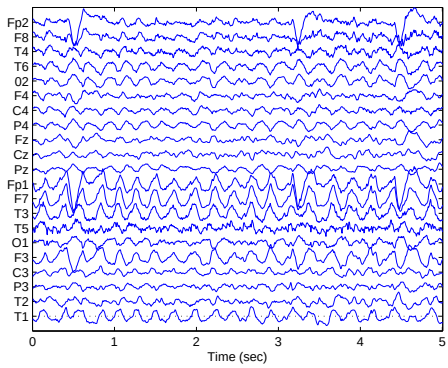


Figure 5.9: *The original EEG containing an eye blink around time 0.5, 3.3 and 4.5 s.*

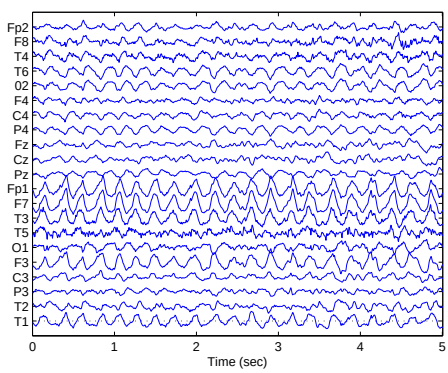


Figure 5.10: *The EEG after eye artifact removal with scICA.*

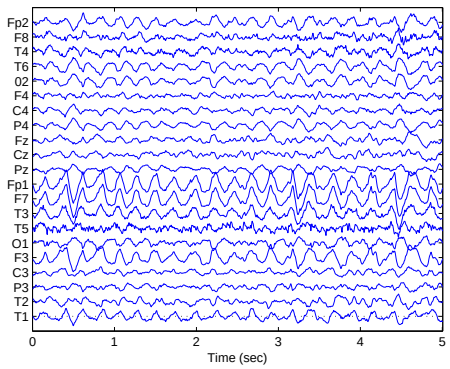


Figure 5.11: *The EEG after eye artifact removal with SOBI.*

same person. On this dataset, we computed standard ICA (SOBI). One source of such decomposition will contain only the eye blink (see *e.g.*, [126]). This source and corresponding mixing vector can be easily visually selected. This mixing vector gives a robust estimation of the spatial distribution of an eye blink. There is very little variation between artifacts from one person. So we can use scICA, based on SOBI, to remove other artifacts. In many cases, standard ICA will give a good separation between eye artifacts and brain activity. However, when the spatial distribution of prominent brain activity, *e.g.*, an epileptic seizure, overlaps with the eye blink and the time course of the blink is synchronised with the seizure activity, standard ICA can fail as we illustrate now.

Fig. 5.9 shows the original EEG dataset containing an epileptic seizure (mainly present on electrodes Fp1, F7 and F3) eye blinks around time $= 0.5, 3.3$ and $4.5s$. Fig. 5.10 shows the EEG when the artifact is removed in an automatic way by imposing during the computation of ICA that the first component is related to eye-blinks, and afterwards subtracting this component. Fig. 5.11 shows the EEG, from which the eye-related component, estimated with SOBI, is subtracted. In the latter figure, the blink is not fully removed at channels Fp1, F7 and F3.

5.6 Discussion on ICA with fixed constraint

Standard ICA blindly separates a linear mixture of statistically independent source signals. In several applications, one has prior knowledge about sources and/or mixing vectors. Hence, the development of semi-blind methods that take this prior knowledge into account is of interest. In this chapter, we proposed a new

algorithm that incorporates exact prior knowledge as a constraint: it is a semi-blind version of JADE and SOBI. The method works for both real- and complex-valued data, and an arbitrary number of spatial constraints.

To the best of our knowledge, we are the first to perform a study to assess the improvement in separation by incorporating prior knowledge, compared to the fully blind decomposition. We showed (Fig. 5.6) that in the case of a standard linear mixture, scICA performs better than ordinary ICA. When the mixing matrix is worse conditioned (Fig. 5.7), scICA largely outperforms standard ICA. This is partially due to a better estimation of the signal subspace. It is also interesting to note that we obtain a better estimate of the sources than when the full mixing matrix is known. In Fig. 5.8, we could also see that in the presence of colored noise, knowing more mixing vectors improves the estimation. Again, this is partially due to a better estimation of the signal subspace.
scICA has the additional advantage over blind ICA that the sources for which spatial information is available, do not appear in a random order.
An interesting application is eye artifact removal from EEG recordings. A full validation study was outside the scope of this study. However, the given example illustrates a limitation of standard ICA, that can be overcome with the incorporation of available prior knowledge. The method can even be further extended. In this study, we assumed that the mixing vector related to the eye blinks was known from a standard ICA decomposition of one person. Alternatively, one could also isolate several eye artifacts from many recordings, and concatenate these EEG segments. From the SVD of the resulting matrix, a very reliable estimate can be obtained of the subspace in which the mixing vector related to the artifact lies. That is, instead of incorporating the prior knowledge under the form of a single vector, one could also resort to a subspace that contains the mixing vector with high probability. This prior knowledge could be exploited to develop a new artifact removal method.

Our concept of scICA is more generally useful. EEG measures the potentials on a limited number of scalp electrodes. When one is interested in brain activity generated from one specific region, scICA can be used to compute the independent source signal from that region. This can be used for monitoring *e.g.*, the irritative zone. Another possible neuro-scientific application of scICA is situated in multimodal imaging. Nowadays, it is possible to record simultaneously different modalities. With scICA, activated regions seen in one modality can be used as prior knowledge in the decomposition of another modality.

5.7 Conclusion

This chapter discussed two topics. First, BSS-based decomposition of EEG was illustrated and discussed. Afterwards, spatially constrained ICA algorithms were developed.
Attempts have been made to quantify in an objective way the performance of artifact removal techniques, *e.g.*, in [352], but it should be emphasised that analysis of artifact removal is inherently difficult, since no true signal is known based on which methods can be compared. Simulation studies can be and are often biased to the strong points of a newly proposed method.

Chapter 6

Removing artifacts prior to ictal source localisation

The contamination of muscle and eye artifacts during an ictal period of the EEG significantly obscures EEG reading. In former chapters, it has been shown that recent blind source separation (BSS) techniques based on canonical correlation (BSS-CCA) and ICA with spatial constraints (scICA) have shown much promise in the removal of these artifacts. In this study we want to study the effect of BSS-CCA and scICA as a preprocessing step for automatic source localisation of the ictal period. This research was done in collaboration with MEDISIP from the University of Ghent. Hans Hallez was the expert in source localisation and performed the dipole localisation analysis.

6.1 Introduction

The objective of epilepsy surgery is the complete resection of the epileptogenic area, i.e., the region of the cortex that has to be removed in order to render the patient seizure-free. Detection of ictal epileptic activity is one of the diagnostic methods in order to pinpoint the epileptogenic area during a presurgical evaluation [102, 304].

The ictal EEG gives time information about the voltage distribution over the electrodes on the scalp during a seizure, and is the most commonly used method to localise the ictal onset zone [117]. Interrater variability in the visual interpretation of ictal scalp EEG recordings is considerable, and seizure activity is often obscured

by muscle and other artifacts [326]. The sensitivity of ictal EEG to localise the seizure onset zone is only 40–70% [117].

Visual analysis of EEG recordings aims to determine which lobe or which electrodes are activated. A challenging problem in neuroscience is to estimate in a more objective and precise way the regions of the brain that are active, given only the measured potential distributions [309]. Estimating the electrical source in the brain from the scalp EEG is a difficult problem since an infinite number of internal electrical currents can generate the same potential distribution on the scalp. Several approaches to solve this source localisation or inverse problem exist, based on different assumptions [11, 247]. One assumption is that the surface potentials are generated by a dense set of dipolar sources distributed on the cortical surface. The most popular method from this "distributed source" family is Loreta [285]. In a second approach, which is the most common, a limited number of "equivalent dipoles" are assumed to generate the measured potential distribution [309]. Dipole modeling is a well-established technique for localising interictal spikes, see *e.g.*, [93, 246, 196] and references herein. Dipolar models can be characterized by three spatial coordinates and three components, which have to be estimated by EEG dipole analysis. Ictal EEG recordings have been subjected to dipole modeling much less often than interictal spikes. The seizure discharge is a very complex pattern. Artifacts, such as electromyogram, movement, eye blinks and eye movements artifacts, render modeling difficult [130]. Even visual analysis of the seizure onset can be significantly improved by removing muscle artifacts [350]. Moreover, the low signal to noise ratio of the seizure signal can render the correct localisation very diffuse. However, when source localisation of seizure onset would be possible, it can reduce the need for invasive intracranial EEG recordings. So far, the results of ictal EEG source localisation have been discouraging. One study reports that the used "inverse solution" ([197]) is not useful at all for localising seizure onsets [359]. Some studies were restricted to temporal lobe seizures [249, 98]. One reason to select temporal lobe seizures is that source analysis is most reliable during periods of relative signal stationarity in order to average repetitive ictal waveforms, which is more common in temporal than in extratemporal lobe seizures. Another reason for selecting only temporal lobe seizures is that extratemporal lobe seizures are much more frequently contaminated by severe artifacts. Two other studies were not restricted to temporal lobe seizures. Gotman [130] obtained reliable models for seizure onset in 6 out of 15 patients (40%) and Boon and colleagues [24] in 31 out of 100 patients (31%). In the latter study, the ictal EEG was filtered with a narrow band filter (1–14 Hz), while ictal seizure activity is known to consist of rhythmical waves with a frequency between 3 and 29 Hz [129]. Also in [7] a very narrow bandpass filter (between 2 and 20 Hz) was used to remove the muscle and movement artifacts in spikes in order to obtain a signal with high signal to noise ratio. It was shown that the source estimation correlated well with the temporal seizure onset zone. Filtering should be avoided because these filters suppress all high frequency activity, including

electrical brain activity [128]. Moreover, muscle artifacts filtered by a low pass filter can resemble cerebral activity [186]. All these studies illustrate how difficult it is to reliably estimate ictal sources, and indicate that the current ictal scalp EEG source analysis tools cannot be used for a reliable localisation of the ictal onset zone during presurgical evaluation. A recent study on source analysis developed a novel integrative approach to characterise the structure of seizures in the space, time and frequency domains and showed some promising results [94].

Blind source separation (BSS) techniques have shown promising results in the field of artifact removal [89, 178, 340]. BSS-CCA is used in [49], which assumes mutually uncorrelated sources that are maximally autocorrelated. In the former chapter, we developed a spatially constrained ICA technique (scICA) to remove eye blinks and movements. The potential of these techniques to preprocess noisy EEG data prior to EEG source estimation is still unknown. In this study, we want to investigate how much the accuracy can improve by artifact removal.

Besides the EEG, an other useful technique used in the presurgical evaluation to determine the ictal onset zone is SISCOM [284, 278], as introduced in chapter 2. The time resolution of the perfusion SPECT is in the order of 30–60 seconds and thus very low compared to EEG, but in spatial resolution, SPECT (5-10 mm) outperforms EEG source estimation (2 – 4 cm according to [291]).

The ultimate goal would be to combine the benefits from both techniques and add time information from the EEG to the SPECT activations. In this study we applied BSS-CCA, reviewed in section 4.2 and scICA, explained in section 5.3 as a preprocessing step for source analysis of contaminated EEG. As for this purpose, there is no absolute gold standard available, we used “interpreted” SISCOM as the gold standard. We want to investigate how much source localisation of ictal EEG would benefit from artifact removal.

6.2 Materials

The EEG was measured in a clinical setup at the University hospital U.Z. Gasthuisberg in Leuven, Belgium. The electrodes were placed in a 10-20 standard system with two extra temporal electrodes and the sample rate was 250 Hz. From eight patients where the SISCOM revealed correctly the epileptogenic focus, the EEG was collected. The EEG segments were about two minutes in length and the start of the seizure was indicated by an experienced neurologist. Information about the patients used in this study is shown in table 6.1. Four out of eight patients underwent a surgical procedure. One patient was almost rendered seizure

free (Engel outcome II) and three patients were rendered seizure free completely (Engel outcome I).

	Age	Gender	Age at onset	FrCPS	MRI MRI	Latera-lization	Ictal EEG	duration of seizure	time of injection	SISCOM	surgery	Engel outcome
1	32	M	4	5	HS	L	LT	67	27	LT	Yes	I
2	59	M	39	20	Nl	R	RT	96	17	RT	Yes	II
3	35	F	25	15	HS	L	LT	94	12	LT	Yes	I
4	36	F	15	4	Nl	L	LFT	33	17	LT	No	NA
5	20	F	2	30	Nl	L	NA	93	25	LP	No	NA
6	30	M	6	8	CD	L	FC	70	18	LF	Yes	I
7	38	F	5	4	PS	L	LFC	36	3	LFT	No	NA
8	41	F	20	10	Nl	R	O	160	26	O	No	NA

Table 6.1: *The patient data used in the study. Gender: M = Male, F = Female. FrCPS = frequency of complex partial seizures per month. MRI: HS = Hippocampal sclerosis, Nl = Normal, CD = cortical dysplasia, PS = Posttraumatic scar. EEG localisation and SISCOM: L = left, R = right, T = temporal, F = frontal, C = central, O = occipital, NA = not applicable. Engel outcome Scores: I = seizure free, II = almost seizure free, III = 75% seizure reduction, IV = no improvement. The duration of the seizure is in seconds and time of injection is in seconds after the seizure onset.*

6.3 Estimating the sources by RAP-MUSIC

The EEG dipole localisation problem is twofold: the forward and the inverse problem. First, the forward problem determines the potential values at the scalp electrodes, given a dipolar source at a specific location by solving Poisson's equation in a specified geometry. The head model we used, was a realistic head model derived from a T1 Magnetic Resonance (MR) image. As an individual T1 MR image was not available, we used one MR image to construct a head model for all patients. We used Statistical Parameter Mapping (SPM) to perform a segmentation on the anatomical T1 MR images [121]. The head model was segmented in 4 compartments: scalp tissue, skull, cerebrospinal fluid (CSF) and brain tissue. The forward problem was then solved by using a finite difference method [138]. The conductivities of the compartments were isotropic and were equal to 0.33 S/m, 0.020 S/m, 1 S/m and 0.33 S/m for scalp tissue, skull, CSF and brain tissue, respectively [127]. The 21 electrode positions were set in a standard 10-20 system and were hence not adjusted to the patient specific positions.

Second, the inverse problem searches the six parameters of the dipole given the scalp potentials. A number of methods have been developed to estimate the location and orientation of the dipolar source, depending on the assumptions on the input EEG potentials. In this study we used a multidipolar estimation technique, namely, RAP-MUSIC [262], which is an improvement of the traditional MUSIC algorithm [310] by recursively applying MUSIC. MUSIC divides the measurement space into two subspaces representing the signal and noise in the EEG using an EVD. Then a search grid of 10 mm resolution is defined over the brain. Over this grid, a scanning procedure is performed. The position in the search grid whose forward solution has a maximal correlation with the signal subspace is chosen as a starting point. After the optimal position of a dipole is found, its forward solution is projected away from the signal subspace in RAP-MUSIC and the procedure is repeated iteratively for the next dipole source.

RAP-MUSIC was applied in moving windows of 62 samples (~ 0.25 s) with a step size of 6 samples (~ 0.025 s). In each window at most 2 moving dipole sources were estimated. Furthermore, a dipole was only considered a valid solution if the correlation between the forward solution of the dipole source and the signal subspace was at least 95%. This yields sources that are dominant in the EEG. If one of the sources did not meet the requirements, only one dipole was estimated. In addition to the traditional RAP-MUSIC algorithm described in [262], we restricted the dipole estimation in such a way that the distance of the 2 dipole estimates was strictly larger than 1 cm, in order to avoid dipole estimations that have a high magnitude but cancel each other.

The relative residual energy is a measure indicating the "goodness-of-fit" and is defined as the relative amount of energy in the EEG that cannot be explained by

the p dipoles, found by the RAP-MUSIC algorithm:

$$RRE = \frac{\|(\sum_i^p \mathbf{G}(\mathbf{r}_i, \mathbf{d_i})) - \mathbf{V}\|}{\|\mathbf{V}\|} \tag{6.1}$$

where $\mathbf{G}(\mathbf{r}_i, \mathbf{d}_i)$ are the scalp potentials caused by the i-th dipole at location $\mathbf{r}_i$ and orientation $\mathbf{d}_i$.

Used in a moving window, the RRE gives an inverse goodness-of-fit in function of time. The lower the RRE, the better the dipoles represent the EEG segment. Hence, a lower RRE and thus a higher "goodness-of-fit" is a characterization of the focality of the sources. To identify focal sources, we selected the dipole estimations that resulted in an RRE within the first quantile of the total set of RREs in the 21 second time window: 1 second before till 20 seconds after the start of the seizure. In this way we selected the sources that gave the best goodness-of-fit as the goodness-of-fit is different for each patient and depends on the noise in the EEG. Hence, in [136] it was shown that the RRE decreased when filtering the EEG with BSS.

6.4 Evaluation of EEG dipole source localisation

Starting from an EEG fragment of about two minutes for each of the 8 patients, the EEG was processed using BSS-CCA and scICA with fixed constraint to filter out the muscle and eye blink artifacts. The scICA algorithm can be applied automatically as the topography of the eye blink artifacts is known. The muscle artifact removal by BSS-CCA is done semi-automatically as follows. An expert was presented 10 s segments of EEG. Using a graphical user interface the expert could increment or decrement the number of components removed from the presented segment. When the muscle artifact was removed, the expert continued to the next 10 s segment. For each patient we obtained a raw and a filtered EEG. Both the filtered and unfiltered EEG segments were processed by RAP-MUSIC as described above. The electroencephalographic seizure onset was depicted by a neurologist. The source estimations were done in moving windows from 1 second before the seizure onset to 20 seconds after the seizure onset. In this way we want to compensate for the difference in timing of the SPECT activation and the EEG dipole source estimations.

In order to quantify the accuracy of the dipole estimation, the localisation was compared to SISCOM. The SISCOM image is then coregistered to the MR image, from which the head model was made using SPM5 [121]. The SPECT image of the SISCOM was normalized and a 95% threshold with respect to the maximum was set, considering only the activity above that threshold. We assumed that the edge of the 95% surface confined the seizure onset zone. Note that the activity

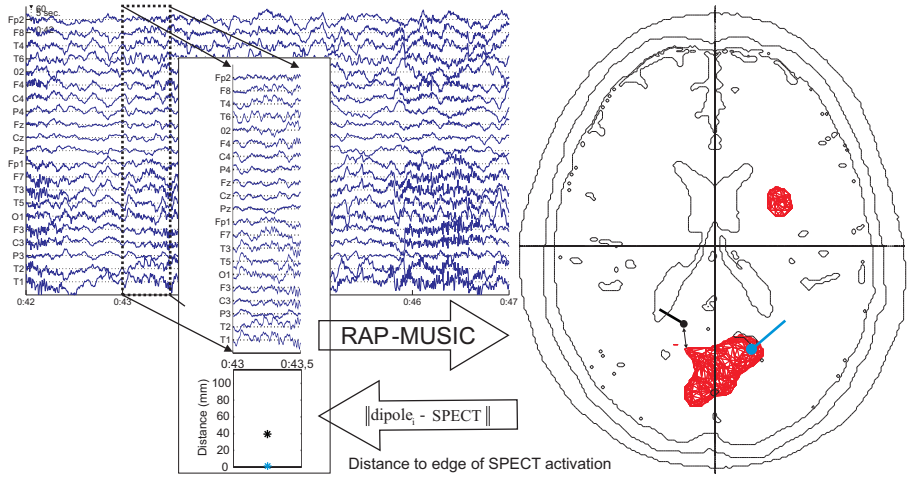


Figure 6.1: *A schematic representation of the calculation of the distance between the dipole locations and the edge of the 95% surface of the ictal SPECT activation. In a moving window of 62 samples (≈ 0.2 sec) at most 2 dipoles are estimated using the RAP-MUSIC algorithm. The distance of each dipole to the edge of the 95% interval of the ictal SPECT activation is then calculated. [137]*

delineated by the 95% surface could also be multifocal. In that case, the epileptic seizure had already propagated into other regions.

By combining the ictal SPECT activity with the dipole estimates from the RAP-MUSIC algorithm, we can evaluate the proximity of the dipole estimates before and after the artifact removal (see Fig. 6.1). The distances of the estimated dipoles with respect to the SPECT activations during the time period of 21 seconds are calculated. The distance is considered zero inside the SPECT activation. The result is a set of Euclidean distances for each of the at most 2 dipoles in that window. A boxcar plot is shown to compare the distributions of the distances before and after filtering (Fig. 6.2).

The distances with respect to the SPECT activations of the estimated dipoles during that time period are calculated as illustrated in Fig. 6.1.

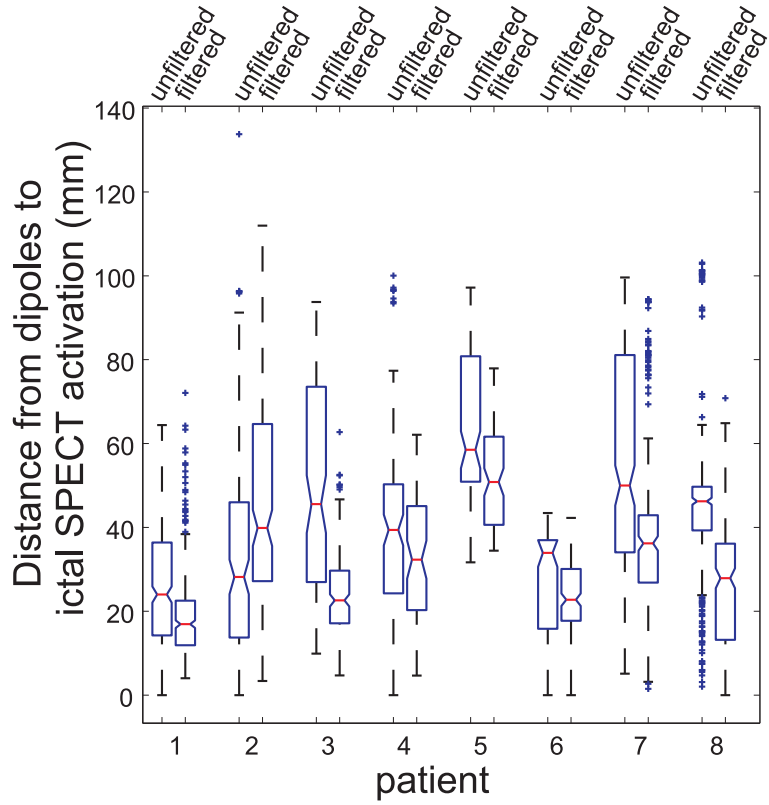


Figure 6.2: *A boxcar plot of the distance between the dipole locations and the edge of the ictal SPECT (iSPECT) activation for each patient.*

6.5 Results and discussion

In Fig. 6.3 and 6.4, two examples are shown of the unfiltered and filtered EEG. Below each EEG fragment, the distances of the dipoles to the thresholded ictal SPECT region are shown. The black and blue asterisks indicate the distance of the first and second dipole to the ictal SPECT activation respectively. Fig. 6.5 and 6.6 show the dipole estimations during the 21 second interval of the unfiltered and filtered EEG for these patients. The red blob indicates the contours of the ictal SPECT activity.

The first recording (Fig. 6.3a) contained mainly muscle artifact. In the unfiltered EEG, the distance between the dipoles and the SPECT border has a high variability. Applying the muscle and eye artifact removal techniques results in the EEG shown in Fig. 6.3b. We can visually appreciate from the figures that the

muscle artifact has been reduced. This is reflected in the distance measure. It is clearly reduced at the beginning of the seizure compared to the unfiltered EEG fragment. The dipole estimates from this data set are visualised in Fig. 6.5.

The second seizure recording (Fig. 6.4a) was severely obscured by frequent eye artifacts. While the ictal onset zone is located in the occipital area of the brain, most of the dipoles in the unfiltered case are located in the frontal area at the physical position of the eyeball (see Fig. 6.6). When the eye blink artifacts are removed, the dipoles are located more occipitally. In the vicinity of the start of the seizure the dipoles were estimated within the enclosed surface of SPECT activity. This example also shows that the removal of artifacts is necessary for dipole estimation.

Although these examples are promising, only in four patients source localisation really improved after artifact removal, meaning that the clusters became more concentrated or that new clusters emerged near SPECT activation. In Fig. 6.2 a boxcar plot of the distances is shown for each patient in the unfiltered and filtered case. Patients 1, 3, 7 and 8 correspond to the patients with improved localisation. Three patients did not show any improvement. In one patient, the artifact removal techniques removed so much of the signal, that not enough signal was left for a reliable estimation.

The failure of accurate source estimation, even after artifact removal, can be due to several reasons. In one recording, a huge electrode artifact was present. A second reason is inaccurate muscle artifact removal. The removal was done by a clinician to visually read the EEG. This does not guarantee an optimal cleaning. When the spatial distribution of ictal activity and muscle artifacts are overlapping, the separation can be suboptimal, and ictal activity can also be removed. A third reason can be the impossibility to model seizure activity with the current dipole model. It has never been proven in the literature that this model is fully valid. A seizure is a complex phenomenon that involves interactions between different brain regions.

We also have to mention that in the cases where we consider source estimation after muscle removal reliable, we do not obtain a perfect correlation between the dipole estimates and ictal SPECT is not obtained. Because a SPECT image gives an integrated estimate of the uptake of the blood tracer, sometimes several activated blobs were visible. We hoped to derive time information from the EEG. The ultimate goal was even that we could visualize the propagation patterns of the seizure, by correlating this time information with the ictal SPECT. This was however not possible.
It is also important to keep in mind that we used a thresholded SPECT image as a reference to compute an objective validation measure of distances. This is a rather rough measure.

To conclude this discussion, some critical points for improvement can be

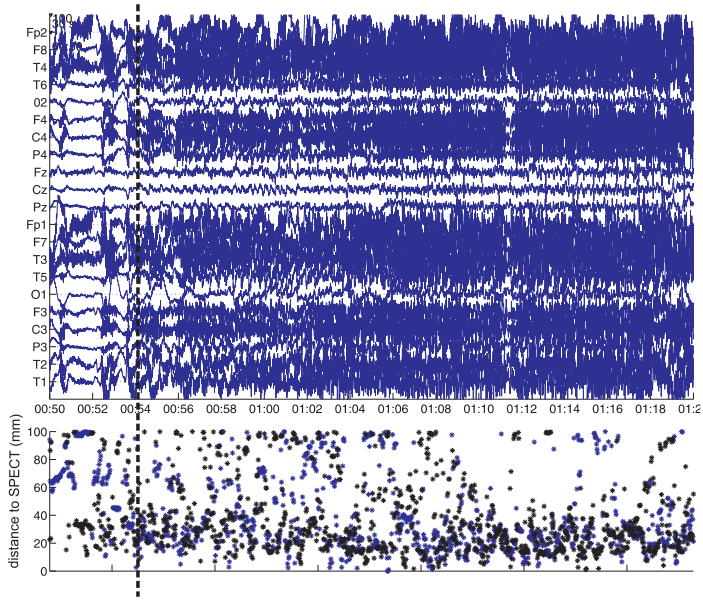
mentioned. The imperfections in dipole localisation are due to the fact that source estimation depends on many parameters. One set of parameters is the head model. A more personalized head model can reduce estimation errors. Further, source localisation is also very sensitive to noise, so every imperfection in artifact removal can deteriorate the results. Finally, we want to point out the necessity of using a large number of electrodes to have an accurate dipole estimation. In our study we used a clinical setup of 21 electrode channels. In this case the brain activity and artifacts are captured in 21 channels. By using subspace techniques as a preprocessing step (BSS-CCA and scICA), from 21 channels only 21 sources can be estimated. This can *e.g.* result in one source containing muscle activity as well as ictal activity. If a larger number of electrodes is used, the number of sources that can be estimated, increases. This can yield a better separation between sources representing the muscle activity and sources representing the ictal activity. Therefore, by using high resolution EEG (64 or 128 channels), a more accurate elimination of the artifact and thus a more accurate estimation of the source can be expected. However, this setup is less practical in the clinic practice.

6.6 Conclusion

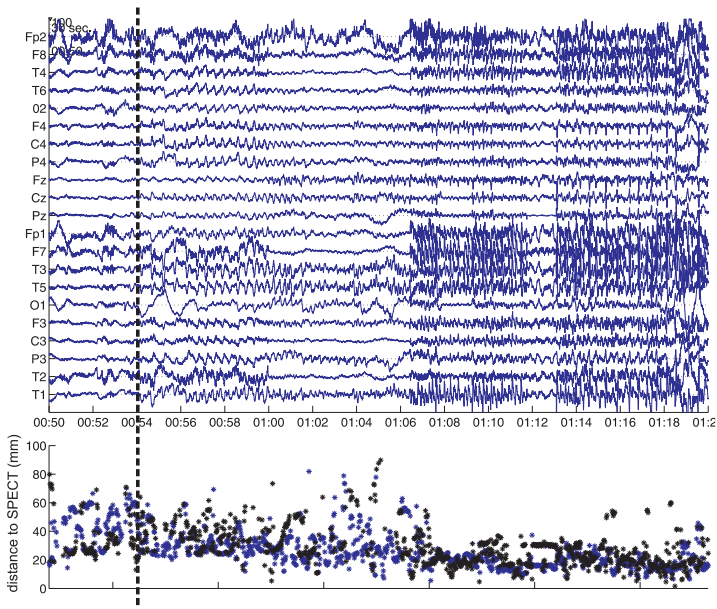
Ictal EEG is mainly contaminated by muscle and eye artifacts. This disturbs automatic source localisation methods. In this small pilot study, we used blind source separation techniques (BSS-CCA and scICA) to remove these artifacts and investigated if a better localisation of the epileptogenic focus could be obtained. We examined 8 patients suffering from epilepsy.

In 4 out of 8 patients a real improvement of the EEG dipole estimation could be observed. In these patients tighter clusters or new clusters were observed near or overlapping with the ictal SPECT activity. This suggests that EEG source estimation is promising to correctly depict the ictal onset zone if artifacts are removed.

However, some things are to be kept in mind. First, the dimension of the signal subspace can become too small if too many artifacts are removed. Also the objectivity in removing sources can be improved by a more careful application of BSS-CCA. Last, research should aim at developing methods that take better into account the complementarity between EEG and SPECT.

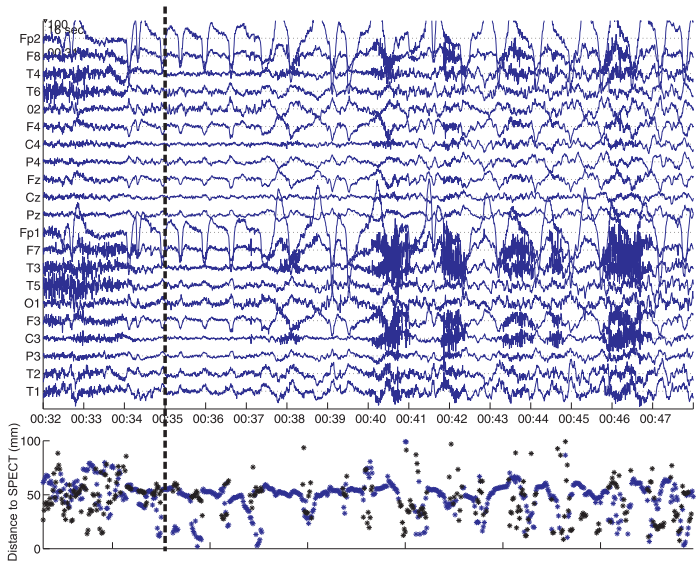


(a)

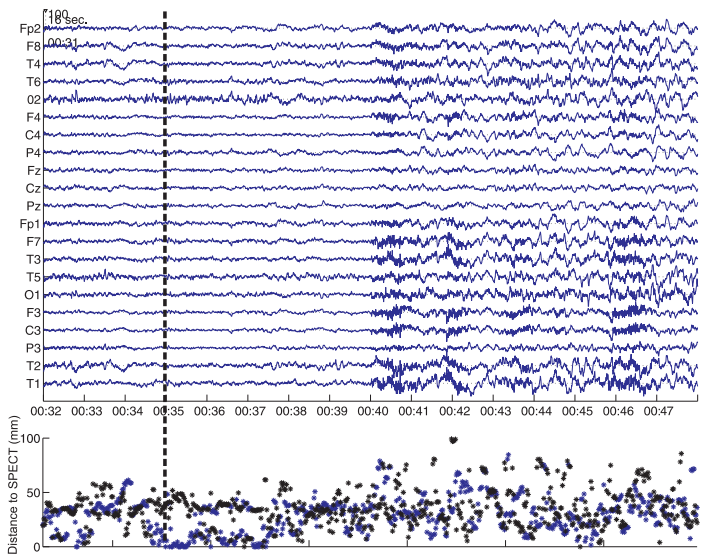


(b)

Figure 6.3: *An EEG fragment of 30 seconds is shown: (a) the unfiltered case and (b) the filtered case. Below the EEG fragments the distance measure (in mm) is shown. The seizure onset, as indicated by the neurologist, was between 53 and 54 seconds. In the unfiltered EEG, we can see that the distance from the dipole to the ictal SPECT activation is very high ($\geq$ 100 mm) and is very irregular. The black and blue asterisks indicate respectively the distance of the first and second dipole to the ictal SPECT activation.*



(a)



(b)

Figure 6.4: *Figure (a) and (b) show the raw and the filtered EEG. Also the distance of the dipoles to the edge of the ictal SPECT activation is shown in function of time. The results show an EEG window of 15 seconds. The seizure onset, as indicated by the neurologist, was at second 35. In the unfiltered EEG, we can see that the distance from the dipole to the ictal SPECT activation is very high due to eye artifacts. In the filtered EEG, the dipoles from the early seizure onset are situated in or near the SPECT activation. The black and blue asterisks indicate respectively the distance of the first and second dipole to the ictal SPECT activation.*

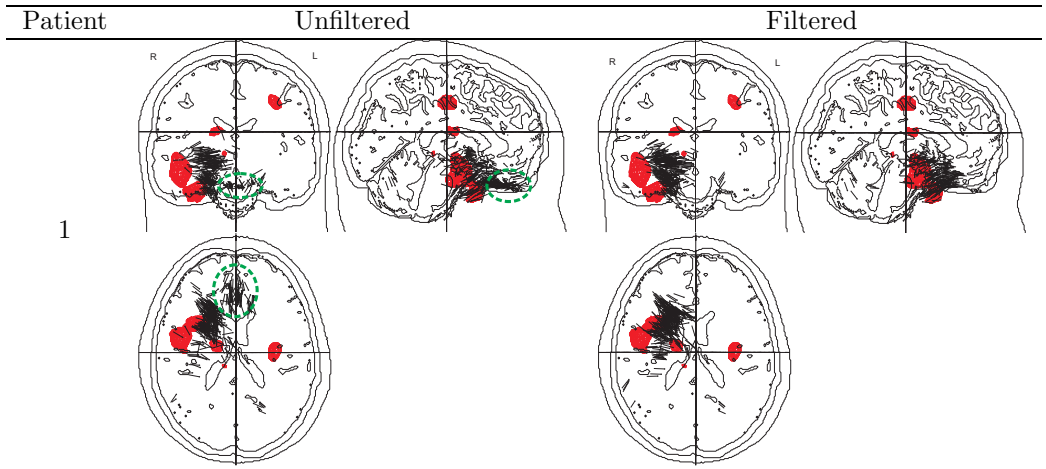


Figure 6.5: *Patient 1: The dipole estimations during a 21 second interval starting one second prior to the start of the seizure in the case of the unfiltered EEG data (left) and filtered EEG data (right), shown in Fig. 6.3. Each figure depicts 3 views of the dipole estimates in a realistic head model: axial, sagittal and coronal. The red contours denote the thresholded ictal spect activity coregistered with the realistic head model. For the sake of visualization, the threshold of the ictal spect was chosen to be 85% instead of the 95% used in the calculations (a 95% surface resulted in a very small area, which could be obscured by the dipole estimates). In patient 1 the green circle indicates a region of interest depicting the dipole estimates during an eye blink artifact. This results in dipole estimates estimated very close to the eyes.*

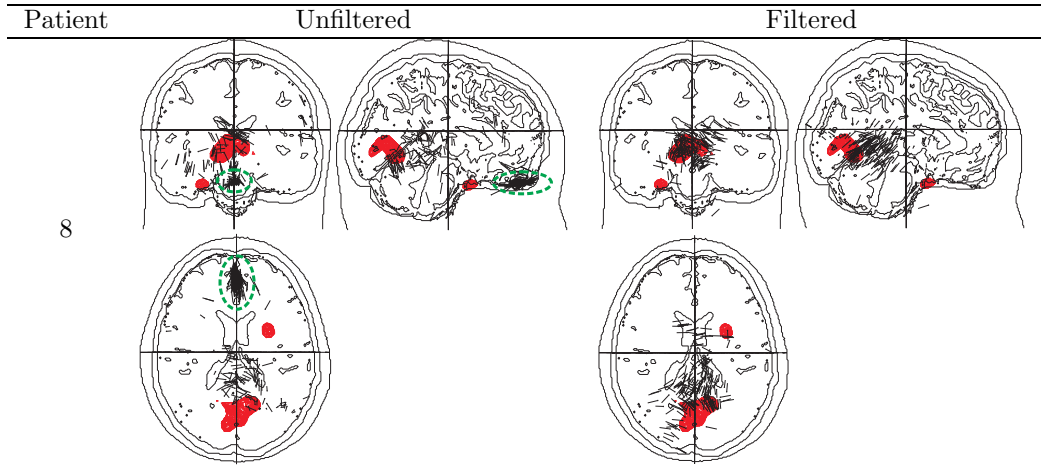


Figure 6.6: *Patient 8: The dipole estimations during a 21 second interval starting one second prior to the start of the seizure in the case of the unfiltered EEG data (left) and filtered EEG data (right), shown in Fig. 6.4. Each figure depicts 3 views of the dipole estimations in the realistic head model: axial, sagittal and coronal. The red contours denote the thresholded ictal SPECT activity coregistered with the realistic head model. The green ROI depicts the dipoles during source analysis during eye blink artifacts.*

Chapter 7

Extracting ictal activity with CPA

In the previous chapter, we aimed at localising the epileptic focus by removing the most prominent artifacts from ictal recordings and afterwards estimating dipole localisations. The localising value of dipole modeling of ictal EEG could be improved in several cases in this way. However, instead of decomposing the dataset in order to accurately estimate the artifacts, the optimal way would be to accurately estimate the ictal activity itself.

In this chapter, we provide a way to decompose the measured EEG in a sum of individual contributions of distinct brain sources and identify one source as epilepsy-related. This source will provide accurate localising information about the epileptic focus.

7.1 Introduction

Our aim was to establish a robust and automatic localisation method for ictal EEG activity. We show that CPA provides such a method. With this decomposition, multichannel time-varying EEG is decomposed into a series of distinct "atoms" (defined as rank-one arrays), which represent in an ideal situation distinct brain sources. CPA can be considered as the higher-order variant of factor analysis. The aim of factor analysis is to relate the different "atoms" to the different "physical mechanisms". However, it is often not a priori known whether an atom obtained with factor analysis truly reveals the underlying phenomenon.

Atomic decompositions of the EEG in matrix-form have been shown to be useful as an exploratory tool. With Principal Component Analysis (PCA) and Independent Component Analysis (ICA), space-time decompositions of multichannel EEG can be used for artifact removal [340, 223] or for extracting interesting activities [365, 195]. Factor analysis of matrices is underdetermined, and assumptions like orthogonality or independence, which may not be adequate in the particular application, have to be imposed in order to obtain a 'unique' solution. Factor analysis of higher order datasets is unique under mild conditions. The main advantage of CPA is that no extra assumptions have to be imposed. We will show that the CP decomposition reliably identifies one epileptical seizure onset atom. The CP model was used before for the simultaneous space-time-frequency decomposition of the power of wavelet transformed EEG with dimensions channels x time x scales during a mathematics tasks [251, 243]. It was shown that CPA allowed identifying frontal theta atoms during a cognitive task and occipital alfa atoms during resting conditions. In [261], CPA was used as an exploratory tool in the analysis of inter-trial phase coherence of event-related EEG. Another tensor decomposition, the Tucker model, does not give an atomic decomposition, but it has recently been used for seizure analysis [2]. Separation of the ictal source has not been reported.

Ictal seizure activity is known to consist of rhythmical waves often with a frequency between 3 and 29 Hz [129]. As the CP decomposition of wavelet transformed EEG is especially useful in the description of oscillatory phenomena, the goal of this study was to investigate the localising value of the epileptical atom, identified after CPA. In contrast to previous studies, where the CP decomposition was computed on the power of wavelet transformed EEG, we compute the decomposition on the 'pure' wavelet transform of the ongoing EEG recording, which makes the decomposition more robust with respect to artifacts. After the decomposition, the potential distribution over the electrodes of the epilectical activity is obtained, and can be displayed as a 2D image. Electrodes with large potential amplitudes can be considered as close to the focus. In this study, we assess the localising accuracy of this 2D image with a realistic simulation study. Furthermore, we want to investigate whether it is possible to localise the ictal onset zone in the 3D head by applying dipole source localisation after CPA of ictal EEG recordings. We assessed the accuracy of the 3D method with realistic simulations under different conditions. We are especially interested in the influence of the frequency of the seizure activity on the localisation, in how the dipole localisation would be influenced by changes in frequency, and if the dipole estimation accuracy could be improved by increasing the number of electrodes.
We validated the 2D method clinically on a set of ictal EEG recordings, and showed that the CPA method was more sensitive than visual interpretation of the ictal EEGs in localising the ictal onset zone.

7.2 Method

CPA was introduced in chapter 3. We define now how to construct the third-order tensor from the EEG recording and show how CPA will be used for seizure localisation.

Wavelet transform

Fourier Transform (FT) of EEG data can estimate the frequency content of the signal [92], but does not reveal time-varying frequency changes in the signal. Even the Short-Time Fourier Analysis (STFA) is not the most accurate and efficient method for time-frequency localisation because the interval in which the frequency content is computed, is fixed. Since the time and frequency resolutions are reciprocal, good time resolution means poor frequency resolution and vice versa. A more optimal time-frequency analysis is given by the wavelet transform. Wavelets resolve high frequency components within small time windows and low frequencies in larger time windows. The continuous wavelet C at scale a and time t of a signal $x(t)$ is defined by

$$C(a, time) = \int_{-\infty}^{\infty} x(t)\phi(a, time, t)dt \tag{7.1}$$

with ϕ the chosen wavelet. We used in this study a biorthogonal wavelet with decomposition order 3 [62]. The exact choice of the wavelet does not really affect the result in our study as long as a real wavelet is chosen. From the scale a of the wavelet, the frequency f of the signal can be estimated as:

$$f = f_c/(a\Delta t) \tag{7.2}$$

with f_c the center frequency of the wavelet and Δt the sampling period.

CPA for seizure localisation

After wavelet transformation of every channel of the original EEG at the seizure onset with dimensions (channels x time), a three-way array with dimensions (channels x time x scales) was obtained. We chose 2 seconds as length of the time-window as an optimal trade-off. A smaller time-window will give a less stable decomposition. A larger time-window would enlarge the risk that the seizure activity is already propagating in a realistic setting. Smaller time-windows can also avoid the presence of eye blink artefacts (see for example Fig. 7.16). The seizure onset was defined as the time point when an ictal EEG discharge was

observed without going back in time. It should be stressed that it is sufficient that a neurophysiologist is able to observe an ictal EEG change, even if the neurophysiologist is not able to localise the ictal onset zone at that point in time (*e.g.*, generalised attenuation of EEG signal, ictal EEG changes obscured by muscle artifact, etc.). Before computing the CP decomposition, the data had to be carefully preprocessed. The data were standardised by centering the EEG, *i.e.*, subtracting the mean, and dividing each channel by its standard deviation. After the decomposition, the spatial distribution of the scalp potential was again multiplied with this standard deviation in order to preserve the amplitude information. This three-way array was decomposed with CPA. There are several ways of determining the correct number of atoms [320]. By checking the Core Consistency Diagnostic (Corcondia), we found that the optimal number of atoms was usually two ($R = 2$).

When the goal is to localise the seizure onset region, the seizure atom has to be identified after computing the decomposition. The atom can be selected visually. The spatial component of the identified atom gives the spatial distribution of the epileptical activity. The electrodes with a resulting potential above a predefined threshold (we used 75% of the maximal value of amplitudes across all channels for fixed atom) define the epileptic focus. In addition, it can be verified that the maximal frequency of the frequency distribution corresponded to the frequency of the rhythmical seizure activity and that the topography does not represent symmetric background activity.

Illustration

CPA of ictal EEG is illustrated on our tutorial example in Fig. 7.1. We illustrate this approach with another example, this time decomposing only the advised 2 seconds. EEG containing clear ictal activity in the right temporal lobe is given in Fig. 7.2. Each atom has a component in the space ($\mathbf{a}_i$), time ($\mathbf{b}_i$) and frequency domain ($\mathbf{c}_i$). The seizure atom(s) can be selected based on characteristic signatures in the different domains. At the ictal onset, seizure activity is visually recognised by rhythmical activity that is well localised in space and frequency. This was also described in [1]. Another possibility is to reconstruct the EEG signals that correspond to the individual atoms in EEG settings by means of the inverse continuous wavelet transform (ICWT) [201]. Going back to the example, the seizure starts at second 3, and the EEG between second 3 and 5 is wavelet transformed and decomposed with CPA (Fig. 7.3). Corcondia indicated that a decomposition in two atoms would be appropriate. The first atom is recognised as seizure atom. The frequency component peaks around 3 Hz and the time component is a rhythmical waveform that increases in amplitude. When this

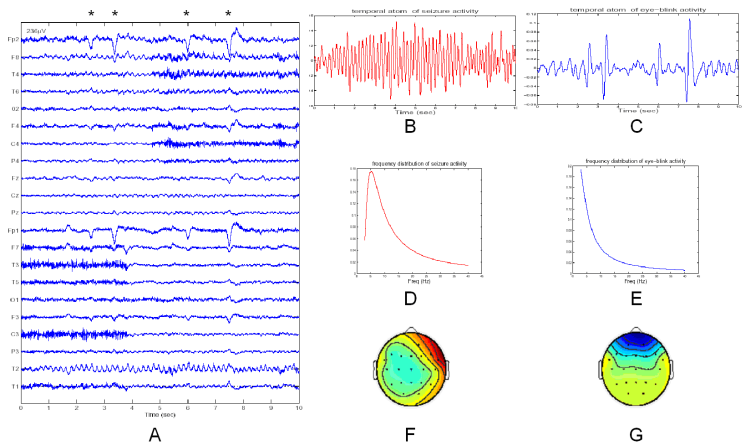


Figure 7.1: *CP decomposition of a pure wavelet transformed ictal EEG. A 21-channel average reference montage EEG showed 10 seconds of the beginning of a right temporal lobe complex partial seizure, characterised by a recruiting theta rhythm of around 5 Hz over the right temporal derivations (T2 (sphenoidal electrode), F8, T4, and also T6) (A). There were muscle (e.g., first three seconds in T3 and C3, and last five seconds in C4) and eyeblink artifacts (FP1 and FP2, asterisks). The EEG data matrix was wavelet transformed to obtain a third-order tensor without any artifact removal. The three modes of the CP decomposition of the 10 seconds of ictal EEG are shown: temporal mode (B and C), frequency mode (D and E), and spatial mode (F and G). The blue atom (C, E, G) was related to the eyeblink artifacts. The temporal mode (C) showed high peaks at the same time of the eyeblinks in the original EEG (asterisks in A). The frequency mode (E) showed a peak at low frequencies, consistent with low-frequency eye-blinks. The spatial component (G) showed the typical distribution of eyeblinks at FP1 and FP2. The red atom (B, D, F) contained all the information related to the ictal activity. The temporal mode (B) reflected the rhythmicity of the ictal wave. The frequency mode (D) displayed a peak around 5.7 Hz, which corresponded to the frequency of the recruiting theta waves in the ictal EEG (A). The spatial mode (F) showed that the potential distribution of the ictal activity had maximal values (red area) in the antero- and midtemporal and spenoidal electrodes of the right temporal lobe, which was also readily apparent on visual inspection of the ictal EEG (A). Notice that the spatial mode (F and G) is a two-dimensional display, with the right side corresponding to the right side of the patient, and the black dots representing the electrode positions according to the 10-20 system. Red regions indicated highly positive potentials, dark blue regions indicate highly negative potentials.*

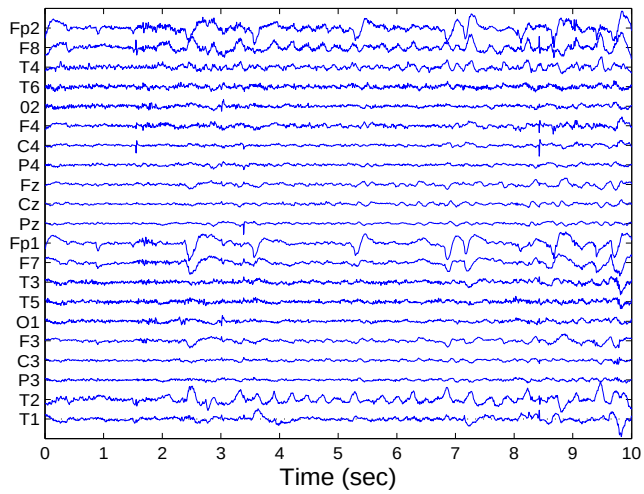


Figure 7.2: *10 seconds of EEG containing the start of a seizure.*

component is reconstructed as an EEG signal, pure ictal activity can be seen (Fig. 7.2). The second component contains an eye blink. Note however that the eye blink is not as well separated as with SOBI (chapter 5).

Dipole estimation after CP decomposition

With CPA, the potential distribution that corresponds to the ictal source is extracted. If we want to obtain a 3D position of this source, we can directly apply known source localisation techniques to this potential distribution [66]. Because the atoms in the Canonical Decomposition have a very simple, trilinear structure, we propose to fit only 1 dipole for every atom. We expect that, when a patient suffers from multifocal epilepsy, different atoms will be related to activity generated by the different dipoles. Dipole estimation then determines the dipole's coordinates and orientation that best generates the given potential distribution in a least squares sense. For computational simplicity, we used a spherical head model in this study.

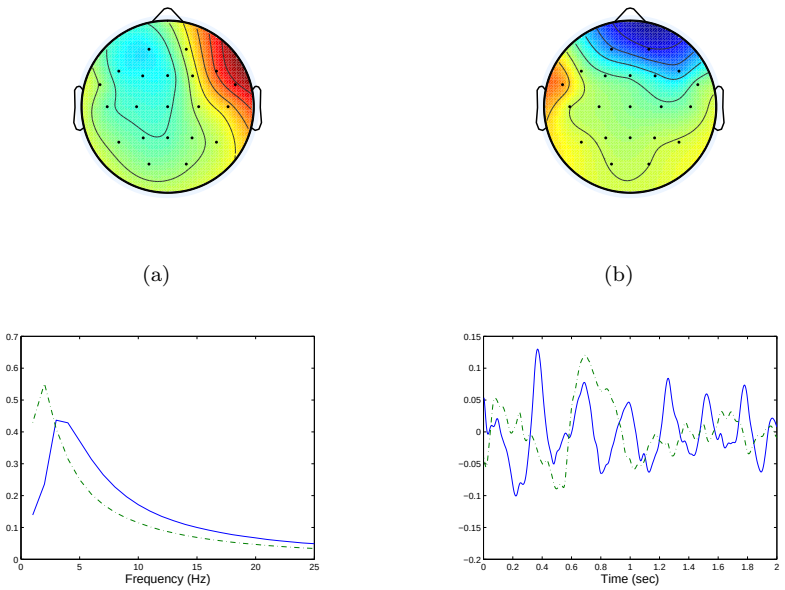


Figure 7.3: *Seconds 3 to 5 of the seizure shown in Fig. 7.2 are decomposed with the Canonical Decomposition with 2 atoms. (a) (b) the spatial potential distributions of the two atoms. (c) The frequency content of the atoms. (d) The time course of the atoms. First atom drawn in solid line corresponds with (a). Dash-dotted line corresponds with (b). First atom is seizure atom. Second atom is related to eye blink.*

7.3 Simulation study

7.3.1 Simulation setup

Consider a matrix $\mathbf{X}$ of dimension 500-by-21 representing a 21 channel EEG section of 2.0 s long. Each column $\mathbf{x_s}, s = 1, \ldots, 21$ of $\mathbf{X}$ contains the time course of an EEG channel:

$$\mathbf{X} \quad = \quad [\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_{21}}]. \tag{7.3}$$

In this simulation study $\mathbf{X}$ includes both seizure activity, and superimposed noise. Both signals are described below.

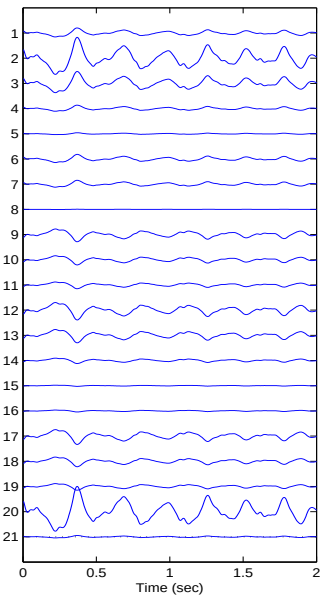


Figure 7.4: *The seizure atom from Fig. 7.3 is reconstructed in EEG coordinates after CPA.*

The EEG of the ictal activity (Fig. 7.5A) was generated using a fixed dipole in a three-shell spherical head model with a moment having a 5.7 Hz sinusoidal waveform (for example typical in patients with mesial temporal lobe epilepsy (MTLE) [267]). The amplification factors at each electrode were computed by solving the forward problem for a dipole in a three-shell spherical head model consisting of a brain, a skull and a scalp compartment ([306]). Each compartment had a specific conductivity with a ratio equal to 1:$\frac{1}{16}$:1 for the brain, skull and scalp compartment respectively [281]. The brain and scalp conductivity was $3.3 \times 10^{-4}/\Omega mm$ [58]. Radii of the outer boundary of the brain, skull and scalp region equal to respectively 8 cm, 8.5 cm and 9.2 cm were used [281]. Unless otherwise stated, dipole coordinates x (left ear to right ear), y (posterior to anterior) and z (up, through the Cz electrode) were [-0.5 0 0.1] and the dipole orientations d_x, d_y and d_z were [1 0 0]. Twenty-one electrodes were used according to the 10-20 system for electrode placement and additional electrodes T1 and T2 on the temporal region. The time course of the scalp potentials was stored in a 500-by-21 dimensional matrix A, representing 2 seconds of EEG with sample frequency of 250 Hz.

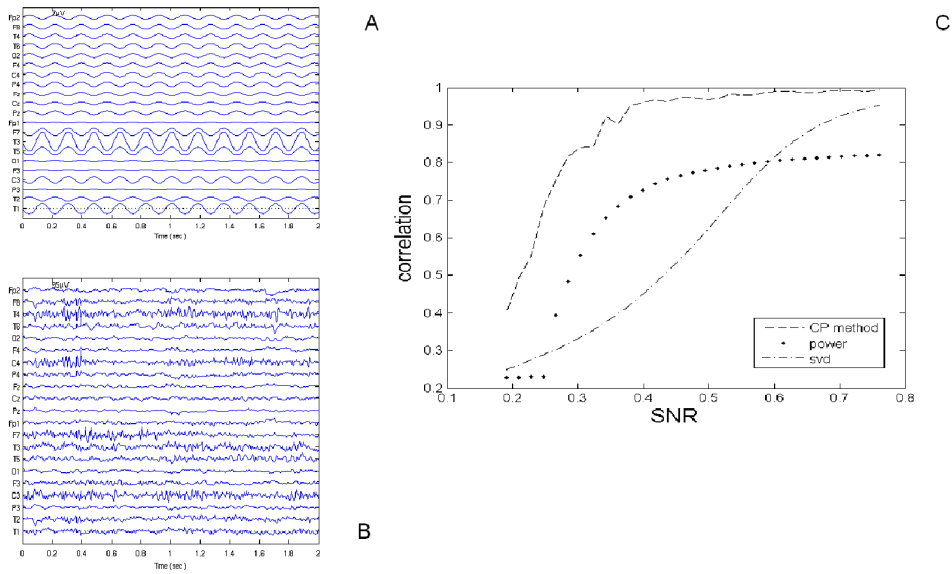


Figure 7.5: *Noise-free simulated ictal activity (A). The 2 seconds time course of the scalp potentials showed 5.7 Hz rhythmic activity in all electrodes with highest amplitude at T3 and T1. A simulated data matrix was obtained by mixing the noise-free simulated ictal activity (A) with noise resembling muscle artifact. In the example of (B), the SNR was equal to 0.42. The correlation between the simulated distribution and the spatial distribution, extracted with CPA, is shown in C. The 5.7 Hz seizure activity was largely embedded in noise. The correlation for wavelet transformed EEG remained high up to a SNR of around 0.3. At a SNR of 0.42 in (B), the correlation was 0.98, which means that the correct spatial distribution of the seizure activity was obtained with the CP decomposition, i.e. the method appeared insensitive to noise resembling muscle artifact. Also notice that the performance of CPA for power of wavelet transformed EEG data was worse, compared to wavelet transformed EEG signals. Figure C also shows the correlation between the simulated distribution and the spatial distribution, extracted with CPA and with the SVD on the same simulated data. Introducing the third, frequency-related dimension clearly improves the localisation.*

A 500-by-21 noise matrix B contained 2 seconds of awake background EEG activity, recorded with the same electrode configuration, from a normal subject. On this matrix B, muscle artifacts were superimposed. These muscle artifacts were separated from contaminated background activity using BSS-CCA [49]. In the simulation study, the noise matrix B was superimposed on the signal matrix A containing the epileptical activity:

$$\mathbf{X}(\lambda) = \mathbf{A} + \lambda \cdot \mathbf{B}, \tag{7.4}$$

with $\lambda \in \mathbb{R}$ (Fig. 7.5B). The data was digitally filtered by a band pass filter (0.3-35Hz). The Root Mean Squared (RMS) value of the signal is then equal to

$$RMS(\mathbf{A}) = \sqrt{\frac{1}{S \cdot N} \sum_{s=1}^{S} \sum_{n=0}^{N-1} \mathbf{A}^2(n,s)}, \tag{7.5}$$

and the RMS value of the noise is equal to

$$RMS(\lambda \cdot \mathbf{B}) = \sqrt{\frac{1}{S \cdot N} \sum_{s=1}^{S} \sum_{n=0}^{N-1} (\lambda \cdot \mathbf{B}(n,s))^2}. \tag{7.6}$$

An important measure is the signal-to-noise ratio (SNR), which is defined as follows:

$$SNR = \frac{RMS(\mathbf{A})}{RMS(\lambda \cdot \mathbf{B})}. \tag{7.7}$$

Changing the parameter λ alters the noise level of the simulated signal. We performed several simulations with different noise levels.

Following seizure characteristics were simulated:

A Seizure activity in patients with mesial temporal lobe epilepsy (MTLE) is typically expressed by a low frequency sinusoidal waveform [267]. The aim of the first simulation was to evaluate the performance of the proposed method in extracting the potential distribution of the epileptiform activity, for different noise levels. The performance was evaluated by the correlation coefficient between the true potential distribution and the extracted potential distribution of the seizure activity (Fig. 7.5C). We compared this performance with results obtained with CPA on the power of wavelet transformed EEG, as is described in previous publications [251]. We also compare our method with the localising information that can be

obtained from the SVD. We compare our results with the first left singular vector of the EEG (Fig. 7.5C). Instead of only looking at the 2D potential distribution, we further estimated the dipole localisation error when seizure activity was represented by the same waveform at different noise levels (Fig. 7.6a). As seizures will never be perfectly rhythmical, we also investigated the influence of the specific waveform and estimated the localisation error when seizure activity was represented by a 4 Hz sawtooth, instead of a sinusoidal wave, at different noise levels.

B Ictal EEG activity can have a frequency in the delta, theta, alpha or beta range. In a second simulation, therefore, we estimated the influence of the frequency of the seizure signal on ictal scalp EEG source localisation at a fixed noise level. We were particularly interested if the possible overlap in frequency content between faster ictal activity and muscle activity would bias the decomposition and thus the dipole estimate.

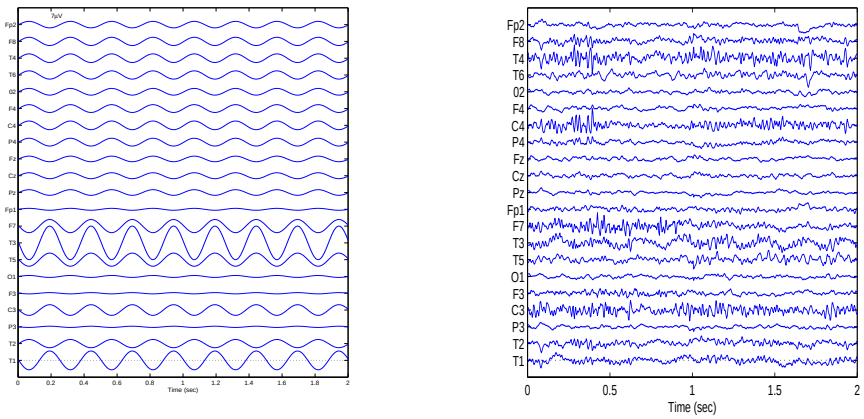
C Epileptic seizure activity can rapidly change in frequency. Ictal EEG activity is often characterised by low voltage fast activity in the beta range that gradually slows down to alpha or theta frequencies with increasing amplitude. CPA exploits frequency information during the decomposition. In order to test possible shortcomings of CPA of ictal EEG, we wanted to estimate the accuracy when the model is violated. In a third simulation, we assessed the dipole localisation error when the frequency changed during the 2 seconds under investigation. This does not give a trilinear signal after wavelet transformation. We simulated a chirp that linearly changed in frequency from 8 Hz at the start to 4 Hz at the end of the considered 2 seconds. The signal also doubled in amplitude.

D Normally, two atoms are obtained after the decomposition of *in vivo* seizures and a distinction could be made between a seizure and a non-seizure atom. An interesting question is how well different dipoles generating similar ictal signals will be distinguished from each other. Such activity can be measured in the case of multifocal epilepsy. In a fourth simulation, we considered two rhythmical sources firing at the same frequency separated from each other by about 1 cm: the second dipole had coordinates [-0.4 0 0.1]. These dipoles generated similar potential distributions at the scalp.

E In a fifth simulation, the influence of the dipole was investigated. Deeper sources generate a weaker signal at the electrodes and are possibly less accurately separated from background EEG. We varied the z-coordinate of the dipole between 0 and 0.8. x and y were kept fixed at -0.5 and 0 respectively.

F 21-channel EEG does not have an optimal spatial resolution due to the low spatial sampling. In a last simulation, we investigated how much the dipole localisation error could be improved by using dense array EEG [282]. We

used 148 electrodes, uniformly distributed over the realistic domain of the same spherical head model.



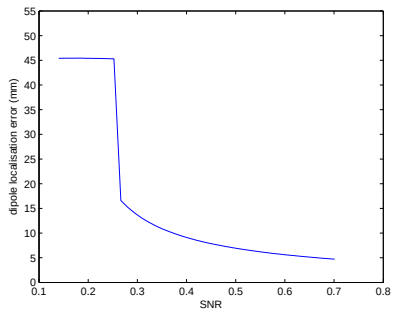
(a) (b)

Figure 7.6: *Simulated data. (a) The time course of the scalp potentials reflecting the 4 Hz epileptiform activity on each electrode. (b) The simulated data matrix for an SNR equal to 0.7.*

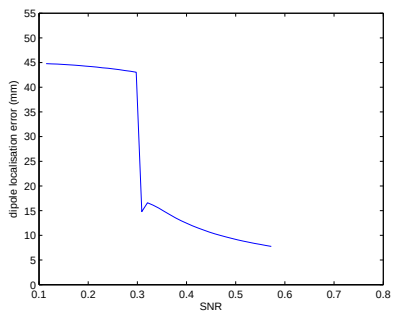
7.3.2 Results of simulation study

The correlation between the simulated distribution and the spatial distribution, extracted with CPA, is shown as a function of SNR in Fig. 7.5C. The figure shows that the CPA correctly localised epileptic activity despite severe contamination by noise. One can expect that real ictal EEG has an SNR much larger than 0.42. But even at this low noise level (Fig. 7.5B), the correct potential distribution is extracted. The figure also shows that using pure wavelet-transformed EEG, as we did, was better than using the power of the wavelet-transformed EEG as it was described in previous publications. A comparison of the localising value of CPA and the SVD on the same simulated data is also given in Fig. 7.5C. Adding frequency information and decomposing the array in a unique, unconstrained way clearly improves the localisation.

Fig. 7.7a shows the dipole localisation error in function of the SNR when one dipole was fitted on the potential distribution extracted with CPA. At an SNR of 0.4, the localisation error became smaller than 1 cm and at an SNR of 0.7, the error between the simulated and the fitted dipole was only 5 mm. At SNRs lower than 0.26, none of the 2 estimated atoms corresponded to the seizure activity as



(a)



(b)

Figure 7.7: *(a) The dipole localisation error in function of the noise level when a sinusoidal waveform was used as epileptic signal. (b) Idem as (a) but a sharp wave was used as epileptic signal.*

can be seen from the large localisation error. At higher noise levels, one atom contained pure rhythmical activity, as can be seen by the sudden improvement in dipole localisation error. Fig. 7.7b shows the dipole fit error when a sawtooth was used to simulate ictal EEG. The error was slightly larger compared to the perfect sinusoidal signal, but still in the same range.
Fig. 7.8 shows the dipole localisation error for different frequencies of the simulated epileptic signal at an SNR of 0.7 (Fig. 7.6b). From this figure, it can be seen that the accuracy of the separation of ictal EEG, does not depend on the frequency of the signal. At all frequencies, a dipole is fitted with an error smaller than 1 cm. This means that even when the frequency content of ictal activity overlaps with frequency content of muscle artifacts, a good separation is obtained.
Fig. 7.9 shows the dipole localisation error as a function of the SNR when the simulated epileptic signal changed in frequency and amplitude during the considered 2 seconds. The figure strongly resembles Fig. 7.7a. This means

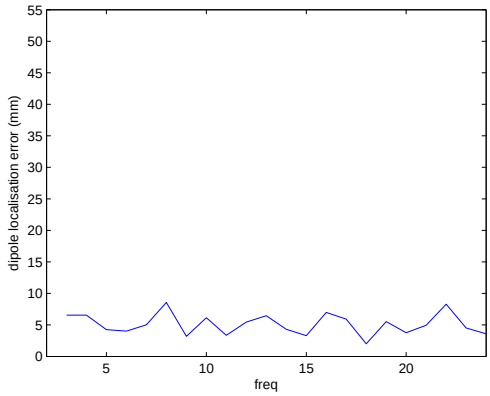


Figure 7.8: *The dipole localisation error as a function of the seizure frequency.*

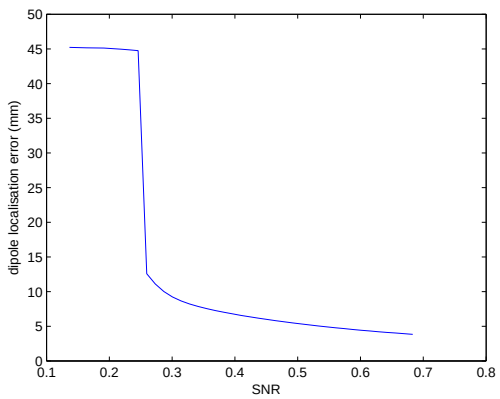
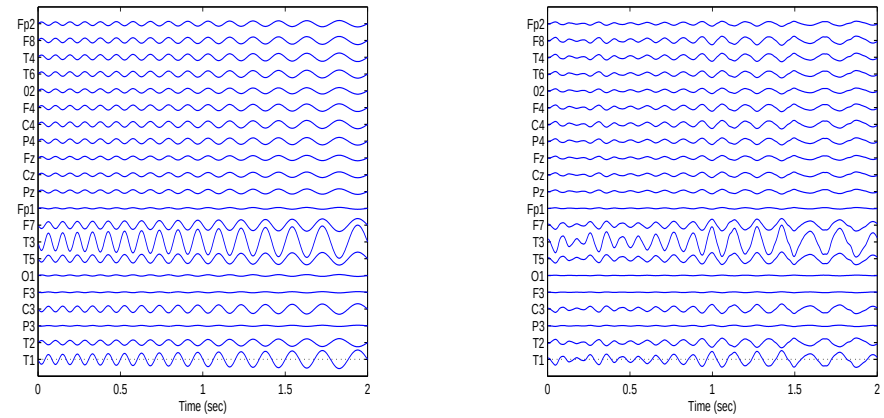


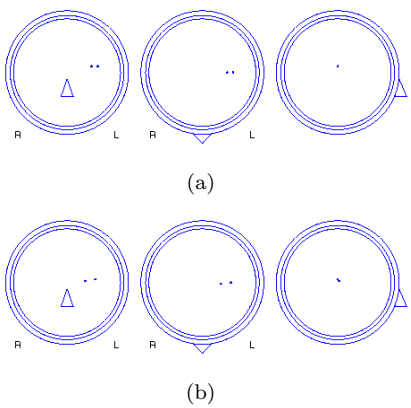
Figure 7.9: *The dipole localisation error as a function of the noise level, when the seizure activity is changed in frequency during the time interval under investigation.*

that, although the signal is not well localised in frequency, the decomposition still reliably detects the correct location. This does not mean that the seizure activity is fully assigned to one atom. When we looked at the frequency component of the epileptic atom, this component had maximal values around 6 Hz, *i.e.,* the average of the start (8 Hz) and end frequency (4 Hz), while the frequency component in simulation 1 peaked around 4 Hz. When the epileptic atom is reconstructed (Fig. 7.10b), the change in frequency is not captured and the reconstruction is poor in the beginning and at the end. This is also reflected by a lower CPA fit percentage. In the first simulation, the fit percentage was about 75%, while in this



(a) (b)

Figure 7.10: *(a) The simulated frequency-modulated signal used in figure 7.9 (b) The reconstructed atom.*



(a)

(b)

Figure 7.11: *(a) The original dipole location of two simulated dipoles. (b) The dipole localisation when three atoms were estimated with CPA.*

simulation only 58% of the activity could be modeled. However, the best trilinear approximation provides a good localisation.
Fig. 7.11 shows in (a) the simulated location of two close dipoles and in (b) the estimated localisation with the proposed method at an SNR of 0.7. The Corcondia [320] indicated that three atoms was the correct number of atoms for this simulated EEG. Two of them corresponded to the 2 dipolar foci. The localisation error

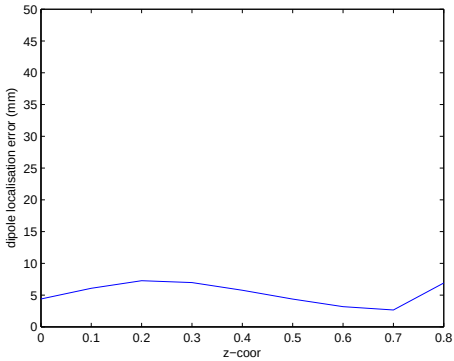


Figure 7.12: *The dipole localisation error as a function of the z - coordinate of the dipole, in order to assess the influence of the depth of the dipole location. The figure shows that the localisation does not influence the accuracy.*

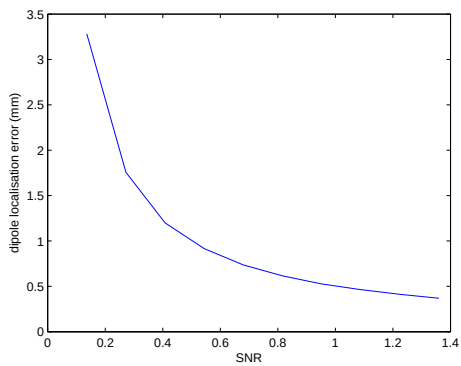


Figure 7.13: *The dipole localisation error as a function of the noise level, when the EEG is recorded with 148 electrodes.*

was for both sources about 5mm, which indicates that a reliable separation and localisation was obtained.
The dipole localisation error as a function of the position of the dipole are shown in Fig. 7.12.
The last figure, Fig. 7.13, shows the dipole estimation error when 148 electrodes are used to acquire the EEG. It can be seen that with a high spatial sampling, the estimation accuracy became about 1 mm.

7.4 Clinical validation on ictal EEGs

7.4.1 Materials

EEG acquisition

Video-EEGs were recorded on 21-channel OSG EEG recorders (Rumst, Belgium). Electrodes were placed according to the International 10-20 System with additional sphenoidal electrodes. Sampling frequency was 250 Hz and an average reference montage was used. The EEG was digitally filtered by a band pass filter (0.3-35Hz). A notch filter was applied to suppress the 50 Hz power-line interference.

Data selection

Patients, who underwent a full presurgical evaluation for refractory partial epilepsy, were included when seizure semeiology, structural MRI, interictal EEG, subtraction ictal SPECT co-registered with MRI (SISCOM) and neuropsychological assessment were concordant, and reliably defined the epileptogenic zone. Ictal EEG findings were not an inclusion criterion. SISCOM that was concordant with other data was a selection criterion. SISCOM was considered the gold standard with which we compared the localisation of ictal EEG as determined by visual assessment and CPA. SISCOM images were displayed on a standard MRI in MNI (Montreal Neurological Institute) space using a minimum threshold of z=2 (for details on this procedure, see [264]). An epileptologist, who disposed of all the data of the presurgical evaluation, selected the ictal EEG recordings. The selected EEG was of the seizure during which the ictal SPECT injection was given. One ictal EEG per patient was used in this study. Thirty-seven EEGs were included in the study. The ictal EEGs were also presented to a clinical neurophysiologist/epileptologist, who was blinded to all other clinical and localising data, for visual assessment of the ictal EEGs and determination of the localisation and lateralisation of the ictal onset. The validation study is performed on the same data set as described in [350]. In this paper, all details concerning the optimal conditions under which visual reading was performed are described (*e.g.*, band pass filtering as well as the amplitude could be freely adjusted by the human reader). Localisation of ictal EEG (both visual assessment and CPA) was considered correct when lobe of onset was concordant with SISCOM findings.

Thirty-seven patients fulfilled the inclusion criteria. Twenty-four were women. Median age was 33 years (range: 14-62). Median age at seizure onset was 14 years (0-62). The median seizure frequency per month was 6 (1-600). MRI showed unilateral hippocampal sclerosis (n=10), focal cortical dysplasia (n=5), scar tissue

(n=5), a tumor (n=3), dual pathology (n=2), a cavernous angioma (n=1) or no abnormality (n=10). Seventeen patients underwent an operation. With a follow-up of more than 2 years, 12 have remained seizure free, and 5 almost seizure free. Five patients refused surgery, and 15 were not offered surgery, because the epileptogenic zone involved eloquent cortex (n=6) or the MRI did not reveal an abnormality (n=9). The ictal SPECT injection was given during a complex partial seizure (n=32), simple partial seizure (n=2) or secondarily generalised seizure (n=3). The median duration of the injected seizure was 70 seconds (range: 11-389) and the median time of injection was 26 seconds after seizure onset (range: 3-109).

7.4.2 Statistics

In order to estimate the statistical significance of the improvement of this automatic localisation compared with blinded human reading, a non-parametric signed rank test was performed.

7.4.3 Results of validation study

CPA correctly localised the seizure onset zone in 34 of 37 cases (92%) while a human reader, blinded to all other information, was able to localise 21 of 37 cases (57%). This improvement is significant (p= 0.00024). In 3 EEGs no automatic localisation was possible. In one seizure, there was an electrode artifact during the whole seizure, which impeded a reliable decomposition. Two spatial distributions did not clearly reveal the epileptical focus, because there were high potentials at both sides of the brain, which indicates that no perfect separation of ictal activity from background EEG was obtained with the CPA. Localisation can only be achieved when all the active electrodes are adjacent. Otherwise the decomposition is not meaningful. The human reader did not recognise ictal activity in 3 cases, was not able to localise the seizure onset in 7 cases, and wrongly localised the seizure onset in 6 cases. Figures 7.14-7.16 illustrate our results with 3 cases. Fig. 7.14A shows a 10 s epoch of EEG with the beginning of a frontal lobe seizure in the last two seconds, which could not be localised by visual inspection of the EEG. However, the CPA localisation (Fig. 7.14B), obtained with 2 seconds of EEG starting from the red line, was concordant with the SISCOM findings (Fig. 7.14C). Fig. 7.14D shows the first left singular vector, which does not provide a clear localisation. Figures 7.14E1-2 show the atoms of the decomposition of the same EEG when the power of the wavelet-transformed EEG was used in the decomposition. None of the images revealed the epileptic focus. The figure illustrates that the pure wavelet transformed EEG is much less influenced by the muscle artifacts than the power of the wavelet transformed EEG. When computing the power, the underlying structure is not trilinear anymore. Fig. 7.15 illustrates

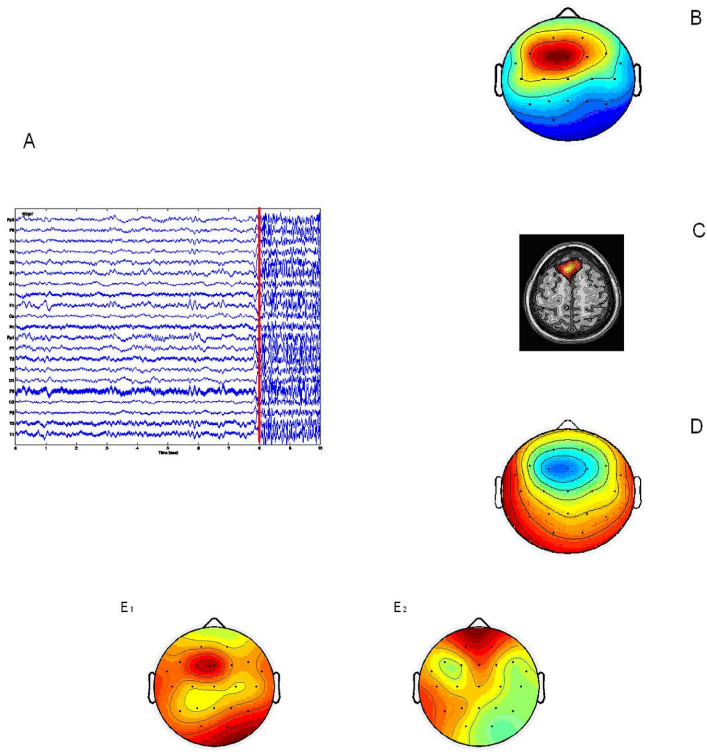


Figure 7.14: *(A) 10 s EEG recording of a left frontal lobe seizure, the red line indicates the start of the 2 seconds used in the CPA (B) The spatial distribution of the epileptical atom (red area) and (C) SISCOM. The ictal EEG (last two seconds) was severely contaminated by muscle artifacts, and localisation by visual assessment was impossible. CPA, however, showed a potential distribution concordant with SISCOM. (D) The first singular vector did not give clear localising information. (E1-2) Spatial distributions of the 2 atoms of the decomposition of the power of wavelet transformed EEG did not reveal the epileptic focus.*

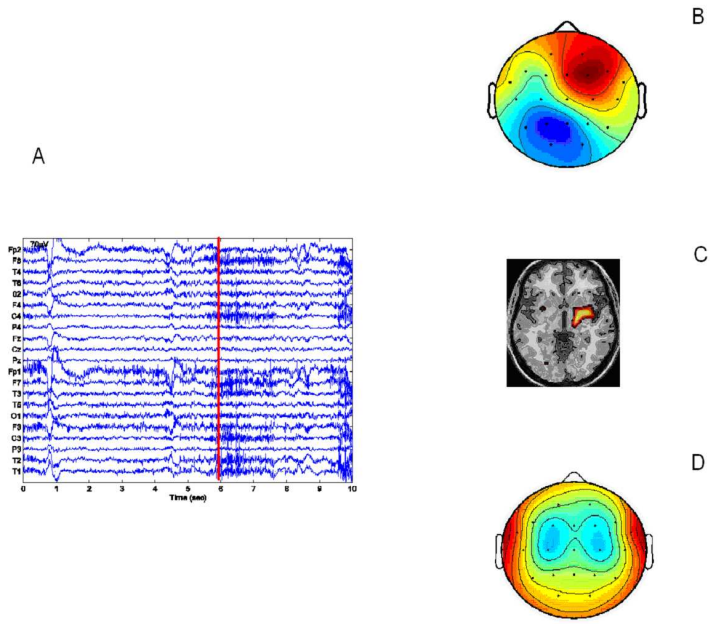


Figure 7.15: *(A) 10 seconds of EEG during a seizure without clear ictal EEG changes, the red line indicates the start of the 2 seconds used in CPA. (B) The spatial distribution of the epileptic atom (red area) and (C) SISCOM. The spatial distribution of the identified atom was concordant with SISCOM. (D) The first singular vector was not informative.*

the sensitivity of CPA: the ictal activity was not visible for a human reader, while the decomposition findings were quite in agreement with SISCOM. Fig. 7.16 shows the data of a patient with a right posterior focus, which was not detected with CPA using a time window of 2 seconds due to the rhythmic eye artifacts. When removing the eye artifacts, or computing the decomposition between 2 eye blinks, the correct focus was obtained (Fig. 7.16B). The SVD detected remaining eye artifacts (Fig. 7.16D_1) and did not reveal the focus in other singular vectors (Fig. 7.16D_2). Note that this is patient 8 in the previous chapter.

7.5 Discussion

We introduced a new method for extracting the potential distribution of ictal EEG activity. The method allows for a fast and sensitive visualisation of the ictal onset zone. The method is based on the multi-way CPA of wavelet-transformed EEG in distinct 'atoms'. After the decomposition, one atom can be identified as the

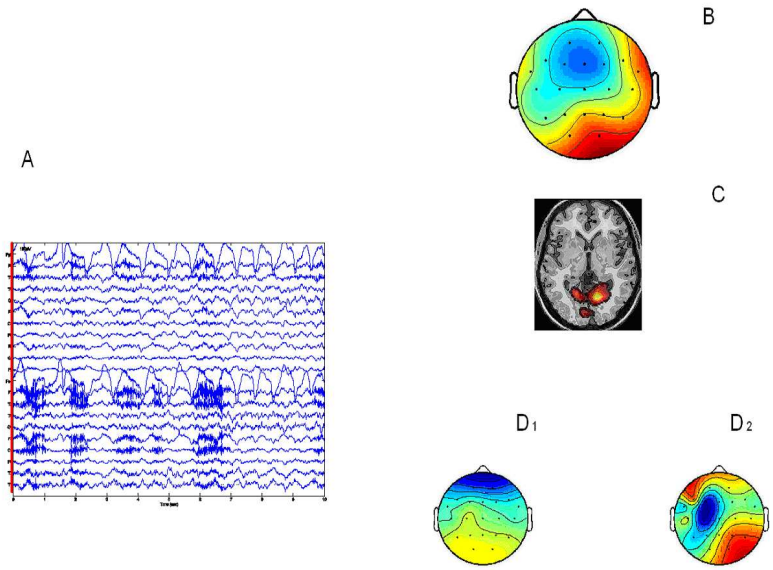


Figure 7.16: *(A) 10 s EEG recording of a right occipital seizure showed frequent eyeblinks (Fp1 and Fp2), but no clear ictal activity, the red line indicates the start of the 2 seconds used in the CPA (B) The spatial distribution of the epileptical atom and (C) SISCOM showed a concordant focus in the right occipital lobe. (D1-2) Two singular vectors did not give clear localising information.*

epileptical atom, and the spatial component of this atom reveals the focus. To the best of our knowledge, we were the first to propose a decomposition method that separates ictal activity from background EEG into one component. We claim that CPA exploits the fact that seizure activity remains more localised in space and frequency during a short time-interval than background EEG signals. The components in CPA should model most of the variation of the array, *i.e.*, model the dominant activity. This modeled activity will be a signal with high amplitude or a signal with stable components in frequency and spatial distribution. The background EEG and random noise (*e.g.*, muscle artifacts) cannot be modeled with this simple CP structure. Hence, CPA will be insensitive to such random artifacts. This is an advantage, as there is no need to first remove these artifacts. All EEG signals that cannot be modeled contribute to the residuals.

Seizure activity is often obscured by artifacts, which complicates visual analysis. These artifacts can be removed with different techniques, see *e.g.*, [340, 89] and references herein. Our BSS-CCA method for muscle artifact removal made visual analysis of ictal EEG more sensitive [350]. Our current CP decomposition is also more sensitive than visual assessment, with the additional advantage that the

technique is automatic and more objective, *i.e.*, not dependent on a subjective visual interpretation. By means of simulations, we showed that the method is able to extract the correct potential distribution, even at very low SNRs. The validation study demonstrated that the method may be a valuable and objective localising tool for the epileptologist in clinical practice. As opposed to *e.g.*, [251], in which the complex wavelet was used to obtain a third order tensor, we used a real wavelet. In addition, we did not decompose the power of the wavelet-transformed EEG. We demonstrated that by decomposing 'pure' wavelet-transformed EEG, the method became very robust for muscle artifacts and low SNRs because random "noise" cannot be modeled with this restricted CP decomposition.

With simulations, we also tried to assess the potential of applying dipole localisation after CP decomposition. It is known that an infinite number of internal electrical currents correspond with exactly the same potential distribution on the scalp. The discussion if dipolar sources are superior to distributed sources is beyond the scope of this study. We chose the dipolar source because it is most popular. It is known that the generator of ictal activity can be an extended area, and that a dipole situated in a certain region should be considered as the center of mass of a larger activated brain region [246]. In [251], source densities were computed after CPA.

We focused in the study on seizure activity, but the method can also be used to localise other origins of oscillatory activity [345]. We have shown that in a spherical head model with realistically simulated EEG, our algorithm correctly localised the seizure-related atom with an accuracy of about 5 mm, even at SNR ratios that are lower than one encounters during real ictal recordings. SNRs below 1 mean that the signal contains more noise than signal (*e.g.*, Fig. 7.6). Although the shape of seizure activity will not be perfectly sinusoidal, we have shown that the exact shape of the seizure signal did not really influence the localisation accuracy. In a second simulation, we have shown that the localisation error does not depend on the frequency of the epileptic signal, and that overlapping frequency content of signal and noise, representing muscle artifacts, does not lower the reliability of the decomposition. The third simulation investigated a more challenging, but maybe more realistic situation in which the frequency of the seizure changed during the considered time interval. The resulting atom could not fully capture the exact frequency-varying signal, as indicated by a lower fit-percentage of CPA and the reconstructed epileptic signal. However, the best tri-linear approximation still reliably localised the signal. We should emphasise that CPA is an interesting decomposition method due to its uniqueness properties. However, the trilinear decomposition in space - time -frequency components really restricts the activity that can be fully captured. When the frequency content changes in time at a fixed position, the exact signal will not be fully extracted. However, the best tri-linear approximation will find a rhythmical signal at the correct location. A similar result is observed when a moving dipole was simulated. Moving activity can not

be captured with a trilinear model, but the best approximation will result in an "average" localisation. The fourth simulation showed that the localisation error is quite insensitive to dipole location. In [155], it was observed that dipoles closer to the scalp, are slightly better estimated due the higher SNR associated with higher dipoles. However, in our simulation this effect is negligible.

We investigated also the situation in which two dipolar sources generating the same signal were placed near each other. This simulates multifocal epilepsy. The Corcondia [320] indicated that three atoms was the correct number of atoms for this simulated EEG. Two of them corresponded to the two dipolar sources. This example illustrates the interesting uniqueness property of CPA [320] for EEG source localisation. When matrix decomposition techniques like SVD or Independent Component Analysis (ICA) would have been used to decompose the EEG, the two simulated sources would not have been separated because they are not independent neither uncorrelated. It would then not be obvious to determine the correct number of dipoles. In our approach, CPA determines the optimal number of components and only one dipole will correspond to each atom. Tensor decomposition techniques offer clearly advantages over matrix decomposition techniques as preprocessing technique for EEG source localisation. The last simulation assessed the accuracy when more electrodes are used. It is known that dipole localisation based on 21 electrode measurements gives only an approximate indication of source localisation. However, using 148 electrodes can reduce the dipole estimation error to less than one mm at the same low SNRs. So we think it is worth to record the EEG with denser spatial sampling.

We have used SISCOM as a gold standard for the ictal onset zone. It is well known that the time resolution of ictal SPECT is poor and that areas of hyperperfusion often represent a combination of ictal onset zone and propagated activity, even when ictal SPECT injections were given early during a seizure. Selection of patients who were rendered seizure free after surgery, would have made our sample size rather small. The aim of the study was to compare the sensitivity of two methods of analysis of the same EEG data set (visual analysis versus CPA) to detect the ictal onset zone. We believe that our current inclusion criteria were sufficient to reliably indicate the lobe of seizure onset and thus sufficient to validate our method.

Nowadays, dipole modeling or other source localisation techniques can rarely be applied to ictal EEGs, because the EEGs are too much contaminated by artifacts. The advantage of the extraction of the whole potential distribution of the epileptical activity in one component opens perspectives for source localisation techniques of seizure onset activity in all patients. This would be an extension of the work done by other groups concerning source localisation after signal decomposition, see *e.g.*, [192, 193]. It would open the way for a multimodality imaging platform in which the epileptic lesion is displayed together with two methods to localise the ictal onset zone, namely ictal SPECT and the ictal onset

zone detected by CPA of the ictal scalp EEG, which is routinely and easily obtained during long-term video-EEG monitoring.

The length of the time window in which a CPA is computed, is not a critical issue as long as the seizure activity remains well localised. When analysing a seizure, it is important that the CP decomposition is computed in a time window at the seizure onset in order to localise seizure onset zone. It will be interesting to compute different decompositions with a short time window at consecutive time points during a seizure. Each decomposition will have localising information for the considered time interval, and may highlight seizure propagation patterns. Our CPA could be important in the presurgical evaluation of frontal lobe seizures that are characterised by fast propagation and ictal EEGs that are difficult to interpret by visual assessment. Because of potential fast propagation, we used a time window of two seconds in our study, which is probably the smallest window in which our algorithm will give reliable information. We postulate that our CPA, with high temporal resolution, may improve the interpretation of SISCOM, which has an excellent spatial resolution, but notoriously poor temporal resolution (*i.e.*, it shows a combination of ictal onset zone and seizure propagation pathways).

We implemented the inverse continuous wavelet transform in order to reconstruct the ictal atom in the standard EEG matrix. Now, the clinician is also able to look at the seizure activity on EEG, separated from artifacts and background signal.

The aim of epilepsy surgery is the complete resection or complete disconnection of the epileptogenic zone, which is defined as the area of cortex indispensable for the generation of clinical seizures. We use a variety of diagnostic tools, such as analysis of seizure semiology, EEG recordings, functional testing and neuroimaging techniques to define the location and boundaries of the epileptogenic zone. It should be stressed that a priori knowledge of the exact timing of the clinical seizure was essential in the application of our algorithm, i.e., it is not a seizure detection algorithm. We showed that our algorithm was more sensitive in the delineation of the ictal onset zone than visual assessment of the ictal EEG according to current standards. The exact relationship, however, between "activated electrodes" using our algorithm and the ictal process in the brain will need further validation studies.

In this study, we only evaluated one EEG recording from each patient. It will be interesting to analyze all seizures recorded from the same patient during a presurgical evaluation using our CPA. When all epileptical potential distributions obtained with CPA at seizure onset are very similar, it can be expected that there is only one epileptogenic focus. However, when different distributions are obtained, there might be different foci and surgery may not be a good option. Although this is current practice in the presurgical evaluation of patients with refractory partial epilepsy, we anticipate that the higher sensitivity of CPA as compared with visual assessment of the ictal EEGs will improve and streamline the non-invasive presurgical evaluation of patients with refractory partial epilepsy.

7.6 Conclusion

We proposed a method based on CPA to localise the seizure onset. The current simulation study is the most reliable validation of our method. It is not yet possible to validate the 3D method on *in vivo* seizures for several reasons. First, we have to define a gold standard. This gold standard can be intracranial EEG, ictal SPECT, or the site of epilepsy surgery in patients who were rendered seizure free. Comparing the estimated dipole localisation with localising information derived from other data, like ictal SPECT or MR–visible lesions, however, will be biased by the accuracy of the onset delineation with these diagnostic tools. Second, we speculate that a single dipole is a too simple model for modeling the complex phenomenon of a seizure. So further research has to be done to clarify this. At this moment, we anticipate that the higher sensitivity and objectivity of the CPA as compared with visual assessment of the ictal EEGs will improve and streamline the non-invasive presurgical evaluation of patients with refractory partial epilepsy. We can mention that in clinical practice, CPA is already sometimes used to confirm seizure onset localisation.

Independently of us, [1] proposed also a CP-based method for localising the seizure onset. However, that method has two pitfalls: it uses the power of wavelet-transformed EEG, and it does not take into account that seizures can propagate. Our method has also shown to be useful for localising neonatal seizures [87]. That CPA will become of increasing importance in EEG research is also reflected by more recent publications, *e.g.*, [260, 171]. It is interesting that space–time–frequency methods are also proposed without CPA [342]. With the development of new algorithms [275], it will also become possible to track the CP decomposition over time. This can become useful for visualising propagation patterns in the EEG.

Finally, the introduction of this thesis showed a representative example of ictal EEG (Fig. 1.1). In the different chapters, we showed decompositions of this EEG obtained by different methods (CCA, ICA, CPA). In each case, different source signals were obtained. We mentioned in the beginning that specific assumptions have to be made in order to obtain a decomposition. If the assumptions are valid, meaningful sources can be estimated. Increasing the number of electrodes can make a technique more reliable, because the signal subspace can be estimated more reliably. However, it is very unlikely that in the future, there will be one optimal all-round decomposition method. Decomposition methods can be very useful in very different situations, but they have to be applied with knowledge. As a general recommendation, everybody facing a decomposition problem should be very careful to spend some time on checking the properties of the sources that one wants to extract (statistically independent, temporally colored, sparse, ...) and the properties of the mixture (linear, instantaneous, ...). Only the

selection of the appropriate BSS method will give a good decomposition for the data under investigation. Without the matching of the mathematical properties of the method and the physical properties of sources of the mixture, no good solution can be obtained!

Chapter 8

A combination of Parallel Factor and Independent Component Analysis

Although CPA provides a unique trilinear decomposition, it can be useful to impose extra constraints. In this chapter, we will impose independence constraints, and show that our method outperforms in certain situations both standard ICA and standard CPA.

8.1 Introduction

In this last chapter, the research was inspired by a fMRI application that is intrinsically higher dimensional. One typically records tensors with dimensions (voxels, time, trial) or (voxels, time, session) or (voxels, time, subject) or combinations thereof. In [6], CPA was proposed to analyse fMRI data in its higher-dimensional form. Afterwards, it was noticed that the original CPA could suffer from interference between components [14]. Based on a simulation study in which artificial fMRI activations (Fig. 8.1) were mixed in different strenghts (related to different persons), it was claimed that a combination of ICA and CPA was more robust than CPA as such.

This is an interesting new concept. In fact, the CP model is unique. But in real-world applications it can still be useful to impose extra constraints to increase the robustness of the estimation. One could for instance impose non-negativity [47]. The authors of [14] proposed a trilinear decomposition with independence

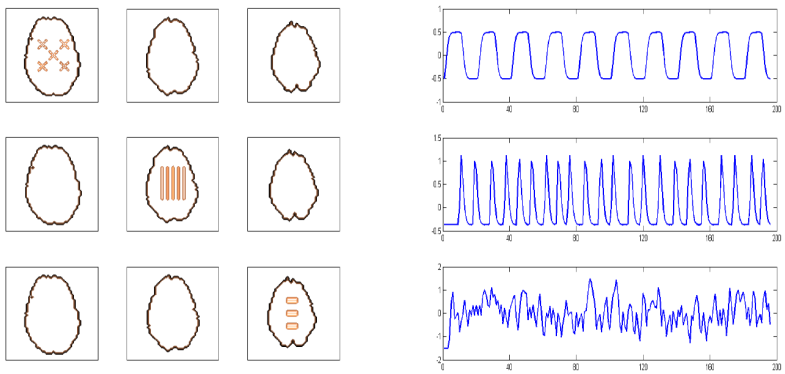


Figure 8.1: *Artificial spatial maps and time-courses used for the generation of artificial group fMRI data.*

imposed in one mode. However, the multi-way structure was imposed after the computation of the independent components. A discussion of [14] can be found in [327]. In this chapter, we propose an algorithm that imposes the CP structure during the ICA computation. We will refer to it as the ICA-CP algorithm.

The results are potentially useful for many applications. CP has found numerous applications in wireless communication [313, 312]. In these applications, it is natural to impose statistical independence of the transmitted signals. A biomedical application is discussed in [14].

The basic concepts of ICA and CP are known from chapter 3. We start with giving an ICA algorithm based on soft whitening. In section 8.2 we derive a new algorithm for computing ICA-CP. We perform a simulation study to investigate when the proposed method outperforms ordinary ICA and CP (section 8.3). We conclude with a discussion of the results (section 8.4). Our formulation is in terms of real data. The derivation also holds for complex data, with trivial modifications.

ICA based on Soft whitening

We start again from the model in Equation 3.18:

$$\mathbf{X} = \mathbf{M} \cdot \mathbf{C}. \tag{8.1}$$

A balanced way to combine equations (3.20) and (3.21), is called soft whitening [361, 215]. The problem of soft whitening can be formulated as the estimation of a matrix $\mathbf{M} \in \mathbb{R}^{L \times R}$ that minimises the cost function

$$\begin{aligned} f(\mathbf{M}, \mathbf{C}_\mathbf{C}, \mathcal{C}_\mathbf{C}^{(N)}) = \quad & (1-w) \frac{\|\mathbf{C}_\mathbf{X} - \mathbf{C}_\mathbf{C} \times_1 \mathbf{M} \times_2 \mathbf{M}\|^2}{\|\mathbf{C}_\mathbf{X}\|^2} \\ + w \quad & \frac{\|\mathcal{C}_\mathbf{X}^{(N)} - \mathcal{C}_\mathbf{C}^{(N)} \times_1 \mathbf{M} \times_2 \mathbf{M} \ldots \times_N \mathbf{M}\|^2}{\|\mathcal{C}_\mathbf{X}^{(N)}\|^2}, \end{aligned} \tag{8.2}$$

in which $\mathbf{C}_\mathbf{C} \in \mathbb{R}^{R \times R}$ is an unknown diagonal matrix and $\mathcal{C}_\mathbf{C}^{(N)} \in \mathbb{R}^{R \times R \times \ldots \times R}$ an unknown diagonal tensor, and in which $\mathbf{C}_\mathbf{X}$ and $\mathcal{C}_\mathbf{X}^{(N)}$ are the covariance matrix and the cumulant tensor of $\mathbf{X}$. A positive value $0 \leqslant w \leqslant 1$ is used to weigh the confidence in equations (8) and (9). A small value of w reflects that one has much confidence in the estimate $\mathcal{C}_\mathbf{X}$ and little confidence in the estimate $\mathcal{C}_\mathbf{X}^{(N)}$ (e.g., in the case of short datasets). When one considers $\mathcal{C}_\mathbf{X}^{(N)}$ as more reliable than $\mathbf{C}_\mathbf{X}$ (e.g., in the presence of colored Gaussian noise), one will increase the value of w.

An efficient way to minimise eq. (8.2), is by rewriting the problem as a weighted simultaneous matrix diagonalisation [319, 180]. In [37], it was derived that the matrix $\mathbf{M}$ that minimises the second part of eq. (8.2), can be estimated as the matrix that simultaneously diagonalises the R dominant eigenmatrices $\mathbf{U}_r, r = 1, \ldots, R$ of the fourth-order cumulant $\mathcal{C}_\mathbf{X}^{(4)}$. These eigenmatrices are the matricised eigenvectors of the corresponding matrix $\mathbf{C}_\mathbf{X}^{(4)} \in \mathbb{R}^{L^2 \times L^2}$, defined as

$$(\mathbf{C}_\mathbf{X}^{(4)})_{i+I(j-1),k+K(l-1)} = (\mathcal{C}_\mathbf{X}^{(4)})_{ijkl}. \tag{8.3}$$

This can be written in another way as

$$\begin{aligned} \mathbf{U}_1 \quad &= \mathbf{M} \cdot \mathbf{D}_1 \cdot \mathbf{M}^T \\ &\vdots \\ \mathbf{U}_R \quad &= \mathbf{M} \cdot \mathbf{D}_R \cdot \mathbf{M}^T. \end{aligned} \tag{8.4}$$

In order to minimise also the first part of eq. (8.2), we add to this set of matrices the covariance matrix

$$\mathbf{U}_{R+1} = \mathbf{C}_\mathbf{X} = \mathbf{M} \cdot \mathbf{D}_{R+1} \cdot \mathbf{M}^T. \tag{8.5}$$

When we normalise the set of equations (8.4) and (8.5) based on their Frobenius norm, and weigh them with respectively a factor w and $1 - w$, the resulting simultaneous diagonalisation will provide an estimate of the mixing matrix $\mathbf{M}$.

8.2 Combination of ICA and CPA: ICA-CPA.

In this section, we review the tensorial extension of ICA, called 'tensor pICA', as it was introduced in [14]. Then we present a new algorithm to compute CPA of a tensor $\mathcal{X}$ where independence is imposed on the factors in one mode. For notational convenience, we will restrict the exposition to the third-order case, but the generalization to higher orders is straightforward. In the following, we always assume that the components in the third mode are independent. Due to the symmetric structure of CPA, similar equations can be derived for the other two modes.
In equation (3.18), we assume that we have K realizations of the observation vector, represented by a matrix $\mathbf{X} \in \mathbb{C}^{L\times K}$. We also assume that the mixing matrix $\mathbf{M}$ has the Khatri-Rao structure $(\mathbf{A}\odot\mathbf{B})$. Alternatively, seen from equation (3.47), matrix $\mathbf{S}$ contains the independent sources and $(\mathbf{A}\odot\mathbf{B})$ corresponds to the mixing matrix $\mathbf{M}$. The rank of $\mathbf{X}$ corresponds to the number of sources R if $(\mathbf{A}\odot\mathbf{B})$ and $\mathbf{S}$ have both full column rank.

8.2.1 Tensor pICA

In [14], the following algorithm was proposed.

1. Perform an iteration of the two-dimensional probabilistic ICA algorithm for the decomposition of a data matrix $\mathbf{X} \in \mathbb{C}^{IJ\times K}$ in the product of a mixing matrix $\mathbf{M} \in \mathbb{C}^{IJ\times R}$ and associated source matrix $\mathbf{S} \in \mathbb{C}^{K\times R}$: $\mathbf{X} = \mathbf{M}\mathbf{S}^T + \tilde{\mathbf{E}}_1$.

2. Decompose the estimated mixing matrix $\mathbf{M}$ such that $\mathbf{M} = \mathbf{A}\odot\mathbf{B} + \tilde{\mathbf{E}}_2$. Each column in $\mathbf{A}\odot\mathbf{B}$ is formed by J scaled repetitions of a single column from $\mathbf{A}$, the scaling factors being given by the corresponding column of $\mathbf{B}$. In order to obtain $\mathbf{A}$ and $\mathbf{B}$, we define matrices $\mathbf{G_1}, \ldots, \mathbf{G_R} \in \mathbb{C}^{I\times J}$ as

$$(\mathbf{G_r})_{ij} = m_{(i-1)J+j,r} \qquad \forall i, j, r \tag{8.6}$$

   $\mathbf{a}_r$ and $\mathbf{b}_r$ can then be computed as the dominant left and right singular vector of $\mathbf{G_r}, 1 \leqslant r \leqslant R$.

3. Iterate the decomposition over steps 1 and 2 until convergence, i.e., until $||\mathbf{A}^{\text{new}} - \mathbf{A}^{\text{old}}||_F + ||\mathbf{B}^{\text{new}} - \mathbf{B}^{\text{old}}||_F + ||\mathbf{C}^{\text{new}} - \mathbf{C}^{\text{old}}||_F < \epsilon$. This third step is in reality redundant, as the decompositions in step 1 and 2 are deterministic and always give the same result.

This algorithm was called tensor pICA, but, roughly speaking it is ordinary ICA, with afterwards the best rank-1 approximation of the mixing vectors, interpreted as $(I \times J)$ matrices.

8.2.2 Imposing independence constraints in CPA: ICA-CPA

Formulation as a fifth-order CPA with partial symmetry

Recall that in soft whitened ICA, in order to estimate the mixing matrix, the R dominant eigenmatrices $\mathbf{U}_r$ of $\mathcal{C}_{\mathbf{x}}^{(4)}$ have to be jointly (approximately) diagonalized together with the covariance matrix. After diagonalization, we can write:

$$\mathbf{U}_{\tilde{r}} = \mathbf{M} \cdot \mathbf{D}_{\tilde{r}} \cdot \mathbf{M}^H, \qquad 1 \leqslant \tilde{r} \leqslant R+1 \tag{8.7}$$

with $\mathbf{D}_{\tilde{r}}$ a diagonal matrix. Stack the diagonal elements of the matrices $\mathbf{D}_{\tilde{r}}$ in a matrix $\mathbf{D} \in \mathbb{C}^{(R+1)\times R}$ as follows:

$$(\mathbf{D})_{wr} = (\mathbf{D}_w)_{rr} \quad \forall r, w. \tag{8.8}$$

We stack the $(R+1)$ matrices $\mathbf{U}_{\tilde{r}}$ one after the other in a tensor $\mathcal{U} \in \mathbb{C}^{IJ\times IJ\times(R+1)}$ as follows:

$$(\mathcal{U})_{pqw} = (\mathbf{U}_w)_{pq} \quad \forall p, q, w. \tag{8.9}$$

Equation (8.7) is a standard formulation of the CPA of $\mathcal{U}$ [320]:

$$\mathcal{U} = \sum_{r=1}^{R} \mathbf{m}_r \circ \mathbf{m}_r^* \circ \mathbf{d}_r. \tag{8.10}$$

Equation (8.10) can be written in an element-wise manner as follows:

$$u_{pqw} = \sum_{r=1}^{R} m_{pr} m_{qr}^* d_{wr}. \tag{8.11}$$

Now stack $\mathcal{U}$ in a fifth-order tensor $\mathcal{U}^{(5)} \in \mathbb{C}^{I\times J\times I\times J\times(R+1)}$ as follows:

$$(\mathcal{U}^{(5)})_{ijkln} = \mathcal{U}_{(i-1)J+j,(k-1)L+l,n}. \tag{8.12}$$

Due to the Khatri-Rao structure of the mixing matrix, namely $\mathbf{M} = \mathbf{A} \odot \mathbf{B}$, $\mathcal{U}^{(5)}$ has the following CPA structure:

$$\mathcal{U}^{(5)} = \sum_{r=1}^{R} \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{a}_r^* \circ \mathbf{b}_r^* \circ \mathbf{d}_r. \tag{8.13}$$

Indeed, equation (8.13) can be written in an element-wise manner as follows:

$$(\mathcal{U}^{(5)})_{ijkln} = \sum_{r=1}^{R} a_{ir} b_{jr} a_{kr}^* b_{lr}^* d_{nr}, \tag{8.14}$$

which is equivalent to equation (8.11). After obtaining estimates $\hat{a}_r$ and $\hat{b}_r$, the mixing matrix is estimated as $\hat{\mathbf{M}} = \hat{\mathbf{A}} \odot \hat{\mathbf{B}}$ and $\mathbf{S}$ is estimated as

$$\hat{\mathbf{s}} = \hat{\mathbf{M}}^{\dagger} \cdot \mathbf{x}. \tag{8.15}$$

The question is now how decomposition (8.13) can be computed, taking into account the partial symmetry.

Computing the fifth-order CPA with partial symmetry

It is still largely an open problem how CPA can be computed in the case of full or partial symmetry. In this section we present a solution. Namely, we explain that symmetry is naturally preserved in line search schemes.

We want to fit the decomposition in (8.13) in least squares sense, *i.e.*, we want to minimize the following cost function:

$$\phi(\mathbf{A}, \mathbf{B}, \mathbf{D}) = \|\mathcal{U}^{(5)} - \sum_{r=1}^{R} \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{a}_r^* \circ \mathbf{b}_r^* \circ \mathbf{d}_r\|_F^2.$$

In a line search scheme, $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{D}$ are updated in the following manner:

$$\begin{aligned} \mathbf{A}_{p+1} &= \mathbf{A}_p + \lambda \boldsymbol{\Delta}\mathbf{A}_p, \\ \mathbf{B}_{p+1} &= \mathbf{B}_p + \lambda \boldsymbol{\Delta}\mathbf{B}_p, \\ \mathbf{D}_{p+1} &= \mathbf{D}_p + \lambda \boldsymbol{\Delta}\mathbf{D}_p, \end{aligned} \tag{8.16}$$

in which $\boldsymbol{\Delta}\mathbf{A}_p$, $\boldsymbol{\Delta}\mathbf{B}_p$, $\boldsymbol{\Delta}\mathbf{D}_p$ are the search directions after the pth iteration step for $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{D}$, respectively, and in which λ is the step size. If $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{D}$ are updated as in (8.16), then (partial) symmetry is naturally preserved. This is an advantage of line search schemes that is not fully appreciated in the literature. We mention here that the popular ALS scheme [141, 320] breaks the symmetry since different estimates of $\mathbf{A}$ are obtained in the first and third mode, and different estimates of $\mathbf{B}$ are obtained in the second and fourth mode.

The line search approach yields two subproblems: (i) the determination of good search directions and (ii) the determination of a good step size. These aspects will now subsequently be discussed.

It makes sense to search along the line that connects the current estimate of $\mathbf{A}$ ($\mathbf{B}$, $\mathbf{D}$) with the estimate $\mathbf{A}_{ALS}$ ($\mathbf{B}_{ALS}$, $\mathbf{D}_{ALS}$) that would be obtained after one ALS step. That is, the search directions could be defined as follows:

$$\begin{aligned} \mathbf{\Delta A}_p &= \mathbf{A}_{ALS} - \mathbf{A}_p, \\ \mathbf{\Delta B}_p &= \mathbf{B}_{ALS} - \mathbf{B}_p, \\ \mathbf{\Delta D}_p &= \mathbf{D}_{ALS} - \mathbf{D}_p. \end{aligned} \tag{8.17}$$

In this way, the current estimate is obtained for $\lambda = 0$ and $\mathbf{A}_{ALS}$, $\mathbf{B}_{ALS}$, $\mathbf{D}_{ALS}$ are obtained for $\lambda = 1$. Since the line contains the current estimate, the cost function can be monotonically decreased if suitable values for λ are chosen. The Appendix contains the expressions for $\mathbf{A}_{ALS}$, $\mathbf{B}_{ALS}$ and $\mathbf{D}_{ALS}$ that go with decomposition (8.13).

In [141] the step size λ is given a fixed value between 1.2 and 1.3. In [27] λ is set to $p^{1/3}$ and the step along the line is accepted only if the cost function decreases. In [295] it was shown that the optimal real-valued step size can be computed by rooting a polynomial. This procedure was generalized to complex-valued λ in [274]. The technique was called Enhanced Line Search with Complex Step (ELSCS). In the Appendix we explain how one can work along the lines of [274, 295] in the specific case of decomposition (8.13).

8.3 Numerical experiments

In this section, we illustrate the performance of our algorithm by means of numerical experiments and compare it to standard ICA (as a variant of tensor pICA), and standard CPA. Our ICA-CPA algorithm is soft whitening based. To allow for a fair comparison, we report the performance of an ICA method that is also soft whitening based. The latter amounts to the computation of the CP decomposition (8.10) by means of the ALS+ELS algorithm [295]. (Tensor pICA itself was computed with pre-whitened fastICA.) For standard CPA, we fit decomposition (3.5) by means of the ALS+ELS algorithm, without taking into account that the vectors $\mathbf{s}_r$ represent statistically independent signals.

Rank-R tensors $\tilde{\mathcal{X}}$, of which the components in the different modes will be estimated afterwards, are generated in the following way:

$$\tilde{\mathcal{X}} = \sigma \cdot \frac{\mathcal{X}}{||\mathcal{X}||_F} + \frac{\mathcal{N}}{||\mathcal{N}||_F}, \tag{8.18}$$

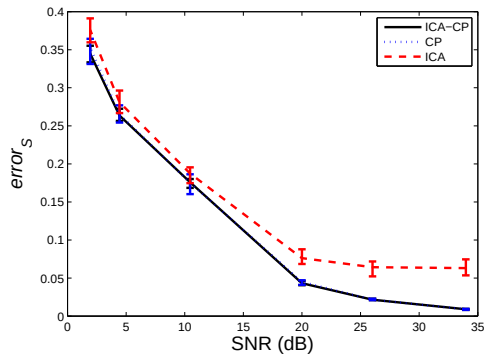


Figure 8.2: *The median value and interquartile range of $error_S$ as a function of the SNR for the algorithms ICA-CPA (solid), CPA (dotted) and soft whitened ICA (dashed). The 4 sources were uniformly distributed. The mixture ($\in \mathbb{R}^{5\times4\times400}$)was optimally conditioned. The noise was colored Gaussian.*

in which $\mathcal{X}$ exactly satisfies the CPA model with $R = 4$ independent sources in the third mode ($\mathbf{S}$), in which $\mathcal{N}$ represents noise, of which the characteristics will be specified later, and in which σ controls the signal-to-noise ratio.

We conduct 5 different experiments, covering very different conditions, with Monte Carlo simulations consisting of 250 runs.

In a first simulation, we generate $\mathcal{X} \in \mathbb{R}^{5\times4\times400}$ with all source distributions (in $\mathbf{S}$) zero-mean and uniformly distributed between $-\sqrt{3}$ and $\sqrt{3}$. The matrices $\mathbf{A}$ and $\mathbf{B}$ have mutually orthonormal columns, so these matrices are optimally conditioned. The entries from these matrices are drawn from a zero-mean unit-variance Gaussian distribution, and afterwards orthogonalized. The noise is colored. It was generated by multiplying the Gaussian noise with a matrix, of which the entries were randomly taken from a uniform distribution. We estimate 4 sources, and stack the entries in $\hat{\mathbf{S}}$. The weighting factor w for soft whitening was chosen equal to 0.3. After estimation, the columns of $\hat{\mathbf{S}}$ are optimally ordered and scaled. We evaluate the performance of the different algorithms by means of the relative Frobenius norm of the difference between the estimated and the real sources:

$$error_S = \frac{||\mathbf{S} - \hat{\mathbf{S}}||_F}{||\mathbf{S}||_F}. \tag{8.19}$$

In Fig. 8.2, we plot the median value and the interquartile range of $error_C$ as a function of the SNR.

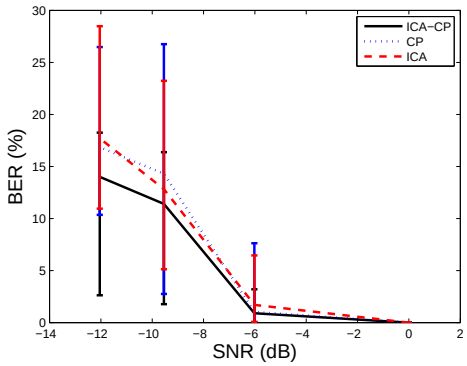


Figure 8.3: *The median value and interquartile range of BER as a function of the SNR for the algorithms ICA-CPA (solid), CPA (dotted) and ICA (dashed). The 4 sources were binary. The mixture ($\in \mathbb{C}^{5\times 4\times 1000}$) was not optimally conditioned. The noise was colored Gaussian.*

The solid, dashed and dash-dotted line represent respectively the accuracy of ICA-CPA, ICA and CPA. It can be seen that the estimation accuracy of standard ICA is worse than that of the two other algorithms.

In a second simulation, we assume that all source distributions are binary (1 or -1), with an equal probability of both values. Tensors $\mathcal{X} \in \mathbb{C}^{5\times 4\times 1000}$ are generated. The entries of the other two modes ($\mathbf{A}$ and $\mathbf{B}$) are drawn from a complex zero-mean unit-variance Gaussian distribution. The noise is colored in the same way as in the previous simulation. The decomposition with correct number of components ($R = 4$) is estimated. The weighting factor w for soft whitening was chosen equal to 0.5. We evaluate the performance of the different algorithms by means of the Bit Error Rate (BER), defined as the percentage of wrongly estimated bits. Fig. 8.3 shows the result of this experiment. It can be seen that for higher noise levels, ICA-CPA outperforms the other two algorithms. Also a high variability of the BER can be observed.

In a third simulation, we generate the 4 sources in $\mathbf{C}$ and the mixing matrix $\mathbf{A} \odot \mathbf{B}$ in the same way as in the second simulation, but we consider tensors $\mathcal{X} \in \mathbb{C}^{5\times 2\times 1000}$. Note that for these dimensions, the CPA decomposition is still unique. We consider white Gaussian noise. The results are shown in Fig. 8.4. The figure shows that when one dimension of the tensor is smaller than the number of sources, a small improvement in accuracy can be obtained by incorporating the rank-1 constraint in the ICA decomposition instead of estimating ordinary CPA.

The fourth simulation (Fig. 8.5) shows the result when the mixture ($\in \mathbb{C}^{5\times 2\times 1000}$) contained 4 sources and colored noise was present. The noise was generated in the

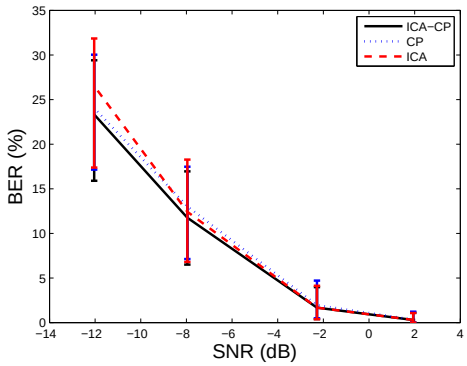


Figure 8.4: *The median value and interquartile range of BER as a function of the SNR for the algorithms ICA-CPA (solid), CPA (dotted) and ICA (dashed). The sources were binary. The noise was white Gaussian. The mixture ($\in \mathbb{C}^{5\times 2\times 1000}$) contained 4 sources.*

same way as in the second experiment. It can be seen that a substantial increase in accuracy can be obtained by ICA-CPA compared to ordinary ICA and CPA.

The last simulation validates the robustness of the method to overestimation of the rank. Tensors $\mathcal{X} \in \mathbb{C}^{5\times 4\times 1000}$ are generated. $\mathbf{A}, \mathbf{B}$ and $\mathbf{C}$ are generated in the same way as the second simulation. We assume again white Gaussian noise and we estimate 5 sources. So we overestimate the number of sources by 1. At the considered high noise levels, it is not always straightforward to estimate the correct number of sources as there is no gap between the contributions. The weighting factor w for soft whitening was again equal to 0.5. Fig. 8.6 shows the BER as a function of the SNR. It can be seen that ICA-CPA is most robust to overestimation of the rank. Ordinary ICA is not robust at all.

8.4 Discussion

In [14], the 'tensor pICA' method was introduced to impose independence constraints in one mode of the CP model. The method boils down to standard ICA, with afterwards the mixing vectors mapped to the closest Kronecker product. In this chapter, we developed a new and improved algorithm for the same problem, which we call ICA-CP. The algorithm simultaneously imposes the independence and the rank-1 structure constraints. We performed several simulations, covering different situations. The simulations were constructed to investigate the accuracy of the new method in (a) an ideal mixture and (b) less ideal mixtures. The latter mixtures were constructed by changing each time only one parameter: adding

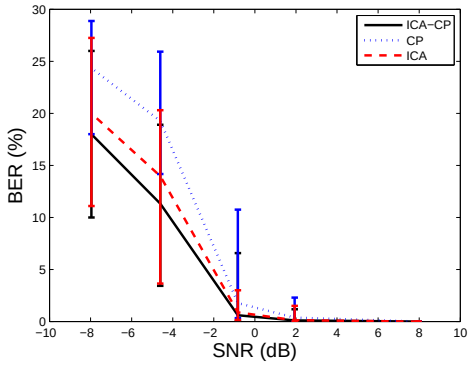


Figure 8.5: *The median value and interquartile range of BER as a function of the SNR for the algorithms ICA-CPA (solid), CPA (dotted) and ICA (dashed). The sources were binary. The noise was colored Gaussian. The observed tensor* $\mathcal{X}$ *(*$\in \mathbb{C}^{5\times 2\times 1000}$*) contained contributions of 4 sources.*

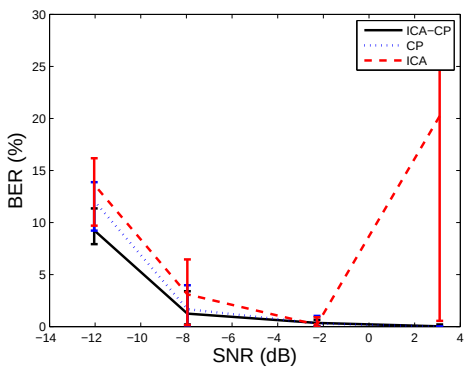


Figure 8.6: *The median value and interquartile range of BER as a function of the SNR for the algorithms ICA-CPA (solid), CPA (dotted) and ICA (dashed). The sources were binary. The noise was white Gaussian. The observed tensor* $\mathcal{X}$ *(*$\in \mathbb{C}^{5\times 4\times 1000}$*) contained contributions of 4 sources. The number of components was overestimated by 1.*

colored noise, considering a tensor with one dimension lower than the number of sources, and overestimating the rank.

In all the experiments, our method outperformed standard ICA. This shows that adding the trilinear constraint improves the accuracy of the decomposition. In a mixture deviating from an ideal case, the increase in accuracy can be very substantial.

It is also interesting to look at the differences between ICA-CP and standard CP. In an ideal mixture or at low noise levels, the use of standard CP is advisable, because it is faster than ICA-CP (because no cumulant has to be computed) and it has comparable accuracy. Although there is a high variability in the results, the simulations show clearly that ICA-CP is more robust than CP with respect to colored noise. Also when one of the dimensions of the tensor is smaller than the number of components, fitting an ICA-CP model should be advised. In [14, 327], it was shown that CP suffered under specific conditions from interference between different components. In these papers, it was claimed that this could be caused by overestimating of the number of components. We performed a simulation with white gaussian noise that confirmed their statement.

Our algorithm is based on soft whitening, which means that second and higher order statistics are taken into account with a predetermined relative weight. It was shown in literature that this approach is sometimes better than pre-whitened based algorithms [36, 319, 361]. The weighting factor w can be optimised for a specific application.

In this paper, we considered an implementation of ICA that was based on the fourth-order cumulant. A similar algorithm can be developed for ICA based on sets of covariance matrices (SOBI) [18], see also [70]. A similar reasoning can also be followed to impose independence constraints in the higher order block component model (BCM), which is more general than CP [203, 204, 214, 207, 273].

8.5 Conclusion

Throughout this PhD, we discussed many decomposition methods. This last chapter combined two of them: ICA and CPA. The result was a trilinear model with independent sources in one mode. We developed an algorithm to compute it, and showed with simulations that it outperformed both standard ICA and CPA in specific situations.

Chapter 9

Conclusions and open problems

In this final chapter, findings and conclusions of the presented work are summarized and some directions for further research are pointed out.

9.1 General conclusions of the thesis

This thesis discussed a broad range of signal decompositions. We showed some applications in neuroscience where such decomposition methods can be useful. We divide our conclusions in main achievements concerning applications and novel mathematical algorithms. The latter section discusses newly proposed algorithms that can be used in different contexts. The former section discusses how existing decomposition methods are integrated into automatic analysis tools to improve the analysis of *e.g.*, EEG data.

9.1.1 Applications

- MRSI data can be used to classify different types of brain tissue, it e.g., for the discrimination of normal and tumor tissue. In a preliminary phase of this PhD, we developed a method based on CCA to segment short echo time MRSI data and to classify the segments [68, 77, 78, 79]. We validated it on simulations and showed some segmentations of real-life recordings. The results were promising, although only a few *in vivo* images could be analysed.
- Throughout the whole thesis, we highlighted differences between decomposition methods, and illustrated when the different methods can provide meaningful results. These findings are recapitulated in [73]. The presented

blind source separation techniques are powerful tools to decompose EEG recordings into contributing sources. The following contributions illustrate this in more detail.

- Muscle artifacts are a frequent disturbance in EEG recordings. We adapted a method based on BSS-CCA to clean muscle artifacts from EEG recordings of spoken language production. In such data, the brain response is always contaminated by the electrical activity from the cheek muscles. In order to analyse the data, removal of the artifacts is a must. We showed in different ways that our method both removed the artifacts and preserved brain activity [80, 347].

- Muscle and eye artifacts are frequently obscuring ictal EEG recordings. We used accurate BSS techniques as a preprocessing step for ictal source localisation [139, 136, 137]. We showed that in some examples, source localization became more accurate after preprocessing. However, we also pointed out the limitations of the approach.

- The main challenge in epileptic EEG analysis is still to find a robust method for localising the ictal EEG. We developed a method for separating the pure ictal brain signal from background EEG and artifacts, based on the third order CPA method [82, 74, 81, 83]. After the decomposition, the spatial component provided accurate information on the localisation of the epileptic focus. We showed both for simulated and real-life data that the proposed method largely outperforms existing methods.

9.1.2 Mathematical algorithms

- ICA decomposes a linear mixture in statistically independent components. Often prior knowledge is available about one or more mixing vector(s). We developed a method to impose that one mixing vector is equal to the known one [72]. Also extensions to the situation when more mixing vectors are known, are given. We also showed that the method outperforms standard ICA in accuracy in several cases.

- In some situations, the available prior knowledge will not be exact, but only known with a certain accuracy. For this goal, we developed a method to incorporate a constraint that one mixing vector is similar to a known vector [75, 76]. A weighting factor weighs the confidence in the independence versus the prior knowledge constraint.

- Although the CP decomposition is in general a unique decomposition, in practice, it can be useful to impose extra constraints for the computation of the model. We provided an algorithm for imposing independence constraints in CPA and showed in simulations that it outperforms in several cases (*e.g.*,

colored noise, one dimension lower than rank of the tensor, ...) the trilinear estimation with the standard CPA [71, 70, 69].

9.2 Future work and open problems

Decomposition techniques (ICA, CPA, ...) are a powerful tool for decomposing datasets in very different situations. We think that these decompositions will become very important computational tools for data-driven analysis, certainly in the growing field of neuroscience. We mention some possible research tracks to further refine these tools and we mention possible applications.

Improving source estimation in ictal EEG

There is not yet a reliable 3D localization technique available for localising ictal EEG. It will be useful to improve existing source imaging techniques in such a way that they exploit the complementarity of EEG and SPECT for visualizing the ictal onset zone. One possibility is to constrain the standard source localisation algorithm with prior information from *e.g.*, SISCOM. Another possibility is to develop a good source imaging technique that can be applied to the spatial distribution of the ictal activity.

Combined EEG/MEG

Recent research focuses on combined measurement of different modalities. One of the challenges is to develop a general method for simultaneous decomposition of *e.g.*, EEG/MEG [9, 286, 174] that exploits again the complementarity of these techniques.

Multilinear decomposition methods in multimodal signal processing

[243] proposed a multilinear framework for simultaneous decomposition of EEG and fMRI in which the correlation between the sources in the different modalities was exploited to extract the sources. These models can be further explored and improved in order to obtain an optimal analysis of the multimodal recording.

Models for connectivity studies

Brain networks, better known as brain connectivity, refer to a pattern of anatomical links ("anatomical connectivity"), statistical dependencies ("functional connectivity") or causal interactions ("effective connectivity") between distinct units within a nervous system. Formally, brain connectivity patterns can be represented in graph or matrix format [157, 176]. Most brain activation is dynamic in the sense that the brain connectivity changes over time. To describe these spatiotemporal changes, interactions, driving forces, ..., new data structures and new mathematical methodologies are needed. In this way, interactions can be unraveled. A CP-based method is already proposed in [242].

Extension of weighted constraint for the case when multiple mixing vectors are known with a limited accuracy

We developed an algorithm that weighs available prior knowledge on one mixing vector and the independence constraint. It is not straightforward to extend our scICA algorithm with one weighted constraint to the case of multiple mixing vectors approximately known. However, this is an interesting problem that will be investigated in the near future.

Orthogonality constrained CPA

We developed a CPA-algorithm with imposed independence constraints. Instead of imposing independence constraints in CPA, also orthogonality constraints can be imposed to make the decomposition more robust at a low computational cost. The start to this approach has been given in [147, 334, 324].

Higher order CCA

CCA is originally a tool to compare second-order data. Higher-order generalisations of CCA have been proposed in [144]. There is a broad spectrum of such generalisations (considering correlations in different dimensions) that can be further explored and probably applied to neuroscience data.

Appendix

In this appendix, we provide the formulas for the updates of $\mathbf{A}_{ALS}$, $\mathbf{B}_{ALS}$, $\mathbf{D}_{ALS}$ and the stepsize λ in section 8.2.2.

Expressions for $\mathbf{A}_{ALS}$, $\mathbf{B}_{ALS}$, $\mathbf{D}_{ALS}$

Define the following matrix representations of $\mathcal{U}^{(5)}$:

$$\begin{aligned}(\mathbf{U}_1^{(5)})_{(j-1)IJ(R+1)+(k-1)J(R+1)+(l-1)(R+1)+n,i} &= (\mathcal{U}^{(5)})_{ijkln},\\ (\mathbf{U}_2^{(5)})_{(k-1)IJ(R+1)+(l-1)I(R+1)+(n-1)I+i,j} &= (\mathcal{U}^{(5)})_{ijkln},\\ (\mathbf{U}_5^{(5)})_{(i-1)IJ^2+(j-1)IJ+(k-1)J+l,n} &= (\mathcal{U}^{(5)})_{ijkln}.\end{aligned}$$

Given $\mathbf{A}_p$, $\mathbf{B}_p$ and $\mathbf{D}_p$, an ALS update in the first mode yields:

$$\mathbf{A}_{ALS} = (\mathbf{U}_1^{(5)})^T \cdot (\mathbf{B}_p \odot \mathbf{A}_p^* \odot \mathbf{B}_p^* \odot \mathbf{D}_p)^{T,\dagger}. \tag{1}$$

An ALS update in the second mode yields:

$$\mathbf{B}_{ALS} = (\mathbf{U}_2^{(5)})^T \cdot (\mathbf{A}_p^* \odot \mathbf{B}_p^* \odot \mathbf{D}_p \odot \mathbf{A}_p)^{T,\dagger}. \tag{2}$$

An ALS update in the fifth mode yields:

$$\mathbf{D}_{ALS} = (\mathbf{U}_5^{(5)})^T \cdot (\mathbf{A}_p \odot \mathbf{B}_p \odot \mathbf{A}_p^* \odot \mathbf{B}_p^*)^{T,\dagger}. \tag{3}$$

Computation of λ

After vectorization, cost function ϕ can be written as

$$\begin{aligned}&\phi(\mathbf{A}_{p+1}, \mathbf{B}_{p+1}, \mathbf{D}_{p+1})\\ &= \|(\mathbf{A}_{p+1} \odot \mathbf{A}_{p+1}^* \odot \mathbf{B}_{p+1} \odot \mathbf{B}_{p+1}^* \odot \mathbf{D}_{p+1}) \cdot \mathbf{1}_R - \mathbf{u}^{(5)}\|_F^2,\end{aligned} \tag{4}$$

$$\phi(\lambda) = \|\lambda^3(\lambda^2)^*\mathbf{t}_0+\lambda^2(\lambda^2)^*\mathbf{t}_1+\lambda^3\lambda^*\mathbf{t}_2+\lambda^3\mathbf{t}_3+\lambda^2\lambda^*\mathbf{t}_4+\lambda(\lambda^2)^*\mathbf{t}_5+\lambda^2\mathbf{t}_6+(\lambda^2)^*\mathbf{t}_7+\lambda\lambda^*\mathbf{t}_8+\lambda\mathbf{t}_9+\lambda^*\mathbf{t}_{10}+\mathbf{t}_{11}\|_F^2, \tag{5}$$

$$\begin{aligned}
\mathbf{t}_0 &= (\boldsymbol{\Delta}\mathbf{A} \odot \boldsymbol{\Delta}\mathbf{A}^* \odot \boldsymbol{\Delta}\mathbf{B} \odot \boldsymbol{\Delta}\mathbf{B}^* \odot \boldsymbol{\Delta}\mathbf{D}) \cdot \mathbf{1}_R, \\
\mathbf{t}_1 &= (\mathbf{A} \odot \boldsymbol{\Delta}\mathbf{A}^* \odot \boldsymbol{\Delta}\mathbf{B} \odot \boldsymbol{\Delta}\mathbf{B}^* \odot \boldsymbol{\Delta}\mathbf{D} + \boldsymbol{\Delta}\mathbf{A} \odot \boldsymbol{\Delta}\mathbf{A}^* \odot \mathbf{B} \odot \boldsymbol{\Delta}\mathbf{B}^* \odot \boldsymbol{\Delta}\mathbf{D}) \cdot \mathbf{1}_R \\
&\quad +(\boldsymbol{\Delta}\mathbf{A} \odot \boldsymbol{\Delta}\mathbf{A}^* \odot \boldsymbol{\Delta}\mathbf{B} \odot \boldsymbol{\Delta}\mathbf{B}^* \odot \mathbf{D}) \cdot \mathbf{1}_R, \\
\mathbf{t}_2 &= (\boldsymbol{\Delta}\mathbf{A} \odot \boldsymbol{\Delta}\mathbf{A}^* \odot \boldsymbol{\Delta}\mathbf{B} \odot \mathbf{B}^* \odot \boldsymbol{\Delta}\mathbf{D} + \boldsymbol{\Delta}\mathbf{A} \odot \mathbf{A}^* \odot \boldsymbol{\Delta}\mathbf{B} \odot \boldsymbol{\Delta}\mathbf{B}^* \odot \boldsymbol{\Delta}\mathbf{D}) \cdot \mathbf{1}_R, \\
\mathbf{t}_3 &= (\boldsymbol{\Delta}\mathbf{A} \odot \mathbf{A}^* \odot \boldsymbol{\Delta}\mathbf{B} \odot \mathbf{B}^* \odot \boldsymbol{\Delta}\mathbf{D}) \cdot \mathbf{1}_R \\
\mathbf{t}_4 &= (\boldsymbol{\Delta}\mathbf{A} \odot \boldsymbol{\Delta}\mathbf{A}^* \odot \boldsymbol{\Delta}\mathbf{B} \odot \mathbf{B}^* \odot \mathbf{D} + \boldsymbol{\Delta}\mathbf{A} \odot \boldsymbol{\Delta}\mathbf{A}^* \odot \mathbf{B} \odot \mathbf{B}^* \odot \boldsymbol{\Delta}\mathbf{D} + \mathbf{A} \odot \boldsymbol{\Delta}\mathbf{A}^* \odot \boldsymbol{\Delta}\mathbf{B} \odot \mathbf{B}^* \odot \boldsymbol{\Delta}\mathbf{D}) \cdot \mathbf{1}_R \\
&+ (\boldsymbol{\Delta}\mathbf{A} \odot \mathbf{A}^* \odot \boldsymbol{\Delta}\mathbf{B} \odot \boldsymbol{\Delta}\mathbf{B}^* \odot \mathbf{D} + \boldsymbol{\Delta}\mathbf{A} \odot \mathbf{A}^* \odot \mathbf{B} \odot \boldsymbol{\Delta}\mathbf{B}^* \odot \boldsymbol{\Delta}\mathbf{D} + \mathbf{A} \odot \mathbf{A}^* \odot \boldsymbol{\Delta}\mathbf{B} \odot \boldsymbol{\Delta}\mathbf{B}^* \odot \boldsymbol{\Delta}\mathbf{D}) \cdot \mathbf{1}_R \\
\mathbf{t}_5 &= (\boldsymbol{\Delta}\mathbf{A} \odot \boldsymbol{\Delta}\mathbf{A}^* \odot \mathbf{B} \odot \boldsymbol{\Delta}\mathbf{B}^* \odot \mathbf{D} + \mathbf{A} \odot \boldsymbol{\Delta}\mathbf{A}^* \odot \boldsymbol{\Delta}\mathbf{B} \odot \boldsymbol{\Delta}\mathbf{B}^* \odot \mathbf{D} + \mathbf{A} \odot \boldsymbol{\Delta}\mathbf{A}^* \odot \mathbf{B} \odot \boldsymbol{\Delta}\mathbf{B}^* \odot \boldsymbol{\Delta}\mathbf{C}) \cdot \mathbf{1}_R, \\
\mathbf{t}_6 &= (\boldsymbol{\Delta}\mathbf{A} \odot \mathbf{A}^* \odot \boldsymbol{\Delta}\mathbf{B} \odot \mathbf{B}^* \odot \mathbf{D} + \boldsymbol{\Delta}\mathbf{A} \odot \mathbf{A}^* \odot \mathbf{B} \odot \mathbf{B}^* \odot \boldsymbol{\Delta}\mathbf{C} + \mathbf{A} \odot \mathbf{A}^* \odot \boldsymbol{\Delta}\mathbf{B} \odot \mathbf{B}^* \odot \boldsymbol{\Delta}\mathbf{D}) \cdot \mathbf{1}_R, \\
\mathbf{t}_7 &= (\mathbf{A} \odot \boldsymbol{\Delta}\mathbf{A}^* \odot \mathbf{B} \odot \boldsymbol{\Delta}\mathbf{B}^* \odot \mathbf{D}) \cdot \mathbf{1}_R, \\
\mathbf{t}_8 &= (\boldsymbol{\Delta}\mathbf{A} \odot \boldsymbol{\Delta}\mathbf{A}^* \odot \mathbf{B} \odot \mathbf{B}^* \odot \mathbf{D} + \mathbf{A} \odot \boldsymbol{\Delta}\mathbf{A}^* \odot \boldsymbol{\Delta}\mathbf{B} \odot \mathbf{B}^* \odot \mathbf{D} + \mathbf{A} \odot \boldsymbol{\Delta}\mathbf{A}^* \odot \mathbf{B} \odot \mathbf{B}^* \odot \boldsymbol{\Delta}\mathbf{D}) \cdot \mathbf{1}_R \\
&+ (\boldsymbol{\Delta}\mathbf{A} \odot \mathbf{A}^* \odot \mathbf{B} \odot \boldsymbol{\Delta}\mathbf{B}^* \odot \mathbf{C} + \mathbf{A} \odot \mathbf{A}^* \odot \boldsymbol{\Delta}\mathbf{B} \odot \boldsymbol{\Delta}\mathbf{B}^* \odot \mathbf{D} + \mathbf{A} \odot \mathbf{A}^* \odot \mathbf{B} \odot \boldsymbol{\Delta}\mathbf{B}^* \odot \boldsymbol{\Delta}\mathbf{D}) \cdot \mathbf{1}_R, \\
\mathbf{t}_9 &= (\boldsymbol{\Delta}\mathbf{A} \odot \mathbf{A}^* \odot \mathbf{B} \odot \mathbf{B}^* \odot \mathbf{D} + \mathbf{A} \odot \mathbf{A}^* \odot \boldsymbol{\Delta}\mathbf{B} \odot \mathbf{B}^* \odot \mathbf{C} + \mathbf{A} \odot \mathbf{A}^* \odot \mathbf{B} \odot \mathbf{B}^* \odot \boldsymbol{\Delta}\mathbf{D}) \cdot \mathbf{1}_R, \\
\mathbf{t}_{10} &= (\mathbf{A} \odot \mathbf{A}^* \odot \mathbf{B} \odot \Delta\mathbf{B}^* \odot \mathbf{D} + \mathbf{A} \odot \boldsymbol{\Delta}\mathbf{A}^* \odot \mathbf{B} \odot \mathbf{B}^* \odot \mathbf{D}) \cdot \mathbf{1}_R, \\
\mathbf{t}_{11} &= (\mathbf{A} \odot \mathbf{A}^* \odot \mathbf{B} \odot \mathbf{B}^* \odot \mathbf{D}) \cdot \mathbf{1}_R - \mathbf{u}.
\end{aligned} \tag{6}$$

where $\mathbf{1}_R = [1, 1, \ldots, 1]^T \in \mathbb{R}^R$ and $\mathbf{u}^{(5)} \in \mathbb{C}^{I^2J^2(R+1)}$ is defined by

$$(\mathbf{u}^{(5)})_{(i-1)IJ^2(R+1)+(k-1)J^2(R+1)+(j-1)J(R+1)+(l-1)(R+1)+n} = (\mathcal{U}^{(5)})_{ijkln}.$$

Substitution of (8.16) in (4) yields equation (5), given on top of the next page, with the vectors $\mathbf{t}_0$ to $\mathbf{t}_{11}$ given in equation (6), also given on the next page. (We dropped the subscripts for notational convenience.)

Let $\mathbf{h} = [\lambda^3(\lambda^2)^*, \lambda^2(\lambda^2)^*, \lambda^3\lambda^*, \lambda^3, \lambda^2\lambda^*, \lambda(\lambda^2)^*, \lambda^2, (\lambda^2)^*, \lambda\lambda^*, \lambda, \lambda^*, 1]^T \in \mathbb{C}^{12}$ and $\mathbf{T} = [\mathbf{t}_0, \ldots, \mathbf{t}_{11}] \in \mathbb{C}^{I^2J^2(R+1)\times 12}$. From (5) we have now that

$$\phi = \mathbf{h}^H \cdot \mathbf{T}^H \cdot \mathbf{T} \cdot \mathbf{h}. \tag{8}$$

$$\phi(\lambda) = \|\lambda^5 \mathbf{t}_0 + \lambda^4(\mathbf{t}_1 + \mathbf{t}_2) + \lambda^3(\mathbf{t}_3 + \mathbf{t}_4 + \mathbf{t}_5) + \lambda^2(\mathbf{t}_6 + \mathbf{t}_7 + \mathbf{t}_8) + \lambda(\mathbf{t}_9 + \mathbf{t}_{10}) + \mathbf{t}_{11}\|_F^2. \quad (7)$$

From the parameterization $\lambda = a + jb$, we get

$$\begin{bmatrix} \lambda^3(\lambda^2)^* \\ \lambda^2(\lambda^2)^* \\ \lambda^3\lambda^* \\ \lambda^3 \\ \lambda^2\lambda^* \\ \lambda(\lambda^2)^* \\ \lambda^2 \\ (\lambda^2)^* \\ \lambda\lambda^* \\ \lambda \\ \lambda^* \\ 1 \end{bmatrix} = \begin{bmatrix} j & a & 2ja^2 & 2a^3 & ja^4 & a^5 \\ 0 & 1 & 0 & 2a^2 & 0 & a^4 \\ 0 & -1 & 2ja & 0 & 2ja^3 & a^4 \\ 0 & 0 & -j & -3a & 3ja^2 & a^3 \\ 0 & 0 & j & a & ja^2 & a^3 \\ 0 & 0 & -j & a & -ja^2 & a^3 \\ 0 & 0 & 0 & -1 & 2ja & a^2 \\ 0 & 0 & 0 & -1 & -2ja & a^2 \\ 0 & 0 & 0 & 1 & 0 & a^2 \\ 0 & 0 & 0 & 0 & j & a \\ 0 & 0 & 0 & 0 & -j & a \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} b^5 \\ b^4 \\ b^3 \\ b^2 \\ b \\ 1 \end{bmatrix} \stackrel{def}{=} \mathbf{P}_a \cdot \mathbf{b}$$

$$= \begin{bmatrix} 1 & jb & 2b^2 & 2jb^3 & b^4 & jb^5 \\ 0 & 1 & 0 & 2b^2 & 0 & b^4 \\ 0 & 1 & 2jb & 0 & 2jb^3 & -b^4 \\ 0 & 0 & 1 & 3jb & -3b^2 & -jb^3 \\ 0 & 0 & 1 & jb & b^2 & jb^3 \\ 0 & 0 & 1 & -jb & b^2 & -jb^3 \\ 0 & 0 & 0 & 1 & 2jb & -b^2 \\ 0 & 0 & 0 & 1 & -2jb & -b^2 \\ 0 & 0 & 0 & 1 & 0 & b^2 \\ 0 & 0 & 0 & 0 & 1 & jb \\ 0 & 0 & 0 & 0 & 1 & -jb \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a^5 \\ a^4 \\ a^3 \\ a^2 \\ a \\ 1 \end{bmatrix} \stackrel{def}{=} \mathbf{P}_b \cdot \mathbf{a}.$$

Substitution in (8) yields that the cost function can be written as

$$\begin{aligned} \phi &= \mathbf{b}^T \cdot (\mathbf{P}_a^H \cdot \mathbf{T}^H \cdot \mathbf{T} \cdot \mathbf{P}_a) \cdot \mathbf{b}, \\ &= \mathbf{a}^T \cdot (\mathbf{P}_b^H \cdot \mathbf{T}^H \cdot \mathbf{T} \cdot \mathbf{P}_b) \cdot \mathbf{a}. \end{aligned}$$

For a fixed, ϕ is a real polynomial of degree 10 in the real variable b. Hence, given a, optimization over b is easy. One just has to compute the roots of the derivative of $\phi(b)$ and select the real root that corresponds to the global minimum. Similarly, for b fixed, optimization over a amounts to computing the roots of the derivative of $\phi(a)$ and selecting the real root that corresponds to the global minimum. We alternate between such conditional updates of a and b, starting with a search over the real line ($b = 0$). The purpose of this iteration is just to increase the convergence speed. It is not theoretically guaranteed that by alternating between updates of a and b, one will find the global optimum for λ. However, since we start from a search over the real line, the alternating updates in the complex plane can only improve the fit. Since the goal is only to increase the convergence speed, just a few updates suffice.

The matrix $\mathbf{T}^H \cdot \mathbf{T}$ can be computed in a relatively efficient way. Since it is Hermitian, only the coefficients of its upper triangular part have to be computed. The following property of the Khatri-Rao product can be used to compute these coefficients efficiently:

$$(\mathbf{A}_1 \odot \ldots \odot \mathbf{A}_N)^H \cdot (\mathbf{B}_1 \odot \cdots \odot \mathbf{B}_N) = (\mathbf{A}_1^H \cdot \mathbf{B}_1) \diamond \cdots \diamond (\mathbf{A}_N^H \cdot \mathbf{B}_N),$$

where $\diamond$ denotes the Hadamard (element-wise) product.

Finally, in the case of real-valued data, expression (5) simplifies to equation (7) given on top of this page. Let $\tilde{\mathbf{h}} = [\lambda^5, \lambda^4, \lambda^3, \lambda^2, \lambda, 1]^T \in \mathbb{R}^6$ and $\tilde{\mathbf{T}} = [\mathbf{t}_0, \mathbf{t}_1 + \mathbf{t}_2, \mathbf{t}_3 + \mathbf{t}_4 + \mathbf{t}_5, \mathbf{t}_6 + \mathbf{t}_7 + \mathbf{t}_8, \mathbf{t}_9 + \mathbf{t}_{10}, \mathbf{t}_{11}] \in \mathbb{R}^{I^2J^2(R+1)\times 6}$. From (7) we have now that

$$\phi = \tilde{\mathbf{h}}^T \cdot \tilde{\mathbf{T}}^T \cdot \tilde{\mathbf{T}} \cdot \tilde{\mathbf{h}}. \tag{9}$$

The cost function is a real polynomial of degree 10 in the real variable λ and can easily be minimized.

Bibliography

[1] E. Acar, C.A. Bingol, H. Bingol, R. Bro, and B. Yener. Multiway analysis of epilepsy tensors. *Bioinformatics*, 23:i10–i18, 2007.

[2] E. Acar, C.A. Bingol, H. Bingol and B. Yener. Computational analysis of epileptic focus localization. *Proc. of the IASTED Conference on Biomedical Engineering*, pages 519–107, 2006.

[3] F.-X. Alario and L. Ferrand. A set of 400 pictures standardized for french: Norms for name agreement, image agreement, familiarity, visual complexity, image variability, and age of acquisition. *Behavior Research Methods, Instruments & Computers*, 31(3):531 – 552, 1999.

[4] P.J. Allen, G. Polizzi, K. Krakow, D.R. Fish, and L. Lemieux. Identification of EEG events in the MR scanner: The problem of pulse artifact and a method for its subtraction. *NeuroImage*, 8:229–239, 1998.

[5] O. Alter, P.O. Brown, and D. Botstein. Generalized singular value decomposition for comparative analysis of genome-scale expression data sets of two different organisms. *Proc. Natl. Acad. Sci. USA*, 100:3351–3356, 2003.

[6] A.H. Andersen and W.S. Rayens. Structure-seeking multilinear methods for the analysis of fMRI data. *Neuroimage*, 22:728–739, 2004.

[7] B.A. Assaf and J.S. Ebersole. Continuous source imaging of scalp ictal rhythms in temporal lobe epilepsy. *Epilepsia*, 38:1114–1123, 1997.

[8] D.P. Auer, C. Gössl, T. Schirmer, and M. Czisch. Improved lipid analysis of ^{1}H-MR spectra in the presence of mobile lipids. *Magnetic Resonance in Medicine*, 46:615–618, 2001.

[9] F. Babiloni, C. Babiloni, F. Carducci, G.L. Romani, P.M. Rossini, L. Angelone, and F. Cincotti. Multimodal integration of EEG and MEG data: A simulation study with variable signal-to-noise ratio and number of sensors. *Human Brain Mapping*, 22:52–62, 2004.

[10] B.W. Bader and T. Kolda. Algorithm 862: Matlab tensor classes for fast algorithm prototyping. *ACM Transactions on Mathematical Software*, 32:635–653, 2006.

[11] S. Baillet, J. Mosher, and R. Leahy. Electromagnetic brain mapping. *IEEE Signal Proc. Mag.*, 18:14–30, 2001.

[12] J. Bancaud, O. Henriksen, F. Rubio-Donnadieua, M. Seino, F.E. Dreifuss, and J. Kiffin Penry. Proposal for revised clinical and electroencephalographic classification of epileptic seizures. From the Commission on Classification and Terminology of the International League Against Epilepsy. *Epilepsia*, 22:489–501, 1981.

[13] S.J. Barton, F.A. Howe, A.M. Romlins, S.A. Cudlip, J.K. Nicholson, B.A. Bell, and J.R. Griffiths. Comparison of ^{1}H MRS of human brain tumours with ^{1}H HR-MAS spectroscopy of intact biopsy samples in vitro. *Magn. Res. Mat. in Phys., Biol. and Med.*, 8:121–128, 1999.

[14] C.F. Beckmann and S.M. Smith. Tensorial extensions of independent component analysis for multisubject fmri analysis. *Neuroimage*, 25:294–311, 2005.

[15] K.L. Behar, D.L. Rothman, D.D. Spencer, and O.A. Petroff. Analysis of macromolecule resonances in ^{1}H-MR spectra spectra of human brain. *Magnetic Resonance in Medicine*, 32:294–302, 1994.

[16] A.J. Bell and T.J. Sejnowski. An information-maximisation approach to blind source separation and blind deconvolution. *Neural Computing*, 7:1003–1034, 1995.

[17] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines. A blind source separation technique using second order statistics. *IEEE Trans. Signal. Proc.*, 45:434–444, 1997.

[18] A. Belouchrani, K. Abed-Meraim, and Y. Hua. Jacobi-like algorithms for joint block diagonalization: Applications to source localization. In *Proc. IEEE Int Work on Intelligent Sig Proc and Comm Sys (ISPACS)*, Melbourne, Australia, 1998.

[19] H. Berger. Über das electroenkephalogramm des menschen. *Archiv für Psychiatrie und Nervenkrankheiten*, 87:527–570, 1929.

[20] M. Bernstein and A.G. Parrent. Complications of CT-guided stereotactic biopsy of intra-axial brain lesions. *Journal of Neurosurgery*, 81:165–168, 1994.

[21] C. Bledowski, D. Prvulovic, K. Hoechstetter, M. Scherg, M. Wibral, R. Goebel, and D.E. Linden. Localizing P300 Generators in Visual Target and Distractor Processing: A Combined Event-Related Potential and Functional Magnetic Resonance Imaging Study. *The Journal of Neuroscience*, 24:9353–9360, 2004.

[22] W. Blume, H. Lüders, E. Mizrahi, C. Tassinari, B.W. van Emde, and J. Engel. Glossary of descriptive terminology for ictal semiology: report of the ILAE task force on classification and terminology. *Epilepsia*, 42:1212–1218, 2001.

[23] H. Blumenfeld. Functional MRI studies of animal models in epilepsy. *Epilepsia*, 48:18–26, 2007.

[24] P. Boon, M. D'Havé, B. Vanrumste, G. Van Hoey, K. Vonck, Ph. Van Walleghem, J. Caemaert, E. Achten, and J. De Reuck. Ictal source localisation in presurgical patients with refractory epilepsy. *J. of Clin. Neurophysiol.*, 19:461–468, 2002.

[25] M. Borga and H. Knutsson. A canonical correlation approach to blind source separation. Technical Report LiU-IMT-EX-0062, Dept. of Biomedical Engineering, Linköping University, Sweden, 2001.

[26] A. Bragin, J. Engel, C.L. Wilson, I. Fried, and G.W. Mathern. Hippocampal and entorhinal cortex high-frequency oscillations (100- 500 hz) in human epileptic brain and in kainic acid - treated rats with chronic seizures. *Epilepsia*, 40:127–137, 1999.

[27] R. Bro. *Multi-way analysis in the food industry: models, algorithms, and applications*. PhD thesis, University of Amsterdam, 1998.

[28] B. H. Brooker and M. W. Donald. Contribution of the speech musculature to apparent human eeg asymmetries prior to vocalization. *Brain and Language*, 9:226 – 245, 1980.

[29] R.J. Broughton. Polysomnography: principals and applications in sleep and arousal disorders. In E. Niedermeyer and F. Lopes da Silva, editors, *Electroencephalography: Basic Principels, Clinical Applications and Related Fields*, chapter 44. Urban and Swarzenberg, Baltimore, second edition, 1987.

[30] V.D. Calhoun and T. Adali. Unmixing fMRI with independent component analysis. *IEEE Engineering in Medicine and Biology Magazine*, 25:79–90, 2006.

[31] V.D. Calhoun, T. Adali, M. Stevens, K. A. Kiehl, and J. J. Pekar. Semi-blind ICA of fMRI: A method for utilizing hypothesis-derived time courses in a spatial ICA analysis. *NeuroImage*, 35:527–538, 2005.

[32] J.-F. Cardoso. Iterative techniques for blind source separation using only fourth-order cumulants. In *Proc. EUSIPCO '92*, pages 739–742, Brussels, Belgium, 1992.

[33] J.-F. Cardoso. Multidimensional independent component analysis. In *Proc. of ICASSP*, pages 1941–44, Seattle, USA, 1998.

[34] J.-F. Cardoso. On the performance of orthogonal source separation algorithms. In *Proc. EUSIPCO '94*, pages 776–779, Edinburgh, 1998.

[35] J.-F. Cardoso. The three easy routes to independent component analysis: contrasts and geometry. In *Proceeding of ICA 2001*, San Diego, USA, 2001.

[36] J.-F. Cardoso, S. Bosc, and S. Friedlander. On optimal source separation based on second and fourth order cumulants. In *Proceedings of the 8th IEEE Signal Processing Workshop on Statistical Signal and Array Processing*, 1996.

[37] J.-F. Cardoso and A. Souloumiac. Blind beamforming for non-gaussian signals. *IEE Proc. F*, 140:362–370, 1994.

[38] J.-F. Cardoso and A. Souloumiac. Jacobi angles for simultaneous diagonalization. *SIAM J. Matrix Anal. Appl.*, 17:161–164, January 1996.

[39] J.D. Carroll and J. Chang. Analysis of individual differences in multidimensional scaling via an n-way generalization of 'Eckart-Young' decomposition. *Psychometrika*, 35:283–319, 1970.

[40] G.D. Cascino. Advances in neuroimaging: surgical localization. *Epilepsia*, 42:3–12, 2001.

[41] G.D. Cascino. Video-EEG monitoring in adults. *Epilepsia*, 43(Suppl.3):80–93, 2002.

[42] M. Castella, A. Chevreuil, and J.-C. Pesquet. Mélanges convolutifs. In P. Comon and C. Jutten, editors, *Séparation de sources*. Hermès-Lavoisier, 2007.

[43] M. Castillo, J.K. Smith, and L. Kwock. Correlation of myo-inositol levels and grading of cerebral astrocytomas. *American Journal of Neuroradiology*, 21:1645–1649, 2000.

[44] R. Cattell. Parallel proportional profiles and other principles for determining the choice of factors by rotation. *Psychometrika*, 9:267–283.

[45] G.G. Celesia. Visual Evoked Potentials. In E. Niedermeyer and F. Lopes da Silva, editors, *Electroencephalography: Basic Principels, Clinical Applications and Related Fields*, chapter 47. Urban and Swarzenberg, Baltimore, second edition, 1987.

[46] G.G. Celesia and M.H. Grigg. Auditory Evoked Potentials. In E. Niedermeyer and F. Lopes da Silva, editors, *Electroencephalography: Basic Principels, Clinical Applications and Related Fields*, chapter 48. Urban and Swarzenberg, Baltimore, second edition, 1987.

[47] A. Cichocki, R. Zdunek, A.H. Phan, and S.-I. Amari. *Nonnegative matrix and tensor factorizations.* John Wiley & Sons, 2009.

[48] W. De Clercq. *Advanced preprocessing techniques and nonlinear signal analysis applied to scalp electroencephalograms for the prediction of epileptic seizures.* PhD thesis, Faculty of Engineering, K.U.Leuven, 2005.

[49] W. De Clercq, A. Vergult, B. Vanrumste, W. Van Paesschen, and S. Van Huffel. Canonical correlation analysis applied to remove muscle artifacts from the encephalogram. *IEEE Trans. Biomed. Eng.*, 53:2583–2587, 2006.

[50] P. Comon. Independent component analysis. In J.L. Lacoume, editor, *Higher-Order Statistics*, pages 29–38. Elsevier, 1992.

[51] P. Comon. Independent component analysis, a new concept? *Signal Process.*, 36:287–314, 1994.

[52] P. Comon. Blind identification and source separation in 2×3 underdetermined mixtures. *IEEE Trans. Signal Process.*, 52:11–22, 2004.

[53] P. Comon. Canonical tensor decompositions. Technical report, Tech. Report I3S, RR-2004-17, 2004.

[54] P. Comon and C. Jutten. *Séparation de sources.* Hermès-Lavoisier, 2007.

[55] P. Comon, B. Mourrain, L.-H. Lim, and G. Golub. Symmetric tensors and symmetric tensor rank. 30:12541279, 2008.

[56] M. Congedo, C. Gouy-Pailler, and C. Jutten. On the blind source separation of human electroencephalogram by approximate joint diagonalization of second order statistics. *Clinical Neurophysiology*, 119:2677–2686, 2008.

[57] R.J. Croft and R.J. Barry. Removal of ocular artifact from the EEG: a review. *Clinical Neurophysiology*, 49:5–19, 2000.

[58] B.N. Cuffin, D. Cohen, K. Ynokuchi, R. Maniewski, C. Purcell, G.R. Cosgrove, J. Ives, J. Kennedy, and D. Schomer. Tests of EEG localization accuracy using implanted sources in the human brain. *Ann. Neurol.*, 29:132–138, 1991.

[59] J.S. Damoiseaux, S.A.R.B. Rombouts, F. Barkhof, P. Scheltens, C.J. Stam, S M. Smith, and C.F. Beckmann. Consistent resting-state networks across healthy subjects. *Proceedings of the National Academy of Sciences*, 103:13848–13853, 2006.

[60] G. Darmois. Analyse générale des liaisons stochastiques. *Rev. Inst. Intern. Stat.*, 21:2–8, 1953.

[61] S. Das and P. Sen. Restricted canonical correlations. *Linear Algebra and Its Applications*, 210:29–47, 1994.

[62] I. Daubechies. *Ten Lectures on Wavelets*. SIAM, Philadephia, sixth edition, 1992.

[63] I. Daubechies, E. Roussos, S. Takerkart, M. Benharrosh, C. Golden, K. D'Ardenne, W. Richter, J. D. Cohen, and J. Haxby. Independent component analysis for brain fMRI does not select for independence. *Proceedings of the National Academy of Sciences*, 106:10415–10422, 2009.

[64] M.E. Davies and C.J. James. Source separation using single channel ICA. *Signal Processing*, 87:1819–1832, 2007.

[65] L. De Lathauwer, B. De Moor, and J. Vandewalle. Fetal electrocardiogram extraction by source subspace separation. In *Proc. of IEEE Signal Processing / Athos Workshop on Higher-Order Statistics*, pages 134–138, Girona, Spain, 1995.

[66] J.C. de Munck. The potential distribution in a layered anisotropic spheroidal volume conductor. *J. Applied Physics*, 64:464–470, 1988.

[67] J.C. de Munck, B.W. van Dijk, and H. Spereijse. Mathematical dipoles are adequate to describe realistic generators of human brain activity. *IEEE Trans. Biomed. Eng.*, 35:960–966, 1988.

[68] M. De Vos. Nosologische beeldvorming als hulp bij kankerdiagnose. *Het Ingenieursblad*, 8:28–32, 2006.

[69] M. De Vos, L. De Lathauwer, and S. Van Huffel. Imposing independence constraints in the CP model. In *Proceedings of the 7th international conference on Independent Component Analysis (ICA)*, London, England, 2007.

[70] M. De Vos, L. De Lathauwer, and S. Van Huffel. Second order algorithm for imposing independence constraints to the CP model. In *Proc. of IEEE International Symposium on Circuits and Systems*, pages 1344–1347, Seattle, USA, 2008.

[71] M. De Vos, L. De Lathauwer, and S. Van Huffel. A combination of Parallel Factor and Independent Component Analysis. *Technical Report 09-04, ESAT-SISTA, K.U.Leuven (Leuven, Belgium), submitted*, 2009.

[72] M. De Vos, L. De Lathauwer, and S. Van Huffel. Spatially constrained ICA with an application in EEG processing. *Technical Report 09-02, ESAT-SISTA, K.U.Leuven (Leuven, Belgium), submitted*, 2009.

[73] M. De Vos, L. De Lathauwer, and S. Van Huffel. Decomposition methods in neuroscience. In *Recent Advances in Biomedical Signal Processing*. Bentham publishers, 2010.

[74] M. De Vos, L. De Lathauwer, B. Vanrumste, Sabine Van Huffel, and Wim Van Paesschen. Canonical decomposition of ictal scalp EEG and accurate source localisation: Principles and simulation study. *Computational Intelligence and Neuroscience: Special issue on EEG/MEG signal processing*, doi:58253:1–10, 2007.

[75] M. De Vos, L. De Lathauwer, A. Vergult, W. De Clercq, and S. Van Huffel. Jacobi iterations for spatially constrained independent component analysis. In *Proceedings of the 7th international conference on methematics in signal processing (IMA 2006)*, pages 38–41, Cirencester, 2006.

[76] M. De Vos, L. De Lathauwer, A. Vergult, W. De Clercq, W. Van Paesschen, and S. Van Huffel. Spatially constrained independent component analysis for real-time eye artifact removal from the electroencephalogram. In *Proc. of the The first Annual Symposium of the IEEE/EMBS Benelux Chapter. (IEEE EMBS)*, pages 159–163, Brussel, Belgium, 2006.

[77] M. De Vos, T. Laudadio, A.W. Simonetti, A. Heerschap, and S. Van Huffel. Fast nosologic imaging of the brain. *Journal of Magnetic Resonance*, 184:292–301, 2006.

[78] M. De Vos, T. Laudadio, A.W. Simonetti, A. Heerschap, and S. Van Huffel. Fast nosologic imaging of the brain. In *Proceedings of the 7th International Society of Magnetic Resonance in Medicine (ISMRM)*, 2006.

[79] M. De Vos, P. Pels, T. Laudatio, S. Van Huffel, J. Kurhanewicz, and S. Nelson. Automation of spectroscopic scoring of prostate MRSI by a canonical analysis based classification. In *Proceedings of the European Society of Magnetic Resonance in Medecin and Biology (ESMRMB)*, 2006.

[80] M. De Vos, S. Riès, K. Vanderperren, B. Vanrumste, F. Alario, S. Van Huffel, and B. Burle. Removal of muscle artifacts from EEG recordings of spoken language production. *Technical Report 09-03, ESAT-SISTA, K.U.Leuven (Leuven, Belgium), submitted*, 2009.

[81] M. De Vos, A. Vergult, L. De Lathauwer, W. De Clercq, S. Van Huffel, P. Dupont, A. Palmini, and W. Van Paesschen. Canonical decomposition of ictal EEG reliably detects the seizure onset zone. In *Proceedings of the American Epilepsy Society conference (AES)*, San Diego, US, 2006.

[82] M. De Vos, A. Vergult, L. De Lathauwer, W. De Clercq, S. Van Huffel, P. Dupont, A. Palmini, and W. Van Paesschen. Canonical decomposition of ictal scalp EEG reliably detects the seizure onset zone. *Neuroimage*, 37:844–854, 2007.

[83] M. De Vos, A. Vergult, L. De Lathauwer, W. De Clercq, S. Van Huffel, and W. Van Paesschen. An apparatus for decomposing electroencephalogram recordings into a set of distinct atoms. Patent application, GB0610795.7, KULeuven, 2006.

[84] S. Debener, A. Strobel, B. Sorger, J. Peters, C. Kranczioch, and A.K. Engel. Improved quality of auditory event-related potentials recorded simultaneously with 3-T fMRI: Removal of the ballistocardiogram artefact. *Neuroimage*, 34:587–597, 2007.

[85] S. Debener, M. Ullsperger, M. Siegel, and A.K. Engel. Single-trial EEG/fMRI reveals the dynamics of cognitive function. *Trends in Cognitive Sciences*, 10:558–563, 2006.

[86] W. Deburchgraeve, P.J. Cherian, M. De Vos, R.M. Swarte, J.H. Blok, G.H. Visser, P. Govaert, and S. Van Huffel. Automated neonatal seizure detection mimicking a human observer reading EEG. *Clinical Neurophysiology*, 119:2447–2454, 2008.

[87] W. Deburchgraeve, P.J. Cherian, M. De Vos, R.M. Swarte, J.H. Blok, G.H. Visser, P. Govaert, and S. Van Huffel. Neonatal seizure localization using PARAFAC decomposition. *Clinical Neurophysiology, accepted*, 2009.

[88] S. Dégerine and R. Malki. Second Order Blind Separation of Sources Based on Canonical Partial Innovations. *IEEE Trans. Signal. Proc.*, 48:629–641, 2000.

[89] A. Delorme, T. Sejnowski, and S. Makeig. Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *Neuroimage*, 34:1443–1449, 2007.

[90] M.R. DeLucci, B. Garoutte, and R.B. Aird. The scalp as an electroencephalographic averager. *Electroencephalogr. Clin. Neurophysiol.*, 14:191–196, 1962.

[91] A. Devos, A.W. Simonetti, M. van der Graaf, L. Lukas, J.A. Suykens, L. Vanhamme, L.M. Buydens, A. Heerschap, and S. Van Huffel. The use of multivariate MR imaging intensities versus metabolic data from MR spectroscopic imaging for brain tumour classification. *J. Magnetic Resonance*, 173:218–282, 2005.

[92] G. Dietsch. Fourier analyse van elektroenzephalogrammen des menschen. *Pflugers Arch Ges Physiol*, 230:106–112, 1932.

[93] L. Ding, G.A. Worrell, T.D. Lagerland, and B. He. 3D source localization of interictal spikes in epilepsy patients with MRI lesions. *Physics in Medicine and Biology*, 51:4047–4062, 2006.

[94] L. Ding, G.A. Worrell, T.D. Lagerland, and B. He. Ictal source analysis: localization and imaging of causal interactions in humans. *Neuroimage*, 34:575–586, 2007.

[95] D. Donoho. On minimum entropy deconvolution. In J.L. Lacoume, editor, *Applied time-series analysis II*, pages 565–609. Academic Press, 1981.

[96] R. Duda, P. Hart, and D. Storkn. *Pattern classification.* John Wiley & Sons, 2001.

[97] P. Durka. *Matching Pursuit and Unification in EEG Analysis.* Artech House, 2007.

[98] J.S. Ebersole. Noninvasive Localization of Epileptogenic Foci by EEG Source Modeling. *Epilepsia*, 41:S24–S33, 2000.

[99] C. Eckart and G. Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1:211–218, 1936.

[100] F. Szabo De Edelenyi, C. Rubin, F. Estve, G. Sylvie, M. Dcorps, V. Lefournier, J-F. Le Bas, and C. Rmy. A new approach for analyzing proton magnetic resonance spectroscopic images of brain tumors: nosologic images. *Nature in Medicine*, 6:1287–1298, 2000.

[101] T. Eichele, V. Calhoun, M.W. Moosmann, K. Specht, M. Jongsma, R.Q. Quiroga, H. Nordby, and K. Hugdahl. Unmixing concurrent EEG-fMRI with parallel independent component analysis. *International Journal of Psychophysiology*, 67:222–234, 2008.

[102] J. Jr. Engel. Update on surgical treatment of the epilepsies. Summary of the second international palm desert conference on the surgical treatment of the epilepsies (1992). *Neurology*, 43:1612–1617, 1993.

[103] C.W. Erwin, M.P. Rozear, R.A. Radtke, and A.C Erwin. Somatosensory Evoked Potentials. In E. Niedermeyer and F. Lopes da Silva, editors, *Electroencephalography: Basic Principels, Clinical Applications and Related Fields*, chapter 48. Urban and Swarzenberg, Baltimore, second edition, 1987.

[104] F. Estienne, N. Matthijs, D.L. Massart, P. Ricoux, and D. Leibovici. Multi-way modelling of high-dimensionality electroencephalographic data. *Chemometrics and Intelligent Laboratory Systems*, 58:59–72, 2001.

[105] G. Fan, B. Sun, Z. Wu, Q. Guo, and Y. Guo. In vivo single-voxel proton MR spectroscopy in the differentiation of high-grade gliomas and solitary metastases. *Clinical Radiology*, 59:77–85, 2004.

[106] S. Fauser, H.J. Huppertz, T. Bast, K. Strobl, G. Pantazis, D.M. Altenmueller, B. Feil, S. Rona, C. Kurth, D. Rating, R. Korinthenberg, B.J. Steinhoff, B. Volk, and A. Schulze-Bonhage. Clinical characteristics in focal cortical dysplasia: a retrospective evaluation in a series of 120 patients. *Brain*, 129:1907–1916, 2006.

[107] C.E. Fernandes, G. Favier, and J.C. Mota. Blind channel identification algorithms based on the parafac decomposition of comulant tensors: The single and multiuser cases. *Signal Processing*, 88:1382–1401, 2008.

[108] A.S. Field and D. Graupe. Topographic component (parallel factor) analysis of multichannel evoked potentials: Practical issues in trilinear spatiotemporal decomposition. *Brain Topography*, 3:407–423, 1991.

[109] D. Field. What is the goal of sensory coding? *Neural Computation*, 6:559–601, 1994.

[110] B.J. Fisch. Artifacts. In *Fisch and Spelhmann's EEG Primer*, chapter 6, pages 107–121. Elsevier, third edition, 1999.

[111] B.J. Fisch. Generalized epileptiform patterns. In *Fisch and Spelhmann's EEG Primer*, chapter 18, pages 285–305. Elsevier, third edition, 1999.

[112] B.J. Fisch. Localized epileptiform patterns. In *Fisch and Spelhmann's EEG Primer*, chapter 17, pages 261–283. Elsevier, third edition, 1999.

[113] B.J. Fisch. The normal EEG of wakefull resting adults of 20-60 years of age. In *Fisch and Spelhmann's EEG Primer*, chapter 11, pages 185–198. Elsevier, third edition, 1999.

[114] B.J. Fisch. The normal sleep EEG adults over 20 years. In *Fisch and Spelhmann's EEG Primer*, chapter 11, pages 185–198. Elsevier, third edition, 1999.

[115] B.J. Fisch. The source of the EEG. In *Fisch and Spelhmann's EEG Primer*, chapter 1, pages 3–17. Elsevier, third edition, 1999.

[116] B.J. Fisch. Spatial analysis of the EEG. In *Fisch and Spelhmann's EEG Primer*, chapter 4, pages 73–106. Elsevier, third edition, 1999.

[117] N. Foldvary, G. Klem, J. Hammel, W. Bingaman, I. Najm and H. Lüders. The localizing value of ictal EEG in focal epilepsy. *Neurology*, 57:2022–2028, 2001.

[118] M. Franceschi, F. Triulzi, L. Ferini-Strambi, et al. Focal cerebral lesions found by magnetic resonance imaging in cryptogenic non-refractory temporal lobe epilepsy patients. *Epilepsia*, 30:540–546, 1989.

[119] B. H. Friedman and J. F. Thayer. Facial muscle activity and EEG recordings: redundancy analysis. *Electroencephalography and clinical Neurophysiology*, 79:358 – 360, 1991.

[120] O. Friman. *Adaptive Analysis of Functional MRI Data.* PhD thesis, Linköping University, Sweden, 2003.

[121] K.J. Friston. *Statistical Parametric Mapping: The Analysis of Functional Brain Images.* Academic Press Inc., US, 2006.

[122] K.J. Friston, A.P. Holmes, J.B. Poline, P.J. Grasby, S.C. Williams, R.S. Frackowiak, and R. Turner. Analysis of fMRI time-series revisited. *Neuroimage*, 2:45–53, 1995.

[123] L. Y. Ganushchak and N. O. Schiller. Motivation and semantic context affect brain error-monitoring activity: An event-related brain potentials study. *NeuroImage*, 39:395 – 405, 2008.

[124] P. Gloor. Neuronal generators and the problem of localization in electroencephalography: applications of volume conductur theory to electroencephalography. *J. Clin. Neurophysiol.*, 2:327–354, 1985.

[125] G.H. Golub and C.F. Van Loan. *Matrix computations.* Third Edition. The Johns Hopkins University Press, 1996.

[126] G. Gómez-Herrero, W. De Clercq, H. Anwar, O. Kara, K. Egiazarian, S. Van Huffel, and W. Van Paesschen. Automatic removal of ocular artifacts in the EEG without a reference EOG channel. In *7th Nordic Signal Processing Symposium (NORSIG)*, Reykjavik, Iceland, 2006.

[127] S. Goncalves, J. de Munck, J. Verbunt, F. Bijma, R. Heethaar, and F. Lopes da Silva. In vivo measurement of the brain and skull resistivities using an EIT-based method and realistic models for the head. *IEEE Trans. Biomed. Eng.*, 50:754–767, 2003.

[128] I. Goncharova, D. McFarland, T. Vaughan, and J. Wolpaw. EMG contamination of EEG: Spectral and topographical characteristics. *Clinical Neurophysiology*, 114:1580–1593, 2003.

[129] J. Gotman. Automatic recognition of epileptic seizures in the EEG. *Electroencephalogr. Clin. Neurophysiol.*, 54:530–540, 1982.

[130] J. Gotman. Noninvasive methods for evaluating the localization and propagation of epileptic activity. *Epilepsia*, 44:21–29, 2003.

[131] J. Gotman, E. Kobayashi, A.P. Bagshaw, C.G. Benar, and F. Dubeau. Combining EEG and fMRI. a multimodal tool for epilepsy research. *Journal of Magnetic Resonance Imaging*, 23:906–920, 2006.

[132] G. Gratton, M.G.H. Coles, and E. Donchin. A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, 55:468–484, 1983.

[133] O. Grellier and P. Comon. Blind separation of discrete sources. *IEEE Signal Processing Letters*, 5:212–214, 1998.

[134] G. Hagberg, A.P. Burlina, I. Mader, W. Roser, E.W. Radue, and J. Seelig. In vivo proton MR spectroscopy of human gliomas: definition of metabolic coordinates for multi-dimensional classification. *Magnetic Resonance in Medicine*, 34:242–252, 1995.

[135] S. Halder, M. Bench, J. Mellinger, M. Bogdan, A. Kbler, N. Birbaumer, and W. Rosenstiel. Online Artifact Removal for Brain-Computer Interfaces Using Support Vector Machines and Blind Source Separation. *Computational Intelligence and Neuroscience*, DOI:10.1155/82069, 2007.

[136] H. Hallez, M. De Vos, R. Phlypo, B. Vanrumste, P. Van Hese, Y. D'Asseler, W. Van Paesschen, S. Van Huffel, and I. Lemahieu. Can dipole modelling be improved by removing muscular and ocular artifacts from ictal scalp EEG? In *Proc. of the EMBC neural engineering conference*, pages 252–255, Kohala Coast, Hawaii, 2007.

[137] H. Hallez, M. De Vos, B. Vanrumste, P. Van Hese, S. Assecondi, K. Van Laere, P. Dupont, W. Van Paesschen, S. Van Huffel, and I. Lemahieu. Removing muscle and eye artifacts using blind source separation techniques in ictal EEG source imaging. *Clinical Neurophysiology*, 120:1262–1272, 2009.

[138] H. Hallez, B. Vanrumste, P. Van Hese, Y. DAsseler, I. Lemahieu, and R. Van de Walle. A finite difference method with reciprocity used to incorporate anisotropy in electroencephalogram dipole source localization. *Physics in Medicine and Biology*, 50:3787–3806, 2005.

[139] H. Hallez, A. Vergult, R. Phlypo, W. De Clercq, Y. D'Asseler, R. Van de Walle, B. Vanrumste, W. Van Paesschen, S. Van Huffel, and I. Lemahieu. Muscle and eye movement artifact removal prior to EEG source localization. In *Proceedings of IEEE EMBS*, pages 2002–2005, New York, 2006.

[140] P. C. Hansen and S. H. Jensen. Fir filter representations of reduced-rank noise reduction. *IEEE Trans. Signal Proc.*, 46:1737–1741, 1998.

[141] R.A. Harshman. Foundations of the Parafac procedure: models and conditions for an 'explanatory' multi-modal factor analysis. *UCLA Working Papers in Phonetics*, 16, 1970.

[142] R.A. Harshman. Parafac2: Mathematical and technical notes. *UCLA working papers in phonetics*, 22:30–44, 1972.

[143] R.A. Harshman. Models for analysis of asymmetrical relationships among n objects or stimuli. In *First Joint Meeting of the Psychometric Society and the Society for Mathematical Psychology*, Hamilton, Ontario, 1978.

[144] R.A. Harshman. Generalization of Parafac and Tucker models to canonical correlation models. In *TRICAP.*, 2006.

[145] R.A. Harshman, S. Hong, and M. Lundy. Shifted factor analysis - part I: Models and properties. *J. Chemometrics*, 17:363–378, 2003.

[146] R.A. Harshman and M.E. Lundy. Data preprocessing and the extended Parafac model. In H.G. Law, C.W. Snyder Jr, J.A. Hattie, and R.P. McDonald, editors, *Research Methods for Multimode Data Analysis*. 1995.

[147] R.A. Harshman and M.F. Lundy. PARAFAC: Parallel factor analysis. *Computational Statistics and Data Analysis*, 18:39–72, 1994.

[148] K. Hashiguchi, T. Morioka, F. Yoshida, Y. Miyagi, S. Nagata, A. Sakata, and T. Sasaki. Correlation between scalp-recorded electroencephalographic and electrocorticographic activities during ictal period. *Seizure*, 16:238–247, 2007.

[149] B. He and Z. Liu. Multimodal functional neuroimaging: integrating functional MRI and EEG/MEG. *IEEE Reviews in Biomedical Engineering*, 1:23–40, 2008.

[150] K. Hermus, W. Verhelst, P. Lemmerling, P. Wambacq, and S. Van Huffel. Perceptual audio modeling with exponentially damped sinusoids. *Signal Processing*, 85:163–176, 2004.

[151] C.W. Hesse and James C.J. On semi-blind source separation using spatial constraints with applications in EEG analysis. *IEEE Trans. Biomed. Eng.*, 53:2525–2534, 2006.

[152] D. Hirtz, D.J. Thurman, K. Gwinn-Hardy, M. Mohamed, A.R. Chaudhuri, and R. Zalutsky. How common are the 'common' neurologic disorders? *Neurology*, 68:326–327, 2007.

[153] F.L. Hitchcock. Multiple invariants and generalized rank of a p-way matrix or tensor. *J. Math. Phys.*, 7:39–79.

[154] F.L. Hitchcock. The Expression of a Tensor or a Polyadic as a Sum of Products. *J. Math. Phys.*, 6:164–189.

[155] G. Van Hoey, B. Vanrumste, M. D'Havé, R. Van de Walle, I. Lemahieu, and P. Boon. Influence of measurement noise and electrode mislocalisation on EEG dipole source localisation. *Med. Biol. Eng. Comp.*, 38:2878–296, 2000.

[156] L. Hofmann, J. Slotboom, C. Boesch, and R. Kreis. Characterization of the macromolecule baseline in localized ^{1}H-MR spectra of human brain. *Magnetic Resonance in Medicine*, 46:855–863, 2001.

[157] C.J. Honey, R. Kötter, M. Breakspear, and O. Sporns. Network structure of cerebral cortex shapes functional connectivity on multiple time scales. *Proc. Natl. Acad. Sci. USA*, 104:10240–10245, 2007.

[158] H. Hotelling. Relations between two sets of variates. *Biometrika*, 28:321–371, 1936.

[159] http://azizu.uab.es/INTERPRET/.

[160] http://spot.colorado.edu/ dubin/talks/brodmann/brodmann.html.

[161] http://www.etumour.net/.

[162] N.E. Huang, Z. Shen, S.R. Long, M.L. Wu, and H.H. Shih. The empirical mode decomposition and Hilbert spectrum for nonlinear and nonstationary time series analysis. In *Proc. R. Soc. London*, pages 903–995, 1998.

[163] A. Hyvärinen. Fast and robust fixed-point algorithm for independent component analysis. *IEEE Transactions on Neural Networks*, 10:419–433, 1999.

[164] A. Hyvärinen, J. Karhunen and E. Oja. *Independent Component Analysis*. John Wiley & Sons, 2001.

[165] N. Ille, P. Berg, and M. Scherg. Artifact correction of the ongoing EEG using spatial filters based on artifact and brain signal topographies. *Journal of Clinical Neurophysiology*, 19:113–124, 2002.

[166] P. Indefrey and W. J. M. Levelt. The spatial and temporal signatures of word production components. *Cognition*, 92:101–144, 2004.

[167] M. Ishteva, L. De Lathauwer, P.-A. Absil, and S. Van Huffel. Dimensionality reduction for higher-order tensors: algorithms and applications. *International Journal of Pure and Applied Mathematics*, 42:337–343, 2008.

[168] M. Ishteva, L. De Lathauwer, P.-A. Absil, and S. Van Huffel. Differential-geometric Newton method for the best rank-(R1,R2,R3) approximation of tensors. *Numerical Algorithms, Tributes to Gene H. Golub Part II*, 51:179–194, 2009.

[169] M. Iwasaki, Ch. Kellinghaus, A.V. Alexopoulos, R.C. Burgess, A.N. Kumar, Y.H. Hand and H.O. Lüders. Effects of eyelid closure, blinks and eye movements on the electroencephalogram. *Clinical Neurophysiology*, 116:878–885, 2005.

[170] C.J. James and O. J. Gibson. Temporally constrained ICA: An application to artifact rejection in electromagnetic brain signal analysis. *IEEE Trans. Biomed. Eng.*, 50:1108–1116, 2003.

[171] D. Jannek, F. Roemer, M. Weis, M. Haardt, and P. Husar. Identification of Signal Components in Multi-Channel EEG Signals via Closed-Form PARAFAC Analysis and Appropriate Preprocessing. In *Proceedings of the 4th European Conference of the International Federation for Medical and Biological Engineering*, pages 1226–1230, 2008.

[172] S. Pruzansky J.D. Carroll and J.B. Kruskal. CANDELINC: A general approach to multidimensional analysis of many-way arrays with linear constraints on parameters. *Psychometrika*, 45:3–24, 1980.

[173] J. D. Jescheniak, H. Schriefers, M. F. Garrett, and A. D. Friederici. Exploring the activation of semantic and phonological codes during speech planning with event-related brain potentials. *Journal of Cognitive Neuroscience*, 14(6):951 – 964, 2002.

[174] A.M. Jiménez. Multimodal integration of EEG and MEG data using minimum L2-norm estimates. Master's thesis, MIT, 2007.

[175] M. Jing and S. Sanei. A new constrained topographic independent component analysis for separation of epileptic seizure signals. *Computational Intelligence and Neuroscience*, id 21315, 2007.

[176] V.K. Jirsa and A.R. McIntosh. *Handbook of Brain Connectivity.* Springer, 2007.

[177] C.A. Joyce, I.F. Gorodnitsky, and M. Kutas. Automatic removal of eye movement and blink artifacts from EEG data using blind component separation. *Psychophysiology*, 41:313–325, 2004.

[178] T.P. Jung, S. Makeig, C. Humphries, T.W. Lee, M.J. McKeown, V. Iragui, and T.J. Sejnowski. Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, 37:163–178, 2000.

[179] T.P. Jung, S. Makeig, M. Westerfield, J. Townsend, E. Courchesne, T.J. Sejnowski. Analysis and visualization of single-trial event-related potentials. *Human Brain Mapping*, 14:166–185, 2001.

[180] S.M. Kay. *Fundamentals of Statistical Signal Processing.* Prentice-Hall, 1993.

[181] H. Kiers. Towards a standardized notation and terminology in multiway analysis. *Journal of Chemometrics*, 14:105–122, 2000.

[182] H.A.L. Kiers and R.A. Harshman. Relating two proposed methods for speedup of algorithms for fitting two- and three-way principal component and related multilinear models. *Chemometrics and Intelligent Laboratory Systems*, 36:31–40, 1997.

[183] H.A.L. Kiers and I. Van Mechelen. Three-way component analysis: Principles and illustrative application. *Psychological Methods*, 6:84–110, 2001.

[184] L.G. Kiloh, A.J.Mc Comas, and J.W. Osselton. *Clinical Electroencephalography.* Butterworths, 1981.

[185] J. Kim and C.W. Mueller. *Introduction to factor analysis: What it is and how to do it.* Sage Publications, Newbury Park, CA, 1987.

[186] D.W. Klass. The continuing challenge of artifacts in the EEG. *Am. J. EEG Technol.*, 35:239–269, 1995.

[187] G.H. Klem. Artifacts. In J.S. Ebersole and T.A. Pedley, editors, *Current practice of clinical electroencephalography.*, pages 271–287. Philadelphia: Lippincott Williams and Wilkins, 2003.

[188] U. Klose. In vivo proton spectroscopy in presence of eddy currents. *Magnetic Resonance in Medicine*, 14:26–30, 1990.

[189] J.N. Knight. Signal fraction analysis and artifact removal in eeg. Master's thesis, Department of Computer Science, Colorado State University, Fort Collins, CO., 2003.

[190] D. Knuth. *The art of computer programming, 2.* Addison-Wesley, Reading, MA, 1997.

[191] K.H. Knuth. Informed source separation: A Bayesian tutorial. In *Proc. 13th EUSIPCO '05*, Turkey, 2005.

[192] K. Kobayashi, T. Akiyama, T. Nakahori, H. Yoshinaga, and J. Gotman. Systematic source estimation of spikes by a combination of independent component analysis and RAP-MUSIC. I: Principles and simulation study. *Clinical Neurophysiology*, 113:713–724, 2002.

[193] K. Kobayashi, T. Akiyama, T. Nakahori, H. Yoshinaga, and J. Gotman. Systematic source estimation of spikes by a combination of independent component analysis and RAP-MUSIC. II: Preliminary clinical application. *Clinical Neurophysiology*, 113:725–734, 2002.

[194] K. Kobayashi, C.J. James, T. Nakahori, T. Akiyama, and J. Gotman. Isolation of epileptiform discharges from unaveraged EEG by independent component analysis. *Clinical Neurophysiology*, 110:1755–1763, 1999.

[195] K. Kobayashi, I. Merlet, and J. Gotman. Separation of spikes from background by independent component analysis with dipole modeling and comparison to intracranial recording. *Clinical Neurophysiology*, 112:405–413, 2001.

[196] K. Kobayashi, H. Yoshinaga, Y. Ohtsuka, and J. Gotman. Dipole modeling of epileptic spikes can be accurate or misleading. *Epilepsia*, 46:397–408, 2005.

[197] T. Krings, K.H. Chiappa, B.N. Cuffin, B.R. Buchbinder, and G.R. Cosgrove. Accuracy of EEG dipole localisation of epileptiform activities associated with focal brain lesions. *Ann. Neurol.*, 44:76–86, 1998.

[198] P.M. Kroonenberg. *Applied Multiway Data Analysis.* John Wiley & Sons, 2008.

[199] J.B. Kruskal. Three-way arrays: rank and uniqueness of trilinear decompositions, with applications to arithmetic complexity and statistics. *Psychometrika*, 18:95–138, 1977.

[200] J.B. Kruskal, R.A. Harshman, and M.E. Lundy. How 3-MFA data can cause degenerate parafac solutions, among other relationships. In R. Coppi and S. Bolasco, editors, *Multiway Data Analysis*, pages 115–122. 1989.

[201] M.A. Kulesh, M.S. Diallo, and M. Holschneider. Wavelet analysis of ellipticity, dispersion, and dissipation properties of Rayleigh waves. *Acoustical Physics*, 51:425–434, 2005.

[202] L. De Lathauwer. A link between the canonical decomposition in multilinear algebra and simultaneous matrix diagonalization. *SIAM J. Matrix Anal. Appl.*, 28:642–666, 2006.

[203] L. De Lathauwer. Decompositions of a higher-order tensor in block terms — part I: Lemmas for partitioned matrices. *SIAM J. Matrix Anal. Appl.*, 30:1022–1032, 2008.

[204] L. De Lathauwer. Decompositions of a higher-order tensor in block terms — part II: Definitions and uniqueness. *SIAM J. Matrix Anal. Appl.*, 30:1033–1066, 2008.

[205] L. De Lathauwer and J. Castaing. Blind identification of underdetermined mixtures by simultaneous matrix diagonalization. *IEEE Trans. Signal. Proc.*, 56:1096–1105, 2008.

[206] L. De Lathauwer, J. Castaing, and J.-F. Cardoso. Fourth-order cumulant based underdetermined independent component analysis. *IEEE Trans. on Signal Processing*, 55:2965–2973, 2007.

[207] L. De Lathauwer and A. de Baynast. Blind deconvolution of DS-CDMA Signals by Means of Decomposition in Rank-$(1, L, L)$ Terms. *IEEE Trans. on Signal Processing*, 56:1562–1571, 2008.

[208] L. De Lathauwer, B. De Moor, and J. Vandewalle. Independent component analysis based on higher-order-only ica. In *Proc. of 8th IEEE SP workshop on Statistical Signal and Array Processing*, pages 356–359, Corfu, Greece, 1996.

[209] L. De Lathauwer, B. De Moor, and J. Vandewalle. An introduction to independent component analysis. *Journal of Chemometrics*, 14:123–149, 2000.

[210] L. De Lathauwer, B. De Moor, and J. Vandewalle. A multilinear singular value decomposition. *SIAM J. Matrix Anal. Appl.*, 21:1253–1278, 2000.

[211] L. De Lathauwer, B. De Moor, and J. Vandewalle. On the best rank-1 and rank-$(r_1, r_2, \ldots, r_n)$ approximation of higher-order tensors. *SIAM J. Matrix Anal. Appl.*, 21:1324–1342, 2000.

[212] L. De Lathauwer, B. De Moor, and J. Vandewalle. Computation of the canonical decomposition by means of a simultaneous generalized Schur decomposition. *SIAM J. Matrix Anal. Appl.*, 26:295–327, 2004.

[213] L. De Lathauwer, B. De Moor, and J. Vandewalle. A prewhitening-induced bound on the identification error in independent component analysis. *IEEE Trans. on Circuits and Systems*, 52:546–554, 2005.

[214] L. De Lathauwer and D. Nion. Decompositions of a higher-order tensor in block terms — part III: Alternating least squares algorithms. *SIAM J. Matrix Anal. Appl.*, 30:1067–1083, 2008.

[215] L. De Lathauwer and J. Vandewalle. Dimensionality reduction in higher-order signal processing and rank-$\{R_1, R_2, \ldots, R_N\}$ reduction in multilinear algebra. *Linear Algebra and its applications*, 391:31–55, 2004.

[216] M.A. Latif, S. Sanei, J. Chambers, and L. Shoker. Localization of abnormal EEG sources using blind source separation partially constrained by the locations of known sources. *IEEE Signal Processing Letters*, 13:117–120, 2006.

[217] T. Laudadio, M.C. Martinez-Bisbal, B. Celda, and S. Van Huffel. Fast nosological imaging of 2D TSI brain data using canonical correlation analysis. *NMR in Biomedicine*, 21:311–321, 2008.

[218] T. Laudadio, N. Mastronardi, L. Vanhamme, P. Van Hecke, and S. Van Huffel. Improved Lanczos algorithms for blackbox MRS data quantitation. *Nature in Medicine*, 157:292–297, 2002.

[219] T. Laudadio, P. Pels, L. De Lathauwer, P. Van Hecke, and S. Van Huffel. Tissue segmentation and classification of MRSI data using Canonical Correlation Analysis. *Magnetic Resonance in Medicine*, 54:1280–1284, 2005.

[220] D. Lee and H. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401:788–791, 1999.

[221] P. Lemmerling, N. Mastronardi, and S. Van Huffel. Efficient implementation of a structured total least squares based speech compression method. *Linear Algebra and its Applications*, pages 295–315, 2003.

[222] S.E. Leurgans, R.T. Ross, and R.B. Abel. A decomposition for three-way arrays. *SIAM J. Matrix Anal. Appl.*, 14:1064–1083, 1993.

[223] P. LeVan, E. Urrestarazu, and J. Gotman. A system for automatic artifact removal in ictal scalp EEG based on independent component analysis and bayesion classification. *Clinical Neurophysiology*, 117:912–27, 2006.

[224] H.L. Lew, J.H. Poole, J.Y. Chiang, E.H. Lee, E.S. Date, and D. Warden. Event-related potential in facial affect recognition: Potential clinical utility in patients with traumatic brain injury. *Journal of Rehabilitation Research and Development*, 42:29–34, 2005.

[225] Y. Li, A. Cichocki, and S. Amari. Blind estimation of channel parameters and source components for EEG signals: A sparse factorization approach. *IEEE Trans. on Neural Networks*, 17:419–431, 2006.

[226] J. Lindon, E. Holmes, and J. Nicholson. Pattern recognition methods and applications in biomedical magnetic resonance. *Prog. Nucl. Magn. Res. Spectr.*, 39:1–40, 2001.

[227] F. Lopes da Silva and A. Van Rotterdam. Biophysical Aspects of EEG and Magnetoencephalogram Generation. In E. Niedermeyer and F. Lopes da Silva, editors, *Electroencephalography: Basic Principels, Clinical Applications and Related Fields*. Lippincott Williams & Wilkins, New York, fifth edition, 2005.

[228] F.H. Lopes da Silva. Event-Related Potentials: methodology and quantification. In E. Niedermeyer and F. Lopes da Silva, editors, *Electroencephalography: Basic Principels, Clinical Applications and Related Fields*, chapter 46. Urban and Swarzenberg, Baltimore, second edition, 1987.

[229] W. Lu and J.C. Rajapakse. Approach and applications of constrained ICA. *IEEE Trans. on neural networks*, 16:203–212, 2005.

[230] S. L. Luck. *An Introduction to the Event-Related Potential Technique*. MIT Press, 2005.

[231] J. Luts, A. Heerschap, J.A.K. Suykens, and S. Van Huffel. A combined MRI and MRSI based Multiclass System for Brain Tumour Recognition using LS-SVMs with Class Probabilities and Feature Selection. *Artificial Intelligence in Medicine*, 40:87–102, 2007.

[232] J. Luts, T. Laudadio, A.J. Idema, A.W. Simonetti, A. Heerschap, D. Vandermeulen, J.A.K. Suykens, and S. Van Huffel. Nosologic imaging of the brain: segmentation and classification using MRI and MRSI. *NMR in Biomedicine*, 22:374–390, 2009.

[233] M. Atienza M. Crespo-Garcia and J.L. Cantero. Muscle artifact removal from human sleep EEG by using independent component analysis. *Journal of Biomedical Engineering*, 36:467–475, 2008.

[234] A. M.-Djafari. Bayesian Source Separation: beyond PCA and ICA. In *Proc. ESANN 2006*, Bruges, Belgium, 2006.

[235] M. W. Mahoney and P. Drineas. Cur matrix decompositions for improved data analysis. In *Proc. Natl. Acad. Sci. USA*, volume 106, pages 697–702, 2009.

[236] C. Majos, M. Julia-Sape, J. Alonso, M. Serrallonga, C. Aguilera, J.J. Acebes, C. Arus, and J. Gili. Brain tumor classification by proton MR spectroscopy: Comparison of diagnostic accuracy at short and long TE. *American Journal of Neuroradiology*, 25:1696–1704, 2004.

[237] S. Makeig, A.J. Bell, T.-P. Jung, and T.J. Sejnowski. Independent component analysis of electroencephalographic data. *In M. Mozer and M. Hasselmo, editors. Advances in Neural Information Processing Systems*, pages 145–151, 1996.

[238] G. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, pages 3397–3415, 1993.

[239] J. Malmivuo and R. Plonsey. *Bioelectromagnetism: Principles and Applications of Bioelectric and Biomagnetic Fields*. Oxford University Press, New York, 1995.

[240] D. Mantini, M.G. Perrucci, C. Del Gratta, G.L. Romani, and M. Corbetta. Electrophysiological signatures of resting state networks in the human brain. *Proc. Natl. Acad. Sci. USA*, 104:13170–13175, 2007.

[241] A. Marchal. *La production de la parole (Speech production)*. Lavoisier, Hermes Science, 2007.

[242] E. Martínez-Montes, R. Sarmiento-Prez, J.M. Snchez-Bornot, and P.A. Valdés-Sosa. Information Entropy-Based Penalty for PARAFAC Analysis of Resting EEG. In *Advances in Cognitive Neurodynamics ICCN 2007*. Springer Netherlands, 2007.

[243] E. Martínez-Montes, P.A. Valdès-Sosa, F. Miwakeichi, R.I. Goldman, M.S. Cohen. Concurrent EEG/fMRI analysis by multiway partial least squares. *Neuroimage*, 22:1023–1034, 2004.

[244] H. Masaki, H. Tanaka, N. Takasawa, and K. Yamazaki. Error-related brain potentials elicited by vocal errors. *Neuroreport: For Rapid Communication of Neuroscience Research*, 12(9):1851 – 1855, 2001.

[245] D. W. McAdam and H. A. Whitaker. Language production: Electroencephalographic localization in the normal human brain. *Science*, 172(3982):499 – 502, 1971.

[246] I. Merlet and J. Gotman. Reliability of dipole models of epileptic spikes. *Clinical Neurophysiology*, 110:1013–1028, 1999.

[247] C. Michel, M. Murray, G. Lantz, S. Gonzales, L. Spinelli, and R. Grave de Peralta. EEG source imaging. *Clinical Neurophysiology*, 115:2195–2222, 2004.

[248] S. Mierisova, A. van den Boogaart, I. Tkae, P. Van Hecke, L. Vanhamme, V. Mlynarik, and T. Liptaj. AMARES: a new approach for Quantitation of Short Echo Time in vivo ^{1}H Spectra of Rat Brain. *NMR in Biomedicine*, 11:32–39, 1998.

[249] S. Mine, H. Iwasa, Y. Kasagi and A. Yamaura. Ictal dipole source analysis based on a realistic scalp-skull-brain head model in localizing the epileptogenic zone. *Neuroscience Research*, 51:453–461, 2005.

[250] B. Mirkin. *Clustering for data mining.* Chapmann & Hall/CRC, Florida, USA, 2005.

[251] F. Miwakeichi, E. Martínez-Montes, P.A. Valdès-Sosa, N. Nishiyama, H. Mizuhara and Y. Yamaguchi. Decomposing EEG data into space-time-frequency components using parallel factor analysis. *Neuroimage*, 22:1035–1045, 2004.

[252] J. Möks. Topographic components model for event-related potentials and some biophysical considerations. *IEEE Trans. Biomed. Eng.*, 35:482–484, 1988.

[253] B. De Moor and H. Zha. A tree of generalizations of the ordinary singular value decomposition. *Linear Algebra and Its Applications*, 147:469–500, 1998.

[254] M. Moosmann, T. Eichele, H. Nordby, K. Hugdahl, and V. Calhoun. Joint independent component analysis for simultaneous EEG-fMRI: Principle and simulation. *International Journal of Psychophysiology*, 67:212–221, 2008.

[255] E. Moreau and J.C. Pesquet. Generalized contrasts for multichannel blind deconvolution of linear systems. *IEEE Signal Processing Letters*, 4:182–183, 1997.

[256] L. K. Morrell, D. A. Huntington, D. W. McAdam, and H. A. Whitaker. Electrocortical localization of language production. *Science*, 174(4016):1359 – 1360, 1971.

[257] G. Morren, P. Lemmerling, S. Van Huffel, G. Naulaers, H. Devlieger, and P. Casaer. Detection of autoregulation in the brain of premature infants using a novel subspace-based technique. In *Proceedings 23rd Annual International Conference IEEE Engineering in Medicine and Biology Society*, pages 2064–2067, 2001.

[258] G. Morren, S. Van Huffel, I. Helon, G. Naulaers, H. Daniels, H. Devlieger, and P. Caesar. Effects of non-nutritive sucking on heart rate, respiration and oxygenation: a model-based signal processing approach. *Comp. Biochem. Physiol. A*, 132:97–106, 2002.

[259] M. Mørup, L. K. Hansen, and S. M. Arnfred. Algorithms for sparse nonnegative tucker decomposition. *Neural Computation*, 20:2112–2131, 2008.

[260] M. Mørup, L.K. Hansen, S.M. Arnfred, L.-K. Lim, and K.M. Madsen. Shift invariant multilinear decomposition of neuroimaging data. *Neuroimage*, 42:1439–1450, 2008.

[261] M. Mørup, L.K. Hansen, Ch.S. Herrmann, J. Parnas and S.M. Arnfred. Parallel factor analysis as an exploratory tool for wavelet transformed event-related EEG. *Neuroimage*, 29:938–947, 2006.

[262] J. Mosher and R. Leahy. Source localization using recursively applied and projected (RAP) MUSIC. *IEEE Transactions on Signal Processing*, 47:332–340, 1999.

[263] C. Navasca, L. De Lathauwer, and S. Kindermann. Swamp reducing technique for tensor decomposition. In *Proceedings 16th European Signal Processing Conference (EUSIPCO 2008)*, Lausanne,Switzerland, 2008.

[264] N. Nelissen, W. Van Paesschen, K. Baete, K. Van Laere, A. Palmini, H. Van Billoen, and P. Dupont. Correlations of interictal FDG-PET metabolism and ictal SPECT perfusion changes in human temporal lobe epilepsy with hippocampal sclerosis. *Neuroimage*, 32:684–95, 2006.

[265] E.J. Nelson. Minireview: Multivoxel magnetic resonance spectroscopy of brain tumors. *Molecular Cancer Therapy*, 2:497–507, 2003.

[266] E. Niedermeyer. Abnormal EEG patterns (epileptic and paroxysmall). In E. Niedermeyer and F. Lopes da Silva, editors, *Electroencephalography: Basic Principels, Clinical Applications and Related Fields*, chapter 14. Urban and Swarzenberg, Baltimore, second edition, 1987.

[267] E. Niedermeyer. Epileptic seizure disorder. In E. Niedermeyer and F. Lopes da Silva, editors, *Electroencephalography: Basic Principels, Clinical Applications and Related Fields*, chapter 27. Urban and Swarzenberg, Baltimore, second edition, 1987.

[268] E. Niedermeyer. Sleep and EEG. In E. Niedermeyer and F. Lopes da Silva, editors, *Electroencephalography: Basic Principels, Clinical Applications and Related Fields*, chapter 10. Urban and Swarzenberg, Baltimore, second edition, 1987.

[269] E. Niedermeyer and F. Lopes da Silva. In *Electroencephalography: Basic Principels, Clinical Applications and Related Fields*. Urban and Swarzenberg, Baltimore, second edition, 1987.

[270] J. Nieuwenhuijse, R. Heusdens, and E.F. Deprettere. Robust exponential modeling of audio signals. In *Proc. of International Conference on Acoustic, Speech and Signal Processing (ICASSP 98)*, volume 6, pages 3581–3584, 1998.

[271] C.L. Nikias and J. Mendel. Signal processing with higher-order spectra. *IEEE Signal Processing Magazine*, pages 10–37, 1993.

[272] C.L. Nikias and A.P. Petropulu. *Higher-Order Spectra Analysis. A Nonlinear Signal Processing Framework.* Englewood Cliffs, N.J.: Prentice-Hall, 1993.

[273] D. Nion and L. De Lathauwer. Block component model based blind DS-CDMA receivers. *IEEE Trans. on Signal Processing*, 56:5567–5579, 2008.

[274] D. Nion and L. De Lathauwer. An enhanced line search scheme for complex-valued tensor decompositions. application in DS-CDMA. *Signal Processing*, 88:749–755, 2008.

[275] D. Nion and N.D. Sidiropoulos. Adaptive Algorithms to Track the PARAFAC decomposition of a Third-Order Tensor. *IEEE Trans. Signal. Proc.*, 2009.

[276] M.R. Nuwer, G. Comi, R. Emmerson, and et al. IFCN standards for digital recording of clinical eeg. *Electroencephalogr. Clin. Neurophysiol*, 106:259–261, 1998.

[277] R.D. Oades, A. Dittmann-Balcar, R. Schepker, C. Eggers, and D. Zerbin. Auditory event-related potentials (ERPs) and mismatch negativity (MMN) in healthy children and those with attention-deficit or Tourette/tic symptoms. *Biological Psychology*, 43:163–185, 1996.

[278] T.J. O'Brien, E.L. So, B.P. Mullan, M.F. Hauser, B.H. Brinkmann, N.I. Bohnen, D. Hanson, G.D. Cascino, C.R. Jack, and F.W. Sharbrough. Subtraction ictal spect co-registered to mri improves clinical usefulness of spect in localizing the surgical seizure focus. *Neurology*, 50:445–454, 1998.

[279] T.R. Olson and W. Pawlina. *A.D.A.M. Atlas of Anatomy.*

[280] M. Omladic and V. Omladic. More on restricted canonical correlations. *Linear Algebra and its Applications*, 321:285–293, 2000.

[281] T.F. Oostendorp, J. Delbeke, and D.F. Stegeman. The conductivity of the human skull: Results of in vivo measurements. *IEEE Trans. Biomed. Eng.*, 47:1487–1492, 2000.

[282] R. Oostenveld and P. Praamstra. The five percent electrode system for high-resolution EEG and ERP measurements. *Clinical Neurophysiology*, 112:713–719, 2001.

[283] P. Paatero. The multilinear engine — a table-driven, least squares program for solving multilinear problems, including the n-way parallel factor analysis model. *Journal of Computational and Graphical Statistics*, 8:854–888, 1999.

[284] W. Van Paesschen, P. Dupont, S. Sunaert, K. Goffin, and K. Van Laere. The use of SPECT and PET in routine clinical practice in epilepsy. *Opinion in Neurology*, 20:194202, 2007.

[285] R. Pascual-Marqui, C. Michel, and D. Lehman. Low resolution electromagnetic tomography: a new method for localizing the electrical activity in the brain. *Int. J. Psychophysiol.*, 18:49–65, 1994.

[286] E. Pataraia, G. Lindinger, L. Deecke, D. Mayer, and C. Baumgartner. Combined MEG/EEG analysis of the interictal spike complex in mesial temporal lobe epilepsy. *Neuroimage*, 24:607–614, 2004.

[287] K. Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2:559–572, 1901.

[288] J.J. Pekar. A brief introduction to functional MRI. *IEEE Engineering in Medicine and Biology Magazine*, 25:24–26, 2006.

[289] D.T. Pham. Blind Separation of Instantaneous Mixture of Sources via an Independent Component Analysis. *IEEE Trans. Signal. Proc.*, 44:2768–2779, 1996.

[290] R. Plonsey and D.B. Heppner. Considerations of quasistationarity in electrophysiological systems. *Bulletin of Mathematical Biophysics*, 29:657–664, 1967.

[291] C. Plummer, A.S. Harvey, and M. Cook. EEG source localization in focal epilepsy: Where are we now? *Epilepsia*, 20:201–218, 2008.

[292] J-B. Poullet, D.M. Sima, A.W. Simonetti, B. De Neuter, L. Vanhamme, P. Lemmerling, and S. Van Huffel. An automated quantitation of short echo time mrs spectra in an open source software environment: AQSES. *NMR Biomedicine*, 20:493–504, 2007.

[293] M.C. Preul, Z. Caramanos, R. Leblanc, J.G. Villemure, and D.L. Arnold. Using pattern analysis of *in vivo* proton MRSI data to improve the diagnosis and surgical management of patients with brain tumors. *NMR in Biomedicine*, 11:192–200, 1998.

[294] S.W. Provencher. Estimation of metabolite concentrations from localized *in vivo* proton NMR spectra. *Magnetic Resonance in Medicine*, 30:672–679, 1993.

[295] M. Rajih, P. Comon, and R.A. Harshman. Enhanced line search: a novel method to accelerate PARAFAC. *SIAM J. Matrix Anal. Appl.*, 30:1128–1147, 2007.

[296] C.R. Rao and S. Mitra. *Generalized Inverse of Matrices and Its Applications.* Wiley, New York, 1971.

[297] H. Ratiney, Y. Coenradie, S. Cavassila, D. van Ormondt, and D. Graveron-Demilly. Time domain quantitation of ^{1}H short echo-time signals: background accomodation. *MAGMA*, 16:284–296, 2004.

[298] E.L. Reilly. EEG recording and operation of the apparatus. In E. Niedermeyer and F. Lopes da Silva, editors, *Electroencephalography: Basic Principels, Clinical Applications and Related Fields*, pages 122–142. Williams and Wilkins, 1999.

[299] S. Ries, N. Janssen, S. Dufau, F.-X. Alario, and B. Burle. Electroencephalographic evidence for general purpose monitoring during speech production. *submitted*, 2009.

[300] M.W. Risinger. Electroencephalographic strategies for determining the epileptogenic zone. In H. Lüders, editor, *Epilepsy Surgery*, pages 337–347. Raven Press, 1991.

[301] P. Ritter, R. Becker, C. Graefe, and A. Villringer. Evaluating gradient artifact correction of EEG data acquired simultaneously with fMRI. *Magnetic Resonance Imaging*, 25:923–932, 2007.

[302] R. Rodionov, M.F. De, H. Laufs, D.W. Carmichael, E. Formisano, M. Walker, J.S. Duncan, and L. Lemieux. Independent component analysis of interictal fMRI in focal epilepsy: comparison with general linear model-based EEG-correlated fMRI. *Neuroimage*, 38:488–500, 2007.

[303] S. Romero, M.A. Mananas, and M.J. Barbanoj. A comparative study of automatic techniques for ocular artifact reduction in spontaneous EEG signals based on clinical target variables: a simulation case. *Computers in Biology and Medicine*, 38:348–360, 2008.

[304] F. Rosenow and H. Lüders. Presurgical evaluation of epilepsy. *Brain*, 124:1683–1700, 2001.

[305] M. Sahmoudi and K. Abed-Meraim. Robust blind separation algorithms for heavy-tailed sources. In *Proc. of ISSPIT '04*, pages 56–59, Rome, Italy, 2005.

[306] Y. Salu, L.G. Cohen, D. Rose, S. Sato, C. Kufta, and M. Hallet. An improved method for localizing electric brain dipoles. *IEEE Trans. Biomed. Eng.*, 37:699–708, 1990.

[307] E. Sanchez and B.R. Kowalski. Tensorial resolution: a direct trilinear decomposition. *J. Chemometrics*, 4:29–45, 1990.

[308] N. Schaul. The fundamental neural mechanisms of electroencephalography. *Electroencephalography and clinical neurophysiology*, 106:101–107, 1998.

[309] M. Scherg and D. von Cramon. Two bilateral sources of the late AEP as identified by a spatio-temporal dipole model. *Electroencephalogr. Clin. Neurophysiol.*, 62:32–44, 1985.

[310] R.O. Schmidt. Multiple emitter location and signal parameter estimation. 34:276–280, 1986.

[311] U. Seeger, U. Klose, I. Mader, W. Grodd, and T. Nägele. Parameterized evaluation of macromolecules and lipids in proton MR spectroscopy of brain diseases. *Magnetic Resonance in Medicine*, 49:19–28, 2003.

[312] N. D. Sidiropoulos, R. Bro, and G. B. Giannakis. Parallel factor analysis in sensor array processing. *IEEE Trans. Signal Process.*, 48:2377–2388, 2000.

[313] N. D. Sidiropoulos, G. B. Giannakis, and R. Bro. Blind PARAFAC receivers for DS-CDMA systems. *IEEE Trans. Signal Process.*, 48:810–823, 2000.

[314] M. Signoretto, K. Pelckmans, L. De Lathauwer, and J.A.K. Suykens. Data-Dependent Norm Adaptation for Sparse Recovery in Kernel Ensembles Learning. Technical Report 09-97, ESAT-SISTA, K.U.Leuven (Leuven, Belgium), 2009.

[315] V. De Silva and L.-H. Lim. Tensor rank and the ill-posedness of the best low-rank approximation problem. *SIAM J. Matrix Anal. Appl.*, 30:1084–1127, 2006.

[316] A. Simonetti, W. Melsen, F. Szabo de Edelenyi, J. van Asten, A. Heerschap, and L. Buydens. Combination of feature-reduced MR spectroscopic and MR imaging data for improved brain tumor classification. *NMR Biomedicine*, 18:34–43, 2005.

[317] A. Simonetti, W. Melsen, M. van der Graaf, G. Postma, A. Heerschap, and L. Buydens. A chemometric approach for brain tumor classsification using magnetic resonance imaging and spectroscopy. *Anal. Chem.*, 75:5352–5361, 2003.

[318] N.D. Siridopoulos and R. Bro. On the uniqueness of multilinear decomposition of n-way arrays. *J Chemometrics*, 14:229–239, 2000.

[319] A. Smekhov and A. Yeredor. Optimization of JADE using a novel optimally weighted joint diagonalization approach. In *Proc. of EUSIPCO*, 2004.

[320] A. Smilde, R. Bro, and P. Geladi. *Multi-way Analysis with applications in the Chemical Sciences*. John Wiley & Sons, 2004.

[321] D. Smith and D. Chadwick. The management of epilepsy. *J. Neurosurg. Psychiatry*, 70:ii15–ii21, 2001.

[322] H. Snoussi and A.M. Djafari. Bayesian unsupervised learning for source separation with mixture of gaussians prior. *J. VLSI Signal Process. Syst.*, 37:263–279, 2004.

[323] E.L. So. Integration of EEG, MRI and SPECT in localizing the seizure focus for epilepsy surgery. *Epilepsia*, 41:S48–S54, 2000.

[324] M. Sørensen, L. De Lathauwer, and L. Deneire. Parafac with orthogonality in one mode and applications in DS-CDMA systems. In *Proceedings of ICASSP*, Dallas, Texas, 2009.

[325] E.J. Speckmann and C.E. Elgar. Introduction to the neurophysiology, basis of the eeg and dc potentials. In E. Niedermeyer and F. Lopes da Silva, editors, *Basic principles, clinical applications and related fields*, chapter 1, pages 1–13. 2nd edition, 1987.

[326] S.S. Spencer, P.D. Williamson, S.L. Briders, R.H. Mattson, D.V. Cicchetti and D.D. Spencer. Reliability and accuracy of localization by scalp ictal eeg. *Neurology*, 30:1567–1575, 1985.

[327] A. Stegeman. Comparing independent component analysis and the parafac model for artificial multi-subject fmri data. *Technical Report, Heymans Institute, University of Groningen, the Netherlands*, 2007.

[328] A. Stegeman. Degeneracy in candecomp/parafac explained for $p \times p \times 2$ arrays of rank $p + 1$ or higher. *Psychometrika*, 72:601–619, 2007.

[329] A. Stegeman and N.D. Siridopoulos. On Kruskal's uniqueness condition for the Candecomp/Parafac decomposition. *Lin. Algebra and Applications*, 420:540–552, 2007.

[330] B. Stemmer and H. Whitaker. *Handbook of the Neuroscience of Language*. Academic Press, 2008.

[331] J.T. Sun, H.J. Zeng, H. Liu, Y. Lu, and Z. Chen. Cubesvd: A novel approach to personalized web search. In *Proc. of the 14th International World Wide Web conference*, pages 382–390, 2005.

[332] A.R. Tate, C. Majós, A. Moreno, F.A. Howe, J.R. Griffiths, and C. Arus. Automated classification of short echo time spectra in in vivo ^{1}H brain spectra: a multicenter study. *Magnetic Resonance in Medicine*, 49:29–36, 2003.

[333] D.C. Taylor, D. McMackin, H. Staunton, N. Delanty, and J. Phillips. Patients' aims for epilepsy surgery: desires beyond seizure freedom. *Epilepsia*, 42:629–633, 2001.

[334] J.M.F. ten Berge. Least squares optimization in multivariate analysis. Technical report, Leiden: DSWO, 1993.

[335] G. Tomasi and R. Bro. A comparison of algorithms for fitting the parafac model. *Computational Statistics & Data Analysis*, 50:1700–1734, 2006.

[336] Z. Tong, T. Yamaki, K. Harada, and K. Houkin. In vivo quantification of metabolites in normal brain and brain tumors by proton MR spectroscopy using water as an internal standard. *Magnetic Resonance Imaging*, 22:1017–1024, 2004.

[337] L.R. Tucker. The extension of factor analysis to three-dimensional matrices. In Gulliksen H Frederiksen N, editor, *Contributions to Mathematical Psychology*, pages 110–182. Holst, Rinehart Winston, New York, 1964.

[338] L.R. Tucker. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31:279–311, 1966.

[339] E. Urrestarazu, R. Chander, F. Dubeau, and J. Gotman. Interictal high-frequency oscillations (100-500 Hz) in the intracerebral EEG of epileptic patients. *Brain*, 130:2354–2366, 2007.

[340] E. Urrestarazu, J. Iriarte, M. Alegre, M. Valencia, C. Viteri, and J. Artieda. Independent component analysis removing artifacts in ictal recordings. *Epilepsia*, 45:1071–1078, 2004.

[341] P.P. Vaidyanathan and I. Djokovic. Wavelet transforms. In W.K. Chen, editor, *The Circuits and Filters Handbook*, chapter 6. 1995.

[342] P.A. Valdés-Sosa, M. Vega-Hernndez, J.M. S'anchez-Bornot, E. Martínez-Montes, and M.A. Bobes. EEG source imaging with spatio-temporal tomographic nonnegative independent component analysis. *Human Brain Mapping*, 30:1898–1910, 2009.

[343] S. Van Huffel, H. Chen, C.P. Decanniere, and P. Van Hecke. Algorithm for time-domain NMR data fitting based on total least squares. *J. Magn. Res. A*, 110:228–237, 1994.

[344] M. van Turennout, P. Hagoort, and C. M. Brown. Brain activity during speaking: From syntax to phonology in 40 milliseconds. *Science*, 280:572 – 574, 1998.

[345] K. Vanderperren, B. Hunyadi, M. De Vos, M. Mennes, H. Wouters, B. Vanrumste, P. Stiers, L. Lagae, and S. Van Huffel. Reduction of alpha distortion in event related potentials. In *Proc. of the 4th European Medical and Biomedical Engineering congress (eMBEC)*, pages 1298–1301, Antwerp, Belgium, 2008.

[346] K. Vanderperren, J. Ramautar, N. Novitski, M. De Vos, M. Mennes, B. Vanrumste, P. Stiers, B. Van den Bergh, J. Wagemans, L. Lagae, S. Sunaert, and S. Van Huffel. Ballistocardiogram artifacts in simultaneous eeg-fmri acquisitions. *International Journal of Bioelectromagnetism, Special Issue on Methods for the estimation of brain activity I*, 9:146–150, 2007.

[347] K. Vanderperren, S. Riés, M. De Vos, B. Vanrumste, F. Alario, B. Burle, and S. Van Huffel. Removal of muscle artifacts from EEG recordings of spoken language production. In *Proc. of the Neuromath Workshop, Advanced Methods for the Estimation of Human Brain Activity and Connectivity (Neuromath)*, Jena, Germany, 2008.

[348] L. Vanhamme, A. van den Boogaart, and S. Van Huffel. Improved method for accurate and efficient quantification of MRS data with use of prior knowledge. *Journal of Magnetic Resonance*, 129:35–43, 1997.

[349] A.-J. Van Der Veen and A. Paulraj. An analytical constant modulus algorithm. *IEEE Trans. Signal. Proc.*, 44:1136–1155, 1996.

[350] A. Vergult, W. De Clercq, A. Palmini, B. Vanrumste, P. Dupont, S. Van Huffel, and W. Van Paesschen. Improving the interpretation of ictal scalp EEG: BSS - CCA algorithm for muscle artifact removal. *Epilepsia*, 48:950–958, 2007.

[351] A. Vergult, M. De Vos, P. LeVan, J. Gotman, and S. Van Huffel. Automatic muscle artifact removal as a preprocessing technique for seizure detection. In *Proc. of the The first Annual Symposium of the IEEE/EMBS Benelux Chapter. (IEEE EMBS)*, 2006.

[352] L. Vigon, M.R. Saatchi, J.E.W. Mayhew, and R. Fernandes. Quantitative evaluation of techniques for ocular artefact filtering of eeg waveforms. *IEE Proc.-Sci. Meas. Technol.*, 147(5), 2000.

[353] J. Volpe. *Neurology of the newborn.* Philadelphia, 4th edition, 2001.

[354] S.A. Vorobyov, Y. Rong, N.D. Sidiropoulos, and A.B. Gershman. Robust iterative fitting of multilinear models. *IEEE Trans. on Signal Processing*, 53:2678–2689, 2005.

[355] H. Wang and N. Ahuja. Facial expression decomposition. In *Proceedings of the 9th IEEE International Conference on Computer Vision*, pages 958–965, Nice, France, 2003.

[356] A. Wauquier. EEG and pharmacology. In E. Niedermeyer and F. Lopes da Silva, editors, *Basic principles, clinical applications and related fields*, chapter 33, pages 689–698. 1987.

[357] Z. Weidong and J. Gotman. Removal of EMG and ECG artifacts from EEG based on wavelet transform and ICA. In *26th Annual International Conference of the Engineering in Medicine and Biology Society*, pages 392 – 395, San Fransisco, USA, 2004.

[358] E.M. Whitham, K.J. Pope, S.P. Fitzgibbon, Lewis T., C.R. Clark, S. Loveless, M. Broberg, Wallace A., D. DeLosAngeles, P. Lillie, A. Hardy, R. Fronsko, Pulbrook A., and J.O. Willoughby. Scalp electrical recording during paralysis: quantitative evidence that EEG frequencies above 20 Hz are contaminated by EMG. *Clinical Neurophysiology*, 118:1877–1888, 2007.

[359] T.B.J. Wiederin, K.H. Chiappa, T. Krings, B.N. Cuffin, D.B. Hoch, A.J. Cole, and G.R. Cosgrove. The utility of dipole source analysis of seizure onsets in the localization of epileptogenic zones as assessed by postsurgical outcome. *J. Contemp. Neurol.*, 1A:2–11, 1999.

[360] B. Wilken, P. Dechent, J. Herms, C. Maxton, E. Markakis, F. Hanefeld, and J. Frahm. Quantitative proton magnetic resonance spectroscopy of focal brain lesions. *J. Pediatr. Neurol.*, 23:22–31, 2000.

[361] A. Yeredor. Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation. *IEEE Trans. Signal. Proc.*, 50:1545–1553, 2002.

[362] V. Zarzoso and P. Comon. Robust independent component analysis for blind source separation and extraction with application in electrocardiography. In *Engineering in Medicine and Biology Society (EMBS). 30th Annual International Conference of the IEEE*, pages 3344–3347, 2008.

[363] V. Zarzoso, R. Phlypo, and P. Comon. A contrast for Independent Component Analysis with priors on the source kurtosis signs. *IEEE Signal Processing Letters*, 15:501–504, 2005.

[364] Y. Zhang and M. Amin. Blind separation of nonstationary sources based on spatial time-frequency distributions. *EURASIP J Appl. Signal Processing*, ID 64785, 2006.

[365] L. Zhukov, D. Weinstein, and C. Johnson. Independent component analysis for EEG source localization. *IEEE Eng. Med. Biol. Mag.*, 19:87–96, 2000.

Publication List

Papers in international journals

1. **M. De Vos**, T. Laudadio, A.W. Simonetti, A. Heerschap and S. Van Huffel. Fast nosologic imaging of the brain. Journal of Magnetic Resonance, vol. 184, no. 2, Jan. 2007, pp. 292-301.
2. **M. De Vos**, A. Vergult, L. De Lathauwer, W. De Clercq, S. Van Huffel, P. Dupont, A. Palmini and W. Van Paesschen. Canonical decomposition of ictal scalp EEG reliably detects the seizure onset zone. Neuroimage, vol. 37, no. 3, Sep. 2007, pp. 844-854.
3. **M. De Vos**, L. De Lathauwer, B. Vanrumste, S. Van Huffel and W. Van Paesschen. Canonical Decomposition of ictal scalp EEG and accurate source localisation: Principles and simulation study. Computational Intelligence and Neuroscience, Special issue on EEG/MEG , vol. 2007, no. Article ID 58253, Dec. 2007, pp 1-10.
4. K. Vanderperren, J. Ramautar, N. Novitski, **M. De Vos**, M. Mennes, B. Vanrumste, P. Stiers, B. Van den Bergh, J. Wagemans, L. Lagae, S. Sunaert and S. Van Huffel. Ballistocardiogram artifacts in simultaneous EEG-fMRI acquisitions. International Journal of Bioelectromagnetism, Special Issue on Methods for the estimation of brain activity I, vol. 9, no. 3, Dec. 2007, pp. 146-150.
5. W. Deburchgraeve, P.J. Cherian, **M. De Vos**, R.M. Swarte, J.H. Blok, G.H. Visser, P. Govaert and S. Van Huffel. Automated neonatal seizure detection mimicking a human observer reading EEG. Clinical Neurophysiology, vol. 119, no. 11, Nov. 2008, pp. 2447-2454.
6. H. Hallez, **M. De Vos**, B. Vanrumste, P. Van Hese, S. Assecondi, K. Van Laere, P. Dupont P., W. Van Paesschen, S. Van Huffel and I. Lemahieu. Removing muscle and eye artifacts using blind source separation techniques in ictal EEG source imaging. Clinical Neurophysiology, vol 120, Jun. 2009, pp. 1262-1272.
7. W. Deburchgraeve, P.J. Cherian, **M. De Vos**, R.M. Swarte, J.H. Blok, G.H. Visser, P.Govaert, S.Van Huffel. Neonatal seizure localization using PARAFAC decomposition. Clinical Neurophysiology, Nov. 2009, pp. 1787-1796.

Papers in national journals

8. **M. De Vos** Nosologische Beeldvorming als hulp bij kankerdiagnose. Het Ingenieursblad, no. 8, Sep. 2007, pp. 28-32.

Chapters in books

9. **M. De Vos**, L. De Lathauwer and S. Van Huffel. Decomposition methods in neuroscience. chapter in J.M. G'orriz, E.W. Lang, J. Ramrez: "Recent Advances in Biomedical Signal Processing" (Bentham Publishers), 2009. *invited*

Internal reports (submitted papers)

10. **M. De Vos**, L. De Lathauwer and S. Van Huffel. Spatially Constrained ICA algorithm with an application in EEG processing. Technical Report 09-02, ESAT-SISTA, K.U.Leuven (Leuven, Belgium), IEEE TSP, submitted, 2009.
11. **M. De Vos**, S. Riès, K. Vanderperren, B. Vanrumste, F. Alario, S. Van Huffel and B. Burle. Removal of muscle artifacts from EEG recordings of spoken language production. Technical Report 09-03, ESAT-SISTA, K.U.Leuven (Leuven, Belgium), Neuroinformatics, revised, 2009.
12. **M. De Vos**, L. De Lathauwer and S. Van Huffel. A combination of Parallel Factor and Independent Component Analysis. Technical Report 09-04, ESAT-SISTA, K.U.Leuven (Leuven, Belgium), IEEE TSP, submitted, 2009.
13. N. Novitskiy, J. Ramautar, K. Vanderperren, **M. De Vos**, B. Vanrumste, P. Stiers, B. Van den Bergh, L. Lagae, S. Sunaert, S. Van Huffel and J. Wagemans The BOLD correlates of the visual P1 and N1 in single-trial analysis of simultaneous EEG-fMRI recording during a spatial detection task. Technical Report, 2009.
14. K. Vanderperren, **M. De Vos**, J. Ramautar, N. Novitskiy, S. Assecondi, B. Vanrumste, P. Stiers, B. Van den Bergh, J. Wagemans, L. Lagae, S. Sunaert and S. Van Huffel Removal of BCG artifacts from EEG recordings inside the MR scanner: a comparison of methodological and validation-related aspects. Neuroimage, revised, 2009.
15. W. Deburchgraeve, P.J. Cherian, **M. De Vos**, R.M. Swarte, J.H. Blok, G.H. Visser, P.Govaert, S.Van Huffel. Time varying neonatal seizure localization. Methods of Information in Medicine, minor revision, 2009.
16. B. Mijovic, **M. De Vos**, I. Gligorijevic, J. Taelman and S. Van Huffel. Source separation from single channel recordings. IEEE Transactions on Biomedical Engineering, submitted, 2009.

Papers in proceedings of international conferences

17. **M. De Vos**, L. De Lathauwer, A. Vergult, W. De Clercq, W. Van Paesschen and S. Van Huffel. Spatially Constrained Independent Component Analysis for real-time eye artifact removal from the electroencephalogram. Proc. of the first Annual Symposium of the IEEE/EMBS Benelux Chapter. (IEEE EMBS), Brussel, Belgium, Dec. 2006, pp. 159-163.
18. A. Vergult, **M. De Vos**, P. LeVan, J. Gotman and S. Van Huffel. Automatic muscle artifact removal as a preprocessing technique for seizure detection. Proc. of the The first Annual Symposium of the IEEE/EMBS Benelux Chapter. (IEEE EMBS), Brussel, Belgium, Dec. 2006, on cdrom.
19. W. Deburchgraeve, B. Vanrumste, A. Vergult, **M. De Vos**, W. Van Paesschen and S. Van Huffel. Automatic detection and localization of epilepsy in the electroencephalogram. Proc. of the The first Annual Symposium of the IEEE/EMBS Benelux Chapter. (IEEE EMBS), Brussel, Belgium, Dec. 2006, pp. 321-324.
20. H. Hallez, **M. De Vos**, R. Phlypo, B. Vanrumste, P. Van Hese, Y. D'Asseler, W. Van Paesschen, S. Van Huffel and I. Lemahieu. Can dipole modelling be improved by removing muscular and ocular artifacts from ictal scalp EEG? Proc. of the EMBC neural engineering conference, Kohala Coast, Hawaii, May 2007, pp. 252-255.
21. **M. De Vos**, L. De Lathauwer and S. Van Huffel. Imposing independence constraints in the CP model. Proc. of the 7th international conference on independent component analysis (ICA 07), London, England, Sep. 2007, pp. 33-40.
22. **M. De Vos**, L. De Lathauwer, A. Vergult, W. De Clercq, W. Van Paesschen and S. Van Huffel. Jacobi iterations for spatially constrained Independent Component Analysis. Proc. of the 7th international conference on mathematics in signal processing (IMA 2007), Cirencester, England, Dec. 2007, pp. 38-41.
23. W. Deburchgraeve, P.J. Cherian, **M. De Vos**, R.M. Swarte, J.H. Blok, G.H. Visser, P. Govaert and S. Van Huffel. Automated neonatal spike train detection as part of a neonatal seizure detection system. Proc. Joint Meeting of the 6th International Symposium on Noninvasive Functional Source Imaging of the Brain and Heart and the International Conference on Functional Biomedical Imaging (NFSI), Hangzhou, China, Oct. 2007, pp. 222-225.
24. **M. De Vos**, L. De Lathauwer, B. Vanrumste, W. Deburchgraeve, W. Van Paesschen and S. Van Huffel. Canonical Decomposition of scalp EEG as preprocessing for source localisation. Proc. of the the 6th International Symposium on Noninvasive Functional Source Imaging of the Brain (NFSI 07), Hangzhou, China, Oct. 2007, pp. 377-400.
25. **M. De Vos**, L. De Lathauwer and S. Van Huffel. Second order algorithm for imposing independence constraints to the CP model. Proc. of the IEEE International Symposium on Circuits and Systems, Seattle, USA, pp. 1344-1347. *invited paper*

26. **M. De Vos**, L. De Lathauwer, W. Van Paesschen and S. Van Huffel. Canonical Decomposition of scalp EEG in epileptic seizure localisation. Proc. of the Asilomar conference on signals, systems and computers (asilomar), Asilomar, California, Nov. 2007, pp. 408-411. *invited paper*
27. K. Vanderperren, B. Hunyadi, **M. De Vos**, M. Mennes, H. Wouters, B. Vanrumste, P. Stiers, L. Lagae and S. Van Huffel. Reduction of alpha distortion in event related potentials. Proc. of the 4th European Medical and Biomedical Engineering congress (eMBEC), Antwerp, Belgium, Nov. 2008, pp. 1298-1301.
28. W. Deburchgraeve, P.J. Cherian, **M. De Vos**, R.M. Swarte, J.H. Blok, G.H. Visser, P. Govaert and S. Van Huffel. Exploration of the CP decomposition combined with SVD analysis for neonatal oscillatory seizure localization Proc. of the 6th international Workshop on biosignal interpretation (BSI 2009), New Haven, USA, June 2009, pp. 1-4.
29. J. De Tollenaere, W. Deburchgraeve, **M. De Vos**, S. Van Huffel New Approaches for the Adaptive Segmentation of Neonatal EEG Signals. Proc. of the IEEE Benelux chapter, accepted, 2009

Abstracts in proceedings of international conferences

30. **M. De Vos**, T. Laudadio, A.W. Simonetti, A. Heerschap and S. Van Huffel. Fast nosologic imaging of the brain. Proc. of the 7th International Society of Magnetic Resonance in Medicine (ISMRM), Seattle, US, May 2006.
31. **M. De Vos**, L. De Lathauwer and S. Van Huffel. Imposing independence constraints to the CP model. Proc. of Workshop on Algorithms for Modern Massive Data Sets (MMDS), Stanford, USA, June 2006.
32. **M. De Vos**, P. Pels, T. Laudatio, S. Van Huffel, J. Kurhanewic, and S. Nelson. Automation of spectroscopic scoring of prostate MRSI by a canonical analysis based classification. Proc. of the European Society of Magnetic Resonance in Medecin and Biology (ESMRMB), Warsaw, Poland, Sep. 2006.
33. **M. De Vos**, A. Vergult, L. De Lathauwer, W. De Clercq, S. Van Huffel, P. Dupont, A. Palmini and W. Van Paesschen. Canonical decomposition of ictal EEG reliably detects the seizure onset zone. Proc. of the American Epilepsy Society conference (AES), San Diego, US, Dec. 2006.
34. K. Vanderperren, J. Ramautar, N. Novitski., **M. De Vos**, M. Mennes, B. Vanrumste, P. Stiers, B. Van den Bergh, J. Wagemans, L. Lagae, S. Sunaert and S. Van Huffel. Study of ballistocardiogram artifacts in simultaneous EEG-fMRI acquisitions. Proc. of the Neuromath Workshop, Advanced Methods for the Estimation of Human Brain Activity and Connectivity (Neuromath), Rome, Italy, Dec. 2007, pp. 121-122.
35. W. Deburchgraeve, P.J. Cherian, **M. De Vos**, R.M. Swarte, J.H. Blok, G.H. Visser, P. Govaert and S. Van Huffel. Automated neonatal seizure detection. Proc. of the 3rd International Conference on Brain Monitoring and Neuroprotection in the Newborn, Vienna, Austria, Jan. 2008.

36. K. Vanderperren, S. Ris, **M. De Vos**, B. Vanrumste, F. Alario, B. Burle and S. Van Huffel, Removal of muscle artifacts from EEG recordings of spoken language production. Proc. of the Neuromath Workshop, Advanced Methods for the Estimation of Human Brain Activity and Connectivity (Neuromath), Jena, Germany, April 2008, pp. 121-122.

37. W. Deburchgraeve, P.J. Cherian, **M. De Vos**, R.M. Swarte, J.H. Blok, G.H. Visser, P. Govaert and S. Van Huffel. Automated ECG artefact reduction in EEG as a preprocessing step for neonatal seizure detection. Proc. of the Neuromath Workshop, Advanced Methods for the Estimation of Human Brain Activity and Connectivity (Neuromath), Jena, Germany, April 2008.

38. **M. De Vos**, L. De Lathauwer, W. Deburchgraeve, K. Vanderperren and S. Van Huffel. A wide range of BSS applications in neuroscience. Proc. of the 4th European Medical and Biomedical Engineering congress (eMBEC), Antwerp, Belgium, Nov. 2008.

39. N. Novitski, J. Ramautar J., K. Vanderperren, **M. De Vos**, M. Mennes, B. Vanrumste, P. Stiers, B. Van den Bergh, J. Wagemans, L. Lagae, S. Sunaert and S. Van Huffel. Simultaneous fMRI-EEG study of early cortical processing of lateralized visual stimuli. Proc. of Neuroscience & Cognitive Control, Ghent, Belgium, Dec. 2008.

40. W. Deburchgraeve, P.J. Cherian, **M. De Vos**, R.M. Swarte, J.H. Blok, G.H. Visser, P. Govaert and S. Van Huffel. Automated neonatal seizure detection mimicking a human observer reading EEG. Proc. of he 4th International Conference on Brain Monitoring and Neuroprotection in the Newborn, Feb. 2009.

41. N. Novitski, J. Ramautar, K. Vanderperren, **M. De Vos**, M. Mennes, B. Vanrumste, P. Stiers, B. Van den Bergh, J. Wagemans, L. Lagae, S. Sunaert and S. Van Huffel. Detection of lateralized visual stimuli as an evaluation paradigm for simultaneous EEG/fMRI acquisition and analysis. Proc. of the Neuromath Workshop, Advanced Methods for the Estimation of Human Brain Activity and Connectivity (Neuromath), Leuven, Belgium, March 2009.

42. B. Mijovic, K. Vanderperren, **M. De Vos**, J.R. Ramautar, N. Novitskiy, B. Vanrumste, P. Stiers, B. Van den Bergh, J. Wagemans, L. Lagae, S. Sunaert and S. Van Huffel. Influence of the Stimulus Onset Asynchrony in fMRI experiment to the Independent Component Analysis results. Proc. of Neuromath, Limassol, Cyprus, 2009.

Curriculum vitae

Maarten De Vos was born in Duffel, Belgium on July 16, 1983. He received the degree of electrotechnical-mechanical engineering, specialization biomedical techniques, from the Katholieke Universiteit Leuven, Belgium in July 2005. In 2005, he was laureate of K.V.I.V. engineering prizes for his Master's thesis. Since then, he has been working as a researcher at the Electrical Engineering Department ESAT of the K.U.Leuven. From October 2005, he worked on a K.U. Leuven PhD grant, and since January 2006 on an I.W.T grant (Flemisch Institute for Scientific and Technological Research in Industry). He applied for a patent (GB0610795.7) and was invited speaker at Asilomar '07, Mobilab study day '07 and ISCAS '08. Besides he also studied violin with *o.a.,* Liviu Prunaru and Henry Raudales.

Arenberg Doctoral School of Science, Engineering & Technology
Faculty of Engineering
Department of Electrical Engineering
Research group SISTA-BIOMED
Kasteelpark Arenberg 10, B - 3001 Leuven

KATHOLIEKE UNIVERSITEIT LEUVEN

ASSOCIATIE K.U.LEUVEN