KU LEUVEN

ARENBERG DOCTORAL SCHOOL
Faculty of Engineering Science

Tensor-based blind source separation for structured EEG-fMRI data fusion

Simon Van Eyndhoven

Supervisors:
Prof. dr. ir. S. Van Huffel
Prof. dr. ir. B. Hunyadi
(Delft University of Technology)
Prof. dr. ir. L. De Lathauwer

Dissertation presented in partial fulfillment of the requirements for the degree of Doctor of Engineering Science (PhD): Electrical Engineering

January 2020

Tensor-based blind source separation for structured EEG-fMRI data fusion

Simon VAN EYNDHOVEN

Examination committee:
Prof. dr. A. Bultheel, chair
Prof. dr. ir. S. Van Huffel, supervisor
Prof. dr. ir. B. Hunyadi
(Delft University of Technology)
Prof. dr. ir. L. De Lathauwer
Prof. dr. W. Van Paesschen
Prof. dr. P. Dupont
Prof. dr.ir. F. Maes
Prof. dr.ir. D. Mantini

Prof. dr. P. A. Valdés-Sosa
(Cuban Neuroscience Center, Havana, Cuba)

Dissertation presented in partial fulfillment of the requirements for the degree of Doctor of Engineering Science (PhD): Electrical Engineering

January 2020

© 2020 KU Leuven – Faculty of Engineering Science
Uitgegeven in eigen beheer, Simon Van Eyndhoven, Kasteelpark Arenberg 10 Box 2446, B-3001 Leuven (Belgium)

Alle rechten voorbehouden. Niets uit deze uitgave mag worden vermenigvuldigd en/of openbaar gemaakt worden door middel van druk, fotokopie, 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, electronic or any other means without written permission from the publisher.

Abstract

A complex physical system like the human brain can only be comprehended by the use of a combination of various medical imaging techniques, each of which shed light on only a specific aspect of the neural processes that take place beneath the skull. Electroencephalography (EEG) and functional magnetic resonance (fMRI) are two such modalities, which enable the study of brain (dys)function. While the EEG is measured with a limited set of scalp electrodes which record rapid electrical changes resulting from neural activity, fMRI offers a superior spatial resolution at the expense of only picking up slow fluctuations of oxygen concentration that takes place near active brain cells. Hence, combining these very complementary modalities is an appealing, but complicated task due to their heterogeneous nature.

In this thesis, we devise advanced signal processing techniques which integrate the multimodal data stemming from simultaneously recorded EEG and fMRI. We focus on their application in refractory epilepsy, wherein some brain cells undergo hypersynchronous activity, leading to seizures that cannot be suppressed by medication. In such cases, EEG–fMRI can aid a presurgical evaluation of the patient, to infer the location in the brain where epileptic discharges originate.

In the first part of the thesis, we propose a semi-automatic technique which drastically improves the EEG signal quality, based only on a few examples of recorded epileptic discharges. We use this output to inform an fMRI analysis which searches the brain volume for correlated fluctuations, and demonstrate that this approach is more sensitive than existing methods in detecting fMRI signal changes in epileptic brain regions.

In the second part, we develop data fusion approaches based on representations of the data as tensors ('multidimensional arrays') to capture the rich, complex nature of EEG and fMRI, and to exploit their attractive properties for data mining. We perform simultaneous blind signal separation of both data tensors into distinct 'components', each representing a source of variation in the data. Our experiments show that these customized coupled tensor decompositions are not only able to extract components that model the temporal, spatial and spectral profiles of epileptic discharges, but also to estimate the variable functional relationship between EEG and fMRI, i.e., neurovascular coupling. Clinical validation shows that these novel techniques produce complementary sets of biomarkers, which assist the characterization and diagnosis of epilepsy.

Beknopte samenvatting

Een complex fysisch systeem zoals het menselijk brein kan enkel bevat worden door het gebruik van een combinatie van verschillende medische beeldvormingstechnieken, die elk slechts een specifiek aspect belichten van de neurale processen die zich afspelen onder de schedel. Elektro-encefalografie (EEG) en functionele kernspintomografie (fMRI) zijn twee zulke modaliteiten, die het mogelijk maken om hersen(dys)functie te bestuderen. EEG wordt gemeten met slechts een beperkte set elektroden op de hoofdhuid, die snelle elektrische variaties, afkomstig van neurale activiteit, opnemen. fMRI daarentegen biedt een superieure spatiale resolutie, ten koste van een trage registratie van de schommelingen in zuurstofconcentratie nabij actieve hersencellen. Zodoende is het aanlokkelijk—maar ingewikkeld—om deze zeer complementaire modaliteiten te combineren, omwille van hun heterogene aard.

In dit proefwerk ontwerpen we geavanceerde signaalverwerkingstechnieken, die de multimodale data integreren die afkomstig is van gelijktijdig opgemeten EEG en fMRI. We focussen op hun toepassing in refractaire epilepsie, waarbij sommige hersencellen hypersynchrone activiteit vertonen die leidt tot aanvallen die niet onderdrukt kunnen worden met medicatie. In dergelijke gevallen kan EEG–fMRI een hulpmiddel zijn tijdens de prechirurgische evaluatie, om het hersengebied te bepalen waaruit de epileptische ontlading voortkomen.

In het eerste deel van dit proefwerk stellen we een halfautomatische techniek voor, die de signaalkwaliteit van EEG drastisch verbetert, uitsluitend gebruik makend van enkele voorbeelden van opgemeten epileptische ontladingen. We gebruiken dit resultaat om een fMRI-analyse te sturen die het hersenvolume doorzoekt naar gecorreleerde schommelingen, en verifiëren dat deze aanpak gevoeliger is in het detecteren van fMRI-signaalveranderingen dan bestaande methodes.

In het tweede deel ontwikkelen we datafusiemethodes, gebaseerd op het voorstellen van data als tensoren (‘meerdimensionale matrices’) om de complexe aard van EEG en fMRI te bevatten, en om hun aantrekkelijke eigenschappen voor datamining uit te buiten. We voeren gelijktijdige blinde signaalscheiding uit van beide tensoren, in afzonderlijke ‘componenten’, die elk een bron van variatie in de data voorstellen. Onze experimenten tonen aan dat deze op maat gemaakte, gekoppelde tensorontbindingen niet enkel in staat zijn om componenten te extraheren die de temporele, spatiale, en spectrale profielen

van epileptische ontladingen modelleren, maar ook om het variabele functionele verband tussen EEG en fMRI (d.w.z. neurovasculaire koppeling) te schatten. Klinische validatie toont dat deze vernieuwende technieken complementaire sets biomarkers produceren, die de karakterisatie en diagnose van epilepsie kunnen ondersteunen.

Abbreviations

A

ALS	alternating least squares

B

BCI	brain-computer interface
BOLD	blood oxygen level dependent
BSS	blind source separation
BTD	block term decomposition

C

CMTF	coupled matrix-tensor factorization
CorConDia	core consistency diagnostic
CPD	canonical polyadic decomposition
CSP	common spatial/spectral patterns
CSSP	common spatio-spectral patterns

E

EEG	electroencephalography

F

FIR	finite impulse response
fMRI	functional magnetic resonance imaging
FWE	familywise error

G

GFS	global field synchronization
GLM	general linear model

H

HbO_2	oxygenated hemoglobin
HbR	reduced hemoglobin
HRF	hemodynamic response function

I

ICA	independent component analysis
IED	interictal discharge
IOZ	ictal onset zone

K

KPE	Kronecker product equation

L

LDA	linear discriminant analysis
LFP	local field potential

M

MLE	maximum likelihood estimation
MLSVD	multilinear singular value decomposition
MWF	multi-channel Wiener Filter
MR	magnetic resonance

N

NLS	nonlinear least squares

P

PARAFAC	parallel factor analysis
PCA	principal component analysis
PSI	phase slope index

R

RF	radio frequency
ROI	region of interest
RSN	resting-state network

S

SDF	structured data fusion
SnPM	statistical nonparametric map(ping)
SNR	signal-to-noise ratio
SPM	statistical parametric map(ping)
SVD	singular value decomposition

T

TE	echo time
TLE	temporal lobe epilepsy
TR	repetition time

U

Un	unitary (predictor)

Contents

Abstract i

Beknopte samenvatting iii

Contents ix

List of Figures xv

List of Tables xxix

1 Introduction 1

1.1 Problem statement and scope 1

1.2 Structure of this thesis 4

1.3 Collaboration and funding 7

2 EEG and fMRI imaging in epilepsy 9

2.1 Brain anatomy 9

2.2 Electroencephalography (EEG) 12

2.2.1 Which physiological variable does EEG measure? 12

2.2.2 Physical aspects of EEG 14

2.2.3 Data quality 15

2.3 Functional Magnetic Resonance Imaging (fMRI) 16

2.3.1 Which physiological variable does fMRI measure? 16

2.3.2 Physical foundations of magnetic resonance imaging . . 18

2.3.3 Data quality 22
2.4 Epilepsy as a brain disorder 24
2.4.1 Manifestation and prevalence 24
2.4.2 Diagnosis and presurgical evaluation 25
2.5 Simultaneous EEG–fMRI in epilepsy monitoring 27
2.5.1 Complementary views of interictal events 27
2.5.2 Neurovascular coupling: a double-edged sword 28
2.5.3 Data quality 31
2.6 Summary and prospects 33

3 Signal Processing and Machine Learning for Neural Data 37
3.1 Higher-order data representation 38
3.1.1 Notation 39
3.2 Unsupervised signal processing 41
3.2.1 Unsupervised filtering 41
3.2.2 Unsupervised dimensionality reduction 42
3.2.3 Blind Source Separation (BSS) 46
3.3 Supervised signal processing 55
3.3.1 Supervised filtering 55
3.3.2 Supervised dimensionality reduction 56
3.3.3 Classification 58
3.3.4 Regression 61

4 Improving localization of ictal onset zone in EEG-correlated fMRI via EEG enhancement 67
4.1 Introduction 68
4.2 Methods and materials 70
4.2.1 Patient group 70

4.2.2 Data acquisition and preprocessing 74
4.2.3 EEG-derived BOLD predictors 74
4.2.4 EEG-correlated fMRI analysis 79
4.2.5 Ictal onset zone 80
4.2.6 Model performance 81
4.3 Results for patients and matched controls 81
4.3.1 Multi-Channel Wiener filtering successfully enhances IEDs 81
4.3.2 Sensitivity and specificity of different predictors 84
4.3.3 Activation maps of individual patients 87
4.4 Discussion 91

5 Revealing variable neurovascular coupling via structured matrix-tensor factorization 97

5.1 Introduction 98
5.2 Methods 100
5.2.1 Generative model 100
5.2.2 Structured matrix-tensor factorization 100
5.2.3 Simulation study 102
5.3 Results 105
5.4 Discussion 108

6 Identifying reproducibility of matrix/tensor factorizations via low-rank graph clustering 111

6.1 Introduction 112
6.2 Methods 113
6.2.1 (Non-convex) matrix and tensor factorizations 113
6.2.2 Inter-factorization similarity as a graph 114
6.2.3 Repeated decompositions exhibit cycle consistency . . . 115

6.2.4 Clustering through low-rank graph approximation . . . 116
6.3 Experiments . . . 116
6.3.1 CPD of neuroimaging data . . . 117
6.3.2 Coupled matrix-tensor factorization of synthetic data . 118
6.4 Discussion . . . 120

7 Augmenting interictal mapping with neurovascular coupling biomarkers by structured data fusion 123

7.1 Introduction . . . 124
7.2 Methods and materials . . . 127
7.2.1 Patient group . . . 127
7.2.2 Data acquisition and preprocessing . . . 127
7.2.3 Multi-channel Wiener filtering for spatio-temporal EEG enhancement . . . 128
7.2.4 Higher-order data representation . . . 129
7.2.5 Coupled matrix-tensor factorization of EEG and fMRI . 133
7.2.6 Statistical inference . . . 136
7.2.7 Model performance . . . 140
7.3 Nonlinear fitting of the sCMTF model . . . 140
7.3.1 Accommodating noncausal HRFs . . . 140
7.3.2 Initialization and optimization . . . 141
7.4 Model selection . . . 142
7.4.1 Sign and scale standardization of factor estimates . . . 143
7.4.2 Stability analysis . . . 143
7.4.3 IED component selection . . . 144
7.4.4 Choice of the rank R of the factorization . . . 144
7.5 Experiments . . . 147
7.5.1 Patient-specific model selection . . . 147

7.5.2 Spatio-temporo-spectral profiles of interictal discharges 148
7.6 Discussion 156

8 Conclusion and prospects 165
8.1 Contributions and conclusions of this thesis 165
8.2 Future perspectives 170

A Single-channel EEG classification by multi-channel tensor subspace learning and regression 181
A.1 Introduction 182
A.2 Classifying single-channel data after multi-channel training . . 183
A.2.1 Notation and preliminaries 183
A.2.2 Training phase: subspace learning 184
A.2.3 Testing phase: tensor regression and classification . . . 186
A.3 Experimental results on a mental task 188
A.3.1 Public EEG dataset 188
A.3.2 Preprocessing and analysis 189
A.3.3 Results and discussion 190
A.4 Conclusion 192

B Supplement to: Augmenting interictal mapping with neurovascular coupling biomarkers by structured data fusion 195

List of publications 249

List of Figures

1.1 This thesis is an interdisciplinary endeavor, and focuses on the development of new (tensor-based) signal processing techniques (right block), as well as their application to neuroimaging in epilepsy (left block). The algorithms can further be broadly categorized as supervised (relying on assistance from a human) or unsupervised (fully independent) techniques which 'learn' interesting aspects from data. The diagram shows the logical structure of the thesis; arrows indicate logical coherence between the chapters, and the overlap of chapters with the orange blocks indicate which type of algorithms and experimental validation are involved. (Abbreviations: EEG = electroencephalography; fMRI = functional magnetic resonance imaging; HRF = hemodynamic response function). 5

2.1 A neuron consists of a body (soma), from which several dendrites branch out to receive incoming electrochemical signals from neighbouring neurons at several connecting junctions (synapses). There, neurotransmitter molecules bind to receptor proteins, triggering the opening of ion channels across the cell membrane and hence changing the electrical potential. When the integrated signals exceed a threshold, an action potential is released and forwarded along the axon, towards a synapse with another neuron. Figure taken from (Wikipedia, 2007). 10

2.2 If the integrated potentials from many synapses make the postsynaptic cell membrane potential exceed a threshold, an action potential is fired in the postsynaptic neuron. Excitatory potentials tend to depolarize the cell membrane (i.e., bring the membrane potential closer to the threshold), while inhibitory potentials have a hyperpolarizing effect (i.e., reducing the likelihood of firing). Figure taken from (Queensland Brain Institute, 2019). 11

2.3 On the macroscopic level, the brain consists of several lobes. Figure taken from (WebMD, 2014). 12

2.4 Electroenceophalography (EEG) is the measurement of electrical fields at the scalp. **(a)** Electrodes are spread over the head, and are given a position-dependent name (e.g., F = frontal, P = parietal, T = temporal, C = central, O = occipital lobe, even/uneven number for left/right). **(b)** EEG signals are able to trace brain activity with a high temporal resolution, but are prone to unwanted artifacts that reduce signal quality. This segment contains three large artifacts due to eye blinking. Figure taken from (National Center for Adaptive Neurotechnologies, 2019). 13

2.5 A magnetic resonance (MR) scanner produces 3D images of the human body or brain, by interrogating a quantum-mechanical property known as 'spin' of protons at each position in the volume. **(a)** Most MR scanners are cylindrically shaped because of their large, superconducting coils which generate a very strong magnetic field, which serves to exploit the physical principle of magnetic resonance of atomic nuclei. Gradient coils superimpose small magnetic fields, which are different at each position—this attaches 'labels' to all positions, to tell them apart when making the image. The strong magnetic field initially aligns the spin of all nuclei, until a third coil sends a brief radio frequency (RF) pulse into the volume, which temporarily energizes the spins of resonant nuclei. Before the spins return to baseline, they produce a measurable electromagnetic signal, which is used to make an image. **(b)** A structural or anatomical MR image can highlight different *tissues* in the brain, as shown in cross-sections along three orthogonal axes. In this acquisition mode, the scanner is sensitive to the density of hydrogen nuclei (protons) and the duration of their resonance, which is different in grey and white matter tissue. **(c)** In functional MRI, the location of brain *activity* can be detected. In this acquisition mode, the resonance of protons is altered near regions where a lot of oxygen is consumed due to normal or abnormal (e.g. epileptic) activity in the neurons. Figure **(a)** taken from (Amber Diagnostics, 2016). 17

2.6 Seizures and interictal epileptic discharges (IEDs) or spikes are both types of epileptiform neural activity, which can be measured with electroencephalography (EEG). The left panel shows the development of a focal seizure into a typical oscillatory pattern, which is mainly present over the right temporal lobe, measured by the bottom set of bipolar EEG channels. The right panel shows a brief EEG segment from another patient, containing an IED with a characteristic (sharp) spike and (slow) wave complex in the left temporal lobe. Figure taken from (Van Mierlo et al., 2014; Javidan, 2012). 24

2.7 The natural variation of the neurovascular coupling over the brain can confound fMRI data analysis if it is not accounted for, and lead to erroneous interpretation of the data. **(a–b)** Map of one patient's brain regions which are significantly active (in dark grey , resp. orange) during epileptiform activity which was found on EEG, when adjusting for neurovascular variability. **(c)** The actual regional coupling characteristic (grey line) can differ substantially from the standard model which is often used (orange line). **(d–e)** Same as **(c)**, for two other locations in the brain. Also there, the model does not describe the real (estimated) coupling well. **(f)** The same patient's results, but when using the standard model. Compared to **(a)**, several different regions are now marked as significant. Figure edited from (Lemieux et al., 2008; Salek-Haddadi et al., 2006). 30

2.8 Gradient artifacts heavily corrupt the EEG during the fMRI acquisition. **(a)** EEG recorded outside the MR scanner, with good quality and an observable interictal spike. **(b)** EEG recorded inside the MR scanner, when fast gradient switching for the fMRI acquisition introduces large artifacts in the signals. **(c)** EEG signals recorded inside the MR scanner, after a successful artifact removal procedure. A similar interictal spike as in the left panel can now be observed. Figure taken from (Gotman, 2008). 32

2.9 The ballistocardiogram artifact results from slight movements of the conducting electrodes inside the scanner's magnetic field, which stem from the pulsed motion of blood in the scalp. **(a)** The artifact appears approximately time-locked to the heart beats (cfr. bottom two signals). Also three interictal spikes can be seen. **(b)** The EEG can be cleaned by using an artifact removal procedure, in this case independent component analysis (ICA). Figure taken from (Gotman, 2008). 32

3.1 Any Mth–order tensor $\boldsymbol{\mathcal{T}}$ can be expressed by means of the multilinear singular value decomposition (MLSVD) as the product of an all-orthogonal core tensor $\boldsymbol{\mathcal{G}}$ of the same size as $\boldsymbol{\mathcal{T}}$ and square, orthogonal factor matrices $\mathbf{U}^{(m)}$, $m = 1 \dots M$. Here, we show an example for a tensor with $M = 3$ modes. Truncating the core tensor and factor matrices (leading to a compact core $\bar{\boldsymbol{\mathcal{G}}}$ and thin matrices $\bar{\mathbf{U}}^{(1)}$, $m = 1 \dots M$) can often result in a significant compression of the tensor data with only minor loss of information. 44

3.2 Canonical Polyadic Decomposition (CPD) of a third-order tensor $\boldsymbol{\mathcal{T}}$ into factor matrices $[\![\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \mathbf{A}^{(3)}]\!]$, each of which has a column dimension R. The CPD can equivalently be written as a sum of R rank-1 tensors, each of which is the outer product of 3 signatures. In a blind source separation (BSS) context, every rank-1 term is interpreted as modeling the contribution of one source. 48

3.3 Data fusion can offer a holistic view on a certain phenomenon when complementary data are available. Coupled tensor decompositions reveal sources that underlie the data which are fully or partially shared between the involved tensors. An example is shown of a partially coupled matrix-tensor factorization (CMTF) of a tensor $\boldsymbol{\mathcal{T}}$ and a matrix $\mathbf{M}$, which are measured along a shared first mode. Both $\boldsymbol{\mathcal{T}} \approx [\![\mathbf{C}, \mathbf{A}^{(2)}, \mathbf{A}^{(3)}]\!]$ and $\mathbf{M} \approx [\![\bar{\mathbf{C}}, \mathbf{B}^{(2)}]\!]$ are decomposed into a sum of rank-1 terms, two of which are common to both datasets. A third source (with green mode-1 signature) is not shared and is specific to the decomposition of $\boldsymbol{\mathcal{T}}$. 52

3.4 The general linear model (GLM) in fMRI models all voxels' time courses as the sum of contributions from effects of interests and nuisance time courses, weighted by coefficients which are estimated through regression. By fitting the model (separately) at every voxel, spatial maps of the brain activity of (no) interest can be constructed, and statistical hypotheses on the presence or absence of an effect can be tested in a principled manner. Figure taken from (Monti, 2011). 64

4.1 Example EEG segments of patient 3 illustrate the artifact removal capabilities of the Multi-channel Wiener filter (MWF). Time is indicated in seconds relative to the first spike of the discharge. (a) The MWF properly preserves the signal during annotated periods of interictal discharges, which is evidenced by the strong similarity between the average of the 105 original discharges, and the average of their filtered versions. (b) The MWF successfully attenuates the large artifact, which occurs between 22 and 26 s, because the waveforms do not follow the spatio-temporal statistics of the IEDs, which were used to calibrate the MWF. At the same time, the signals during the discharge around 20 s, which was annotated by the neurologist, stay preserved in the MWF output (compare to the average discharge in (a)). 83

4.2 Example EEG segments of patient 10 showcase the error-correcting capabilities of the Multi-channel Wiener filter (MWF). Time is indicated in seconds relative to the first spike of the discharge. (a) Just as in Figure 4.1, the MWF preserves the signal during annotated periods of interictal discharges, respecting to large extent the annotations that were made by the human annotator. (b) Sometimes, annotated discharges are (partially) attenuated, if the instance does not correspond well to the average statistics that have been used to train the MWF. In this case, this effect is visible in the window of 1.5 seconds around an annotated discharge, on channels O2, C4, T7, ... On the other hand, discharges that were missed by the human annotator may be uncovered by the MWF by enhancing waveforms in the output. In this example, the signal is enhanced between -3...-2 s relative to the annotated spike. Indeed, the waveforms in this period show striking correspondences to the average of the annotated discharges of this patient (in (a)). 85

4.3 The statistical evidence found in the ictal onset zone is highest for especially the MWFpow and also ICApow regressors, and lower for the other regressor types, for most patients (shown by individual markers). This is correlated with a higher sensitivity (i.e., more true positives) for these two regressors, especially at higher (more selective) thresholds to the right of the figure, where synchronization regressors (PSI and GFS) and the unitary regressor Un fail. At different thresholds, cases with clusters of suprathreshold T-values that overlap with the ictal onset zone are counted as true positives (•), other cases are considered false negatives (△). At the lowest threshold ($Z > 3.1$, $k > 0$), all regressors were 0% specific (i.e., produced suprathreshold activations in all healthy controls), and at the three higher thresholds, all of them were 100% specific (no suprathreshold activations in healthy controls), except for one false positive case produced by ICApow at FWE-corrected $p < 0.05$, $k > 0$. . . . 87

4.4 The sensitivity is highest for the MWFpow regressor, whether the overlap with the ictal onset zone of all clusters (left) or only of the maximally active cluster (right) is considered. At the threshold which was found to be most selective in (Tousseyn et al., 2014a), i.e. $Z > 3.4$ with cluster size constraint $k > 350$ voxels or 2.8 cm^3, the ranking of all regressors is the same for both types of sensitivity. The colours for both types of sensitivity correspond to the colours of the clusters in Figure 4.5, 4.6, 4.7 88

4.5 Patient 7's ictal onset zone (blue) was correctly detected by the EEG-correlated fMRI activation maps (red) of all predictors except GFS at $Z > 3.4$ with cluster size constraint $k > 350$ voxels (2.8 cm^3). The maps of MWFpow and ICApow are remarkably similar, while large differences with the other predictors exist. The maximally active cluster for every predictor is marked in green (instead of red). 89

4.6 Patient 9's ictal onset zone (blue) was only correctly detected by the EEG-correlated fMRI activation maps (red) of the MWFpow and ICApow predictors at $Z > 3.4$ with cluster size constraint $k > 350$ voxels (2.8 cm^3). The ICApow predictor produced one additional active cluster outside of the delineated ictal onset zone. The maximally active cluster for every predictor is marked in green (instead of red). 90

4.7 Patient 11's ictal onset zone (blue) was only detected by the EEG-correlated fMRI activation map (red) of the `Un` predictor at Z > 3.4 with cluster size constraint k > 350 voxels (2.8 cm^3). The sparse occurrence of IEDs during the recording may have precluded proper tuning of the `MWFpow` predictor, causing it to miss the ictal onset zone. The maximally active cluster for every predictor is marked in green (instead of red). 92

5.1 Structured coupled matrix-tensor factorization of (a) EEG data and (b) fMRI data. The signatures $\mathbf{a}_r$ along the temporal mode are shared between the EEG factorization and the fMRI factorization, providing the coupling constraint. 101

5.2 Location and extent of the three neural sources in the fMRI domain (a, b, c) and topographic plots of the corresponding EEG spatial signatures, obtained through a forward volume conduction model, based on the anatomical MRI image and the sources' positions (d, e, f). 103

5.3 Temporal (left column) and spectral (right column) EEG signatures, which belong to the neural sources in Figure 5.2 (a.u. = arbitrary units). 104

5.4 True underlying hemodynamic response function (full blue line) in this study, and the canonical hemodynamic response function (dotted red line) as used in the SPM toolbox. The positive (resp. negative) peaks of both HRFs are 3 seconds (resp. 4 seconds) apart. 106

5.5 Mean relative errors δ_a, δ_b, and δ_c between true and estimated sources in the EEG's temporal, spatial, and spectral mode, respectively, in the presence of 2 sources (a) or 3 sources (b). Results are shown for the case where we find the HRF in the optimization procedure ('OPT') and for the case where we fix the HRF to the canonical HRF ('CAN'). 106

5.6 Mean Dice coefficients d between true and estimated fMRI spatial signatures, in the presence of 2 or 3 active sources and WGN (a) or 2 active sources and spatially ('SpCorr') or temporally ('TempCorr') correlated noise (b). Results are shown for the case where we find the HRF in the optimization procedure ('OPT') and for the case where we fix the HRF to the canonical HRF ('CAN'). 107

6.1 Our algorithm successfully clusters spatio-spectral components that appear in nearly all runs of the decomposition and are hence reliable (left column), but also distinguishes components that are specific to the decomposition of each half (light/dark grey) of the data (right column). The mean spectra are superimposed in red. The variability of spatial signatures around their mean (which is plotted) was comparably small as for the spectra. . . 119

6.2 When grouping EEG tensor decomposition components in clusters, our algorithm realizes a favourable tradeoff between high intra-cluster similarity and low maximal inter-cluster similarity. This is thanks to the proper (automatic) estimation of the number of clusters $\tilde{R}$ (darker is higher), in an $\tilde{R}$-dimensional component space, which is not possible with ICASSO. 120

6.3 Also on synthetic multimodal data, the proposed method successfully groups components of a coupled matrix-tensor factorization in clusters which are more coherent than those of ICASSO. 121

7.1 Structured coupled matrix-tensor factorization (sCMTF) of EEG and fMRI data can reveal neural sources that are encoded in both modalities, as well as capture the varying neurovascular coupling between the electrophysiological and BOLD changes. **((a))** The EEG signals vary over time points × frequencies × electrodes. The resulting third-order spectrogram tensor $\boldsymbol{\mathcal{X}}$ is factorized according to (7.8) into R rank-1 components, which each consist of a temporal signature $\mathbf{s}_r$, a spectral signature $\mathbf{g}_r$ and a spatial signature $\mathbf{m}_r$. **((b))** The fMRI data consist of the average BOLD signal in different brain parcels or regions of interest (ROIs), represented in a time points × ROI matrix $\mathbf{Y}$, and are factorized according to (7.11). Neurovascular coupling is modeled as a convolution of the EEG temporal dynamics with a ROI-specific hemodynamic response function (HRF), as in (7.11)–(7.13). In this example, each local HRF is represented as a linear combination (encoded by coefficients $\mathbf{b}_k$) of $K = 3$ optimized basis functions, each populating a Toeplitz matrix $\mathbf{H}_k$ which implements a convolution through matrix multiplication with the temporal signatures $\mathbf{s}_r$. Afterwards, each smoothed component r is spatially weighted by a signature $\mathbf{v}_r$. This is accomplished by the elementwise product $\mathbf{b}_k * \mathbf{v}_r$ of the HRF basis function-specific coefficients $\mathbf{b}_k$ and the component-specific amplitudes $\mathbf{v}_r$. 131

7.2 Interictal EEG and fMRI data are analyzed via structured coupled matrix-tensor factorizations (sCMTF), which reveals both spatial localization of interictal discharges (spikes), and also localized deviations in neurovascular coupling between electrical and BOLD fluctuations. **(a)** fMRI and EEG data are first separately preprocessed into a time points × regions of interest (ROIs) matrix, and a channels × time points × frequencies tensor, after MWF enhancement, respectively. **(b)** The sCTMF of the EEG and fMRI data reveals temporally, spatially and spectrally resolved components, and captures spatially varying hemodynamic response functions (HRFs) (cfr. Figure 7.1). The computation is repeated N times, from different initializations, and the stability of the resulting factors over runs is assessed. **(c)** Statistical images are created for the corresponding sCMTF factors. The spike-related component is picked based on the correlation to the filtered EEG signals's broadband power. An SnPM of this component is created, revealing co-activated ROIs in a pseudo-t-map (red). For every ROI, the entropy of the HRF is computed by assessing its likelihood under the distribution of all other ROIs' HRFs, and a map of this metric is constructed (blue) to reveal localized HRF abnormalities. Both maps can be used to form a hypothesis on the location of the epileptogenic zone., Here, we validate our technique on a set of patients for which the outcome is known. 139

7.3 Different initializations of the HRF basis functions are used in every repetition of the sCMTF fitting procedure. In each repetition, one early-peaking, one mid-peaking and one late-peaking HRF were sampled from probability distributions of the HRF's generating parameters ($\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$ and $\boldsymbol{\theta}_3$, respectively) that was created by applying some multiplicative noise to baseline parameters (shown as the bold waveforms). The support extends to four negative samples, which gives the model the freedom to fit noncausal HRFs (relatively to the synchronized EEG). . . . 142

7.4 Goodness of fit of each patient's EEG tensor $\boldsymbol{\mathcal{X}}$ and fMRI matrix $\mathbf{Y}$, for varying choices of the rank R in the sCMTF. Naturally, the EEG approximation error decreases monotonically for increasing rank (intra-patient). For the fMRI data, the fit already plateaus for very low R. This is due to the presence of additional, uncoupled components $\mathbf{n}_q\mathbf{p}_q^{\mathrm{T}}$ in the fMRI factorization, which can absorb some of the variance when the number of coupled components is low, but which lose their relevance at higher ranks. 147

7.5 **(a)** In the selected solution for patient 3 ($R = 2$), both sources have a temporal signature that correlated strongly to the reference IED time course. The first source modeled the main onset of IEDs and was low-frequency and topographically focal, while the second source was spatially and spectrally diffuse and captured the propagation of IEDs to remote areas (cfr. Figure 7.9). **(b)** Five out of the twenty most deviant HRFs were found inside the ictal onset zone (bold lines, $p < 10^{-4}$). These HRFs had main peaks before 0 s, i.e., they led to BOLD changes before the corresponding EEG correlate of the IED was seen. 150

7.6 The statistical nonparametric maps of the IED-related component (top two rows) and HRF entropy/extremity maps (bottom two rows) of patient 3 show concordance with the ictal onset zone (IOZ). Especially the regions of significant IED activation were accurate, but also five out of the twenty regions with the most deviant (highest entropy) HRFs were found in the IOZ (cfr. Figure 7.5b). The ground truth ictal onset zone is highlighted in dark gray with a white contour. 151

7.7 **(a)** The sCMTF solution with $R = 5$ sources was selected for patient 10. One source's temporal signature is highly correlated with the reference IED time course and is identified as the IED-related source, which has a characteristic low-frequency behaviour and with a frontotemporal topography, consistent with the IOZ location. The second source, which has very narrowband power around ± 33 Hz, likely captured an artifact of the MR acquisition. The fourth source captured α activity in the default mode network (cfr. also Figure 7.10). **(b)** Three out of the twenty most deviant HRFs were found inside the ictal onset zone (bold lines, $p = 0.02$). 153

7.8 The statistical nonparametric maps of the IED-related component (top two rows) and HRF entropy/extremity maps (bottom two rows) of patient 10 show concordance with the ictal onset zone (IOZ). IED occurrences were associated with significantly active regions in and near the IOZ (top row), and at the same time with a deactivation in a part of the default mode network (second row). Three out of the twenty regions with the most deviant (highest entropy) HRFs were found in the IOZ (cfr. Figure 7.7b). The ground truth ictal onset zone is highlighted in dark gray with a white contour. 154

7.9 The second component in patient 3 likely captured the propagation of IEDs from the irritative zone, given its relatively large correlation to the MWF envelope (cfr. Figure 7.5a). The ground truth ictal onset zone is highlighted in dark gray with a white contour. 156

7.10 The fourth component in patient 10 seemed to pick up activity in the Default Mode Network (DMN), predominantly in the α band (cfr. Figure 7.7a). 160

A.1 During training, common spectral pattern (CSP) filters are tuned to frequency bands in which the spectral power is discriminative for the class. For a particular subject, the filterwhich maximizes variance for condition 1 (the task) selects a band around 20 Hz. On the other hand, the filter that focuses on the other class of trials selects complementary bands. 189

A.2 Good single-channel classification accuracy can be attained by the tensor subspace learning and regression method. The performance depends on the selected channel, and differs between subjects (coded by color). This is indicated by black lines at the mean accuracy (over all repetitions) of subjects with the highest, median, and lowest performance. 191

A.3 Support vector machine (SVM) classifiers, especially the linear kernel SVM, are more robust than the K-nearest neighbor (KNN) classifiers. Tree boosting yielded the least reliable classification in the current analysis, followed by the simple `knn1` classifier, used in (Boussé et al., 2018, 2017). Colors and black levels have the same interpretation as in fig. A.2. 192

B.1 Patient 1's estimated sources and neurovascular coupling parameters. **(a)** The IED-related component correlates well with the reference time course, and is mostly a low-frequency phenomenon. The fourth source is most likely an artifact (of unknown origin), picked up mostly by two central channels. **(b)** One of the ROIs with the highest-entropy HRFs belongs to the ictal onset zone (bold line, $p = 0.34$). 200

B.2 Patient 1's statistical nonparametric maps and HRF entropy/extremity maps. The ground truth ictal onset zone is highlighted in dark gray with a white contour. 201

B.3 Patient 2's estimated sources and neurovascular coupling parameters. **(a)** The IED-related component correlates well with the reference time course, and is mostly a low-frequency phenomenon. **(b)** One of the ROIs with the highest-entropy HRFs belongs to the ictal onset zone (bold line, $p = 0.59$). . . . 202

B.4 Patient 2's statistical nonparametric maps and HRF entropy/extremity maps. The ground truth ictal onset zone is highlighted in dark gray with a white contour. 203

B.5 Patient 4's estimated sources and neurovascular coupling parameters. **(b)** Two of the ROIs with the highest-entropy HRFs belong to the ictal onset zone (bold line, $p = 0.32$). . . . 204

B.6 Patient 4's statistical nonparametric maps and HRF entropy/extremity maps. The ground truth ictal onset zone is highlighted in dark gray with a white contour. 205

B.7 Patient 5's estimated sources and neurovascular coupling parameters. **(b)** Six of the ROIs with the highest-entropy HRFs belong to the ictal onset zone (bold line, $p < 10^{-3}$). 206

B.8 Patient 5's statistical nonparametric maps and HRF entropy/extremity maps. The ground truth ictal onset zone is highlighted in dark gray with a white contour. 207

B.9 Patient 6's estimated sources and neurovascular coupling parameters. **(b)** None of the ROIs with the highest-entropy HRFs belong to the ictal onset zone. 208

B.10 Patient 6's statistical nonparametric maps and HRF entropy/extremity maps. The ground truth ictal onset zone is highlighted in dark gray with a white contour. 209

B.11 Patient 7's estimated sources and neurovascular coupling parameters. **(b)** One of the ROIs with the highest-entropy HRFs belongs to the ictal onset zone (bold line, $p = 0.57$). 210

B.12 Patient 7's statistical nonparametric maps and HRF entropy/extremity maps. The ground truth ictal onset zone is highlighted in dark gray with a white contour. 211

B.13 Patient 8's estimated sources and neurovascular coupling parameters. **(b)** None of the ROIs with the highest-entropy HRFs belong to the ictal onset zone. 212

B.14 Patient 8's statistical nonparametric maps and HRF entropy/extremity maps. The ground truth ictal onset zone is highlighted in dark gray with a white contour. 213

B.15 Patient 9's estimated sources and neurovascular coupling parameters. **(b)** None of the ROIs with the highest-entropy HRFs belong to the ictal onset zone. 214

B.16 Patient 9's statistical nonparametric maps and HRF entropy/extremity maps. The ground truth ictal onset zone is highlighted in dark gray with a white contour. 215

B.17 Patient 11's estimated sources and neurovascular coupling parameters. **(b)** Four of the ROIs with the highest-entropy HRFs belong to the ictal onset zone (bold line, $p = 0.01$). . . . 216

B.18 Patient 11's statistical nonparametric maps and HRF entropy/extremity maps. The ground truth ictal onset zone is highlighted in dark gray with a white contour. 217

B.19 Patient 12's estimated sources and neurovascular coupling parameters. **(b)** Seven of the ROIs with the highest-entropy HRFs belong to the ictal onset zone (bold line, $p < 10^{-3}$). . . . 218

B.20 Patient 12's statistical nonparametric maps and HRF entropy/extremity maps. The ground truth ictal onset zone is highlighted in dark gray with a white contour. 219

List of Tables

4.1 List of the number of IEDs that were observed during the fMRI recording of every patient, as well as the degrees of freedom of the fMRI design matrix. 72

4.2 Clinical patient data. Abbreviations: F = female, M = male, L = left, R = right, CNS = central nervous system, DNET = dysembryoplastic neuroepihelial tumor, FCD = focal cortical dysplasia, HS = hippocampal sclerosis, cx = cortex. 73

5.1 Pearson correlation between estimated and true HRF. 108

7.1 For every patient (ID 1–12), the sCMTF model can be fitted for a varying number of sources or rank R. We select a 'good' value for R post hoc, based on four criteria which are checked intra-patient: 1) the core consistency of the EEG tensor decomposition should be high ($> 70\%$); 2) the IED-related source should be found in sufficiently many ($\geqslant 10$) of the 50 repetitions of the estimation procedure; 3) the correlation of the IED-related source's temporal signature with the reference time course, namely the MWF's broadband envelope, is preferably high; 4) the maximal pseudo t-statistic for the IED-related source's spatial signature is preferably high. 146

7.2 The sCMTF leads to three types of spatial information, which can be cross-validated against the ground truth IOZ, as defined in Section 7.2.7 and summarized for all patients in Table 4.2: 1) the EEG topography $\mathbf{v}_{\text{IED}}$ of the IED-related component; 2) the significantly activated and deactivated ROIs in the fMRI spatial signature $\mathbf{v}_{\text{IED}}$; 3) the ROIs with strongly deviating HRF waveforms, as measured with entropy and extremity. Since the EEG topography has a very low spatial resolution, and depends on the attenuation properties of the tissue as well as the orientation of the neural sources in the cortex, we only expect partial similarity to the IOZ's spatial focus; hence, we use the term 'consistent' rather than 'concordant'. 155

Chapter 1

Introduction

The World Health Organization estimates that 50 million people worldwide suffer from epilepsy, making it the most common chronic brain disease (World Health Organization, 2019). Despite a relative lack of public awareness, it poses a major global health issue, affecting people from both developing and developed countries. Nevertheless, many research efforts have been devoted to the development of methods that can aid diagnosis and treatment. Medical imaging techniques such as electroencephalography (EEG) and functional magnetic resonance (fMRI) lie at the heart of these endeavors. Given that we live in an era of rapid technological development, in which increasingly more computing power becomes available to perform complex analyses of ever larger (medical) data, innovation in this field will come from a symbiosis between the neurologist's skill, sophisticated engineering, and sheer processing speed. In this thesis we develop data-driven techniques based on algebraic entities known as tensors, which enable a researcher or clinician to perform advanced analyses of simultaneously recorded EEG and fMRI, in an effort to combat epilepsy in a part of the patient population.

1.1 Problem statement and scope

Epilepsy is characterized by recurrent seizures, which can provoke a variety of symptoms, ranging from muscle spasms and stiffening, to loss of consciousness, imagined sensory sensations, and others. They result from a defect in a part of the brain, causing a critical mass of brain cells to synchronously fire beyond a normal level, resulting in highly regular brain activity patterns which overtake normal brain function by inciting successively more brain cells. Since seizures often occur unpredictably, and are frequently accompanied by loss of body control and/or consciousness, the disorder heavily affects the quality of life of patients, who can e.g. no longer drive a car, or suffer from social stigma. In between seizures ('ictal' periods), the epilepogenic brain region also produces

interictal discharges or spikes, which occur much more frequently than seizures but often go unnoticed, since they do not or barely lead to symptoms. Many patients can be helped by anti-epileptic drugs, which prevent the symptoms, yet do not treat the root cause of the disease. By intervening in the interplay between neurotransmitters and cell receptors in the brain, which are the 'shackles' in the neural communication chain, these drugs temper the initiation or propagation of seizures, but do not resolve the neural defect. Furthermore, an estimated 30% of all patients do not respond to this medication.

In such cases of refractory epilepsy, surgical removal or disconnection of the affected brain tissue can be an adequate alternative treatment. Hence, an accurate delineation of this epileptogenic region is then necessary to ensure that no healthy brain tissue is resected. Nowadays, a large toolbox of medical imaging techniques are at the neurologist's disposal to make a 'map' of the regions that are involved in the generation or propagation of seizures, but unfortunately, no technique in itself is sufficiently accurate to detect the epileptic brain regions. For this reason, doctors preferably measure a patient using several techniques—either simultaneously or in separate recording sessions—whose resulting predictions can be pooled together.

In this thesis, we focus on two of these modalities, namely EEG and fMRI. EEG is a recording of the electrical activity of (large groups of) brain cells by a set of sensor electrodes placed on the patient's scalp. Since seizures and interictal spikes are essentially highly synchronous electrical patterns with a relatively large amplitude, they can be picked up as a time-varying signal by the EEG electrodes. fMRI is a true imaging technique, which can reveal the time-varying consumption of oxygen in minuscule subvolumes of brain cells. The rationale is that brain cells need energy to sustain their electrical activity, which they generate through their metabolism. Hence, when brain regions initiate seizure or spike activity, they are suspected to draw a large oxygen supply from the blood flow, and this physical signal can be tracked via fMRI. Both modalities are also intensively used by neuroscientists, either individually or in conjunction, to conduct research on healthy brain function. However, simultaneous recording of EEG and fMRI is especially attractive for the presurgical evaluation of epilepsy because of three primary reasons:

1. Both modalities are non-invasive, safe, and produce no side effects; they do not require a surgical intervention or the injection of an imaging tracer in the body.

2. Both modalities are sensitive to brain activity from interictal spikes; the patient has no seizure symptoms during recording, and the chance of measuring several spikes is high.

3. They are highly complementary: EEG signals are vital to detect *when* spikes occur, and fMRI images can give a precise answer as to *where* in the brain there are related changes in oxygen concentration.

Out of all possible combinations of imaging techniques which are used to localize epileptic spikes in time and space, simultaneous EEG–fMRI arguably offers the best tradeoff in terms of high temporal resolution (by virtue of the EEG) and spatial resolution (by virtue of the fMRI). The former is absolutely necessary, since spikes occur seemingly at random, and a neurologist does not know a priori at which times the brain enters the state wherein it produces an interictal discharge. The latter is then needed to obtain actionable information for the neurosurgeon, since the goal is to cut away the minimal amount of tissue which still stops seizure generation.

It is nontrivial to integrate the EEG and fMRI measurements to obtain maps of the spatial distribution of brain regions involved in the generation and propagation of spikes and seizures. In the most commonly applied approach, the data are combined asymmetrically: a neurologist or trained EEG reader (or even an automatic algorithm) browses through the EEG signals to find and annotate occurrences of interictal spikes. At each location in the fMRI images, it is then analyzed whether there are concomitant measured fluctuations. When there is a very high concordance between the time courses measured in fMRI and the spike annotations on the EEG, the associated brain regions are deemed to be part of a brain network along which seizures and spikes travel.

This approach is not nearly 100% successful or accurate: some reports estimate that 40–60% EEG–fMRI studies during presurgical evaluation are uninformative because of a variety of reasons, many of which are related to either the low data quality, the inherent physical limitations of the measurement, or the very heterogeneous nature of the pair of modalities (Salek-Haddadi et al., 2006; Grouiller et al., 2011; Gotman, 2008).

The central premise of this thesis is that advanced signal processing techniques can be used to overcome these limitations. Indeed, not only does signal processing offer a wide array of algorithmic tools to improve the quality of digital measurements of physical quantities, it also enables complex modeling of the relationship between data that come from multiple sources, such as EEG and fMRI. The goal of this work is to make contributions in the form of novel signal processing tools, which can be usefully applied to simultaneously recorded EEG and fMRI data from individual patients during the presurgical evaluation of epilepsy. By extension, some of the proposed methods find application in other types EEG–fMRI studies, or signal processing in general. We heavily make use of *tensors*, colloquially referred to as ‘multidimensional arrays’, to preserve and exploit the rich structure which is inherent to neuroimaging data,

and proposed customized blind source separation techniques, to disentangle the complex data into more *interpretable* and *actionable* components.

In the next section, we outline the structure of the thesis, and provide a suggestion on how to read this text.

1.2 Structure of this thesis

Figure 1.1 sketches the different topics which are treated in this text.

Chapter 2 provides elementary background on brain anatomy, and on the disease mechanisms and characteristics of epilepsy. We aim to provide a sufficiently detailed explanation on the why, what, and how of EEG and fMRI, and discuss opportunities and pitfalls that come with the use of these methods. We discuss the merits and shortcoming of EEG-correlated fMRI analysis, the most commonly applied approach to fuse EEG and fMRI data during presurgical evaluation. We postpone the precise statement of the research goals to the end of this chapter, after the problem setting is treated in depth.

Chapter 3 gives an overview of some of the most relevant signal processing techniques and algorithms which have found their way to medical applications, more specifically those that deal with measurements of the healthy or diseased brain. We explain that most of these techniques can be understood in the context of *learning* useful things about the data, with varying degrees of human intervention. I.e., we provide an account of *supervised* and *unsupervised* learning algorithms. The former algorithms exploit various forms of prior knowledge about the data, which is inserted by a human user—these are *trained*, in a certain interpretation. The latter algorithms operate mostly independently, and explore the data without any information which is not conveyed by the raw or preprocessed signals themselves—which gives rise to the adjective 'blind' in *blind source separation*. We spend significant attention to the latter class of techniques, since tensor-based blind source separation will form the backbone of most subsequent chapters.

Chapter 4 proposes a first technique aimed at assisting EEG-correlated fMRI analysis, by semi-automatically improving the quality of interictal EEG signals. Based on prior detection of interictal spikes in the EEG data, our algorithm learns the characteristics of the data in periods with and without spikes. The contrast between these characteristics can then inform a multi-channel Wiener filter, which can be applied to the EEG signals to enhance interictal spikes and spike-like waveforms in the data, and suppress EEG patterns of no interest. The fact that the filter operates in a data-driven, but supervised way, makes

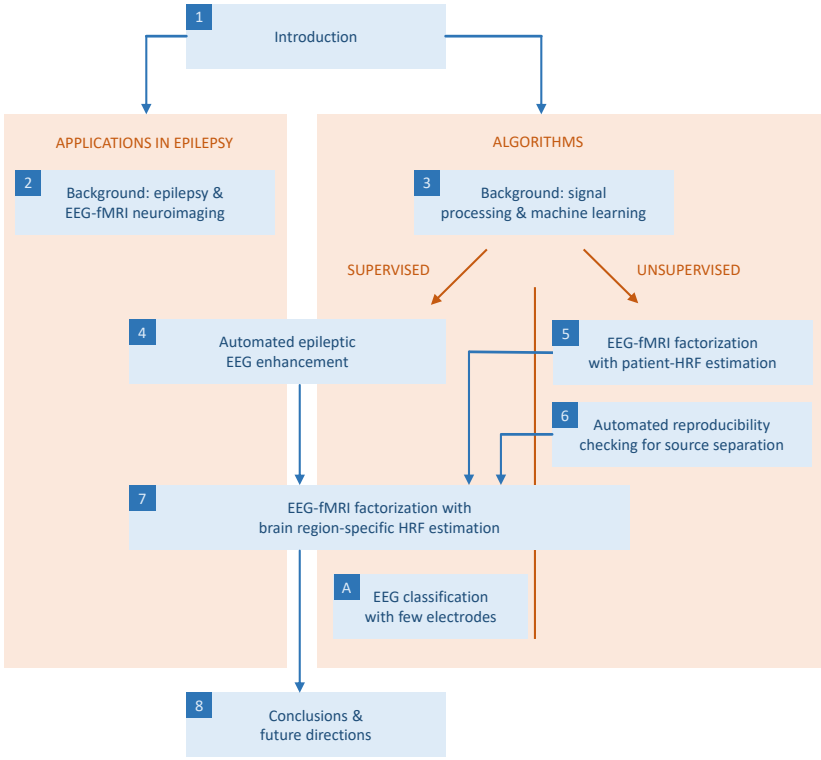


Figure 1.1: This thesis is an interdisciplinary endeavor, and focuses on the development of new (tensor-based) signal processing techniques (right block), as well as their application to neuroimaging in epilepsy (left block). The algorithms can further be broadly categorized as supervised (relying on assistance from a human) or unsupervised (fully independent) techniques which 'learn' interesting aspects from data. The diagram shows the logical structure of the thesis; arrows indicate logical coherence between the chapters, and the overlap of chapters with the orange blocks indicate which type of algorithms and experimental validation are involved. (Abbreviations: EEG = electroencephalography; fMRI = functional magnetic resonance imaging; HRF = hemodynamic response function).

it a powerful tool, which can be adjusted for each patient individually, and increase the chances of a successful EEG-correlated fMRI analysis. We validate this algorithm on real data from twelve epileptic patients.

Chapter 5 introduces the first tensor-based technique for blind source separation of simultaneously recorded EEG and fMRI data. Using a tensor representation for the EEG, and a matrix representation for the fMRI data, we formulate a structured factor model of both data, in which also the relationship or coupling between the time courses in both modalities is a priori unknown. We provide an algorithm which fits the factor model to the data, leading to a set of estimated factor signatures or sources, which can be interpreted as the blocks from which the EEG–fMRI is built, and a patient-specific hemodynamic response function (HRF) as an expression of the neurovascular coupling, which can be interpreted as the bridge between the EEG domain and fMRI domain. The algorithm is subsequently validated on synthetically generated data. The significance of this technique is that it allows the estimation of neural sources *and* unknown neurovascular coupling function simultaneously.

Chapter 6 treats a major hurdle in practical tensor-based (and matrix-based) signal processing, which however mostly remains underexposed: non-convexity of the source separation algorithms. For real-life scenarios, in which the model can never perfectly describe the data, the millstone of non-convexity entails that any such algorithm which estimates sources from the data cannot guarantee to find the optimal solution. Instead, the quality and reliability of the estimated sources' characteristics depend on a good choice for the starting point of the algorithm, which is a hard problem in itself. Here, we propose a technique which can alleviate this burden in a practical setting. Since converting to a convex structure is generally not possible, our approach is to seek reproducible sources among many repetitions of the algorithm, based on a constraint of expected similarity over repetitions.

Chapter 7 builds on the developments in the previous chapters, and introduces a novel structured matrix-tensor factorization which has increased expressive power. We provide an end-to-end data pipeline, which preprocesses, decomposes, and statistically probes the interictal EEG and fMRI data. In the first step, we train data-driven filters as in chapter 4 to enhance epileptic waveforms in the EEG. Subsequently, we apply highly structured blind source separation to a tensor and matrix representation of the EEG and fMRI respectively, in which we account for variations of the HRF over different regions of the brain (in addition to patient-to-patient variations, which are covered by the algorithm in chapter 5). Lastly, we conduct an integrated statistical test to find brain regions which are significantly affected by the occurrence of interictal spikes, either by local increases in fMRI amplitude, or by modulation of the neurovascular coupling. We test this framework on the same patient data as in chapter 4.

Appendix A can be read in isolation from the previous four technical chapters, as it deals with a different application which is relevant in EEG applications, namely the discrimination of brain states using a small set of measurements. We present a technique which enables a user to train a detection algorithm using a 'full' set of measurements (i.e., with many EEG electrodes), and subsequently deploy it using only a reduced set of measurements (i.e., with only one or a few EEG channels). This is potentially attractive for wearable EEG applications, which are constrained to compact, mobile devices in which only few sensors can be accommodated. One example is the use of ambulatory devices for long term monitoring and seizure detection out of the hospital.

Chapter 8 summarizes the thesis's contributions and findings, and casts an eye to potential future developments.

1.3 Collaboration and funding

The work leading to my PhD thesis was done at KU Leuven, Department of Electrical Engineering (ESAT), STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, under the supervision of prof. Sabine Van Huffel, and post-doctoral researcher and later professor Borbála Hunyadi. I profited from the co-supervision of prof. Lieven De Lathauwer, and from informal contacts and collaboration with my former master thesis supervisor, prof. Alexander Bertrand. The new engineering approaches developed in this thesis have been experimentally validated with real data from epileptic patients, and for this reason the collaboration with prof. Wim Van Paesschen and prof. Patrick Dupont from UZ Leuven was of paramount importance. Furthermore, I had cooperation from Nico Vervliet and Martijn Boussé for some of the chapters in this text, as well as informal advice, guidance, and support from many other colleagues (friends) in the BIOMED research group. The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC Advanced Grant: BIOTENSORS (n° 339804).

Chapter 2

EEG and fMRI imaging in epilepsy

This chapter is a primer on two neuroimaging modalities which are relevant for epilepsy monitoring, and which are explored in this thesis: electroencephalography (EEG) and functional magnetic resonance (fMRI). We start by introducing some elementary concepts from brain anatomy, and proceed with an explanation on the physiological and physical foundations of both EEG and fMRI, and discuss practical aspects of their recording. EEG rapidly measures voltages on the scalp, and each EEG sensor can be thought of as a 'microphone' which 'listens' to the electrical activity from brain cells. On the other hand, fMRI slowly scans changes in oxygen concentration of cells throughout the brain, and can be thought of as a 'camera' with a slow shutter speed. Next, we provide some preliminaries on the origin and symptoms of (drug-resistant) epilepsy, and discuss which imaging modalities can play a role in the diagnosis and treatment of this brain disorder. EEG and fMRI provide a non-invasive and complementary view on brain (dys)function, which is the rationale behind combined recording of these two modalities. In the last section, we discuss the opportunities and challenges that arise by recording simultaneous EEG–fMRI for drug-resistant epilepsy.

2.1 Brain anatomy

The brain is the most complex organ in the human body, and deserves a description on many different scales in order to fully appreciate its modus operandi. The elementary building blocks of the brain are neurons or brain cells. The prototypical neuron structure is shown in Figure 2.1. A neuron consists of a body (soma) and several dendrites, which provide the wiring to connect to other neurons.

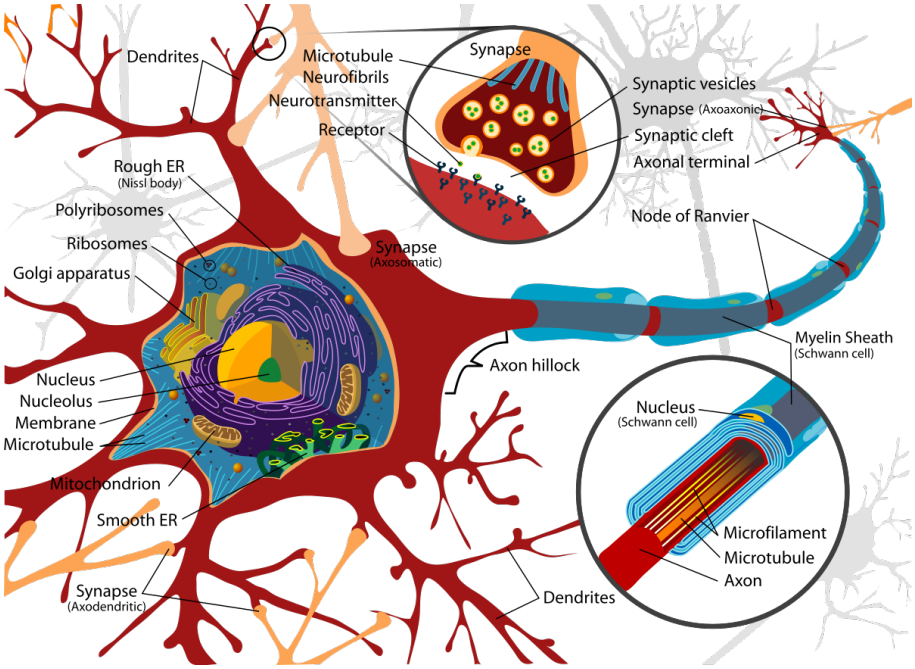


Figure 2.1: A neuron consists of a body (soma), from which several dendrites branch out to receive incoming electrochemical signals from neighbouring neurons at several connecting junctions (synapses). There, neurotransmitter molecules bind to receptor proteins, triggering the opening of ion channels across the cell membrane and hence changing the electrical potential. When the integrated signals exceed a threshold, an action potential is released and forwarded along the axon, towards a synapse with another neuron. Figure taken from (Wikipedia, 2007).

Neurons communicate by propagating signals in the form of electrical action potentials along the axon, a long strand exiting the cell body. The axon is covered with myelin sheaths, consisting of fatty tissue which acts as an insulator to the conducting dendrite. The junction where the axon meets the body of a neighbouring neuron is called a synapse. At the synapse, the arriving action potential initiates an electrochemical reaction by stimulating the release of neurotransmitters, chemical substances which can bind with receptor proteins on the membrane on the other side of the synapse. The binding of neurotransmitters to receptors causes ion channels to open, thereby altering the postsynaptic membrane potential through the flux of ions. Depending on the ions that are exchanged between the inner cell and the extracellular

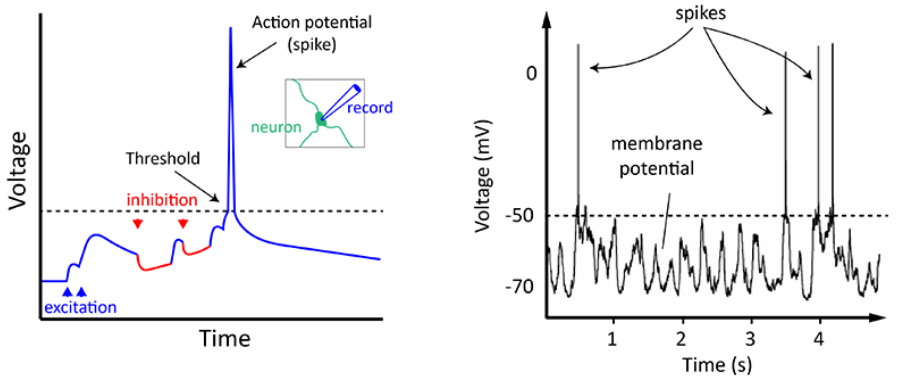


Figure 2.2: If the integrated potentials from many synapses make the postsynaptic cell membrane potential exceed a threshold, an action potential is fired in the postsynaptic neuron. Excitatory potentials tend to depolarize the cell membrane (i.e., bring the membrane potential closer to the threshold), while inhibitory potentials have a hyperpolarizing effect (i.e., reducing the likelihood of firing). Figure taken from (Queensland Brain Institute, 2019).

environment, such events may steer the membrane potential away from its equilibrium value of circa -70 mV, by making it more or less negative. Only when the membrane potential depolarizes sufficiently and reaches a critical value, an action potential is fired, which is sent along the axon towards a next neuron, as displayed in Figure 2.2. Hence, neuronal communication is a highly nonlinear phenomenon, in which the postsynpaptic neuron fires not proportionally to its 'input' membrane potential, but only when a spiking threshold is reached.

The brain is covered with neurons, which are organized on the mesoscopic level in cortical columns, i.e. near the skull, large assemblies of neurons lie perpendicular to the brain's surface, with the dendrites extending inwards. Macroscopically, this leads to a distinction between white matter, which is mostly found in the center of the brain and is composed of myelin, and grey matter (aka cortex), which is a thin layer around the white matter and is the surface of the brain. After birth, the cortex gradually evolves from a flat sheet into a highly folded surface, with gyri (bumps) and sulci (troughs). At the highest level, the brain is organized in several lobes (temporal, occipital, parietal, occipital lobe), as visualized in Figure 2.3.

Figure 2.5b shows a real image of the brain's anatomy. The differences in intensity reveal the distribution of white matter in the center, grey matter mostly in the cortex near the skull, and cerebrospinal fluid (small, dark regions in the center). Cortical folding is also clearly visible.

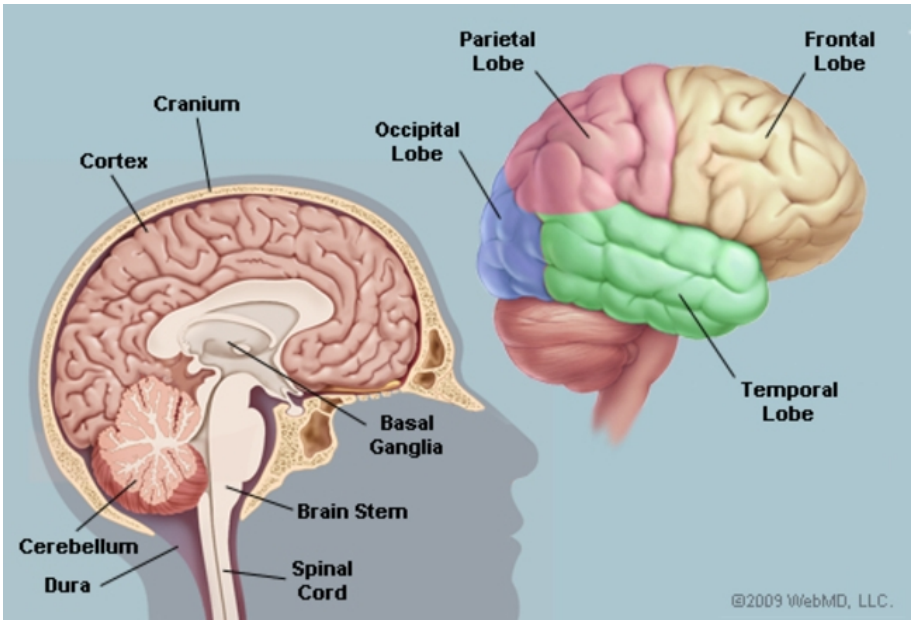


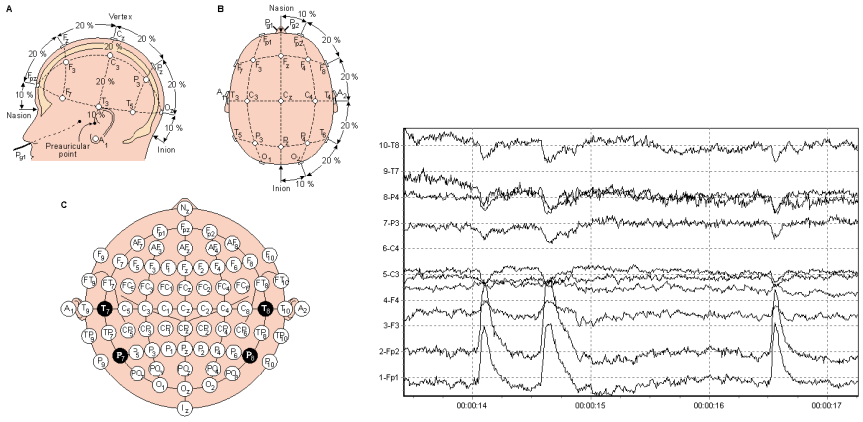
Figure 2.3: On the macroscopic level, the brain consists of several lobes. Figure taken from (WebMD, 2014).

2.2 Electroencephalography (EEG)

Electroencephalography (EEG) is a non-invasive modality to monitor brain activity, in which electrodes are placed on the scalp. It has been one of the earliest modalities to monitor brain function, since its conception in the 1920's by the German doctor Hans Berger. An example EEG setup is shown in Figure 2.4a. The electrodes are sensitive to small electrical fields arising from electrophysiological activity in the neurons, but also pick up non-physiological waveforms, i.e. interference or artifacts, which originate from numerous sources, and which are discussed in more detail below. We discuss the clinical application of EEG in the context of epilepsy monitoring, in sections 2.4.2 and 2.5.

2.2.1 Which physiological variable does EEG measure?

The physiological quantity to which the EEG is sensitive is in fact the superimposed electrical field which arises due to all extracellular ion currents that underlie neuronal signaling (also called local field potentials (LFP)), as described



(a) EEG uses electrodes placed evenly on the scalp.

(b) EEG signals capture time-varying electrical voltages.

Figure 2.4: Electroenceophalography (EEG) is the measurement of electrical fields at the scalp. **(a)** Electrodes are spread over the head, and are given a position-dependent name (e.g., F = frontal, P = parietal, T = temporal, C = central, O = occipital lobe, even/uneven number for left/right). **(b)** EEG signals are able to trace brain activity with a high temporal resolution, but are prone to unwanted artifacts that reduce signal quality. This segment contains three large artifacts due to eye blinking. Figure taken from (National Center for Adaptive Neurotechnologies, 2019).

in the previous section. Thus, the EEG does not reflect the pure activity of neurons that fire action potentials, since only synchronous electrical potentials across many pyramidal neurons produce a measurable signal (Kirschstein and Köhling, 2009). This 'constructive interference' is more easily achieved for slower electrical events like postsynaptic potentials and synaptic currents (also those that do not surpass the spiking threshold), while the very sharp action potentials in different neurons are unlikely to precisely overlap (Buzsáki et al., 2012; Cohen, 2017; Schnitzler and Gross, 2005).

A recurring finding is that the spectral content of the time-varying electrical fluctuations, as measured by EEG, correlates with macroscopic behavior and state of the subject, such as the sleep state, vigilance, attention, consciousness, motor function, ... Typically, several frequency bands of interest are considered: delta (±1–4 Hz), theta (±4–8 Hz), alpha (±8–12 Hz), beta (±12–25 Hz), gamma (±25–150 Hz), each of which have been associated with one or more of such behavioral functions. E.g., one of the most robust effects, which was already discovered in the early days of EEG measurement, is the appearance of alpha

waves when the subject under study closes his or her eyes. More generally, any macroscopic effect requires complex processing in the brain, for which coordination over many different neurons is needed, and oscillatory electrical fluctuations offer a robust means for such communication (Schnitzler and Gross, 2005). Hence, the frequency content of EEG signals is frequently used to summarize the time-varying 'state' of the brain. An application is e.g. the determination of deep and light sleep phases, and awake periods, based on a patient's EEG.

2.2.2 Physical aspects of EEG

The EEG is measured as voltage differences between electrodes that are placed according to the 10–20 system, referring to the angular spacing of 10° or 20° between the sensors (see Figure 2.4a). The sampling rate is in the order of several kHz.

While this allows the EEG to track rapid electrical dynamics on the order of milliseconds, it is a poor spatial indicator of brain function, since the electrodes are placed on the scalp, shielded from the cortex by the skull and other tissue. A main physical limitation is thus that the electrical fields are strongly attenuated by the skull, and as such EEG is only sensitive to neural events near the cortical surface, and the deep neural generators are unlikely to be picked up. As such, the EEG can be thought of as a tomographic snapshot of brain activity, which is however heavily undersampled: the complex cortical neural processes, which require the coordination of brain regions on the millimeter level, are monitored using only 20, 64, 128, or maximally 256 sensors (even setting aside for a moment the increased technical difficulty posed by a high electrode density). Besides by the number of sensors, the spatial resolution of EEG is also limited by volume conduction, i.e. the propagation of electrical fields from localized neuronal activity to extended areas. This conduction phenomenon causes spatial blurring: activity from a well-defined neuronal source is often picked up on several nearby electrodes (Nunez et al., 2006; Hallez et al., 2007). Moreover, interpretation of spatial locus from EEG is further confounded by the fact that the orientation of the electrical dipoles can influence the amplitude of the perceived waveform on nearby electrodes. While the pyramidal neurons lie indeed mostly perpendicular to the cortical surface (leading to an intuitive projection of dipole activity to the scalp electrodes), this is an issue near sulci on the cortex (Van de Steen et al., 2019).

In summary, interpretation—let alone quantification—of spatial features from EEG data alone is tricky and is mostly useful to narrow down the spatial locus of a sought source to one of both hemispheres or potentially one brain lobe.

EEG source imaging refers to a collection of techniques that aim to circumvent this limitation by using a high number of electrodes, and an accurate model of the distribution of the different tissues in the head, and how these tissues affect the propagation of electromagnetic waves (Michel et al., 2004). The idea is to solve an (extremely ill-posed) inverse problem to project measured data from high-density EEG (128–256 electrodes) back to the source space, i.e., to voxelwise estimates of neuronal activation in the cortex (Pascual-Marqui, 1999; Michel et al., 2004).

2.2.3 Data quality

EEG measurements are easily contaminated by artifacts from physiological or external sources. Any nearby source of electrical fields is a potential enemy of the EEG acquisition system, since these fields will superimpose on the 'true' neuronal activity at the sensors. The most easy to recognize artifact is due to eye blinking. The eyeball contains a small electrostatic charge asymmetry, which turns it into an electrical dipole. Blinking accompanies a vertical motion of the eyeballs, which changes the dipoles' orientation with respect to nearby electrodes on the frontal cortex, producing a large waveform which lasts approximately 0.5–1 s (see Figure 2.4b). The same argument holds for horizontal eye movements (saccades), which induce a block-like artifact, whose polarity depends on the direction of motion. Powerline interference from the electrical grid induces sinusoidal artifacts at 50 or 60 Hz, but can easily be removed (often this is already done at the hardware level in commercial EEG equipment). Muscle contraction can lead to high-frequency electromyographic (EMG) artifacts. Various muscles in the forehead (frowning), jaw (chewing, teeth clenching, yawning), neck (head motion, swallowing) can be the culprit, but it is obviously difficult for subjects or patients to completely refrain from using them. The heart is another major muscle which might produce electrical fields at the times of the QRS complexes in the heartbeats that are sufficiently strong to be picked up by the EEG electrodes. Sweating, or drying of the electrolyt gel for wet electrodes, can alter the electrical impedance between the scalp and the sensors, causing a gradual change in the observed EEG signal properties. Perhaps the most complicating artifact of all is due to motion of the electrodes relative to the scalp, which leads to large deviations on the EEG which can completely deteriorate the signal quality.

Despite the challenging data interpretation, and many limitations, EEG is still widely used, because it is well understood, standardized, inexpensive, and relatively easy to set up. Furthermore, a large toolkit of signal processing methods is available to restore or improve EEG signal quality. In the next

chapter, we treat several of such techniques, along with algorithms to perform other useful operations on EEG data.

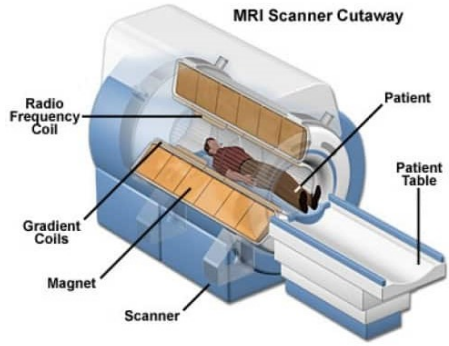
2.3 Functional Magnetic Resonance Imaging (fMRI)

This section is largely based on excellent treatments in (Suetens, 2017; Buxton, 2013; Elster, 2019).

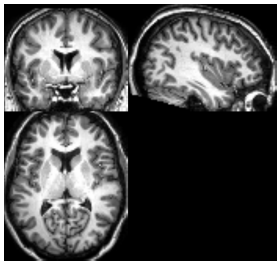
Functional magnetic resonance imaging (fMRI) is a completely non-invasive imaging modality, which is sensitive to local changes in blood oxygen concentrations in the brain. Such changes, coined the *blood oxygen level dependent (BOLD) signal*, occur in response to electrophysiological activity in (large groups of) nearby neurons. This activation, in turn, may be elicited by a visual, auditory or other experimental stimulus, or alternatively may be the result of spontaneous electrical 'events'. It is mostly the latter BOLD generation mechanism which is of interest in the context of epilepsy, since interictal epileptic discharges (IEDs) occur at unpredictable times. We elaborate on the clinical application of fMRI in the context of epilepsy monitoring, in sections 2.4.2 and 2.5.

2.3.1 Which physiological variable does fMRI measure?

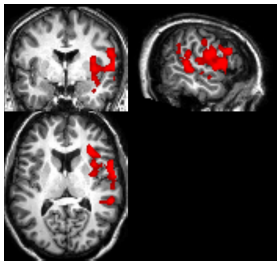
The physical origin of the BOLD signal is best understood as part of neural-metabolic regulation mechanisms involving cerebral blood flow (CBF) and oxygen concentration in the blood. Neurons consume energy when they fire, for which they need oxygen molecules which are delivered through hemoglobin (Hb) molecules in the blood. In response to neuronal firing, two opposing effects take place. The first is a net oxygen transfer from the capillaries to the neurons, to replenish the oxygen which was depleted when generating or propagating the electrical potentials—this tends to decrease the concentration of oxgenated hemoglobin (HbO_2) and increase the concentration of deoxygenated hemoglobin (HbR, i.e. 'reduced' hemoglobin). The second effect is a strong increase in CBF to the activated patch of neurons (mostly in the grey matter), to supply new oxygen for future neuronal firing—this tends to increase (decrease) the concentration of HbO_2 (HbR). The second effect offsets the first effect by a substantial margin, leading to a temporary overshoot in the local concentration of oxygen. The mechanism underlying the adaptive CBF is not yet fully understood, but it is currently suspected that mostly feedforward effects (caused



(a) Layout of a typical magnetic resonance imaging (MRI) scanner.



(b) Anatomical MRI shows white and grey matter.



(c) Functional MRI detects brain activation (in red).

Figure 2.5: A magnetic resonance (MR) scanner produces 3D images of the human body or brain, by interrogating a quantum-mechanical property known as 'spin' of protons at each position in the volume. **(a)** Most MR scanners are cylindrically shaped because of their large, superconducting coils which generate a very strong magnetic field, which serves to exploit the physical principle of magnetic resonance of atomic nuclei. Gradient coils superimpose small magnetic fields, which are different at each position—this attaches 'labels' to all positions, to tell them apart when making the image. The strong magnetic field initially aligns the spin of all nuclei, until a third coil sends a brief radio frequency (RF) pulse into the volume, which temporarily energizes the spins of resonant nuclei. Before the spins return to baseline, they produce a measurable electromagnetic signal, which is used to make an image. **(b)** A structural or anatomical MR image can highlight different *tissues* in the brain, as shown in cross-sections along three orthogonal axes. In this acquisition mode, the scanner is sensitive to the density of hydrogen nuclei (protons) and the duration of their resonance, which is different in grey and white matter tissue. **(c)** In functional MRI, the location of brain *activity* can be detected. In this acquisition mode, the resonance of protons is altered near regions where a lot of oxygen is consumed due to normal or abnormal (e.g. epileptic) activity in the neurons. Figure **(a)** taken from (Amber Diagnostics, 2016).

directly by the neurons' electrical activity and the release of vasoactive ions and neurotransmitters across their membranes), and not feedback effects (set in motion by a lowering oxygen concentration) play a role (Attwell and Iadecola, 2002; Valdes-Sosa et al., 2009). This process can be referred to using the umbrella term *neurovascular coupling*, which will be further explored at a macroscopic level in section 2.5.2.

Key to fMRI is that the two types of hemoglobin molecules have different magnetic properties, i.e., HbR is paramagnetic[1], and as such alters the blood's magnetic susceptibility. A magnetic field, like the one generated in an MR scanner, will be locally distorted by the presence of the HbR molecules during baseline, and when the HbR concentration lowers during the phase of increased CBF, this leads to a measurable signal, i.e. the BOLD effect (Ogawa et al., 1990b,a).

From the previous discussion it is clear that the BOLD signal is not a pure reflection of one underlying physiological parameter, such as CBF or oxygen metabolism. Rather, it is a net imbalance between several complex interacting mechanisms, and as such it is a (very) indirect reflection of neural activity.

2.3.2 Physical foundations of magnetic resonance imaging

BOLD-fMRI is one of several acquisition schemes which can be performed with a magnetic resonance (MR) scanner, aside e.g. structural MRI (to obtain high-resolution anatomical images), diffusion MRI (to map white matter tracts), arterial spin labeling (to measure localized CBF changes), and others. All of these modalities exploit the same quantum mechanical principle, namely nuclear magnetic resonance of (hydrogen) atoms. This effect entails that charged particles behave like a minuscule magnetic dipoles, endowed with a directional angular momentum or 'spin'[2], which can be manipulated only by specific stimuli when they are brought into an external magnetic field. By carefully modifying this external field, a subsequent perturbation of the spins can reveal information about the local tissue's properties.

[1] Paramagnetic materials align their own magnetic dipoles—which result from their atomic nuclei—with an externally applied magnetic field. The opposite effect is diagmagnetism, in which a material's dipoles 'oppose' (contra-align with) the external field.

[2] One can conceive of 'spin' as a rotating charged particle, which produces a tiny magnetic field just like an electric current would. Even though this is a very rough simplification of the quantum mechanical subtleties, it provides a useful mental model.

The magnetic resonance effect

In the MR scanner, a large field $\vec{B_0} = (0, 0, B_{z,0})$ on the order[3] of 1 Tesla (T) (on older scanners), up to 7 T (on some modern scanners) is generated, usually by means of bulky coils which are superconducting through extreme cooling, and which give the MR scanner its characteristic look (see Figure 2.5). In any form of medical MR imaging, the charged particles of interest are the nuclei of hydrogen atoms (i.e. protons), because of the abundance of hydrogen in virtually all tissues of the human body. While in the absence of a strong field the protons have a random magnetic spin orientation due to thermal noise, the dipoles behave like tiny magnets if an external field is applied, and hence they show the tendency to align (on average, and only approximately) with this field. This results in a net equilibrium magnetization $\vec{M_0}$ parallel to $\vec{B_0}$ in every macroscopic volume—such as an imaging voxel. A useful signal can then be measured when this equilibrium is temporarily disturbed. Namely, when a radio-frequency (RF) electromagnetic pulse (at a specific frequency) is sent into the magnetized tissue, $\vec{M_0}$ experiences a torque and is steadily tilted away from its original aligned axis. It will then precess around this axis with an angular frequency $\omega_0 = \gamma B_0$ which is proportional to the external field strength by the gyromagnetic ratio γ, which is atomic nucleus-specific[4].

To relax is to measure

At this point, the magnetization can be described by the vector sum of a still longitudinal component along $\vec{B_0}$ and a rotating transverse component in a plane orthogonal to this axis. The rotating magnetization component is able to induce an electrical signal in a set of nearby coils, which yields the MR measurement. In the subsequent relaxation phase, this transverse component will decay, with a time constant T_2. This happens because the magnetization's comprising dipoles experience slight differences in the external field due to their random and rapidly varying molecular surroundings, causing them to steadily precess out of sync. This leads to the vector sum of transverse magnetization converging to zero due to loss of phase coherence[5]. The time constant of this dephasing

[3]For reference, 1 T is about 20,000 times stronger than the earth's magnetic field, and 200 times stronger than a fridge magnet.

[4]Interestingly, quantum mechanics dictates that the electromagnetic pulse should be matched in frequency to ω_0, hence the name *resonance*. I.e., only photons of frequency $\frac{\omega_0}{2\pi}$—known as the Larmor frequency—have the right energy to make individual nuclei transition between two discrete 'states': the low-energy state (aligned with the external field) and the high-energy state (opposing the external field).

[5]Also the longitudinal magnetization component will tend back to its equilibrium value, aligned with the external field. This occurs because the spins in the unstable, high-energy

phenomenon is further reduced to T_2^* in case of (more static) magnetic field inhomogeneities, which precisely lays the basis for fMRI: as described earlier, the varying concentrations of HbR and HbO_2 locally disturb the magnetic field; this affects the T_2^* relaxation time and hence the measurable signal at different points in time.

Encoding the position of resonant atomic nuclei

This process becomes only useful when the short-lived transverse magnetization and induced electrical signal reveal something about their spatial origin. For this reason, it is necessary to modulate the external magnetic field such that it varies over space, since this will lead to induced signals which are frequency-modulated, allowing to reconstruct where the induced signal came from. This works as follows: in all three directions, a time-varying gradient field is applied:

$$B_z(x, y, z, t) = B_{z,0} + \Delta B_z(x, y, z, t) \tag{2.1}$$

$$= B_{z,0} + G_x(t)x + G_y(t)y + G_z(t)z \ , \tag{2.2}$$

where $B_{z,0}$ is the baseline magnetic field, and G_x, G_y, G_z are spatial gradients with units of Tm^{-1}, which can independently be switched on and off using gradient coils[6]. Imaging proceeds by repeatedly acquiring 2D image slices, at subsequent z positions. Firstly, the scanner briefly applies a gradient G_z, as to change the Larmor frequency throughout the volume to $\omega(x, y, z) = \gamma B_z(x, y, z) = \omega_0 + \gamma G_z z$. At the same time, the RF pulse is applied, and only spins in the plane with a Larmor frequency that is resonant with the pulse are tipped into the transverse plane. Hence, the RF pulse performs 'slice selection', and we denote the slice or plane with resonant spins as $z = z_{RF}$. For ease of exposition, we now adopt a moving reference frame in every voxel, which rotates at the baseline angular frequency ω_0. In the absence of any field distortion or gradient, the transverse magnetization in the excited slice would remain still in such a reference frame. Then, when applying x- and y-gradients, the local Larmor frequency changes to $\omega(x, y, z_{RF}) = \gamma B_z(x, y, z_{RF}) = \omega_0 + \gamma(G_x x + G_y y)$, leading to a transverse magnetization rotating at $\gamma(G_x x + G_y y)$ in the new reference frame, or in mathematical terms, a local multiplication of the transverse magnetization with a complex exponential whose frequency scales with x and y. The signal that is sensed in the coils is thus the summation of the local magnetization,

state can lose energy to their surroundings and transition back to the baseline, low-energy state, until the majority of spins occupies this state. Such 'spin-lattice relaxation' progresses with a time constant T_1, which is longer than T_2 of 'spin-spin relaxation'.

[6]Note that the baseline and resulting fields are always oriented in the z-direction, i.e., the axis which passes through the scanner's bore.

multiplied with a complex exponential, over all voxels in the involved image slice. Beautifully, this process coincides with the 2D spatial Fourier transform of the spatial distribution $\rho(x, y, z_{RF})$ of the net magnetization, which depends on the local proton density and the amount of relaxation that has occurred[7]. In fact, the two coils in quadrature position jointly measure a single complex-valued time series, describing a trajectory through 2D Fourier space in which one Fourier coefficient is sampled at a time. Hence, after sufficient samples are collected, image reconstruction boils down to the inverse 2D Fourier transform of the electrical signals induced in the coils; this is known as the $\vec{k}$-theorem, where $\vec{k} = (k_x, k_y)$ refers to the spatial frequencies of the basis functions in the transform. This process can then be repeated per image slice, each time tailoring the pulse's center frequency so that the Larmor frequency of the new slice falls within the RF bandwidth.

Constraints on the acquisition

Crucially, a sufficient portion of the k-space, i.e. the spatial frequency domain, must be sampled, point per point, to ensure a high-resolution reconstruction in the xy-plane (cfr. the Nyquist theorem). The trajectory through this space is determined by alternately switching on the gradients in the x- and y-directions for short periods of time in a pulse sequence, very much like navigating a cursor on a computer screen by tapping the arrow keys. Ultimately, the imaging resolution (i.e., the voxel size), is then limited by the highest spatial frequencies one can afford to sample under the timing constraints from acquisition and relaxation, and the specificity of slice selection. I.e., in the xy-plane, this depends on the maximal (safe) strength of G_x and G_y ($\sim$50 mTm^{-1}), and the available time window before the magnetization decays away due to T_2^* relaxation and before the next slice needs to be sampled. For the z-direction, slice selection is bounded by the strength of G_z and the narrowness of the RF bandwidth. Typical voxel dimensions are e.g. $2 \times 2 \times 2$ mm^3. In fMRI, multiple 3D images are consecutively recorded, at a rate which is limited by the time it takes to acquire all slices of a whole-brain image. Several constraints in fact do not stem from hardware or time restrictions, but are due to patient safety considerations. For example, the gradients should not switch extremely fast, since such a sudden change in magnetic field strength could induce bio-electrical currents that exceed the nerve stimulation threshold and heat surrounding tissue, or large eddy currents in EEG electrodes that may cause skin burn. The RF power and pulse rate must not be too high, as to prevent tissue heating due to thermal dissipation.

[7] Recall that in fMRI, the T_2^* relaxation time is modulated by the difference in concentration of HbO_2 and HbR.

Parameters for imaging

Fairly rapid k-space sampling is achieved by echo planar imaging (EPI) pulse sequences. Still, the repetition time (TR) between two full-brain images is typically around 2 seconds—which is deemed acceptable, since the relevant hemodynamic fluctuations are sluggish. Using various improvements, TR can be shortened to several hundreds of milliseconds in modern scanners. Besides hardware improvements, developments in MRI have not seldom come from the innovative design of 'smart' pulse sequences, e.g. a spiral trajectory through k-space has been found to offer a range of benefits over simpler trajectories, and methods based on compressed sensing exploit the typically sparse nature of MRI images in k-space to speed up the sampling process without loss of image quality (Lustig et al., 2008). Another innovation comes from parallel imaging, in which multiple measurement coils with different spatial sensitivity profiles can alleviate the Nyquist sampling requirements for the k-space, and from multi-band imaging, in which multiple RF pulses are used to image different slices of the brain simultaneously.

The reconstructed physical quantity $\rho(x, y, z)$ reflects a *weighted* spin density. Namely, the net magnetization not only depends on the number of excited hydrogen nuclei at each position, but also on the degree of T_1 and $T_2^{(*)}$ relaxation that has occurred before the readout time, and on TR and the 'flip angle' induced by the RF pulse. While the first parameters are tissue-dependent and cannot be controlled, the second set of parameters can be tuned by the operator of the MR scanner, to suppress or enhance the effect of a specific physiological parameter (cfr. 'T_1-weighted', 'T_2-weighted', 'density-weighted' images). For anatomical MRI, mostly the fact that T_1, T_2, and the hydrogen density are tissue-dependent is exploited to create images in which either grey matter, white matter, cerebrospinal fluid, tumors, ... are brightest. In fMRI, the time-varying difference in HbO_2 and HbR concentration is a proxy for neural activity and alters T_2^*, which constitutes the main image contrast. There is a confounding effect of activity-independent hydrogen density on the measured fMRI signal, but to a first-order approximation, all voxels in the grey matter have similar density, and it is anyway not so much the absolute BOLD amplitude but the dynamic relative change of the amplitude which is of interest (see section 2.5).

2.3.3 Data quality

Besides picking up true neuronal BOLD signals, the fMRI data are also very susceptible to contamination by artifacts. As explained in the previous section, high-resolution image acquisition relies on a tight control of the magnetic fields

in the scanner, and on the uniformity of the static field. Hence, inhomogeneities in $B_{z,0}$ should be accounted for. Motion of the subject inside the scanner leads to image blurring, since any moving tissue element traverses regions with differing magnetic fields, which leads to an averaging of their spins' precession frequency. Furthermore, in voxels near edges or boundaries between tissues, motion can result in large time-varying fluctuations, since different tissue elements (with potentially very different MR intensities) 'jump' in and out of the voxel's bounds. Hence, fixating the subject's head can minimize the influence of such artifacts, although it can be challenging to eliminate all motion. Physiological influences originate mostly from cardiac activity and respiration (Abreu et al., 2018b). The pulsation of the heart causes a periodic modulation of the CBF, and can cause slight movement of brain tissue or even a small recoil of the head (Marino et al., 2018). Respiration causes the oxygen concentration in the arteries to fluctuate, which confounds the BOLD signal, and also leads to head motion. Scanner drift can be caused by heating of the gradient coils, among other factors, and lead to slow variations in the recorded signals (Vos et al., 2017).

The raw data undergo a long chain of preprocessing steps, to correct for various imaging and physiological artifacts before the data can be appropriately interpreted as successive 3D snapshots of a volume. Firstly, slice timing correction is applied, to correct for the fact that different image slices in the volume were not acquired at precisely the same time (Sladky et al., 2011). Subsequently, realignment of the individual 3D images is necessary to correct for head motion in between and during the repetition time. Coregistration transforms the fMRI data into the same coordinate system as another image, e.g. an anatomical MR image of the subject (Friston et al., 1995). Spatial smoothing is performed to reduce sharp intensity transitions between neighbouring voxels, which are unlikely to have a neuronal origin—at least in the bulk of the brain. Often, fMRI images are also normalized, i.e., nonlinearly warped into a reference anatomical space with associated coordinate system, in order to compare images across different subjects. Correction of the image time series for the physiological artifacts mentioned earlier is also necessary. Since meaningful BOLD signals mainly stem from the grey matter, the regions containing white matter and cerebrospinal fluid can be used as reference measurements of some of these artifacts (Behzadi et al., 2007). E.g., as a fluid, the CSF is particularly sensitive to motion, and thus fluctuations elsewhere in the brain which are correlated to the time courses measured in the CSF can be removed. A similar principle is used for the correction of cardiac artifacts. In the next chapter, we treat several signal processing approaches, which can aid the interpretation of fMRI data.

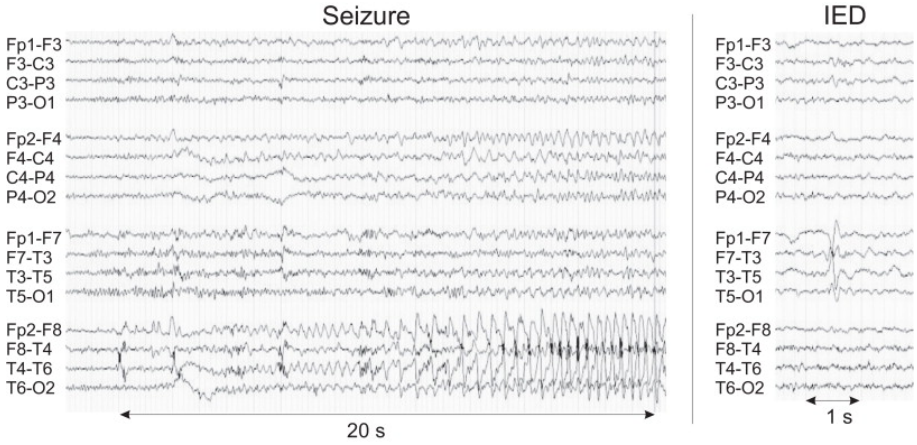


Figure 2.6: Seizures and interictal epileptic discharges (IEDs) or spikes are both types of epileptiform neural activity, which can be measured with electroencephalography (EEG). The left panel shows the development of a focal seizure into a typical oscillatory pattern, which is mainly present over the right temporal lobe, measured by the bottom set of bipolar EEG channels. The right panel shows a brief EEG segment from another patient, containing an IED with a characteristic (sharp) spike and (slow) wave complex in the left temporal lobe. Figure taken from (Van Mierlo et al., 2014; Javidan, 2012).

2.4 Epilepsy as a brain disorder

2.4.1 Manifestation and prevalence

Epilepsy is a neurological disorder which affects approximately 50 million people worldwide (World Health Organization, 2019), and which is caused by hypersynchronous firing of neurons, which overtakes normal brain activity. Patients suffering from epilepsy are struck by recurrent seizures or *ictal* periods, during which they can develop a wide range of symptoms, ranging from muscle spasms or stiffening (tonic–clonic seizures), losing consciousness (absence seizures), or having 'auras', i.e., altered perceptions of sound, light, ... Seizures are clearly distinguishable on EEG recordings by their oscillatory character and large amplitude. In between seizures, interictal epileptic discharges (IEDs) or spikes often occur, but with negligible symptoms. An example of a seizure and an IED are shown in Figure 2.6. While most patients can be helped by anti-epileptic medication that suppresses seizures, a significant part of them (around 30%) suffers from *refractory* (drug-resistant) epilepsy and hence does not

respond to such drugs and needs to be helped using different forms of treatment. One potential mediating solution for some patients is deep brain stimulation via an implanted electronic device, which can alleviate the severity of seizures (Boon et al., 2007). A more permanent treatment is a surgical intervention, in which the brain tissue which is thought to be responsible for seizure generation is removed or disconnected from the surrounding tissue (Rosenow and Lüders, 2001). The preferred treatment depends among others on the locus of this *epileptogenic zone* (defined as the "brain tissue whose removal/resection renders a patient seizure-free"): if this zone coincides or overlaps parts of the eloquent cortex, i.e., brain regions indispensable for e.g. speech, neurologists will decide against surgical removal.

While seizures and interictal spikes initiate at a particular site, they propagate along neural pathways to an extended area (focal epilepsy) or even the whole brain (generalized epilepsy). Epilepsy has therefore been recognized as a network disorder, along with several other brain diseases such as Alzheimer's disease and multiple sclerosis (Stam, 2014; Bullmore and Sporns, 2009; Ponten et al., 2007; Kramer et al., 2008). This makes a full understanding of the disease more complicated, and analysis methods that treat the brain like a network, consisting of hubs, nodes, and links that model their 'connectivity', are increasingly being used (Rubinov and Sporns, 2010; Centeno and Carmichael, 2014; Van Mierlo et al., 2014). For example, based on the observation of reduced connectivity between nodes, it has been postulated that the brain employs inhibition mechanisms to partially protect itself against the propagation of (inter)ictal events from their respective onset zone to surrounding areas (Pedersen et al., 2015; Van Paesschen et al., 2007). Using EEG source imaging techniques in conjunction with functional connectivity metrics and graph theoretical analysis revealed that the seizure onset zone could be found as the node in the network that was the largest hub, with most outgoing links (Staljanssens et al., 2017a,b).

2.4.2 Diagnosis and presurgical evaluation

Before surgery of a refractory epilepsy patient, a neurologist needs to delineate the brain region that is to be resected, based on clinical neuroimaging evidence. There is no imaging modality that can localize the epileptogenic zone directly, as it is a mere abstract concept (Rosenow and Lüders, 2001). Therefore, clinicians are bound to work with surrogates of this region.

The most closely related definition is that of the seizure onset zone, or *ictal onset zone*, which is the region where seizures are actually generated. EEG is routinely recorded from epilepsy patients for several days, and concurrently with video monitoring, to observe seizures in the patients and to diagnose several

characteristics of the pathology. E.g., by analyzing the patient's symptoms jointly with the simultaneously recorded EEG signals, a neurologist can better understand the initiation and development of a seizure over time, and may formulate a hypothesis on the type of epilepsy. Furthermore, (s)he can attempt to coarsely localize the seizure onset zone, by inspecting the topography of the EEG signals recorded during ictal periods, limited by the spatial blurring of EEG and the density of electrode placement. This procedure is not always successful, however: 'EEG-negative' cases are patients for which no ictal or interictal waveforms are observable in the EEG signals, often because of their deeper origin in the brain (cfr. section 2.2.2). When the epileptogenic onset zone can be narrowed down to a specific part of the brain, intracranial EEG electrodes are sometimes inserted, to measure the epileptiform activity directly at or in the cortex, since they offer a much better spatial resolution than scalp EEG. However, radiological imaging methods offer more fine-grained localization. Common techniques are single-photon emission computed tomography (SPECT) imaging, in which a radioactive tracer is injected in the patient's blood stream during a seizure to detect active brain regions with increased CBF, and positron emission tomography (PET) imaging, in which a glucose tracer with a radioactive fluorine isotope allows to track regions with increased or decreased metabolism (Van Paesschen et al., 2007). Although ictal fMRI data have been acquired in some cases, this technique is not often applied, since seizures produce a variety of symptoms[8] which cause patient motion, heavily undermining the image quality (Salek-Haddadi et al., 2002). Furthermore, seizures are rare events, which would result in a low average yield of an fMRI scanning session.

Another surrogate for the epileptogenic zone is the *irritative zone*, where IEDs originate. IEDs are sharp, intermittent, electrophysiological events which are also much more frequent than seizures (cfr. the side-by-side comparison in Figure 2.6). Crucially, and contrary to seizures, they do not induce significant symptoms, which renders them attractive for imaging with fMRI. To detect their presence, EEG can be simultaneously recorded, although fMRI-only approaches for EEG-negative cases exist (Hunyadi et al., 2015). We devote the next section to this technique of combined EEG–fMRI mapping.

Using IEDs as proxies for seizure relies on the premise that these two types of neural events affect the same structures in the brain, a hypothesis which has been explored in various experiments. It has been argued that IEDs can be thought of as 'proto-seizures' or 'sub-threshold' seizures (Rosenow and Lüders, 2001). Measurements with intracranial EEG electrodes have revealed that the onset zones of interictal and ictal activity lie close together in many patients (Hufnagel et al., 2000). IEDs are thought to play a role as precursors of

[8] Absence epilepsy cases are a noteworthy exception, since the symptoms are much more moderate (Moeller et al., 2010).

seizures, by reinforcing axonal connections which eventually form the epileptic network along which seizures may propagate (Rosenow and Lüders, 2001; Staley and Dudek, 2006). In relation to this, it has been found that the spatial distribution of interictal activity corresponds well to that of ictal activity, suggesting that both neural events indeed propagate along and perhaps originate in the same brain structures (Tousseyn et al., 2015). The validity of using IEDs for presurgical evaluation has also been evaluated in a more pragmatical sense, by counting how often the IED-based results correctly predicted the target for resection in a patient that became seizure-free after surgery—we get back to this in section 2.5.

Also non-functional neuroimaging methods often provide helpful diagnostics. Notably, anatomical MR imaging can reveal structural deformations or lesions in the brain, with a high resolution. Many such anatomical abnormalities are linked to epileptogenesis, such as hippocampal sclerosis, tumors, or lesions after traumatic brain injury.

During presurgical evaluation, several of the techniques outlined above are applied, to gather complementary evidence towards the location of the epileptogenic zone.

2.5 Simultaneous EEG–fMRI in epilepsy monitoring

2.5.1 Complementary views of interictal events

The above summary makes clear that EEG and fMRI are very different modalities of measurement, relying on entirely distinct physical principles. Even so, by virtue of neurovascular coupling, electrophysiological activity in the neurons induces changes in the BOLD signals, so neural sources are partially reflected both in EEG and fMRI data. As such, both modalities are indicators of brain function[9], in a non-invasive way. Moreover, while EEG has a high temporal resolution (ms level) but poor spatial resolution (several cm), the opposite is true for fMRI, which has a low temporal resolution (several s), but compensates for it by a high spatial resolution (mm level). Of all potential combinations of non-invasive modalities, EEG–fMRI to date offers the most favorable compromise in terms of these two desiderata. Hence, EEG and fMRI

[9]... albeit they are related to different aspects of neural activity, cfr. sections 2.2.1 and 2.3.1. It has been found that BOLD signals correlate best to LFPs, which represent the overall perisynaptic electrical activity, and not to neuronal spiking activity per se (Logothetis et al., 2001; Ekstrom, 2010).

which have been simultaneously recorded can provide complementary views of the ongoing processes in the brain[10]. This is interesting, not only from the perspective of research, to study neural processes in healthy subjects (e.g. memory, language, visual perception, ...), typically as BOLD correlations to frequency-specific LFP oscillations (Mantini et al., 2007), but also from a clinical perspective, to localize pathological brain activity.

As briefly touched upon in section 2.4.2, EEG–fMRI is especially suited to study interictal epileptiform activity (Mulert and Lemieux, 2009; Gotman, 2008; Gotman et al., 2006; Abreu et al., 2018b). The central idea is that IEDs occur spontaneously, and since they are electrophysiological events, they produce concurrent hemodynamic changes in the vicinity of their origin and the regions they propagate to. One can thus attempt to localize the irritative zone by assessing which voxels have fMRI time courses which are correlated to the occurrences of IEDs seen on the EEG. This is the philosophy behind EEG-correlated fMRI analysis, which we discuss in the next section (and with more rigor in section 3.3.4). Despite important limitations, EEG-correlated fMRI analysis is still the de facto standard for localizing IEDs based on EEG and fMRI data (Lemieux et al., 2001; Salek-Haddadi et al., 2003, 2006; van Houdt et al., 2013)

2.5.2 Neurovascular coupling: a double-edged sword

Thanks to neurovascular coupling, epileptic activity can be studied with both EEG and fMRI. However, since electrical and hemodynamic changes occur at very different time scales, directly correlating EEG and fMRI-BOLD time series is not meaningful. To account for this, advanced biophysical models of the generation mechanisms of EEG and fMRI signals and their mutual coupling exist (Valdes-Sosa et al., 2009), although they are very complex. A far more common approach is to employ a model of the neurovascular coupling in the brain, to predict the approximate temporal shape of an expected IED-related BOLD signal. Often, a *hemodynamic response function* (HRF) is used for this purpose. An HRF describes the change in BOLD signal which is caused by a brief neural event, such as an evoked response which is elicited by an auditory stimulus, or an IED in the case of epilepsy monitoring. EEG-correlated fMRI analysis then proceeds in three stages (Lemieux et al., 2001):

1. the neurologist inspects the EEG signals to detect and mark occurrences of IEDs[11]

[10] Furthermore, they are not susceptible to the same complete set of artifacts, which can make the joint data analysis more robust.

[11] This step might also be automated by detection software, cfr. section 3.3.3.

2. a model for the IED-related BOLD time course is built by convolving the timing of the IEDs with the HRF
3. fMRI software analyzes which brain voxels have a BOLD time course which is (highly) correlated to the model time course

Hence, in the third step, voxels may be found which exhibit a strong correlation to the occurrence of IEDs: this provides evidence that these locations belong to the 'epileptic network' and are involved in IED generation or propagation.

While this is an intuitive approach, it has some pitfalls. One important consideration is that the neurovascular coupling mechanism (aka the HRF) is not set in stone: substantial variability of the HRF has been observed for different (even healthy) subjects (Aguirre et al., 1998), and for different brain regions (Handwerker et al., 2004). The HRF also adapts with age (Jacobs et al., 2008) and even stress level (Elbau et al., 2018), likely because these factors have a physiological counterpart. In the case of epilepsy, the HRFs can be even more deviant (Bénar et al., 2002; Lemieux et al., 2008; Jacobs et al., 2009; van Houdt et al., 2013; Grouiller et al., 2010). This is hardly surprising, since epilepsy is often caused by structural abnormalities and deformations in a part of the brain, which can also have repercussions on the local regulation of metabolism and blood flow—also because the brain itself initiates inhibition mechanisms that may also affect local oxygenation, cfr. section 2.4.1 and (Rigau et al., 2007; Schwartz, 2007). In summary, it is nearly always impossible to accurately know the HRF *a priori*—for healthy and diseased people alike. Nevertheless, a 'canonical' HRF is still often used, i.e. a fixed HRF which is not customized to the patient nor the brain region.

What is the impact of this HRF misspecification? Figure 2.7, demonstrates the effect on EEG-correlated fMRI analysis for an epilepsy patient. It clearly shows how the canonical HRF (orange curves) can be a poor model for the real HRF (which was estimated from the fMRI data itself, in different voxels). Consequentially, the correlations calculated based on this model can be unreliable, and there is a risk of wrong conclusions about the sites which are significantly related to IED activity. The first and last panel of the figure illustrate that an analysis with either location-adjusted HRFs and with a canonical HRF indeed leads to (some) different 'significantly active' regions. The effect may be moderate or detrimental, depending on how strongly the real HRFs deviate from the standard HRF model which is used. Either way, the inevitable variability of the HRF undermines the (statistical) analysis of fMRI data, leading to an increase in Type I errors ('false positives', i.e., brain regions which are estimated to be active but are actually not) and Type II errors ('false negatives', i.e., brain regions which are in fact active, but are not recognized as such) (Calhoun et al., 2004; Lindquist and Wager, 2007; Lindquist

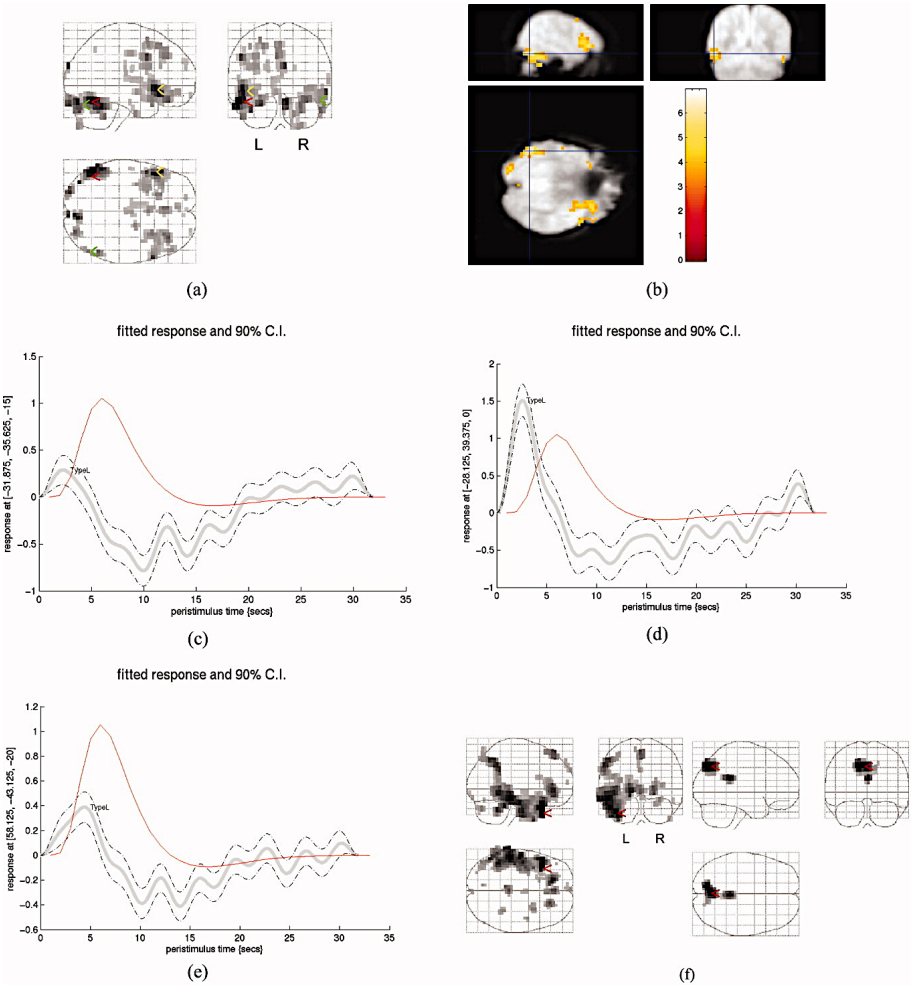


Figure 2.7: The natural variation of the neurovascular coupling over the brain can confound fMRI data analysis if it is not accounted for, and lead to erroneous interpretation of the data. **(a–b)** Map of one patient's brain regions which are significantly active (in dark grey , resp. orange) during epileptiform activity which was found on EEG, when adjusting for neurovascular variability. **(c)** The actual regional coupling characteristic (grey line) can differ substantially from the standard model which is often used (orange line). **(d–e)** Same as **(c)**, for two other locations in the brain. Also there, the model does not describe the real (estimated) coupling well. **(f)** The same patient's results, but when using the standard model. Compared to **(a)**, several different regions are now marked as significant. Figure edited from (Lemieux et al., 2008; Salek-Haddadi et al., 2006).

et al., 2009; Monti, 2011). Especially for epilepsy monitoring, in which the EEG-correlated fMRI results may be used for presurgical evaluation, this is unacceptable. Precautions are therefore needed. Instead of standard EEG-correlated fMRI analysis, either methods that are immune to HRF variability (Ostwald and Bagshaw, 2011; Caballero-Gaudes et al., 2013), or methods that can estimate HRFs from the data (Glover, 1999; Woolrich et al., 2004; Lindquist and Wager, 2007; Bagshaw et al., 2004), can be used.

Furthermore, neurovascular coupling is a rather abstract notion, which stipulates that changes in the EEG, which are indirect reflections of underlying neural activity, have a counterpart in the measured BOLD signals. It is less obvious which feature of the EEG participates most in this coupling mechanism[12]. While the majority of experiments treat IEDs as binary variables ('on' at the times of occurrence, 'off' elsewhere in the EEG time course), this is a very crude model. It may be more accurate to make a distinction between different types of IEDs, or to use other, continuous indicators to represent them. For example, time-varying power of the EEG signal might be more suited. Different representations or features of IED-related EEG have been investigated in (Rosa et al., 2010; Murta et al., 2015; Abreu et al., 2018a). Notably, (Abreu et al., 2018a) observed substantially different outcomes of the EEG–fMRI analysis, depending on whether the simple IED representation, power-based metrics, or metrics related to a network notion of the brain were used. Hence, some features may adhere more closely to the real physical mechanisms behind neurovascular coupling, which can lead to more accurate results.

2.5.3 Data quality

Recording EEG and fMRI simultaneously is a daunting technical task, especially since the large magnetic fields of the MR scanner pose a very hostile environment for electrical measurements like EEG. The fast switching of magnetic gradients during fMRI acquisition leads to a time-varying magnetic field, which generates electrical currents in the electrodes through Faraday's induction principle. These 'gradient artifacts' lead to extreme distortion of the EEG waveforms, due to the involved field strengths (see Figure 2.8). Luckily, the pulse sequence, which dictates the gradients' switching behavior, is exactly known, and the artifacts are very stable/reproducible. Hence, the gradient artifact can in large be mitigated by a time-locked averaging and subtraction procedure (Allen et al., 2000). A

[12]In this thesis, we use the term 'neurovascular coupling' to describe the coupling characteristic between EEG and fMRI temporal dynamics, and make the silent assumption that this characteristic is a proxy/surrogate for 'neurovascular coupling' as it is understood in neuroscience: the model that describes BOLD changes in response to electrical neural 'events', which take the form of local field potentials at the synapses.

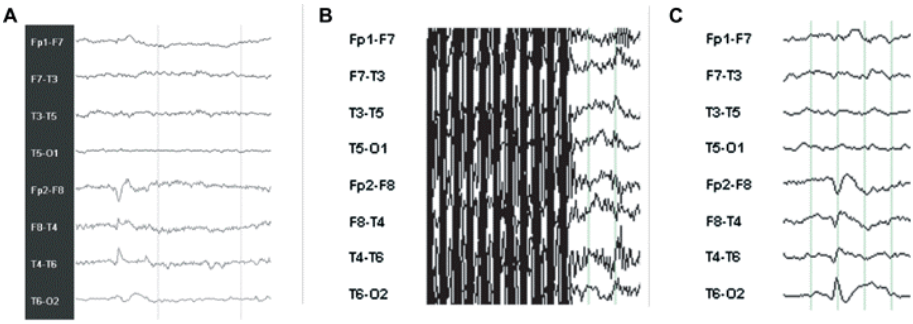


Figure 2.8: Gradient artifacts heavily corrupt the EEG during the fMRI acquisition. **(a)** EEG recorded outside the MR scanner, with good quality and an observable interictal spike. **(b)** EEG recorded inside the MR scanner, when fast gradient switching for the fMRI acquisition introduces large artifacts in the signals. **(c)** EEG signals recorded inside the MR scanner, after a successful artifact removal procedure. A similar interictal spike as in the left panel can now be observed. Figure taken from (Gotman, 2008).

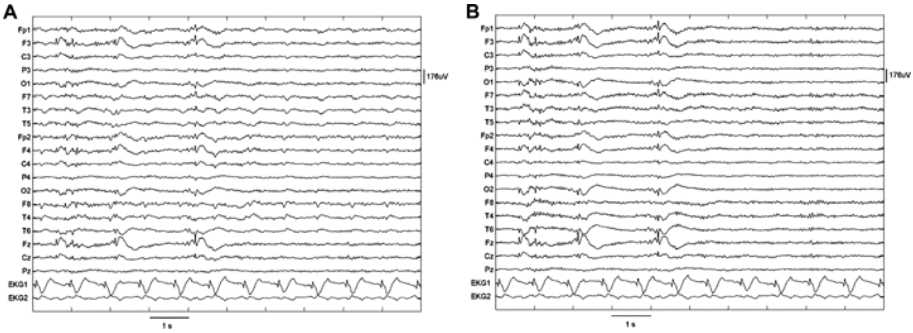


Figure 2.9: The ballistocardiogram artifact results from slight movements of the conducting electrodes inside the scanner's magnetic field, which stem from the pulsed motion of blood in the scalp. **(a)** The artifact appears approximately time-locked to the heart beats (cfr. bottom two signals). Also three interictal spikes can be seen. **(b)** The EEG can be cleaned by using an artifact removal procedure, in this case independent component analysis (ICA). Figure taken from (Gotman, 2008).

greater issue is posed by the ballistocardiogram artifact, which is the combined induction effect ultimately caused by the pulsating heartbeat (Allen et al., 1998) (see Figure 2.9). The pulsed stream of blood to the head's veins causes a subtle, periodic expansion of the scalp on which the electrodes are applied, and causes

the head to slightly rotate with every pulse. The resulting motion of electrodes in the magnetic field contributes artifactual waveforms to the EEG (Allen et al., 1998; Marino et al., 2018). This artifact can be removed using e.g. the 'optimal basis set' method, based on principal component analysis of waveforms that are time-locked to heartbeats detected from a co-recorded electrocardiogram, or by means of independent component analysis (Abreu et al., 2016a; Vanderperren et al., 2010). Motion of the electrodes (and ensuing artifacts) can also result from mechanical vibrations of the scanner's subsystems, such as the cooling system to sustain superconductivity in the coils. The 'usual suspects' are also present, i.e., the artifacts that also occur in standalone EEG (cfr. section 2.2.3).

fMRI data are affected to a lesser extent as a result of the combined recording with EEG. A notable effect is MR signal loss in the vicinity of electrodes; this results from the differences in magnetic susceptibility between the electrodes and the head, which locally distorts the static magnetic field of the scanner. Common fMRI artifacts are still present (cfr. section 2.3.3).

A comprehensive overview of the challenges and available solutions with respect to data quality of simultaneously recorded EEG–fMRI data can be found in (Abreu et al., 2018b).

2.6 Summary and prospects

EEG-correlated fMRI analysis is still the most widely applied method to date that combines EEG and fMRI to map the irritative zone in the brain (Gotman, 2008; Abreu et al., 2018b). It is well understood, and provides a flexible framework to localize interictal epileptiform activity, while accounting for the confounding influence of artifacts (see section 3.3.4). It succeeds moderately well: for the part of the patient population that undergoes surgery and becomes seizure-free afterwards, the EEG–fMRI results often correctly predicted significant BOLD changes within the area that would later be resected (Thornton et al., 2010; An et al., 2013), and EEG–fMRI results either helped or confirmed the predictions made with other modalities (Zijlmans et al., 2007). Nonetheless, it is certainly no 'one-size-fits-all' solution, because 1) a substantial proportion of patients is EEG-negative or no IEDs could be observed during the fMRI scanning session; 2) even for patients with observed IEDs, sometimes no voxels with significantly correlated BOLD changes are found; and 3) even when EEG-correlated analysis does yield IED-related BOLD maps, interpretation can be challenging because of multiple significant clusters of voxels (Salek-Haddadi et al., 2006; Thornton et al., 2010). The first problem can be mediated in a small number of ways. When truly no IEDs are visible on the EEG, even when

recorded outside of the scanner room, with much fewer artifacts, standalone fMRI needs to be used. E.g., in (Hunyadi et al., 2015), a method is proposed that can extract IED-related information and localization, solely from fMRI images. However, IEDs may be present but go unnoticed, due to the highly contaminated EEG signals during fMRI acquisition. One way of circumventing this is to calibrate an IED detection algorithm on 'clean' EEG that was recorded outside the scanner, with the assumption that it may still discover IEDs in the noisy EEG (Grouiller et al., 2011). The second problem may be due to model mismatch: as noted in section 2.5.2, failing to account for HRF variability may reduce statistical power, causing some interesting regions to be missed. Indeed, it has been shown that flexible HRF models can increase the yield of EEG–fMRI studies (Bagshaw et al., 2004). At the same time, misspecified HRFs can lead to false activations as well, which may partially underlie the third problem. Type II errors might also be caused by an incomplete or erroneous marking of IEDs (Flanagan et al., 2009), or due to poorly chosen IED features for modeling IEDs (Abreu et al., 2018a). Since EEG-correlated fMRI analysis is asymmetrical, with the EEG-derived information feeding the fMRI-based mapping, there is no way to correct for the last two categories of errors, at least within the boundaries of this analysis. Other approaches towards analysis, in which EEG and fMRI data are treated on equal footing, can allow interaction between the multimodal data (we touch upon this premise in section 3.2.3).

We therefore identify opportunities to mitigate some of these shortcomings of EEG–fMRI analysis.

1. Within the framework of EEG-correlated fMRI analysis, the **quality of the EEG** signals may be improved, to make IED detection more accurate.
2. More advanced **features for IED representation** may succeed better at modeling BOLD signals.
3. EEG and fMRI data may be combined in a more **symmetrical** way, to emphasize **multimodal discovery**, instead of one-way information flow.
4. Advanced methods to **account for HRF variability** can lead to fewer Type I and Type II errors.

In the next chapters, we aim to grasp these opportunities, and provide solutions to some of the problems outlined above. We briefly reiterate the contributions from each chapter, as outlined in section 1.2. In chapter 4 we provide a solution for item one and two, simultaneously, by applying a semi-automated filtering method that enhances EEG signals based on annotations of IEDs. Chapter 5 proposes a solution for item 3 and a preliminary attempt at solving item 4, by simultaneously analyzing EEG and fMRI by means of a simultaneous

factorization of EEG and fMRI data into shared factors, in which also a patient-specific HRF is co-estimated. This allows to account for patient-by-patient variation, which is a primary source of HRF versatility. Chapter 6 provides a practical method that helps to easily assess the reproducibility of methods like the one introduced in chapter 5, which suffer from a dependency on algorithmic initialization. Chapter 7 extends chapter 5 by proposing a solution that accounts for variation of the HRF over patients *and* over brain regions, and provides an integrated statistical inference framework. This extended method addresses a second fundamental source of intra-patient Type I and Type II errors, which result from HRF misspecification due to its natural variation over brain regions. The 'mixed blessing' of variable neurovascular coupling is even exploited, by utilizing it as a biomarker in IED mapping in the brain. In appendix A, we present a method that enables the detection of brain states from single-channel or few-channel EEG, while still training the system on an extended set of channels. This is potentially relevant for wearable seizure detection devices, which need to distinguish ictal from non-ictal and interictal states, and which may be calibrated in the hospital with more channels than will be available in the device itself, to promote a robust solution.

Chapter 3

Signal Processing and Machine Learning for Neural Data

In all forms of medical imaging, one or more physical variables are measured, such as electrical fields in EEG or blood oxygen concentration in fMRI. Invariably, these measurements are recorded, stored and processed as digital data, whose dimensionality is nowadays often so large that it becomes very difficult for a human to manually inspect and understand the raw data. Besides the sheer size of the data, the diversity of the data—i.e., the fact that the data varies over multiple modes, such as space, time, frequency—forms an obstacle[1]. Therefore, in order to extract *actionable* information from these datasets, one needs to use specialized signal processing techniques which can elicit a certain structure or pattern which is hidden in the raw data. Some of these techniques do not depend solely on manually defined steps, but are tuned for their specific task in a data-driven fashion. Since this latter type of techniques rely on algorithms which discover and/or exploit structure in the available data, they are said to *learn*—in a certain, 'narrow' sense—and are referred to as pattern recognition or *machine learning* techniques. *Unsupervised* learning algorithms operates exclusively on the measured data to perform a task that transforms the data into a more actionable or understandable representation. Conversely, a *supervised* learning algorithm has access to not only the data themselves, but also additional side information provided by the human user. This metadata often take the form of labels, indicating which data samples belong to a certain category or class (e.g., seizure or no seizure, voxel inside or outside of the ictal onset zone) or an a priori data model.

Since the fields of signal processing and machine learning techniques are incredibly vast, this chapter limits itself to a birds' eye view on some of the most relevant techniques to handle data stemming from functional neuroimaging experiments, and more specifically EEG and fMRI. However, we acknowledge the relevance of other fields such as semi-supervised learning (where additional information is available for a subset of the data) or transfer learning (i.e., algorithms which exploit additional, related datasets to improve their task performance on a dataset at hand (Pan and Yang, 2009; Dai et al., 2009).

We first discuss different ways of representing data in a signal-processing context, and then treat a selection of relevant learning algorithms, categorizing them as *supervised* or *unsupervised*. For each category, we highlight how several methods have made their way to application in EEG and/or fMRI analysis. We conclude with a treatment on one of the most widely used methods for brain activation mapping in combined EEG–fMRI analysis, and discuss considerations on hypothesis testing, which backs up learned concepts with statistical rigor, and which is vital in localizing epileptic activity.

3.1 Higher-order data representation

Digital signals can be represented and processed using matrices and extensions thereof. The samples of a single time series (e.g. a single EEG channel's recording during a certain period of time) can be stored into a vector. When multiple time series are available (e.g. a multi-channel EEG recording), the vectors can be collected in a matrix. If, furthermore, one or more additional *modes* of measurement are present (e.g. a multi-channel EEG recording, for multiple subjects), the individual matrices can be stacked into a multiway array or *tensor* (e.g. a channels $\times$ time points $\times$ subjects tensor). Tensors are generalizations of vectors (1st–order tensors) and matrices (2nd–order tensors) to arbitrary order, and allow to preserve higher-order structure that resides in the data (Kolda and Bader, 2009; Cichocki et al., 2015; Sidiropoulos et al., 2017). As such, they provide more expressive power than matrices, which are limited to interactions between two modes. While they arise naturally in many situations through the nature of the data or experiment, tensors can also be constructed in various ways, e.g. by computing higher-order statistics or other *tensorization* strategies (Debals and De Lathauwer, 2015).

3.1.1 Notation

We denote scalars, vectors, matrices and tensors by lower case (e.g. a), lower case boldface (e.g. $\mathbf{a}$), upper case boldface (e.g. $\mathbf{A}$) and bold upper case calligraphic letters (e.g. $\boldsymbol{\mathcal{A}}$), respectively. An Mth–order tensor $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_M}$ is a multiway array which holds data varying over M modes (e.g. sensors, time points, frequencies, ...) with dimensions I_1, I_2, ... , I_M, respectively. Entries from the tensor are indexed as $a_{i_1 i_2 \ldots i_M}$, and the term 'mode-m fiber' is used to refer to any vector that lies along the mth mode (i.e. columns and rows are mode-1 and mode-2 fibers, respectively). Modes of a tensor can be collapsed into one in a process called matricization or unfolding.

Definition 3.1. The mode-m matricization or unfolding of an Mth tensor $\boldsymbol{\mathcal{A}}$ stacks all mode-m fibers of $\boldsymbol{\mathcal{A}}$ into a matrix $\mathbf{A}_{(n)}$, whose column mode contains the collapsed $M-1$ remaining modes.

In many operations involving tensors, dedicated products are being used. We list the most important ones here.

Definition 3.2. The outer product of two tensors $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_M}$ and $\boldsymbol{\mathcal{B}} \in \mathbb{R}^{J_1 \times J_2 \times \cdots \times J_N}$ is denoted as $\boldsymbol{\mathcal{A}} \otimes \boldsymbol{\mathcal{B}} \in \mathbb{R}^{I_1 \cdots \times I_M \times J_1 \cdots \times J_N}$, leading to an $(M+N)$th–order tensor whose entries are given by

$$(\boldsymbol{\mathcal{A}} \otimes \boldsymbol{\mathcal{B}})_{i_1 \ldots i_M j_1 \ldots j_N} = a_{i_1 \ldots i_M} b_{j_1 \ldots j_N} \tag{3.1}$$

Definition 3.3. The Kronecker product of two matrices $\mathbf{A} \in \mathbb{R}^{I_1 \times I_2}$ and $\mathbf{B} \in \mathbb{R}^{J_1 \times J_2}$ is denoted as $\mathbf{A} \otimes \mathbf{B} \in \mathbb{R}^{I_1 J_1 \times I_2 J_2}$, and is computed as

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & \cdots & a_{1I_2}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{I_1 1}\mathbf{B} & \cdots & a_{I_1 I_2}\mathbf{B} \end{bmatrix} \tag{3.2}$$

Definition 3.4. The Khatri-Rao product of two matrices $\mathbf{A} \in \mathbb{R}^{I_1 \times K}$ and $\mathbf{B} \in \mathbb{R}^{J_1 \times K}$ is denoted as $\mathbf{A} \odot \mathbf{B} \in \mathbb{R}^{I_1 J_1 \times K}$, and is computed by columnwise Kronecker products:

$$\mathbf{A} \odot \mathbf{B} = \begin{bmatrix} \mathbf{a}_1 \otimes \mathbf{b}_1 & \ldots & \mathbf{a}_K \otimes \mathbf{b}_K \end{bmatrix} \tag{3.3}$$

Definition 3.5. The Hadamard or elementwise product of two equally sized tensors $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_M}$ and $\boldsymbol{\mathcal{B}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_M}$ is denoted as $\boldsymbol{\mathcal{A}} * \boldsymbol{\mathcal{B}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_M}$, and is computed as

$$(\boldsymbol{\mathcal{A}} * \boldsymbol{\mathcal{B}})_{i_1 \ldots i_M} = a_{i_1 \ldots i_M} b_{i_1 \ldots i_M} \tag{3.4}$$

Definition 3.6. The inner product of two equally sized tensors $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_M}$ and $\boldsymbol{\mathcal{B}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_M}$ is denoted as $\langle \boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}} \rangle \in \mathbb{R}$, and is computed as

$$\langle \boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}} \rangle = \sum_{i_1} \cdots \sum_{i_M} a_{i_1 \ldots i_M} b_{i_1 \ldots i_M} \tag{3.5}$$

Definition 3.7. The mode-m multiplication of a tensor $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{I_1 \times \cdots \times I_m \times \cdots \times I_M}$ and a matrix $\mathbf{U} \in \mathbb{R}^{J \times I_m}$ is denoted as $\boldsymbol{\mathcal{A}} \times_m \mathbf{U} \in \mathbb{R}^{I_1 \times \cdots \times J \times \cdots \times I_M}$, and is computed by multiplying each mode-m fiber in $\boldsymbol{\mathcal{A}}$ with $\mathbf{U}$, i.e.:

$$(\boldsymbol{\mathcal{A}} \times_m \mathbf{U})_{i_1 \cdots j \cdots i_M} = \sum_{i_m=1}^{I_m} a_{i_1 \cdots i_m \cdots i_M} u_{j i_m} \tag{3.6}$$

Application in functional neuroimaging

EEG signals can easily be appreciated as a tensor of higher order, of which the modes depend on the type of experiment. Beyond the obvious modes of channels (sensors or electrodes) and time samples, often a frequency mode is present if the signals have been preprocessed by means of a time-frequency transform (e.g. short-time Fourier transform) or a wavelet decomposition, leading to an array with modes channels × time points × frequencies or scales. Additionally, EEG might be collected over different subjects or patients, or the long time recording may be cut into shorter segments, leading to a windows or segments mode[2]. In fMRI, fewer representations have been used. Typically, the BOLD time series collected over all voxels is stacked into a large time points × voxels matrix, which is very 'fat', as there are often many more voxels ($\sim 10^5$—10^6) than time points ($\sim 10^3$). Sometimes, the time series within well-defined regions of interest (ROIs) are spatially averaged, leading to a much thinner time points × ROIs matrix. These ROIs may stem from a predefined or data-drived parcellation of the brain, either based on an anatomical atlas or a functional atlas (Evans

[2]For experiments with controlled stimulus timing, such a mode is often called a trial mode or epoch mode.

et al., 2012). Various matrix and tensor representations for EEG and fMRI data and their merit are discussed in (Karahan et al., 2015).

3.2 Unsupervised signal processing

3.2.1 Unsupervised filtering

Filtering in the unsupervised regime can hardly be seen as learning, since the operations are often completely manually defined and do not depend on the data themselves[3]. Temporal filtering is applied to enhance or suppress signal components in a certain frequency band and relies on integration of the signals over multiple time samples. Spatial filtering, on the other hand, aims at enhancing or improving signal components with a certain spatial variation by integrating the signals over the recorded variables. Spatio-temporal filtering performs combines both approaches.

Application in functional neuroimaging

EEG data are often analyzed in distinct (temporal) frequency bands that have empirically been linked to certain neural processes. Hence, a common preprocessing step is to pass the signals through a filter bank with filters of different bandwidths, to extract the rhythms of the δ (ca. 1–4 Hz), θ (ca. 4–8 Hz), α (ca. 8–12 Hz), β (ca. 12–30 Hz) and γ ($>$ 30 Hz) bands (Buzsaki, 2006). Alternatively, the data are often processed with a time-frequency transform such as the short-time Fourier transform or a discrete or continuous wavelet transform, to obtain continuous spectra. The selected reference electrode in EEG acquisition implicitly defines a spatial filtering, in which a filter matrix with only the values -1, 0 and 1 acts on the raw, unipolar signals to subtract the signal measured at the reference electrode from all other signals. By extension, any form of rereferencing strategy defines a different spatial filter which is applied to the data. E.g., 'common average referencing', in which the average time series over electrodes is subtracted from all signals, is analogous to mean subtraction in the temporal domain, and removes the 'spatial DC' component from the signals. Laplacian filtering, or bipolar rereferencing, combines the signals at a few (resp. two) neighbouring electrodes, thereby cancelling out signal components which are common to all involved electrodes and thus acting

[3]There are exceptions, such as quadratic variance reduction to remove low-frequency drift (Fasano and Villani, 2014).

as a spatial high-pass filter, sharpening the electrical field distribution over the scalp (Nunez et al., 2006).

fMRI data are often high-pass filtered to remove low-frequency components that are due to scanner drift. The interesting bandwidth for neurally related hemodynamic activity is approximately 0.01–0.1 Hz, as a result of the low-pass filtering effect of the hemodynamic response function. This is lower than typical peak frequencies for breathing (and related body motion) and the beating of the heart (and related ballistocardiogram effects), both of which might distort the fMRI signal quality. Unfortunately, due to the very low sampling rate of fMRI data, artifacts related to breathing and heart beating are often aliased into lower frequencies as well, making simple low-pass filtering ineffective for removing them. fMRI are nearly always spatially smoothed using a 3D low-pass filter (e.g. a Gaussian kernel with a width of several voxels), under the assumption that nearby voxels pick up correlated neural activity. Binning of the voxels into ROIs, thereby averaging the BOLD time series into ROI time series, can be seen as a spatial low-pass filter as well.

3.2.2 Unsupervised dimensionality reduction

Samples of measured signals in raw form live in their respective sensor space, which can be high-dimensional. For example, an EEG recording with 64 electrodes can be seen as many subsequent temporal observations or samples of 64 spatial variables, occupying a 64-dimensional space. Performing operations in such high-dimensional spaces not only pose a computational burden, they can also be ill-posed and may lead to issues like overfitting (see section 3.3.3). On the other hand, the recorded signals may be redundant (oversampled), i.e., while the data are large, they may exhibit sufficient structure so that they can be described with fewer dimensions. Therefore, it is often desirable to find a lower-dimensional representation of the data, which still captures the essential characteristics that one is interested in. This process is referred to as dimensionality reduction, and generally consists of a transformation ϕ as

$$\phi : \mathbf{x} \mapsto \tilde{\mathbf{x}} \,, \tag{3.7}$$

where $\mathbf{x} \in \mathbb{R}^D$ are the original, high-dimensional data observations[4] and $\tilde{\mathbf{x}} = \phi(\mathbf{x}) \in \mathbb{R}^d$ are the transformed observations in the reduced space, i.e., $d \ll D$.

[4]In the most common case, individual observations are vectors, albeit observations can also take the form of entire matrices or higher-order tensors. In this case, one can either vectorize the observations (thereby sacrificing the inherent multiway structure) or employ dimensionality reduction algorihms that can handle matrices and tensors. Several popular dimensionality reduction techniques can indeed be tensorized, see e.g. (Yan et al., 2006).

The transformation $\phi(\mathbf{x})$ may be linear (i.e., a matrix multiplication) or non-linear (in which case the process is sometimes coined 'manifold learning'), and may be explicit (the transformation itself can be known, and be applied to any new observation) or implicit (the transformed observations can be computed in batch, extending to out-of-sample observations is more difficult or impossible). Arguably, the most famous and widely used linear dimensionality reduction method is principal component analysis (PCA) (Pearson, 1901; Hotelling, 1933), which has been used in very diverse fields and whose success may be attributed to its simplicity (involving only a singular value decomposition (SVD)) and its property of capturing as much of the data variance as possible in successive principal components. The paper of (Yan et al., 2006) didactically illustrates that many common dimensionality reduction methods can be derived and understood in a common framework of *graph embedding*. A graph is a mathematical representation of a network, and consists of nodes and edges (links). Any graph with N nodes can be described by its adjacency matrix $S \in \mathbb{R}^{N \times N}$, in which every element s_{ij} is the binary or weighted 'strength' of the connection (the edge) between nodes i and j. In the current context, each data observation may be treated as a node, and relationships between samples may be encoded as edges. For example, the elements in S may be computed by means of a similarity metric between any pair of observations. It can be shown that many unsupervised dimensionality reduction techniques (PCA, Locally Linear Embedding (LLE) (Roweis and Saul, 2000), Laplacian Eigenmap (Belkin and Niyogi, 2003), spectral clustering (Ng et al., 2002; Von Luxburg, 2007), ...) can be seen as an optimization over an adjacency matrix, i.e., a mapping ϕ is sought that approximately preserves similarity relationships (that exist in the original space, encoded by S) in the reduced space. As such, different methods differ solely in how they construct said adjacency matrix[5].

When the data do not come as vector observations, but as matrices or higher-order tensors, the multilinear singular value decomposition (MLSVD) (De Lathauwer et al., 2000a) provides an extension of PCA to tensors of arbitrary order. The MLSVD is a factorization of the Mth–order data tensor into a product of M orthogonal factor matrices $\mathbf{U}^{(m)}$, $m = 1 \dots M$, and an Mth–order, all-orthogonal core tensor $\boldsymbol{\mathcal{G}}$ (generalizing the diagonal matrix in a matrix

[5]In PCA, S takes the form of a matrix in which all elements are equal to one (i.e. an edge exists for all pairs of observations), indicating that neither sample is more important than the other, and that the algorithm only cares about preserving as much accumulated variance as possible. In Locally Linear Embedding (LLE) (Roweis and Saul, 2000) or Laplacian Eigenmap (Belkin and Niyogi, 2003), edges only exist for pairs of observations whose mutual distance is below a threshold. As such, these algorithms care to respect relative distances between close neighbors, and not those between observations that lie far apart in the original space.

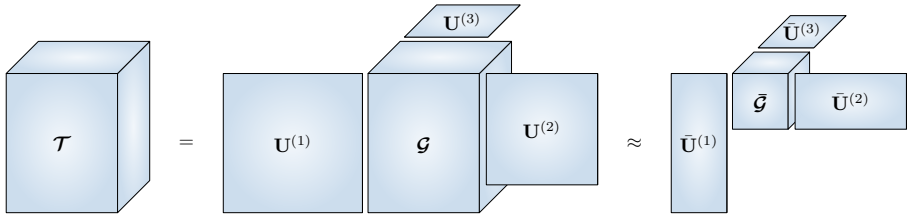


Figure 3.1: Any Mth–order tensor $\boldsymbol{\mathcal{T}}$ can be expressed by means of the multilinear singular value decomposition (MLSVD) as the product of an all-orthogonal core tensor $\boldsymbol{\mathcal{G}}$ of the same size as $\boldsymbol{\mathcal{T}}$ and square, orthogonal factor matrices $\mathbf{U}^{(m)}$, $m = 1 \dots M$. Here, we show an example for a tensor with $M = 3$ modes. Truncating the core tensor and factor matrices (leading to a compact core $\bar{\boldsymbol{\mathcal{G}}}$ and thin matrices $\bar{\mathbf{U}}^{(1)}$, $m = 1 \dots M$) can often result in a significant compression of the tensor data with only minor loss of information.

SVD):.

$$\boldsymbol{\mathcal{T}} = \boldsymbol{\mathcal{G}} \times_1 \mathbf{U}^{(1)} \times_2 \cdots \times_M \mathbf{U}^{(M)} \tag{3.8}$$

$$= \left[\!\left[\boldsymbol{\mathcal{G}}; \mathbf{U}^{(1)}, \dots, \mathbf{U}^{(M)} \right]\!\right] \tag{3.9}$$

The higher-order singular values are defined per mode as the Frobenius norms of the $(N-1)$th–order subtensors ('slabs') of $\boldsymbol{\mathcal{G}}$ in each respective mode. The MLSVD of tensors of order three and higher do not obey a generalization of the Eckart-Young theorem for matrices, which states that the rank-R truncated SVD of a matrix provides the best rank-R approximation, and that the sum of squared residuals is given by the sum of the (squared) discarded singular values. When the core tensor and factor matrices of a higher-order tensor's MLSVD are modewise truncated, the sum of the (squared) high-order singular values associated to the discarded slabs provide only an upper bound to the approximation error of the tensor. Neither is the truncated MLSVD the optimal low multilinear rank (or Tucker) approximation of the original tensor—although in practice, it often provides a very good initialization (De Lathauwer et al., 2000b). For any Tucker decomposition of the form in (3.8), the core tensor is generally non-sparse and encodes the interaction between any M-tuple of factor vectors in the matrices $\mathbf{U}^{(m)}$. It is closely linked to the Kronecker product (which generates interacting terms between any two factor vectors in its arguments) via the identity

$$\mathbf{T}_{(m)} = \mathbf{U}^{(m)} \mathbf{G}_{(m)} (\mathbf{U}^{(M)} \otimes \cdots \otimes \mathbf{U}^{(m+1)} \otimes \mathbf{U}^{(m-1)} \otimes \cdots \otimes \mathbf{U}^{(1)})^{\mathrm{T}} \,. \tag{3.10}$$

Compression of the tensor (and hence dimensionality reduction) can be achieved by eliminating elements of $\boldsymbol{\mathcal{G}}$ which are deemed unimportant by some metric.

This can be done by discarding complete slabs, but also by evaluating each element's merit in a Bayesian framework (Mørup and Hansen, 2009).

Most of the mentioned approaches consider the data in a *stochastic* sense, i.e., all observations are treated as independent realizations of some variables whose joint dimensionality one wishes to reduce. Nonetheless, algorithms which treat the data in a *deterministic* sense do exist. Notably, dynamical component analysis (DyCA) (Seifert et al., 2018) is suited to reduce the dimensionality of data whose dynamics (i.e. temporal observations) are fundamentally governed by a small set of (potentially nonlinear) differential equations of which the form need not be known in advance. Given that it only relies on the solution of a generalized eigenvalue problem, it is an interesting yet simple tool to study a variety of physical phenomena in a data-driven way, since no a priori model is required.

Recently, methods for signal or image denoising using neural network-based autoencoders have been developed (Hinton and Salakhutdinov, 2006; Goodfellow et al., 2016). Autoencoders have a layered architecture, with the number of perceptrons in each layer following hourglass profile from input to output. The task for the autoencoder is to reconstruct the initial data observations as accurately as possible, after they have been propagated through the network. Due to the funnel shape of its first stage (encoder), the network is forced to learn a latent representation of the input which has a much lower dimension, but which still allows to decompress the data with low error in the second stage (decoder).

Lastly, a common form of unsupervised dimensionality reduction is feature extraction (using manually engineered features), typically in the context of a pattern recognition task (see section 3.3.3), or subsampling the data variables.

Application in functional neuroimaging

PCA is sometimes applied as a precursor for Blind Source Separation (BSS) methods (see section 3.2.3), notably for independent component analysis (ICA) of both EEG and fMRI data (Calhoun et al., 2009). Extraction of ad-hoc features is very common in classical machine learning approaches that involve EEG data, e.g. in epileptic seizure detection (Shoeb and Guttag, 2010; Alotaiby et al., 2014). In fMRI, sometimes only data within an a priori defined ROI is analyzed, especially when a strong hypothesis on the localization exists. DyCA has been fruitfully applied to the modeling and detection of absence seizures, which typically evolve in a predictable ('deterministic') manner (Korn et al., 2019; Seifert et al., 2018).

3.2.3 Blind Source Separation (BSS)

Recorded signals are rarely pure and rather often consist of a superposition of multiple *sources* of activity, which are recorded as mixtures at the sensors. The goal is then to recover the individual sources from the measured data, i.e., perform blind source/signal separation (BSS) (Cardoso, 1998). A noteworthy analogy is the cocktail party problem (Cherry, 1953): in a room full of people that are involved in many distinct conversations, a microphone positioned anywhere in the room could only pick up a mixture of all speech sources—yet, the brain's auditory system of every attendee has the remarkable ability to isolate and attend to the voice of the person (s)he is talking to. Unfortunately, we can not (yet?) endow algorithms with the same complex capacities as the human brain, but it turns out that the same result can be achieved using clever signal processing techniques[6], provided that enough microphone signals are recorded, aside some technical conditions. Besides audio and speech signal processing, the BSS problem arises in many contexts, among others telecommunications, environmental signal processing, and in biomedical signal processing (Hyvärinen and Oja, 2000). More formally, the problem is described as

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \epsilon \,, \tag{3.11}$$

where $\mathbf{y} \in \mathbb{R}^N$ is the measured data, $\mathbf{A} \in \mathbb{R}^{N \times R}$ and $\mathbf{x} \in \mathbb{R}^R$ are the unknown mixing matrix and R underlying sources, and ϵ is a residual term. The N variables may represent e.g. sensors or voxels, and may be observed over K time samples, leading to matrices $\mathbf{Y} = \begin{bmatrix}\mathbf{y}[1] & \mathbf{y}[2] & \dots & \mathbf{y}[K]\end{bmatrix}$ and $\mathbf{S} = \begin{bmatrix}\mathbf{s}[1] & \mathbf{s}[2] & \dots & \mathbf{s}[K]\end{bmatrix}$. We aim to find estimates $\hat{\mathbf{A}}$ and $\hat{\mathbf{s}}$ of the mixing matrix and sources, such that $\mathbf{y} \approx \hat{\mathbf{A}}\hat{\mathbf{s}}$ ($\|\epsilon\|_F^2$ is exactly zero, or is minimized). This is generally impossible: by writing

$$\begin{aligned}
\mathbf{y} &= \hat{\mathbf{A}}\hat{\mathbf{s}} + \epsilon && (3.12)\\
&= \hat{\mathbf{A}}\mathbf{Q}\mathbf{Q}^{-1}\hat{\mathbf{s}} + \epsilon && (3.13)\\
&= (\hat{\mathbf{A}}\mathbf{Q})(\mathbf{Q}^{-1}\hat{\mathbf{s}}) + \epsilon && (3.14)\\
&= \tilde{\mathbf{A}}\tilde{\mathbf{s}} + \epsilon \,, && (3.15)
\end{aligned}$$

it can readily be seen that $\tilde{\mathbf{A}} = \hat{\mathbf{A}}\mathbf{Q}$ and $\tilde{\mathbf{s}} = \mathbf{Q}^{-1}\hat{\mathbf{s}}$ are equally good estimates (in terms of fitting $\mathbf{y}$), for an arbitrary invertible matrix $\mathbf{Q} \in \mathbb{R}^{R \times R}$. A rotational ambiguity is said to exist, and hence, additional constraints on $\hat{\mathbf{A}}$ and/or $\hat{\mathbf{s}}$ are needed in order to obtain a well-posed problems in which these quantities

[6] Some of which constituted the candidate's first research experience (Van Eyndhoven et al., 2016).

are identifiable. Independent component analysis (ICA) (Comon, 1994; Bell and Sejnowski, 1995; Hyvärinen and Oja, 2000) makes the assumption that sources are statistically independent[7] (i.e., their joint probability density can be factored as the product of the marginal probability densities). Non-negative matrix factorization (NMF) (Lee and Seung, 1999) imposes that all elements in $\hat{\mathbf{A}}$ and $\hat{\mathbf{s}}$ are greater than or equal to zero, and many NMF algorithms also make certain assumptions on the sparsity of the sources (Cichocki et al., 2007). Canonical correlation analysis-based BSS (CCA-BSS) assumes that the different sources not only are uncorrelated, but also have a different autocorrelation structure (Borga and Knutsson, 2001).

Despite assigning the roles of 'mixing system' and 'sources', $\mathbf{A}$ and $\mathbf{s}$ fulfill symmetrical roles from an algebraic point of view, and it is useful to think of them as factor matrices that contain the sources' *signatures* in the respective modes. This immediately leads to an equivalent way of restructuring expression (3.11):

$$\mathbf{y} = \sum_{r=1}^{R} \mathbf{a}_r s_r + \epsilon \tag{3.16}$$

$$\mathbf{Y} = \mathbf{AS} + \mathbf{E} \tag{3.17}$$

$$= \sum_{r=1}^{R} \mathbf{a}_r \mathbf{s}_r^{\mathrm{T}} + \mathbf{E} \tag{3.18}$$

$$= \sum_{r=1}^{R} \mathbf{a}_r \otimes \mathbf{s}_r + \mathbf{E}\,, \tag{3.19}$$

where $\mathbf{a}_r$ and $\mathbf{s}_r$ are the mode-1 and mode-2 signatures of the rth source, respectively. Hence, BSS can be understood as the decomposition of the original data into a sum of rank-1 terms (outer products of vectors), each of which models the contribution of one source.

When the data exist in more than two modes, a tensor representation and subsequent decomposition is warranted. A tensor $\boldsymbol{\mathcal{T}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_M}$ may be (approximately) decomposed as a sum of rank-1 terms, where every rank-1 term[8] is an outer product of M vectors, which is known as the parallel factor analysis (PARAFAC) or polyadic decomposition model (Harshman et al., 1970;

[7] This is a stronger constraint than uncorrelated sources, which is necessary since the latter property can be preserved for any orthogonal transformation of $\hat{\mathbf{s}}$ and $\hat{\mathbf{A}}$.

[8] ... or "crow's foot", as a close colleague would refer to it.

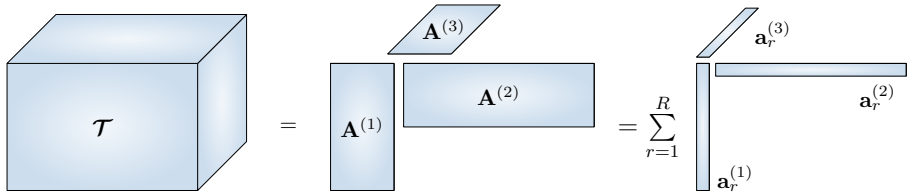


Figure 3.2: Canonical Polyadic Decomposition (CPD) of a third-order tensor $\mathcal{T}$ into factor matrices $[\![\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \mathbf{A}^{(3)}]\!]$, each of which has a column dimension R. The CPD can equivalently be written as a sum of R rank-1 tensors, each of which is the outer product of 3 signatures. In a blind source separation (BSS) context, every rank-1 term is interpreted as modeling the contribution of one source.

Kruskal, 1977; Bro, 1997):

$$\mathcal{T} = \sum_{r=1}^{R} \mathbf{a}_r^{(1)} \otimes \mathbf{a}_r^{(2)} \otimes \cdots \otimes \mathbf{a}_r^{(M)} + \boldsymbol{\mathcal{E}} \tag{3.20}$$

$$= [\![\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \ldots, \mathbf{A}^{(M)}]\!] + \boldsymbol{\mathcal{E}} \,, \tag{3.21}$$

where any column vector $\mathbf{a}_r^{(m)} \in \mathbb{R}^{I_n}$ is the mode-m factor vector of the rth component or term, $\boldsymbol{\mathcal{E}}_x$ is a residual tensor, and $[\![\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \ldots, \mathbf{A}^{(M)}]\!]$ is shorthand notation of the tensor decomposition in terms of the factor matrices in every mode $\mathbf{A}^{(m)} = \begin{bmatrix} \mathbf{a}_1^{(m)} & \mathbf{a}_2^{(m)} & \ldots & \mathbf{a}_R^{(m)} \end{bmatrix}$, $m = 1 \ldots M$. Figure 3.2 shows the example of a rank-R CPD of a third-order tensor. It can readily be seen that the PARAFAC model is a special case of the Tucker model, in which the core tensor is the superdiagonal identity $\boldsymbol{\mathcal{G}} = \boldsymbol{\mathcal{I}}$. Similarly to (3.10), the PARAFAC model is closely linked to the Khatri-Rao product (which only generates interacting terms between corresponding factor vectors and is a 'subset' of the Kronecker product) via the identity

$$\mathbf{T}_{(m)} = \mathbf{U}^{(m)} (\mathbf{U}^{(M)} \odot \cdots \odot \mathbf{U}^{(m+1)} \odot \mathbf{U}^{(m-1)} \odot \cdots \odot \mathbf{U}^{(1)})^{\mathrm{T}} \,. \tag{3.22}$$

The number of sources R is called the *rank* of the PARAFAC decomposition. The minimal R for which the residual in (3.20) is zero leads to the canonical polyadic decomposition (CPD). For real-world data, which are subject to noise, the decomposition is rarely exact (unless with a very high number of terms), and some residual is tolerated besides the low-rank structured approximation (which is colloquially still referred to as CPD). Contrary to matrices, decompositions of tensors of order $M > 2$ are not subject to the type of ambiguity as in (3.12), if the factor matrices are of full column rank and the decomposition rank is not

too high (Kruskal, 1977; Harshman et al., 1970; Sidiropoulos and Bro, 2000). When the data indeed follow a decomposition of the form (3.20), this is an advantage over the matrix case, since no additional constraints need to be imposed on the factors to make the decomposition *unique*, although constraints might further improve the quality of the extracted sources (Sidiropoulos et al., 2017). Hence, identifiability of the source signatures is guaranteed by virtue of the higher-order structure and the subsequent polyadic decomposition, which has no rotational degrees of freedom, principally by virtue of the small number of factor parameters compared to the number of data samples. Uniqueness properties for other, specialized tensor decompositions, potentially including constraints, have been derived (see e.g. (De Lathauwer, 2008; Domanov and De Lathauwer, 2016; Sørensen and De Lathauwer, 2015)). Furthermore, higher-order representations and decompositions allow to incorporate or preserve more of the data's diversity, in the form of additional modes compared to conventional 2-way (matrix) representations. This might be a reason in itself to transform vector or matrix data into a tensor, if justified (Debals and De Lathauwer, 2015).

There is however no free lunch. The tradeoff to be made is that it is typically harder to estimate the optimal number of sources R. While in the matrix case, a drop in the singular value spectrum may indicate an appropriate value for the rank, the same does not hold for higher-order tensors and singular values. Indeed, as also argued in section 3.2.2, the multilinear rank of a tensor is an M-tuple $\{R_1, R_2, \ldots, R_M\}$, which can be estimated by each mode's singular values, and the maximal value of this tuple only provides a lower bound on the CPD rank R (Sidiropoulos et al., 2017; Kolda and Bader, 2009). In many practical situations, the selection of this hyperparameter poses a critical, but tricky problem, and a great deal of research has been undertaken to provide strategies to estimate it. Heuristic methods exist, such as the Core Consistency Diagnostic (CorConDia) metric (Bro and Kiers, 2003; Papalexakis, 2016) which measures how closely a given set of factor matrices demands a superdiagonal core tensor $\boldsymbol{\mathcal{G}}$ to model the data, or metrics based on the Minimum Description Length (MDL) from information theory (Liu et al., 2016; da Costa et al., 2008). Alternatively, cross-validation techniques can provide a practical means of assessing the value of R which is most prolific for the desired task (Bro et al., 2008). Methods grounded in Bayesian principles implement the philosophy of Occam's razor in an automatic relevance determination framework to find an R which trades off data fit and model simplicity (Mørup and Hansen, 2009; Guo and Yu, 2011).

Except in contrived, noiseless cases (in which case the decomposition can be computed via the generalized eigenvalue decomposition), there is no closed-form solution for the factor matrices $\mathbf{A}^{(m)}$ that minimize the Frobenius norm $\|\boldsymbol{\mathcal{T}}\|_F^2$

in (3.21), optimization algorithms are commonly used to iteratively update a cost function, e.g. of the form

$$J(\mathbf{A}^{(1)},\ldots,\mathbf{A}^{(M)}) = \|\boldsymbol{\mathcal{E}}\|_F^2 = \left\|\boldsymbol{\mathcal{T}} - \sum_{r=1}^{R} \mathbf{a}_r^{(1)} \otimes \cdots \otimes \mathbf{a}_r^{(M)}\right\|_F^2 \tag{3.23}$$

Many optimization algorithms to minimize (3.23) exist, notably methods based on the very straightforward alternating least squares (ALS) principle (Harshman et al., 1970; Carroll and Chang, 1970), first order methods based on gradient descent (Acar et al., 2011a), or second order methods based on quasi-Newton or nonlinear (e.g. Gauss-Newton) least squares (NLS) (Sorber et al., 2013; Vervliet, 2018; Vervliet and De Lathauwer, 2018). When the underlying signals or sources of interest indeed admit a rank-R CPD structure, and the additive noise and other residuals are i.i.d. Gaussian distributed, minimizing J in (3.23) corresponds to maximum likelihood estimation (MLE), as J is then proportional to the negative log-likelihood of the observed data. When the residuals follow a different distribution, other cost functions are warranted to find the MLE (Cichocki and Amari, 2010), a consideration which also exists for matrix decompositions (Févotte and Idier, 2011). E.g., when many outliers are present in the data (e.g. due to measurement artifacts), cost functions which do not explode for larger values of the residuals are called for, such as $\|\boldsymbol{\mathcal{E}}\|_1$. Notably, correntropy (Liu et al., 2007; Zhang et al., 2016) is a cost function with a mixed behavior, i.e. it behaves as a quadratic penalty $\|\epsilon_{i_1 i_2 \ldots i_M}\|_2$ for small residuals $\epsilon_{i_1 i_2 \ldots i_M}$, as a linear ($\ell_1$-norm type) penalty $\|\epsilon_{i_1 i_2 \ldots i_M}\|_1$ for moderate residuals and saturates (cfr. the ℓ_0-norm $\|\epsilon_{i_1 i_2 \ldots i_M}\|_0$) for very large residuals. Existing algorithms must be adapted to custom cost functions, but some (e.g. Gauss-Newton methods) require only minor modifications (Vervliet, 2018; Vandecappelle et al., 2019). A difficulty that is encountered during optimization is the that cost function J is non-convex for tensors of order three and higher[9], and hence has many local minima aside the elusive global minimum. Any optimization algorithm can therefore only guarantee to one of these local minima—and not to the global minimum. When interpretability of the individual factor signatures $\mathbf{a}_r^{(m)}$ is a requirement (which is typically the case in BSS applications), this is an issue, since the estimated parameters might

[9]Intuitively, when expanding (3.23) in terms of the individual elements a of the factor matrices, a polynomial of degree equal to twice the tensor order M arises, which is non-convex as soon as $M > 1$. Only for $M = 2$, the Eckart-Young theorem exceptionally provides a closed-form optimal solution (unless special constraints are applicable). Also non-quadratic cost functions such as in (Cichocki and Amari, 2010) suffer from this issue in tensor fitting problems. Non-convexity is a problem for many constrained matrix decompositions alike, such as ICA and NMF. If the parameters in all but one mode are kept fixed, the polynomial is quadratic and easy to minimize, which is the basis of the update rule in the ALS algorithm.

differ, depending on how the algorithm was initialized [10]. Since there is no silver bullet solution, it is therefore wise to repeat the optimization procedure several times, starting from different initializations of the factors. This allows to assess whether the final factor estimates are sufficiently robust, and not exceptional (indicating that the algorithm got stuck in a 'bad' or non-representative local optimum).

When $P > 1$ related data tensors (or matrices) $\boldsymbol{\mathcal{T}}^{(p)}$, $p = 1 \dots P$ which (partially) measure the same sources are available, it is possible and potentially advantageous to perform *data fusion* (Acar et al., 2014; Lahat et al., 2015; Sorber et al., 2015). The rationale is that different measurements might offer complementary views on the same underlying sources, e.g. of some physical process, and analyzing them jointly could generate a holistic understanding. Data fusion through BSS requires that any two of the involved data tensors share at least one mode: datasets which are recorded simultaneously share the 'time' mode, measurements which are performed for the same group of patients share a 'patients' mode, etc. The data tensors can then be simultaneously decomposed in a coupled fashion, by minimizing a cost function of the form $\sum_p w_p J^{(p)}$ with scalar weights w_p, and enforcing that one or more factor matrices are shared between the factor sets in these respective modes (Acar et al., 2011b, 2013; Lahat et al., 2015). Optimization tools are available that are sufficiently generic to handle a wide variety of coupling and other constraints, such as the structured data fusion framework (Sorber et al., 2015; Vervliet, 2018; Vervliet and De Lathauwer, 2018). An example of such a linked decomposition is coupled matrix-tensor factorization (CMTF), in which a matrix and a tensor are simultaneously decomposed using a double CPD structure, and fully (Acar et al., 2013) or partially (Acar et al., 2014) share a factor matrix in a common mode. The CMTF inherits uniqueness properties from the CPD by virtue of the shared factor matrix (De Lathauwer and Kofidis, 2017; Sørensen and De Lathauwer, 2015).

Not every tensor decomposition models sources as rank-1 terms, as in (3.20). Indeed, when sources are more complex, they might be better described by a term of low multilinear rank, leading to a block term decomposition (BTD) (De Lathauwer, 2008). Alternatively, when prior info on the sources is available, it is sometimes meaningful to extract a low-rank term per source, after the data have been appropriately transformed into a tensor with a particular structure in a process coined *tensorization* (Debals and De Lathauwer, 2015). E.g., it can be shown that sources which are exponentially damped sinusoids admit a decomposition in rank-$(2, 2, 1)$ terms, after extending each variable's observations with shifted versions of itself to create a Hankel structure (De Lathauwer, 2011).

[10] Methods that compute a closed-form solution for the noiseless decomposition can provide good initial guesses for the factors when the signal-to-noise (SNR) ratio is sufficiently high.

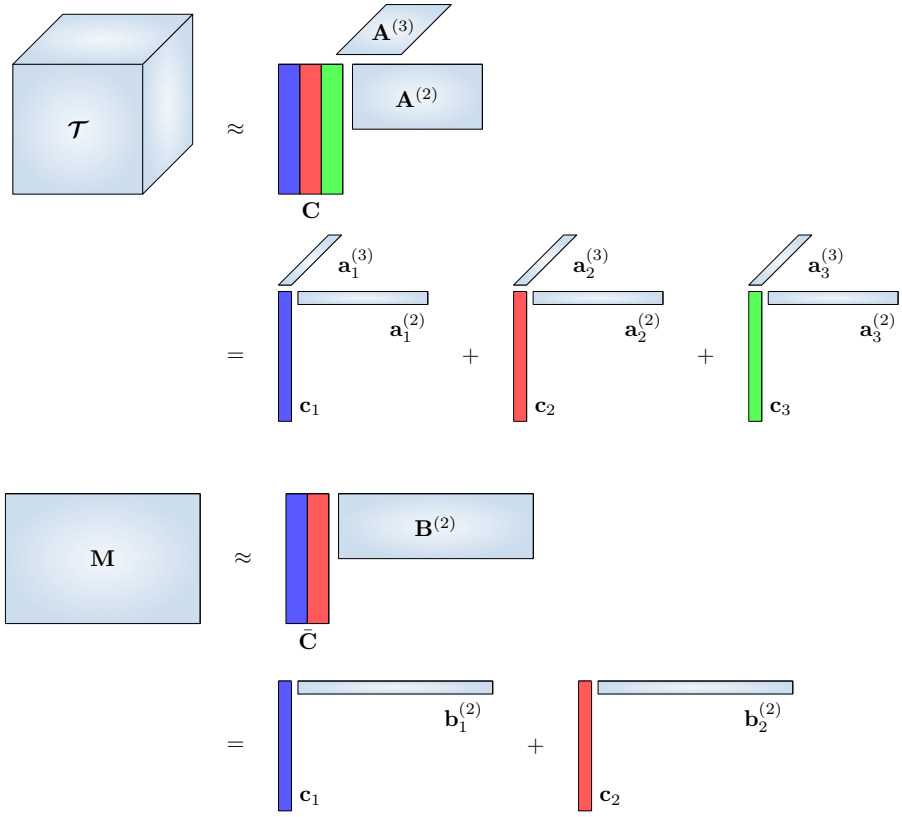


Figure 3.3: Data fusion can offer a holistic view on a certain phenomenon when complementary data are available. Coupled tensor decompositions reveal sources that underlie the data which are fully or partially shared between the involved tensors. An example is shown of a partially coupled matrix-tensor factorization (CMTF) of a tensor $\mathcal{T}$ and a matrix $\mathbf{M}$, which are measured along a shared first mode. Both $\mathcal{T} \approx [\![\mathbf{C}, \mathbf{A}^{(2)}, \mathbf{A}^{(3)}]\!]$ and $\mathbf{M} \approx [\![\bar{\mathbf{C}}, \mathbf{B}^{(2)}]\!]$ are decomposed into a sum of rank-1 terms, two of which are common to both datasets. A third source (with green mode-1 signature) is not shared and is specific to the decomposition of $\mathcal{T}$.

Similarly, when the sources take the shape of rational functions of degree d (which also enclose all polynomials up to degree d), the resulting data mixtures admit a decomposition in rank-$(d, d, 1)$ terms, after transforming each variable's observations into a Löwner matrix (Debals et al., 2015).

Tensors and ICA are closely intertwined. Not only do some ICA algorithms explictly use tensor formulations and CPD-like decompositions under the hood (Belouchrani et al., 1997; Cardoso and Souloumiac, 1993; Pham and Cardoso, 2001; De Lathauwer et al., 2007), the aim of any ICA algorithm is to optimize a criterion in terms of the higher-order statistics of the data, which are formally referred to as 'cumulants', and which can be represented by tensors of the same order (Cardoso, 1999). Indeed, while whitening (i.e., PCA) removes linear relationships between the variables up to the second order—thereby diagonalizing the covariance matrix—the objective in ICA is to find a linearly transformed space in which ideally *all* higher order statistics are devoid from relationships between the variables (Comon and De Lathauwer, 2010). This means that the algorithm either implicitly or explicitly seeks to diagonalize the cumulant tensors, leading to a symmetric CPD structure in the case where this is possible.

Application in functional neuroimaging

In several neuroimaging contexts, a linear mixing mechanism can be assumed to translate underlying neural sources and non-neural variations (e.g. artifacts) to the observed data. E.g., in EEG, electrical activity from the neurons propagates through the brain tissues (grey and white matter, cerebrospinal fluid) to the electrodes, as prescribed by the Maxwell equations. At the frequencies of interest (< 100 Hz), capacitive and inductive effects play no significant role, and the brain acts as a purely resistive medium. Hence, the neural source activity is projected nearly instantaneously and linearly to the sensors, via mixing profiles that depend on the locations of the sources in the brain, and the distribution of the tissues (with specific resistivities) throughout the head. Also physiological (e.g. eyeblink or muscle artifacts) and some non-physiological artifacts and interference (e.g. powerline noise) can be assumed to contribute linearly to the measured EEG signals, as in (3.16)–(3.19). This is known as the 'forward model', which allows to compute for any point in a discretized grid in the brain how strongly an electrical field at that point will be attenuated when propagating to any of the EEG electrodes (Hallez et al., 2007). Linear time-frequency transformations, e.g. the wavelet transform, equip the channels $\times$ time points data with an additional frequency mode, while preserving this

linearity[11] (Debals, 2017). Backed by this physical basis, BSS methods for EEG in both matrix or tensor format have been prolific. ICA and extensions thereof have been a popular tool of choice for many researchers (Makeig et al., 1996; Delorme and Makeig, 2004; Hyvärinen and Oja, 2000; Hyvärinen et al., 2010). For example, (Abreu et al., 2016b) have used ICA to extract an independent component (IC) time course which reflects the interictal discharges in epileptic EEG. ICA has proven very successful in removing artifacts, by capturing them in ICs, which allows to subtract their contribution to the EEG data (Jung et al., 2000; Urigüen and Garcia-Zapirain, 2015). Automated tools exist to categorize estimated ICs as different types of artifacts, e.g., the system in (Swartz Center for Computational Neuroscience, 2020) has been optimized via crowdsourcing on many examples of common EEG artifacts. CCA-BSS is an especially suited algorithm to remove muscle artifacts (which exhibit a much lower temporal autocorrelation than other activity of interest), as showcased in (De Clercq et al., 2006). Tensor decompositions of higher-order EEG representations have also been gaining increasing attention, even though they are not yet applied at the same scale as e.g. ICA. The CPD has been applied to wavelet-transformed, higher-order EEG data to extract sources of interest from healthy patients (see e.g. (Miwakeichi et al., 2004; Mørup et al., 2006; Zink et al., 2016)), from epileptic patients (see e.g. (Acar et al., 2007; De Vos et al., 2007; Becker et al., 2014, 2015)) or from neonatal infants (see e.g. (Deburchgraeve et al., 2009)). The $(L, L, 1)$-BTD has been applied to Hankel-transformed and wavelet-transformed EEG data in (Hunyadi et al., 2014) to extract non-stationary spatial and spectral characteristics of epileptic seizures.

ICA methods have been extensively applied to fMRI data as well (see e.g. (McKeown and Sejnowski, 1998; Beckmann and Smith, 2004; Calhoun and Adali, 2006; Calhoun et al., 2009)), typically as 'spatial ICA' in which the voxels are treated as observations and the time points as variables, which is dual to 'temporal ICA', the form generally used for EEG data[12]. Pseudo-automate denoising by means of ICA and automatic categorization and removal of artifactual ICs exists also for fMRI data (Salimi-Khorshidi et al., 2014). ICA has been used to extract epilepsy-related ICs from fMRI data in patients where the EEG data did not reveal any biomarkers (Hunyadi et al., 2013). For healthy patients, ICA is a suited method to extract spatial and temporal profiles of resting-state networks (RSNs), i.e. spontaneous fluctuations brain activity for which no a priori hypothesis exists (Mantini et al., 2007). An extension of ICA

[11] ... until a nonlinear operation is applied, such as taking the (squared) magnitude of complex valued wavelet or Fourier coefficients.

[12] It has been argued that spatial ICA works well for fMRI data, not because the spatial distributions of active sources are typically independent, but rather because they are sparse, and 'independence' is a trivial byproduct when they are non-overlapping (Daubechies et al., 2009). The jury is still out on a definite conclusion, since the claims in said paper have been rebutted in (Calhoun et al., 2013).

to multi-subject fMRI tensors has been proposed in (Beckmann and Smith, 2005), employing a two-step procedure involving classical matrix ICA and the SVD.

When EEG and fMRI data have been simultaneously acquired for one or more subjects, multimodal data fusion methods based on BSS can be applied. E.g., for groups of patients, JointICA (Calhoun et al., 2006; Mijović et al., 2012) has been applied on concatenated single-channel EEG and fMRI data, effectively forcing a coupling between EEG and fMRI across subjects. Parallel ICA (Eichele et al., 2008) and multi-set CCA (Correa et al., 2010) allow data fusion with looser constraints for EEG–fMRI coupling. CMTF provides a tensor extension of jointICA, allowing to incorporate all EEG channels (Hunyadi et al., 2016; Acar et al., 2017a, 2019). When EEG and fMRI are acquired simultaneously, coupling the temporal mode is valid, and has been explored via multiway partial least squares, a method closely related to CMTF, in which EEG and fMRI temporal signatures with maximal correlation are extracted (Martínez-Montes et al., 2004). Data fusion using tensor representations offer a rich spectrum of possibilities to combine EEG and fMRI data. For excellent reviews of tensor-based integration of EEG and fMRI, we refer to (Karahan et al., 2015; Hunyadi et al., 2017).

3.3 Supervised signal processing

3.3.1 Supervised filtering

When additional prior knowledge on the signal components of interest, or the noise, is available, filters can be tuned in a data-driven fashion, to optimize a criterion which encodes desired filter behavior. In matched filtering, a user provides a template of a waveform of interest, which is correlated to the data in a sliding-window fashion to detect similar waveforms in the recording. Multi-channel Wiener filtering (MWF) (Doclo and Moonen, 2002) is a spatio-temporal filtering technique which aims to reconstruct a desired signal from noisy observations. It hinges on the assumption that the (unknown) desired signal is sparse, i.e., that it is not continuously active, but vanishes during some periods, in which then only noise is present. This division in 'signal + noise' and 'noise-only' periods—which must be provided by the user—allows the filter to learn the second-order spatio-temporal statistics (covariance matrices) of the noise and also those of the desired signal. The MWF is then the optimal filter which minimizes the expected reconstruction error of the desired signal. Rooted in the same principle of disjoint sets of signal and noise observations, other types of filters, for different criteria, may be designed. For example, if

the main concern is not posed by continuous noise, but by short interference bursts, a 'signal-to-peak-interference ratio' filter may proceed by learning the second-order statistics of the interference and incorporating these in the tuning process (Wouters et al., 2018).

Application in functional neuroimaging

Matched filtering is relatively straightforward to apply, by segmenting a few interictal discharges on the EEG and constructing a template from their average. This approach has been successfully applied to (fuzzy) detection of interictal discharges in noisy data (Grouiller et al., 2011; Tousseyn et al., 2014b; Omidvarnia et al., 2019). The MWF has shown to be very effective at removing certain types of artifacts (e.g. eyeblinks, muscle artifacts) from continuous EEG data, if an annotation of artifactual segments in the recording is possible—either by a human observer or an automated algorithm (Somers et al., 2018). Filters that maximize the expected signal-to-peak-interference ratio have been used for neural spike sorting using micro-electrode arrays (Wouters et al., 2018).

3.3.2 Supervised dimensionality reduction

Finding good, lower-dimensional representations of data is a key component in many signal processing tasks, as discussed in section 3.2.2. A non-trivial task in general, especially when the distinct data characteristics are not yet fully understood and no hypothesis is available to support the choice for one or the other approach. For this reason, it may be advantageous to 'let the data speak for themselves', and find the most adequate dimensionality-reducing transformation in a data-driven fashion. While we discussed some data-driven approaches in section 3.2.2 already, additional efficacy of the data representation for a given task may be expected if the process can be guided by a set of available training labels, i.e., in a supervised fashion. For example, when trying to predict the classes to which different data observations belong (see later in section 3.3.3), features that are optimized to exploit class-discriminative information in the data can be expected to do a better job than manually engineered ones. Furthermore, it may be very difficult for a human to discover multivariate interactions in the data, which may be essential for discrimination tasks. This is a very similar philosophy as in supervised filtering 3.3.1, although the emphasis is now on projecting the data to a lower-dimensional space, rather than recovering noise-free or artifact-free full-dimensional signals.

The graph-based framework in (Yan et al., 2006), which we discussed in section 3.2.2 in the context of unsupervised methods, extends also to supervised dimensionality reduction. In the latter case, two main ingredients are used: an 'intrinsic graph', which encodes desirable statistical or geometrical properties of the (reduced) data, and a so-called 'penalty graph', which prescribes statistical or geometrical properties which are unwanted. Both of these need to be constructed by the user (reflecting the supervised aspect), based on the data labels. The penalty graph may e.g. contain nonzero weights for edges between observations that lie close together in the input space, but belong to different classes (i.e., observations that lie dangerously close to a class boundary): this urges the algorithm to not project these observations close together in the reduced space. The lower-dimensional representations $\tilde{\mathbf{x}}$ can be found in a non-parametric, implicit way (computing the features by means of and only for the inputs that were in the training batch, which makes out-of-sample extension more difficult) or in a parametric, explicit way, based on linearization, tensorization, or kernelization of the transformation $\phi(\mathbf{x})$ (Yan et al., 2006). E.g., for linearization, the transformation takes the form $\tilde{\mathbf{x}} = \mathbf{W}^{\mathrm{T}}\mathbf{x}$, and can be found as the minimizer(s) of a generalized Rayleigh quotient:

$$\hat{\mathbf{w}}_i = \arg\min_{\mathbf{w}_i} \frac{\mathbf{w}_i^{\mathrm{T}}(\mathbf{X}\mathbf{L}^{(i)}\mathbf{X}^{\mathrm{T}})\mathbf{w}_i}{\mathbf{w}_i^{\mathrm{T}}(\mathbf{X}\mathbf{L}^{(p)}\mathbf{X}^{\mathrm{T}})\mathbf{w}_i} \,, \tag{3.24}$$

in which $\mathbf{X}$ contains the data observations in the columns, $\mathbf{w}_i$ is the ith column of the transformation matrix $\mathbf{W}$, and $\mathbf{L}^{(i)}$ and $\mathbf{L}^{(p)}$ are the Laplacian matrices of the intrinsic and penalty graph, respectively. The minimization of this quotient favours a low value for the numerator, which expresses that the transformation $\hat{\mathbf{w}}_i$ should respect data relationships/similarities encoded by intrinsic edges, and at the same time a high value for the denominator, which promotes the separation between observations linked by penalty edges. For example, in LDA, the intrinsic graph contains edges between any two samples belonging to the same class (promoting low intra-class scatter and thus compactness), and the penalty graph is a contrast between the data covariance matrix and the intrinsic graph (promoting large inter-class scatter/separation). For a reduction to dimension d, expression (3.24) is optimized for a set of d filters $\mathbf{w}_i$, which can be found by solving a generalized eigenvalue decomposition once (the 'ratio-trace' problem), or in an iterated fashion (the 'trace-ratio' problem), depending on the desired normalization of $\mathbf{W}$ (Wang et al., 2007; Jia et al., 2009; Ngo et al., 2012).

Other linear dimensionality reduction methods include the common spatial patterns (CSP) algorithm (Fukunaga, 2013), which seeks spatial filters[13] $\mathbf{w}_i$

[13]Even though the name 'filters' is used in CSP terminology, we categorize them as a dimensionality reduction method, since the aim is to transform the data to a new space in

that maximize the contrast in signal power of data from two different classes or conditions. Also these filters are found by finding the extrema of a generalized Rayleigh quotient, this time in terms of the covariance matrices of both classes:

$$\hat{\mathbf{w}}_i = \arg \min/\max_{\mathbf{w}_i} \frac{\mathbf{w}_i^{\mathrm{T}} \Sigma^{(A)} \mathbf{w}_i}{\mathbf{w}_i^{\mathrm{T}} \Sigma^{(B)} \mathbf{w}_i} \,, \tag{3.25}$$

where $\Sigma^{(A)}$ and $\Sigma^{(B)}$ refer to the covariance matrices of the data observations in class A and B, respectively.

Multilinear/tensor variants of LDA have been proposed, e.g. to maximize a contrast function between the cores of projected data tensors from different classes (Phan and Cichocki, 2010; Li and Schonfeld, 2014; Frølich et al., 2018).

Application in functional neuroimaging

CSP filters have been very successfully applied as feature extractors to detect oscillatory patterns in EEG, and have become the go-to solution for such tasks in the brain computer interface research field (Lotte et al., 2007, 2018). Extensions exist to exploit temporal/spectral discriminative information, leading to common spatio-spectral patterns (CSSP) filters (Lemm et al., 2005; Blankertz et al., 2007). Multilinear discriminant analysis has been applied to EEG data as well, on a smaller scale (Phan and Cichocki, 2010; Frølich et al., 2018). In medical applications, it is sometimes tedious, costly, or simply infeasible to acquire a large amount of data. In such cases, estimation of the covariance(s) can become ill-posed, and regularization of the sample covariance matrices is warranted to avoid unstable results from algorithms which would otherwise use (nearly-)singular matrices in their optimization. It is then recommended to apply regularization to restore the matrix's condition number. Interestingly, (Ledoit and Wolf, 2004) derived a closed-form expression for the regularization weight, which leads to the asymptotically optimal estimator for the covariance in a minimum squared error-sense, which is even closer to the true, unobservable covariance than the sample covariance itself.

3.3.3 Classification

Transforming (independent) input observations into predictions about a (dependent) categorical variable is a classical pattern recognition task (Fukunaga, 2013; Lemm et al., 2011). The dependent variable encodes class membership

which a contrast between data from different classes is enhanced. In section 3.3.1, we focus on filters that denoise/recover a multivariate signal in its original space.

('label') of the data observations, and can take on two (binary classification) or more (multi-class classification) values. Invariably, a classifier relies on a calibration set of data observations with known labels, in order to learn the decision boundaries between all classes in the data space, a process known as *training*. I.e., a classification algorithm *learns* a rule

$$\psi : (\mathbf{x}; \boldsymbol{\theta}) \mapsto \{c_1, c_2, \ldots, c_P\} \,, \tag{3.26}$$

where $\mathbf{x}$ represents the input data, which is mapped to one of P classes c_p by means of a transformation $\psi(\mathbf{x})$ which relies on a set of parameters $\boldsymbol{\theta}$. The latter need to be learned from the data, typically as the solution of an optimization problem in which e.g. prediction error rate or a proxy thereof is minimized.

The main objective in classification is to gain good generalization performance, i.e., being able to accurately predict the label of observations that were not previously included in the training set. This performance is measured on a holdout or *validation* set, and leads to the notion of *overfitting* and *underfitting*, a tradeoff between bias and variance of the learned classification parameters (Fukunaga, 2013; Friedman, 1997). In the former, the classifier is in a regime of 'low bias, high variance', i.e., the trained parameters $\hat{\boldsymbol{\theta}}$ are highly specialized to the calibration set (low bias), but generalize poorly on unseen examples (the estimated parameters are highly variable on the specific realization of the training set). In the latter, the trained classifier operates in a regime of 'high bias, low variance', i.e., the estimates for the parameters $\boldsymbol{\theta}$ are rather consistent over realizations of the training set, at the price of not learning a fine, custom decision boundary, resulting in underwhelming performance. Even though the basic mechanism governing under/overfitting is well understood, finding the 'sweet spot' between the two remains a challenge. As a rule of thumb, acquiring more data for training is almost always beneficial, but this is a bottleneck in many practical scenarios. Furthermore, severe imbalance of the classes (e.g. when observations of one class are rare) poses a risk of underfitting of the minority class, depending on the training algorithm. Data augmentation can partially alleviate such data scarcity issue, by creating artificial data examples based on extrapolation from real observations (Chawla et al., 2002). Various *regularization* strategies can be used to incorporate prior knowledge on, and promote desirable properties of $\hat{\boldsymbol{\theta}}$, and when justified, will aid to stabilize the optimization in favor of lower variance (see section 3.3.4). Careful tuning of *hyperparameters*, including the degree and type of regularization, is often key, but is a non-trivial and computationally intensive procedure (and requires additional validation data).

Good classification performance often hinges on the adequacy of the data representation. Typically, raw input data offers very poor generalization performance, and is furthermore too high-dimensional typically: more robust

characteristics or *features* of the signals are thus needed. As we discussed in sections 3.2.2 and 3.3.2, many different unsupervised (including manual feature engineering) and supervised dimensionality reduction techniques exist to extract such features, and it is an art in itself to find those who offer the best predictive performance. Many types of linear (linear discriminant analysis (LDA), linear support vector machine (SVM), ...) and nonlinear classifiers (kernel SVM, random forests, gradient boosted trees, K-nearest neighbours, neural nets (NN), ...) can subsequently be trained on data features. Unfortunately, the choice of good features and a good classifier is not entirely independent. However, in the last decades, methods based on deep convolutional neural networks (CNN) have managed to overcome this limitation, by using a classifier architecture consisting of many layers of increasingly higher abstraction, loosely reminiscent of the neural processing cascade in the human brain (Bishop et al., 1995; Goodfellow et al., 2016). In neural nets, there is no fundamental distinction between feature extraction (lower-level layers close to the input) and discrimination/classification (higher-level layers close to the output), both of which are completely data-driven (LeCun et al., 2015). They have caused a revolution (Krizhevsky et al., 2012) in the field of machine learning, thanks to their extreme flexibility and learning potential, but bring additional difficulties, such as the fact that they can be extremely data-hungry, prone to overfitting, and extremely slow and energy-intensive to train (LeCun et al., 2015; Goodfellow et al., 2016).

Application in functional neuroimaging

Classification methods are encountered in the field of epilepsy monitoring for (EEG-based) seizure detection (see e.g. (Gotman, 2008; Alotaiby et al., 2014; Shoeb and Guttag, 2010; Hunyadi et al., 2012)), either with hospital-grade EEG systems, or with compact, wearable devices (Gu et al., 2018). Since seizures are rare events, they constitute a small minority of all training data observations, and this class imbalance should be taken into account. There is a potential for data augmentation methods such as in (Chawla et al., 2002) to build additional representative seizure data, in order to train more robust detection algorithms. Detection of interictal epileptic discharges from EEG recordings, a prerequisite for EEG-correlated fMRI analysis, has also been attempted using automated learning (Wilson et al., 1999; Scheuer et al., 2017; Nonclercq et al., 2012; Hao et al., 2018). Multi-voxel pattern analysis (MVPA) (Norman et al., 2006; Cox and Savoy, 2003) is the name given to the prediction of experimental conditions or 'states' from fMRI images, informally referred to as 'brain decoding' or 'mind reading' (Davatzikos et al., 2005). Many approaches in biomedical engineering still rely heavily on classical machine learning tools (Lemm et al., 2011). CNN-based methods are emerging (e.g. (Hao et al., 2018)) but struggle to take off, mainly due to the shortage of training data—especially in brain monitoring

contexts—and the enormous effort which is required for manual labeling of data. However, efforts are ongoing to overcome both limitations. The creation of large, openly available databases (such as (Obeid and Picone, 2016; Shah et al., 2018)) will likely catalyze the development of sophisticated deep learning methods for epilepsy monitoring. Similarly, crowdsourced labeling of neuroimages can fuel classification algorithms with sufficient training examples: one success story is that of the IClabel tool (Swartz Center for Computational Neuroscience, 2020) which recognizes various EEG artifactual ICs, mentioned in section 3.2.3. Beyond epilepsy monitoring, classification methods are extensively used in EEG-based brain computer interfaces (BCI) (Blankertz et al., 2007; Lemm et al., 2011), or those based on multiple modalities such as EEG combined with near-infrared spectroscopy (Fazli et al., 2012; Dähne et al., 2015).

3.3.4 Regression

Regression refers to the modeling a continuous variable, based on input data or features, and is probably the most commonly applied supervised machine learning technique, next to classification. In regression, a mapping from continuous input variables or predictors $\mathbf{x} \in \mathbb{R}^D$ to a continuous output variable(s) $y \in \mathbb{R}$ is learned:

$$\psi : (\mathbf{x}; \boldsymbol{\beta}) \mapsto y \; . \tag{3.27}$$

In the most commonly used linear regression, input(s) and output(s) are linearly related as $\hat{y} = \psi(\mathbf{x}; \boldsymbol{\beta}) = \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}$. The coefficients/parameters $\boldsymbol{\beta}$ are then optimized to fit the noisy dependent variable as accurately as possible, quantified by e.g. the squared error $\mathrm{E}\{|y - \hat{y}|^2\}$. In the linear case, and when the noise can be assumed independent over observations, the coefficients that minimize the empirical mean squared error can be found in a closed form solution by means of ordinary least squares (OLS):

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} J(\mathbf{X}, \mathbf{y}; \boldsymbol{\beta}) = \arg\min_{\boldsymbol{\beta}} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^{\mathrm{T}}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \tag{3.28}$$

$$= (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y} \; , \tag{3.29}$$

where $\mathbf{X} = \begin{bmatrix}\mathbf{x}_1 & \mathbf{x}_1 & \dots & \mathbf{x}_N\end{bmatrix}^{\mathrm{T}}$ collects all N predictor observations $\mathbf{x}_n$ in its rows, and $\mathbf{y}$ holds the values of the dependent variable corresponding to all observations. When the residuals or noise furthermore follow a Gaussian distribution, the OLS solution corresponds to the maximum likelihood estimator of $\boldsymbol{\beta}$, analogously to matrix and/or tensor factorization in section 3.2.3. When the noise is not white but colored, i.e., correlated over observations, the OLS estimator is still unbiased, but its variance increases. This latter effect should

not be neglected when subsequent statistical inference on the coefficients is done, since underestimating the variance leads to inflated Type I-error rates. Whitening the data (or equivalently, Generalized Least Squares (GLS) methods) can restore the efficiency of the estimator. While the focus in classification lies on generalization performance rather than the coefficients themselves, regression models are also commonly used for the sake of model fitting on calibration data only, e.g. when $\boldsymbol{\beta}$ can be meaningfully interpreted.

As in classification, underfitting and especially overfitting are very real concerns, and similar solution strategies apply. Sufficient calibration data are needed, to ensure that the least squares problem in (3.28) is overdetermined and well-conditioned. At the same time, regularization can be used to promote a meaningful, stable solution $\hat{\boldsymbol{\beta}}$ for the coefficients at the cost of an increased squared error. This is typically done by modifying the cost function to $\tilde{J}(\mathbf{X}, \mathbf{y}; \boldsymbol{\beta}) = J(\mathbf{X}, \mathbf{y}; \boldsymbol{\beta}) + \lambda\Omega(\boldsymbol{\beta})$, where $\lambda \geqslant 0$ and $\Omega(\boldsymbol{\beta})$ is a functional that encodes prior information about the sought coefficients, by acting as an additional penalty term that increases the cost when $\boldsymbol{\beta}$ would take on undesirable characteristics. Choosing $\Omega(\boldsymbol{\beta}) = \|\boldsymbol{\Gamma}\boldsymbol{\beta}\|_2^2$ is known as Tikhonov regularization (Tikhonov, 1943), and the form of the matrix $\boldsymbol{\Gamma}$ can be designed to control which aspect of $\boldsymbol{\beta}$ is penalized during optimization of $\tilde{J}$. E.g., ridge regression takes $\boldsymbol{\Gamma} = \mathbf{I}$, reducing the penalty term to the squared ℓ_2-norm of $\boldsymbol{\beta}$, which helps to prevent excessive amplitude of the fitted coefficients. When the coefficients have a notion of ordering (e.g., when they represent voxels or time points), $\boldsymbol{\Gamma}$ can be chosen as a derivative operator, to promote smoothness of $\hat{\boldsymbol{\beta}}$ by penalizing strong variations between neighboring coefficients. Instead, when $\Omega(\boldsymbol{\beta})$ is proportional to the ℓ_1-norm $\|\boldsymbol{\beta}\|_1$, it is known as the 'least absolute shrinkage and selection operator' (LASSO), wich promotes sparsity of the coefficients Tibshirani (1996); Donoho (2006). When variables are grouped, it can be meaningful to promote sparsity groupwise, using an adapted ℓ_1 penalty over group norms (Yuan and Lin, 2006). Different forms of regularization can be understood from a Bayesian perspective. While OLS provides the MLE for $\lambda = 0$, regularization adds a term which corresponds to the (scaled) negative logarithm of the prior probability distribution of $\boldsymbol{\beta}$. For the right choice of λ, it can thus be seen that solving for $\hat{\boldsymbol{\beta}}$ coincides with maximum a posteriori (MAP) estimation. For example, ridge regression corresponds to a multivariate Gaussian prior distribution of $\boldsymbol{\beta}$ with diagonal covariance, while the LASSO corresponds to a multivariate Laplacian prior distribution with diagonal covariance (having heavier tail regions than the Gaussian distribution, but a larger central peak). However, the strength of regularization λ is a hyperparameter which is often tuned via e.g. cross-validation to tightly regulate the bias-variance tradeoff.

When the input dimensionality D grows larger than the number of observations N, the problem is no longer overdetermined, which undermines a meaningful

and/or unique solution for $\hat{\boldsymbol{\beta}}$. In the matrix setting, one has to resort to dimensionality reduction of $\mathbf{x}$ (see sections 3.2.2 and 3.3.2), or to LASSO-type regularization if sparsity is justified (Donoho, 2006). However, the problem can be also circumvented by imposing a low-rank structure on $\boldsymbol{\beta}$, without altering its extrinsic dimensionality nor imposing penalty terms. I.e., $\boldsymbol{\beta}$ can be assumed to be a vectorized version of a higher-order tensor with a low-rank Tucker or PARAFAC structure (Zhou et al., 2013; Boussé et al., 2018): as discussed in sections 3.2.2 and 3.2.3, these low-rank models employ (much) fewer parameters than the number of elements in the tensor they describe. As such, the coefficients are parametrized as $\boldsymbol{\beta} = \text{vec}(\llbracket \boldsymbol{\mathcal{G}}; \mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(M)} \rrbracket)$ (Tucker) or $\boldsymbol{\beta} = \text{vec}(\llbracket \mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(M)} \rrbracket)$ (PARAFAC), and the process of fitting (3.28) can be referred to as tensor regression. Imposing such structure changes the implicit dimension of $\boldsymbol{\beta}$, which is forced to lie on a lower-dimensional manifold. Since the number of free parameters can be significantly reduced, the regression problem can be effectively overdetermined and well-posed again.

Regression is closely related to factor analysis and BSS. For example, decomposing an Mth order tensor when $M-1$ factor matrices and the core tensor are known (or alternatively, when the core is unknown, all M factor matrices) is equivalent to simple regression. This is precisely the principle behind the ALS algorithm for fitting tensor decompositions (Harshman et al., 1970; Carroll and Chang, 1970; De Lathauwer and Nion, 2008; Sidiropoulos et al., 2017). The same considerations with regard to e.g. regularization which exist for regression are indeed equally applicable to BSS problems.

Application in functional neuroimaging

Mass-univariate linear regression has been ubiquitous in fMRI analysis, nearly since the very inception of fMRI acquisition. Implemented as a general linear model (GLM) in many different open source software, it is the tool of choice in fMRI analysis, since the seminal papers of (Friston et al., 1994; Worsley and Friston, 1995), which fueled two and a half decades of experimental and fundamental research in the field. The GLM considers the time course recorded in each voxel as a dependent variable y, which is modeled as a superposition of effects of interest, such as interictal epileptic discharges or experimental stimuli, and nuisances, such as head motion, all of which are stored in a 'design matrix' (see Figure 3.4).

The ins and outs of the GLM and its pitfalls and strengths in the context of fMRI analysis are by now very well understood[14] (Monti, 2011; Poline and Brett, 2012). It owes its popularity to it relative simplicity and its extreme

[14] Albeit sometimes forgotten or ignored (Monti, 2011; Pernet, 2014).

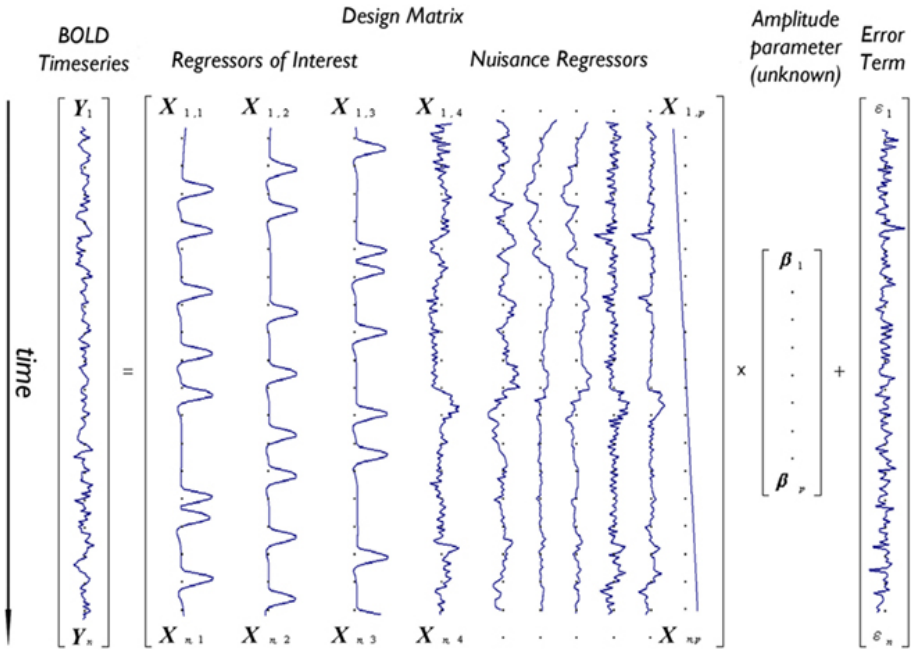


Figure 3.4: The general linear model (GLM) in fMRI models all voxels' time courses as the sum of contributions from effects of interests and nuisance time courses, weighted by coefficients which are estimated through regression. By fitting the model (separately) at every voxel, spatial maps of the brain activity of (no) interest can be constructed, and statistical hypotheses on the presence or absence of an effect can be tested in a principled manner. Figure taken from (Monti, 2011).

flexibility in various modeling efforts in diverse experimental paradigms (Poline and Brett, 2012). Parametric (Friston et al., 1994; Worsley and Friston, 1995; Friston et al., 2002) as well as nonparametric (Nichols and Holmes, 2002) methods for statistical significance testing of the regression coefficients have been built around the GLM, with sophisticated protection mechanisms against false positive[15] inferences (Worsley et al., 1992; Worsley, 1994; Nichols and

[15]The typical way of localizing a certain effect in the brain is by interrogating every voxel for the presence of the corresponding regressor, i.e. by testing whether its regression coefficient is significantly different from zero. Since fMRI data consist of many ($\sim 10^5$—10^6) voxels, this poses a massive multiple testing problem, in which the null hypothesis could be rejected at hundreds or thousands of voxels, even in the absence of any effect in the whole brain, just by sheer chance. Hence, safeguards are needed that can strictly control the false positive rate below a nominal level. The IgNobel neuroscience laureates of 2012 make a strong case for such corrections in their renowned 'dead salmon study' (Bennett et al., 2009).

Hayasaka, 2003).

EEG-correlated fMRI analysis through the GLM is one of the most common ways of fusing both modalities' data (Abreu et al., 2018b). It has been employed to correlate various time-varying EEG features (as columns of the design matrix) to corresponding BOLD time series in the brain, such as signal power in the alpha band (Laufs et al., 2003), IC time signatures (Debener et al., 2006), PARAFAC time signatures of an EEG spectrogram tensor (Marecek et al., 2016), and several other metrics (Rosa et al., 2010; Murta et al., 2015). Specifically for epilepsy monitoring, it is the de facto standard for localizing interictal epileptic discharges with EEG-correlated fMRI (Lemieux et al., 2001; Salek-Haddadi et al., 2003, 2006; van Houdt et al., 2013).

An alternative use of regression in neuroimaging is in denoising. When the time courses of artifacts are known (e.g. by estimation through ICA, or by means of reference measurements such as the electrooculogram or the electrocardiogram), the artifacts can be 'regressed out' of the data. I.e., the columns of $\mathbf{X}$ can be populated with these time courses, and after solving (3.29), the data of every variable can be adjusted as $\tilde{\mathbf{y}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$ to remove any linear contribution from the artifacts[16]. This strategy provides a slightly different result compared to componentwise elimination in ICA. It is very commonly applied in fMRI data, implicitly in the GLM, to 'explain away' data spurious variance that may be due to head motion, or other confounding effects that may otherwise misguide any statistical interpretation about effects of interest (Friston et al., 1994; Poline and Brett, 2012; Monti, 2011). The same approach is also used in EEG (Urigüen and Garcia-Zapirain, 2015; Acar et al., 2007).

Tensor regression has been applied to functional and anatomical MRI images by (Zhou et al., 2013) to detect brain areas related to disease outcome. Also the GLM has been used in conjunction with a rank-1 PARAFAC structure on the coefficients, in fMRI analyses with many experimental conditions and/or HRF basis functions (Pedregosa et al., 2015).

[16]This corresponds to a projection of the data onto the null space of the artifacts. I.e., this step can also be summarized as $\tilde{\mathbf{y}} = \mathbf{P}_{\mathbf{X}^\perp}\mathbf{y}$, with a projection matrix $\mathbf{P}_{\mathbf{X}^\perp} = \mathbf{I} - \mathbf{X}(\mathbf{X}^\mathrm{T}\mathbf{X})^{-1}\mathbf{X}^\mathrm{T}$.

Chapter 4

Improving localization of ictal onset zone in EEG-correlated fMRI via EEG enhancement

ABSTRACT | To improve the accuracy of detecting the ictal onset zone, we propose to enhance the epilepsy-related activity present in the EEG signals, before mapping their BOLD correlates through EEG-correlated fMRI analysis. Based solely on a segmentation of interictal epileptic discharges (IEDs) on the EEG, we train multi-channel Wiener filters (MWF) which enhance IED-like waveforms, and suppress background activity and noisy influences. Subsequently, we use EEG-correlated fMRI to find the brain regions in which the BOLD signal fluctuation corresponds to the filtered signals' time-varying power (after convolving with the hemodynamic response function), and validate the identified regions by quantitatively comparing them to ground-truth maps of the (resected or hypothesized) ictal onset zone. We validate the performance of this novel predictor versus that of commonly used unitary or power-weighted predictors and a recently introduced connectivity-based metric, on a cohort of twelve patients with refractory epilepsy. The findings of this study advocate for the use of the MWF to maximize the signal-to-noise ratio of IED-like events in the interictal EEG, and subsequently use time-varying power as a sensitive predictor of the BOLD signal, to localize the ictal onset zone.

This chapter is a modified version of Van Eyndhoven, S., Hunyadi, B., Dupont, P., Van Paesschen, W., & Van Huffel, S. "Semi-automated EEG enhancement improves localization of ictal onset zone with EEG-correlated fMRI", *Frontiers in Neurology*, 10:805, Aug. 2019, in which changes are made to abstract and layout.

4.1 Introduction

Epilepsy is one of the most common neurological disorders, affecting an estimated 50 million people worldwide (World Health Organization, 2019). Of this group, about 30% of patients suffers from refractory epilepsy, characterized by recurrent seizures that cannot sufficiently be suppressed by anti-epileptic drugs. Hence, these patients are candidates to undergo resective surgery, in which the epileptogenic zone is removed or disconnected from surrounding tissue, with the goal of preventing further seizure generation. However, the epileptogenic zone is a mere abstract concept, defined as the 'brain tissue whose removal / resection renders a patient seizure-free', and there exists no method that measures it (Rosenow and Lüders, 2001). Hence, during the presurgical planning, different imaging modalities provide incomplete, but complementary clues on its exact location and should be consulted by neurologists to detect and segment the target for surgery. In this paper, we focus on the simultaneous recording of electroencephalography (EEG) and functional magnetic resonance imaging (fMRI), which has been an ubiquitous technique in brain research, and epilepsy research in particular (Lemieux et al., 2001; Salek-Haddadi et al., 2006; Zijlmans et al., 2007).

The rationale behind simultaneously recording EEG–fMRI is the excellent temporal resolution of EEG (measuring neurophysiological changes on the order of milliseconds) and the fine spatial resolution of blood oxygen level dependent (BOLD) signals measured in fMRI (representing tissue oxygenation changes in local volumes of a few millimeter in diameter). Neuronal firing, or neural 'events' in general, consumes energy and thus induces metabolic demands, which must be accommodated by increased flow of oxygenated blood to the neural tissue. The temporal occurrence of such neural events can be tracked well using EEG, and by correlating these occurrences to the BOLD signals recorded at every voxel, it is possible to infer which brain regions participate in the generation or propagation of such events (Salek-Haddadi et al., 2003; Leite et al., 2013; Murta et al., 2015). In the context of refractory epilepsy, the events of interest are mostly interictal epileptic discharges (IEDs), as a surrogate of ictal periods (seizures), since they are subclinical and as such do not lead to e.g. movement-related artifacts in the EEG or fMRI data, as most types of seizures do (Rosenow and Lüders, 2001). Hence, this approach crucially relies on the assumption that the regions involved in interictal activity (named the irritative zone) coincide, or at least overlap, with the ictal onset zone (IOZ), which is the actual target for resective surgery. We have studied this hypothesis in (Tousseyn et al., 2015), where it was demonstrated that ictal networks correspond well to interictal networks, by comparing images from single-photon emission computed tomography (SPECT) and fMRI. Beyond the similarity at the network level,

also the site of onset of seizures and IEDs shows remarkable correspondence, based on evidence from intracranial EEG recordings (Hufnagel et al., 2000). The role of IEDs as precursors of seizures, by strengthening or sustaining specific axonal connections that make up the epileptic network, has been discussed in (Staley and Dudek, 2006; Rosenow and Lüders, 2001). Furthermore, there is evidence that the highest BOLD activations associated to IEDs, do indeed coincide with the seizure onset zone (Khoo et al., 2017, 2018). Backed up by this body of evidence, we may expect that IEDs provide a valid alternative to seizures when it comes to mapping of the epileptogenic zone.

The irritative zone or network can then be mapped by means of EEG-correlated fMRI analysis, as explained before. In this workflow, a time-varying representation of the IEDs is constructed, which is used to predict the BOLD time series through a model of the hemodynamic response. Voxels where the predicted time series correlates well to the measured BOLD time series are then considered to be part of the irritative network (Salek-Haddadi et al., 2003). The success of this approach hence depends on the accurate modelling of the neurovascular coupling and the timing of the IEDs. Many heuristic IED representations have been devised, to capture electrophysiological features that are suited to reveal concomitant BOLD changes. Many of these heuristic predictors rely on the manual delineation of IEDs by a neurologist, which is time-consuming and subjective. Recently, EEG synchronization measures were proposed in (Abreu et al., 2018a) to construct a predictor time course, based on the observation that epileptic activity is marked by excessive neuronal synchronization. The authors argue that existing predictors were mostly based on the amplitude of the EEG signals (see e.g. (Leite et al., 2013; Murta et al., 2015)), and as such are susceptible to artifacts and noise, and ignore important connectivity aspects in the EEG data.

In this paper we propose another paradigm to find predictors that are as informative as possible. Our approach is based on the enhancement of IED-like activity in the EEG signals, while suppressing irrelevant background activity, noise and artifacts as much as possible. Specifically, we employ a multi-channel Wiener filter as in (Somers et al., 2018) to maximize this contrast, after which we derive the predictor as the power in the filtered EEG signals. The goal of this paper is to evaluate the feasibility of this approach, and compare its performance to that of several other existing predictors, based on stick models of the IED (Lemieux et al., 2001), on IED amplitude (Rosa et al., 2010), global field synchronization (Jann et al., 2009) and the phase slope index (Mizuhara et al., 2005). To investigate the viability of simultaneous EEG–fMRI recording as one of the presurgical mapping techniques, it is important to assess performance by means of a very direct, clinical metric (van Houdt et al., 2013; Zijlmans et al., 2007; Thornton et al., 2010). Here, we accomplish this by studying a cohort

of refractory focal epilepsy patients for which ground truth information on the IOZ is available, which we described earlier in (Tousseyn et al., 2014a).

4.2 Methods and materials

4.2.1 Patient group

We used data of twelve patients with both left- and right-hemispheral refractory focal epilepsy, whom we previously studied in (Tousseyn et al., 2014a,b, 2015; Hunyadi et al., 2015). These patients were selected based on the following criteria (Tousseyn et al., 2014a): 1) adults who underwent a full presurgical evaluation for refractory focal epilepsy, including seizure history, neurological and physical examination, neuropsychological assessment, interictal and ictal scalp EEG recordings, video analysis of seizures, high-resolution MRI of the brain, and in most patients SISCOM and interictal ^{18}F-FDG PET. In selected cases, intracranial EEG-recordings were performed; 2) concordant data pointing to one epileptic focus using all available presurgical investigations (including a SISCOM) *or else* successful outcome after epilepsy surgery [international league against epilepsy (ILAE) outcome classification 1–3 (1, completely seizure-free; 2, only auras; 3, one to three seizure days per year ± auras; 4, four seizure days per year to 50% reduction of baseline seizure days ± auras; 5, <50% reduction of baseline seizure days to 100% increase of baseline seizure days ± auras; 6, more than 100% increase of baseline seizure days ± auras)]; 3) recording of interictal spikes during EEG–fMRI. Nine patients included in the study had some form of temporal lobe epilepsy (TLE). We present an overview of the clinical data of all included patients in Table 4.2. Note that patients 2 and 6 had a poor outcome after surgery (ILAE classification 4 and 5, respectively), which was likely the result of incomplete surgery, constrained by the cortical location of the IOZ. In the former case, part of the dysplastic lesion was located in the motor cortex, and was not resected. In the latter case, the dysplastic lesion was located in the primary motor cortex, and the surgeon did not remove all cortical tissue constituting the lesion. Therefore, an investigation of the concordance of EEG–fMRI with the other modalities' evidence is still considered meaningful in these cases. To be able to assess the specificity of the evaluated methods (as in (Tousseyn et al., 2014a)), we incorporated fMRI recordings of twelve healthy controls, which were age- and gender-matched to the patients. We discuss the importance of the control group in more detail in subsection 4.2.6. This study was carried out in accordance with the recommendations of the International Conference on Harmonization guidelines on Good Clinical Practice with written informed consent from all subjects. All subjects gave written informed consent

in accordance with the Declaration of Helsinki. The protocol was approved by the Medical Ethics Committee of the University Hospitals KU Leuven.

Table 4.1: List of the number of IEDs that were observed during the fMRI recording of every patient, as well as the degrees of freedom of the fMRI design matrix.

patient	number of fMRI scans	# IEDs during fMRI recording	IED rate (# IEDs per hour) during fMRI	degrees of freedom (DoF) of fMRI design matrix
1	540	15	40	509
2	1620	663	589	1512
3	1080	105	156	1008
4	1620	825	733	1527
5	1080	117	156	1015
6	1080	640	853	1009
7	1080	126	187	1012
8	1080	11	15	1008
9	1620	1815	1613	1530
10	540	226	602	509
11	1080	6	8	1018
12	1350	966	1171	1265

Table 4.2: Clinical patient data. Abbreviations: F = female, M = male, L = left, R = right, CNS = central nervous system, DNET = dysembryoplastic neuroepihelial tumor, FCD = focal cortical dysplasia, HS = hippocampal sclerosis, cx = cortex.

patient	gender	ictal onset zone	etiology	surgery	ILAE outcome	follow-up time after surgery (y)
1	F	L temporal	HS	temporal lobe resection	3	5
2	F	L parietal	FCD	partial lesionectomy	4	5
3	F	R parieto-occipito-temporal	Sturge-Weber			
4	M	R temporal	unknown			
5	F	L anterior temporal	HS	temporal lobe resection	1	8
6	F	R frontal	FCD	partial lesionectomy	5	2
7	F	L anterior temporal	DNET	temporal lobe resection	1	4
8	M	L temporo-parietal	unknown	overlap eloquent cx		
9	F	L occipital	FCD	overlap eloquent cx		
10	F	R temporal	HS	refused		
11	M	L anterior temporal	HS	temporal lobe resection	1	6
12	F	R temporal	CNS infection	refused		

4.2.2 Data acquisition and preprocessing

Functional MRI data were acquired on one of two 3T MR scanners (Achieva TX with a 32-channel head coil and Intera Achieva with an eight-channel head coil, Philips Medical Systems, Best, The Netherlands) with an echo time (TE) of 33 ms, a repetition time (TR) of either 2.2 or 2.5 seconds, and a voxel size of $2.6 \times 3 \times 2.6$ mm^3. EEG data were recorded using MR-compatible caps with 30 to 64 electrodes, sampled at 5 kHz. The fMRI images were realigned, slice-time corrected and normalized to MNI space, resampled to a voxel size of $2 \times 2 \times 2$ mm^3, and smoothed using a Gaussian kernel of 6 mm full width at half maximum (FWHM). All fMRI processing steps were carried out using SPM8 in MATLAB 2015a. The EEG signals were band-pass filtered offline between 1–50 Hz, gradient artifacts were removed and pulse artifacts were subtracted. The signals of all channels were normalized to have unit variance. We refer the reader to (Tousseyn et al., 2014a) for a detailed description of all preprocessing steps. Two neurologists subsequently inspected and annotated the EEG signals for IEDs. Table 4.1 presents the number of annotated IEDs and the duration of the fMRI recording for every subject, along with the number of degrees of freedom for the T-test that are carried out to map the interictal discharges (see section 4.2.4).

4.2.3 EEG-derived BOLD predictors

In this paper, we aim to improve the model of interictal electrophysiological variations which are neurovascularly coupled to BOLD fluctuations. To this end, we focus on the derivation of suitable IED features from the EEG, which serve as *predictors* for the simultaneously recorded BOLD time series, after convolution with a hemodynamic response function (see subsection 4.2.4). To model the EEG-based IED correlate, several time-varying representations have been conceived and used in prior studies (Rosa et al., 2010; Murta et al., 2015; Leite et al., 2013; Abreu et al., 2018a). We discuss some of the most significant heuristics below, without attempting to be exhaustive. It is important to remark that all of the mentioned approaches, besides the Global Field Synchronization (GFS) predictor, in one way or another, require the prior labeling of IEDs on the EEG (either by a neurologist or by an automatic detection algorithm). In our study, all IEDs were manually annotated by a neurologist, and discussed with a second neurologist.

Unitary predictors (Un)

The simplest predictor is constructed by creating a time course with a stick function of unit amplitude at the time of every detected IED, and zeros everywhere else (Lemieux et al., 2001; van Houdt et al., 2013). This 'unitary' (Un) representation treats every IED occurrence in the same way, i.e. neglects differences in amplitude or duration (Zijlmans et al., 2007).

Time-varying power of independent components (ICpow)

As a first logical alternative to the discrete unitary predictors, power-based predictors have been proposed. Different varieties exist, but they all share the continuous aspect since they are computed from the time-varying power of the EEG signal. We employ the approach used in (Abreu et al., 2018a) to allow fair comparison. First, independent component analysis (ICA) is performed on the M-channel EEG data $\mathbf{x}[t] \in \mathbb{R}^M$ at time t:

$$\mathbf{x}[t] = \hat{\mathbf{A}}\hat{\mathbf{s}}[t] \tag{4.1}$$

$$\hat{\mathbf{s}}[t] = \hat{\mathbf{A}}^{-1}\mathbf{x}[t] \tag{4.2}$$

$$\triangleq \hat{\mathbf{W}}^{\mathrm{T}} T\mathbf{x}[t] \tag{4.3}$$

where $\hat{\mathbf{A}} \in \mathbb{R}^{M\times M}$ is a mixing matrix, describing the contribution of every estimated source or independent component (IC) in $\hat{\mathbf{s}} \in \mathbb{R}^M$ to every channel of $\mathbf{x}$. Conversely, the estimated sources can be written as the output of M spatial filters in the columns of an unmixing matrix $\hat{\mathbf{W}}$, applied to the data. From the M ICs, the one which is most related to the IEDs is then kept, according to the PROJIC procedure in (Abreu et al., 2016b), which we summarize here. The EEG data epochs around the times of user-annotated IEDs are averaged, leading to a template IED over M channels and T_w time points, represented in a matrix $\mathbf{Z} \in \mathbb{R}^{M\times T_w}$. Subsequently, the average IED is projected onto the IC space as $\mathbf{P} = \hat{\mathbf{W}}^{\mathrm{T}}\mathbf{Z}$. Under the assumption that IED-related ICs attain a higher power in $\mathbf{P}$ than other, non-IED-related ICs, we can then select one of the IC, namely the one which contributes the most signal power to the average IED, computed as the sum of squares of every row of $\mathbf{P}$. This IC is then considered an estimated time course of interictal activity. Then, a time-frequency decomposition with Morlet wavelets is used to find the spectrogram of the component, and the time-varying power is finally computed by averaging the spectrogram in the band 1-45 Hz (Rosa et al., 2010; Leite et al., 2013). We will use the name *IC*

time-varying power (ICPow) for the final predictor.

EEG Synchronization (PSI and GFS)

The authors of (Abreu et al., 2018a) have prompted the idea to focus on the hypersynchronization associated to epileptic events, and quantify it by means of a time-varying metric of coupling between EEG channels. In their paper, they investigate the Phase Slope Index (PSI) and Global Field Synchronization (GFS) and evaluate their performance versus that of Un and `ICpow` predictors. By comparing the obtained statistical maps to electrical source imaging maps of the delineated IEDs, they concluded that the PSI metric, computed in a narrow band between 3–10 Hz, yielded the best correspondence. Hence, in this paper, we will employ this metric for comparison. The PSI is computed as in (Abreu et al., 2018a; Mizuhara et al., 2005) between a particular pair of EEG channels. Firstly, the PROJIC algorithm is applied, followed by a K-means clustering to detect a set of IED-related ICs, that are used to reconstruct the data while leaving out non-IED-related ICs. The PSI is then computed between every pair of channels. Dimensionality reduction is performed by selecting the channel with the maximal amplitude of the average IED, and subsequently choosing a second channel as the one for which the PSI time course with the first channel had maximal variance. This latter time course was then chosen as the PSI predictor. For completeness, we also include the GFS metric in the same frequency band, computed as in (Jann et al., 2009). The GFS measures, for every time bin, to which degree the signals of all channels in the data are in or out of phase with each other. Contrary to the PSI, which is computed for a specific pair of channels, the GFS is a 'global' metric, in the sense that synchronization between all channels equally is considered.

Time-varying power of enhanced EEG (MWFpow): a new predictor

We propose to use the available information, in the form of annotated IEDs, to enhance the EEG signal prior to extracting a predictor. Ideally, we would like to retain all the relevant, IED-like activity in the EEG data, and discard all other waveforms, such as noise, artifacts, and background activity. We approach this task by means of the multi-channel Wiener filter (MWF). The MWF has long been known in audio signal processing to enhance acoustic sources, and has recently also been demonstrated to be a powerful tool for generic artifact removal in EEG signals (Somers et al., 2018). In the latter paper, the authors have shown the excellent capabilities of the MWF to separate various types of

artifacts from the EEG by means of spatio-temporal filtering. In this paper, we will use it in its dual form, i.e., not to filter out a particular type of activity of interest, but to enhance it. In the current setting we consider M-channel EEG signals $\mathbf{x}[t] \in \mathbb{R}^M$ at time t (notation as in (Somers et al., 2018)):

$$\mathbf{x}[t] = \mathbf{d}[t] + \mathbf{n}[t] \,, \tag{4.4}$$

in which $\mathbf{d}[t] \in \mathbb{R}^M$ represents the IEDs, which are superimposed on the background EEG fluctuations, including artifacts and noise, represented by $\mathbf{n}[t] \in \mathbb{R}^M$. In its basic form, the multi-channel Wiener filter acts as a spatial filter $\mathbf{W}$ that maximizes the objective function

$$\min_{\mathbf{W}} \mathrm{E}\{\|\mathbf{d} - \mathbf{W}^{\mathrm{T}}\mathbf{x}\|^2\} \tag{4.5}$$

and as such tries to extract $\mathbf{d}[t]$ as well as possible (in mean-squared-error sense) from the observations $\mathbf{x}[t]$. The solution to this problem is

$$\mathbf{W} = \mathbf{R}_{xx}^{-1}\mathbf{R}_{dd} \,, \tag{4.6}$$

in which $\mathbf{R}_{xx} = \mathrm{E}\{\mathbf{x}\mathbf{x}^{\mathrm{T}}\}$ is the covariance matrix of the observed signals, and $\mathbf{R}_{dd} = \mathrm{E}\{\mathbf{x}\mathbf{d}^{\mathrm{T}}\}$ is the covariance matrix of the desired signals. If $\mathbf{d}$ and $\mathbf{n}$ can be assumed uncorrelated, then $\mathbf{R}_{xx} = \mathbf{R}_{dd} + \mathbf{R}_{nn}$. By splitting the observations of $\mathbf{x}[t]$ into two subsets - time instants $t \in \mathcal{C}_1$ that cover the occurrences of IEDs, and time instants $t \in \mathcal{C}_0$ in which no IEDs occur - the latter matrices can be estimated via

$$\hat{\mathbf{R}}_{xx} = \frac{1}{|\mathcal{C}_1|} \sum_{t \in \mathcal{C}_1} \mathbf{x}[t]\mathbf{x}[t]^{\mathrm{T}} \tag{4.7}$$

$$\hat{\mathbf{R}}_{nn} = \frac{1}{|\mathcal{C}_0|} \sum_{t \in \mathcal{C}_0} \mathbf{x}[t]\mathbf{x}[t]^{\mathrm{T}} \tag{4.8}$$

$$\hat{\mathbf{R}}_{dd} = \hat{\mathbf{R}}_{xx} - \hat{\mathbf{R}}_{nn} \tag{4.9}$$

Hence, the multi-channel Wiener filter can be found as $\hat{\mathbf{W}} = \hat{\mathbf{R}}_{xx}^{-1}\hat{\mathbf{R}}_{dd}$ and applied to the observed signals to obtain an estimate of clean IED time courses $\hat{\mathbf{d}} = \hat{\mathbf{W}}^{\mathrm{T}}\mathbf{x}$. The MWF can be extended from a purely spatial filter to a spatiotemporal filter, in which not only samples from other channels are used to improve the signal estimation at a particular channel, but also samples from different time instants. In this case, observations are represented in long vectors $\tilde{\mathbf{x}}[t]$ that are constructed by stacking lagged versions of the data $\mathbf{x}[t]$ at time t:

$$\tilde{\mathbf{x}}[t] = \begin{bmatrix} \mathbf{x}[t-\tau] \\ \vdots \\ \mathbf{x}[t-1] \\ \mathbf{x}[t] \\ \mathbf{x}[t+1] \\ \vdots \\ \mathbf{x}[t+\tau] \end{bmatrix}, \tag{4.10}$$

which is followed by the computation of the covariance matrices and the resulting filter as before. This generalization allows finite-impulse-response (FIR) filtering at every channel, and thereby the exploitation of spectral differences between the IEDs and the background EEG. The extra degrees of freedom are described by the number of lags τ, which can be chosen high if enough training data are available. The crucial ingredients for the MWF are two covariance matrices: one covariance matrix $\hat{\mathbf{R}}_{xx} \in \mathbb{R}^{(2M\tau+1)\times(2M\tau+1)}$ of the signal of interest embedded in background EEG (which can be estimated from the EEG during the IED periods) and one covariance matrix $\hat{\mathbf{R}}_{nn} \in \mathbb{R}^{(2M\tau+1)\times(2M\tau+1)}$ of the background EEG itself (which can be estimated from the data during all other periods). Note that these quantities may be estimated over the full length of the EEG recording, or alternatively, over a portion of the recording only[1]. In the latter case, if IED annotation is performed manually, a neurologist may save substantial time by marking IED and non-IED periods in a smaller part of the signal, if these periods can be considered sufficiently representative for the whole signal (Somers et al., 2018).

Hence, the MWF will aim to enhance the waveforms in the EEG with the same covariance structure as the signal of interest, while suppressing waveforms with a covariance structure similar to the signal which was marked as not of interest. The output signals $\hat{\mathbf{d}}[t]$ will then ideally only reconstruct the waveforms at the times of true IEDs, and contain near-zero values everywhere else. Hence, when computing the time-varying power of these filtered signals, and averaging over all channels, a new predictor of IED occurrence is obtained, which we name *multi-channel Wiener filtered EEG time-varying power (MWFpow)*. The advantages of such an approach are that 1) the MWF can find additional IEDs, if they were missed by accident, or were obfuscated by superimposed artifacts, 2) the MWF might correct for erroneous IED annotations, if they indeed follow a difference covariance structure, 3) the MWF allows for a continuous representation in the

[1]Since the inverse of $\hat{\mathbf{R}}_{xx}$ is used in the computation of the filter in (4.5), sufficiently many EEG segments should be marked to ensure that this matrix is full-rank and hence invertible. Conversely, $\hat{\mathbf{R}}_{dd}$ may be rank-deficient, e.g. in case of very stationary IEDs (Somers et al., 2018).

output, differentiating IEDs by amplitude and duration, compared to discrete metrics. At low signal-to-noise ratios, the imperfect estimation of the covariance matrices can lead to 'over-subtraction' in (4.9): while $\hat{\mathbf{R}}_{dd}$ is expected to be a proper, positive semi-definite matrix, the subtraction of two other estimated covariances is not guaranteed to preserve this property. In order to ensure the positive semi-definiteness of $\hat{\mathbf{R}}_{dd}$, one can simultaneously diagonalize $\hat{\mathbf{R}}_{xx}$ and $\hat{\mathbf{R}}_{nn}$ and perform the subtraction of the two diagonal matrices $\mathbf{D}_{xx}$ and $\mathbf{D}_{nn}$ in the common basis, setting negative entries to zero:

$$\hat{\mathbf{R}}_{xx} = \mathbf{V}\mathbf{D}_{xx}\mathbf{V}^{\mathrm{T}} \tag{4.11}$$

$$\hat{\mathbf{R}}_{nn} = \mathbf{V}\mathbf{D}_{nn}\mathbf{V}^{\mathrm{T}} \tag{4.12}$$

$$[\hat{\mathbf{D}}_{\Delta}]_{ii} = \max\{[\mathbf{D}_{xx}]_{ii} - [\mathbf{D}_{nn}]_{ii}, 0\}\ , \forall i = 1 \dots 2M\tau + 1 \tag{4.13}$$

$$\hat{\mathbf{R}}_{dd} = \mathbf{V}\mathbf{D}_{\Delta}\mathbf{V}^{\mathrm{T}} \tag{4.14}$$

In (4.13), $[D]_{ii}$ denotes the element at position (i, i) of a matrix D, i.e., the i-th diagonal element of D. The simultaneous diagonalization in (4.11) and (4.12) can be accomplished via a generalized eigenvalue decomposition (GEVD) of matrices $\hat{\mathbf{R}}_{xx}$ and $\hat{\mathbf{R}}_{nn}$. In the remainder of the paper, we will use GEVD-MWF, as it leads to robust enhancement (Somers et al., 2018).

4.2.4 EEG-correlated fMRI analysis

To map the network of brain regions involved in IED generation and propagation, we relied on the general linear model (GLM) for fMRI (Friston et al., 1994). We conduct a separate analysis per patient/predictor pair. After convolving an extracted predictor with a model of the hemodynamic response function, we obtain a regressor time series to model the IED-related contribution to every voxel's BOLD signal. As covariates of no interest, we include the six motion-correction parameters, and the average time series in the white matter and the lateral ventricles (cerebrospinal fluid). If necessary, also boxcar regressors are added at moments of substantial scan-to-scan head movement (larger than 1 mm based on the translation parameters). We then calculated statistical T-maps for every IED-predictor separately, interrogating every voxel for its participation in IED generation or propagation. We hypothesized that predictors offer differential modeling power, which manifests itself as the (in)ability to correctly localize the IOZ due to a (mis)match between the modeled and measured BOLD signal. As in (Tousseyn et al., 2014a), we evaluated performance metrics for different statistical maps using different threshold criteria. More precisely, we applied

a transformation of the T-map to Z-scores and used different Z-thresholds, familywise error (FWE) correction, and a constraint on the cluster size k for active regions. In (Tousseyn et al., 2014a), we have argued that practitioners are to use the threshold of maximal specificity, i.e., producing the fewest false activations (see 4.2.6). After evaluating a wide range of thresholds in terms of their specificity and sensitivity in localizing the IOZ, we found that a threshold of $Z > 3.4$, combined with a constraint on the cluster size of $k > 350$ voxels, yielded the highest sensitivity (62%) under the condition of maximal specificity (100%)[2]. Aside from this 'optimal' setting, we also evaluate two thresholds that are common in literature: uncorrected $p < 0.001$, and FWE-corrected $p < 0.05$. Lastly, we evaluate a more conservative threshold, i.e. FWE-corrected $p < 0.05$, combined with a cluster constraint of $k > 350$ voxels. In summary, we examined the model's performance at the following thresholds:

1. $p < 0.001$, uncorrected for multiple comparisons, no constraint on cluster size ($Z > 3.1$, $k > 0$)
2. $p < 0.05$, FWE-corrected, no constraint on cluster size ($FWE < 0.05$, $k > 0$)
3. $p < 3.37 \times 10^{-4}$, uncorrected for multiple comparisons, cluster larger than 350 voxels (2.8 cm^3) ($Z > 3.4$, $k > 350$)
4. $p < 0.05$, FWE-corrected, cluster larger than 350 voxels (2.8 cm^3) ($FWE < 0.05$, $k > 0$)

4.2.5 Ictal onset zone

We compared the statistical maps with a ground truth image of the IOZ to assess its predictive power, as in (Tousseyn et al., 2014a). For patients that had undergone surgical treatment with a successful outcome (ILAE classification 1–3), the manually delineated resection zone was considered as IOZ. In other patients (who were ineligible for or refused surgery), the hypothetical resection zone, based on concordant evidence from multiple modalities other than EEG–fMRI was taken as the best possible estimate of the IOZ.

[2]Under the current interpretation of the performance metrics, a predictor may be made arbitrarily sensitive by lowering the threshold appropriately, thereby (often dramatically) sacrificing specificity.

4.2.6 Model performance

To quantitatively assess the performance of the models constructed from different IED predictors, we evaluate sensitivity and specificity. Sensitivity is defined as the fraction of 'true positives', i.e., the percentage of patients for which the statistical map, obtained for a certain model, overlaps with the IOZ (Thornton et al., 2010). Following the reasoning in (Tousseyn et al., 2014a), we do not consider significantly active voxels or regions outside of the delineated IOZ as false positives. Acknowledging epilepsy as a network disorder, such active regions might reflect seizure or IED propagation, despite not being involved in their generation. For this reason, we use the fMRI data from healthy control subjects to calculate the specificity, i.e., the ratio of false positives. Re-using the IED-predictor from the matched patients, we computed T-maps for the twelve control subjects, and considered control maps with significantly active regions as false positive cases. Analogously, true negative cases are healthy controls for which the statistical map had no suprathreshold regions, and false negative cases were those patients for which no overlap with the IOZ was found. To further distinguish the performance obtained with different IED predictors, we also inspected the cumulative statistical evidence that they produced: we summarized the *T-evidence* as the average T-value in a predefined set of voxels. A high value for the T-evidence within the IOZ is then an indication of a good temporal model of the BOLD signal, which is capable of distinguishing IOZ voxels. Note that this metric allows us to rank the performance of different regressors that correctly infer the IOZ (i.e., produce suprathreshold voxels in the IOZ), but even allows to rank the performance of non-sensitive regressors. Hence even when regressors do not produce suprathreshold voxels in the IOZ, they can still be more useful than others if they manage to lift the statistical values, and hence 'hint' towards certain regions. Lastly, we tracked the maximal T-value within the IOZ for every IED-predictor, as a measure of modeling power. Following the 'network disorder' argumentation, we prefer such a metric over other existing metrics such as the Dice coefficient, since these penalize active voxels out of the IOZ, perhaps wrongfully.

4.3 Results for patients and matched controls

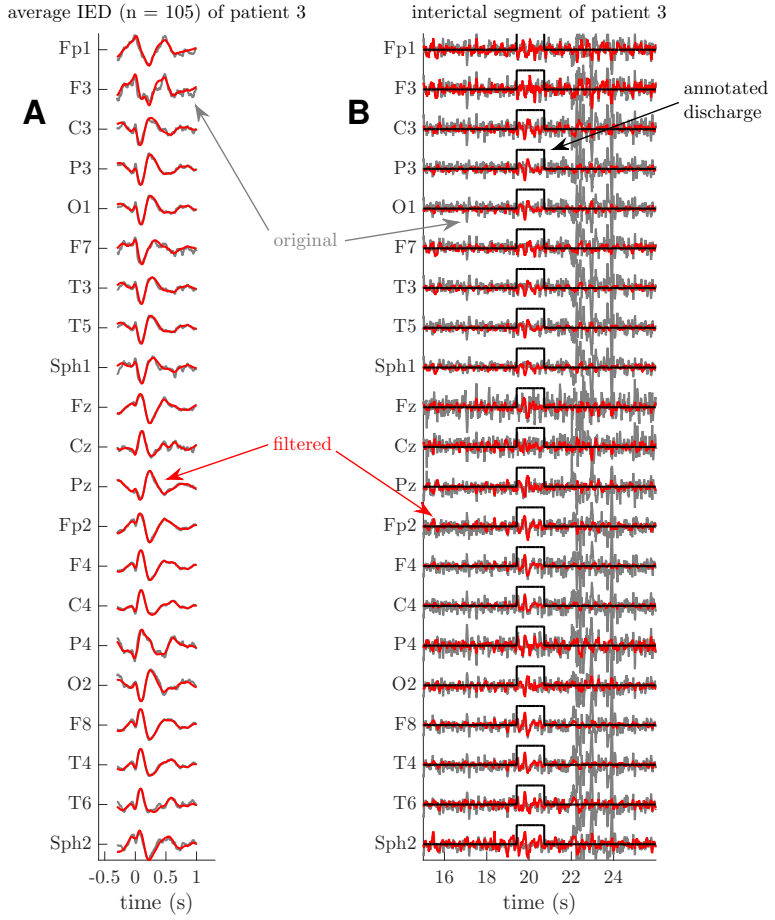
4.3.1 Multi-Channel Wiener filtering successfully enhances IEDs

We illustrate that the MWF succeeds in enhancing the IEDs in the EEG, while suppressing irrelevant (background) EEG activity. In Figure 4.1(a), we show

the average waveform in the annotated IED segments of patient 3, before and after applying the MWF. The excellent correspondence of these two sets of signals on all EEG channels confirms that the MWF leaves – to a large extent – the IEDs intact, respecting the a priori information that was injected by means of the neurologist's annotations. On the other hand, artifacts, such as the one shown in Figure 4.1(b) around 22...26 s, are strongly attenuated in the output of the MWF. This example demonstrates that EEG activity that does not follow the spatiotemporal covariance structure $\hat{\mathbf{R}}_{xx}$ of the IEDs, that was used to calibrate the MWF, is suppressed in the output of the MWF, as desired. Hence, the MWF performs *signal enhancement* (of the annotated IEDs) and *noise reduction*[3], improving the signal-to-noise ratio of IEDs.

As as second example, we show in Figure 4.2 excerpts of the EEG signals of patient 10, before and after filtering. In Figure 4.2(a), it is again clear that the MWF appropriately reconstructs the waveforms of most annotated IED segments. Secondly, we indicate another effect of applying the MWF in Figure 4.2(b). The annotated IED in this portion of the EEG corresponds only partially to the average IED shown in Figure 4.2(a), and hence the signal during the annotation period is also attenuated on several channels. While the waveforms on other channels are preserved when they display a higher similarity to the average IED, this shows that a mild errors during annotation may be rectified by the MWF. E.g., a neurologist may mistakenly annotate some dubious EEG segments as IEDs, while they are not. When these 'false positive' IEDs are sufficiently limited in number, they do not contaminate the covariance $\hat{\mathbf{R}}_{xx}$ in (4.7) significantly, and can probably be suppressed sufficiently by the MWF. At the same time, 'false negative' annotations are conceivable, i.e., when the neurologist overlooks one or more IEDs when inspecting the EEG. The resulting absence of such IEDs in e.g. the `Un` predictor may lead to an underfitting of the BOLD signal and subsequently to a decreased sensitivity when mapping the ictal onset zone. In Figure 4.2(b) we identify such a case. Near the beginning of the shown segment, between -3...-2 s relative to the annotated IED, the MWF output signal is enhanced on nearly all channels, and waveforms which are very similar to those of the average IED can be discerned. This strongly suggests that this segment contains a true IED, which was missed during annotation, but was recognized by the MWF to follow the same spatiotemporal signal characteristics as the other annotated IEDs. Several such segments were found for this patient, and we note that the `MWFpow` predictor was sensitive for this patient, contrary to the `Un` predictor (see infra), which reinforces the hypothesis that the MWF

[3]In this context, with the term 'noise', we mean all waveforms or activity in the EEG that are of no interest to the task of mapping the ictal onset zone. This includes various types of artifacts, oscillatory activity, interference, which are unrelated to interictal epileptiform activity. Conversely, all IEDs jointly constitute the 'signal'.



(a) Average of all (105) interictal discharges of patient 3, before and after applying the MWF

(b) The MWF suppresses a large artifact, while preserving the annotated discharge

Figure 4.1: Example EEG segments of patient 3 illustrate the artifact removal capabilities of the Multi-channel Wiener filter (MWF). Time is indicated in seconds relative to the first spike of the discharge. (a) The MWF properly preserves the signal during annotated periods of interictal discharges, which is evidenced by the strong similarity between the average of the 105 original discharges, and the average of their filtered versions. (b) The MWF successfully attenuates the large artifact, which occurs between 22 and 26 s, because the waveforms do not follow the spatio-temporal statistics of the IEDs, which were used to calibrate the MWF. At the same time, the signals during the discharge around 20 s, which was annotated by the neurologist, stay preserved in the MWF output (compare to the average discharge in (a)).

indeed succeeds in 'discovering' previously undetected IEDs. Hence, the MWF also has *error-correcting capabilities*.

For all data, we assigned all samples within a window of 0.5 s before until 1.0 s after the first peak of any IED to $\mathcal{C}_1$, and subsequently used GEVD-MWF as described by equations (4.11)-(4.14). We chose $\tau = 4$ (corresponding to lags between -16 ms and +16 ms, in increments of 4 ms), since we observed no significant changes to the MWF output signals for higher values of the parameter (i.e. the output signal-to-noise ratio saturates) and wanted to ensure well-conditioned covariance estimation.

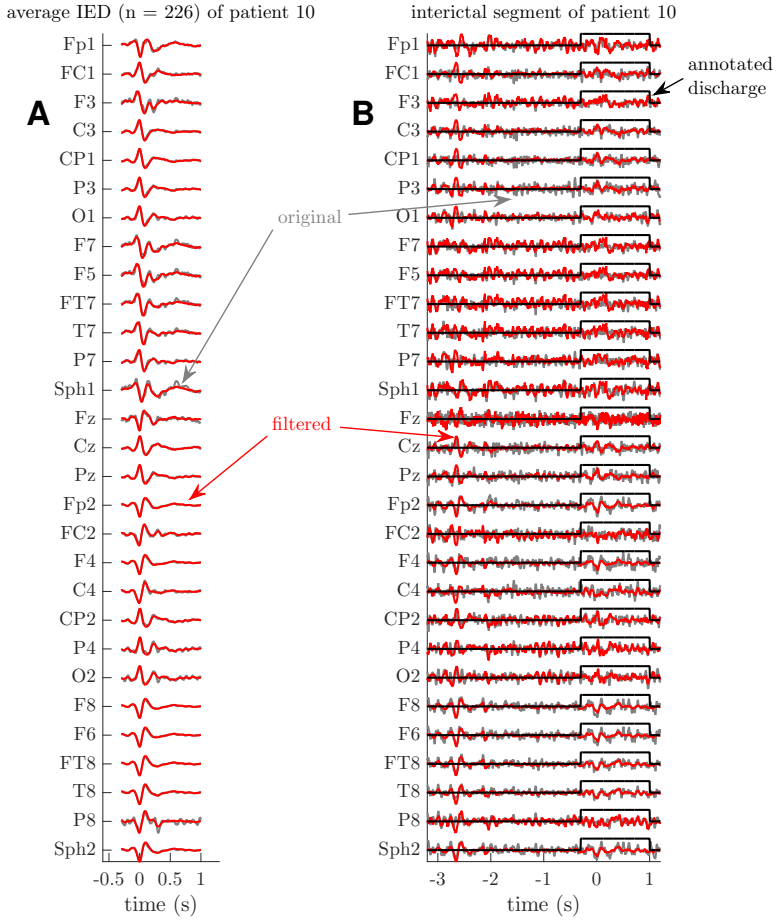
4.3.2 Sensitivity and specificity of different predictors

We found that the choice of predictor largely influences the sensitivity of IOZ detection. Figure 4.3 supports this observation by highlighting true positive detections with filled green markers, and false negative detections with empty red markers, at various combinations of predictor and statistical threshold. To compare different predictors beyond their binary detection performance, we plot the T-evidence within the IOZ[4] for every patient (cfr. section 4.2.6). Note that the specificity cannot be determined from the figure, but followed a simple pattern in our results: at the lowest threshold, all predictors were 0% specific, and at the three higher thresholds, all of them were 100% specific, except for one false positive case produced by `ICApow` at FWE-corrected $p < 0.05$.

At the 'optimal' (according to (Tousseyn et al., 2014a)) threshold of $Z > 3.4$ with cluster size constraint, the `MWFpow` predictor achieves the highest sensitivity (92%), outperforming `ICApow` (83%), `Un` (75%), `PSI` (25%) and `GFS` (0%). At this significance level, all of the predictors led to a specificity of 100%. At FWE-corrected $p < 0.05$, the `Un` predictor is 100% sensitive, better than the `MWFpow` and `ICApow` predictors. At this threshold however, the latter produced one false positive (92% specific, not shown in figure). At the lowest significance level, the sensitivity of most predictors (except `GFS`) is very good, yet all are absolutely non-specific (0%). At the other side of the spectrum, the `MWFpow` predictor was most robust over thresholds, still correctly detecting 75% of IOZs at the highest threshold, whereas the sensitivity of other predictors decreased to 42% and lower.

These results prove the viability of `MWFpow` as a new model in EEG-correlated fMRI analysis for (refractory) epilepsy. In general, the two synchronization

[4]The T-evidence is computed as the average value of unthresholded T-scores in the IOZ, and as such is independent of the imposed threshold.



(a) Average of all (226) interictal discharges of patient 10, before and after applying the MWF

(b) The MWF attenuates an annotated discharge, and finds a new discharge in an unannotated region

Figure 4.2: Example EEG segments of patient 10 showcase the error-correcting capabilities of the Multi-channel Wiener filter (MWF). Time is indicated in seconds relative to the first spike of the discharge. (a) Just as in Figure 4.1, the MWF preserves the signal during annotated periods of interictal discharges, respecting to large extent the annotations that were made by the human annotator. (b) Sometimes, annotated discharges are (partially) attenuated, if the instance does not correspond well to the average statistics that have been used to train the MWF. In this case, this effect is visible in the window of 1.5 seconds around an annotated discharge, on channels O2, C4, T7, ... On the other hand, discharges that were missed by the human annotator may be uncovered by the MWF by enhancing waveforms in the output. In this example, the signal is enhanced between -3...-2 s relative to the annotated spike. Indeed, the waveforms in this period show striking correspondences to the average of the annotated discharges of this patient (in (a)).

metrics performed poorly: for no threshold did GFS reach an acceptable balance of sensitivity/specificity, and PSI was not sensitive in general. We noticed that the case which was wrongfully delineated by MWFpow at the three highest thresholds (the 11th) in the figure, corresponded to a data recording in which particularly few (6) IEDs were observed, compared to more than hundred IEDs in many other recordings. We hypothesize that the low number of IEDs resulted in a 'data deficiency' when training the MWF: with few data segments from one class, it is likely that a good spatio-temporal filter cannot be reliably determined. We investigated the filtered EEG data of this recording and found indeed several remaining, high-amplitude artifacts in the data. After manually masking out these artifact periods (by setting the samples to zero), the resulting MWFpow predictor was sensitive for this case as well (not shown). We therefore advocate for a 'quick and dirty' inspection of high-amplitude periods in the filtered EEG, in order to manually correct artifact segments where needed, especially for datasets with few IEDs.

Besides binary metrics (successful or unsuccessful localization of the IOZ), we also point out the difference in T-evidence among different predictors. The MWFpow predictor is capable of identifying the highest statistical evidence in the ictal onset zone (IOZ), for many patients, as supported by Figure 4.3. The Un and ICApow predictors produce less evidence, and the synchronization metrics underperform with low evidence levels. This corroborates the higher sensitivity of the MWFpow predictor (and also the ICApow predictor) for interictal events, especially at higher (more selective) thresholds, where synchronization regressors (PSI and GFS) fail. Hence, the T-evidence metric confirms that the MWFpow predictor is a robust predictor which allows to find the IOZ with higher statistical significance than the other predictors.

To further investigate the diagnostic relevance of the different regressors, we considered the distribution of activation clusters in the brain. We therefore evaluated the sensitivity in two different ways: 1) by counting a true positive when any activation cluster overlaps with the IOZ or 2) by only counting a true positive when the cluster containing the voxel with the highest T-value overlaps with the IOZ. Evidently, the sensitivity of the maximally active cluster is always upper-bounded by the sensitivity of all clusters. We show these two types of sensitivity, evaluated at the threshold of $Z > 3.4$ with cluster size constraint $k > 2.8$ cm^3, in Figure 4.4. For the alternative type of sensitivity, a similar pattern as before emerges: the performance of the MWFpow is still the highest among all regressors. The interpretation of this figure is that a reader would be able to correctly identify the IOZ in 75% of cases by picking the maximally active cluster after EEG-correlated fMRI analysis using the MWFpow regressors, and with a lower success rate for other types of regressors.

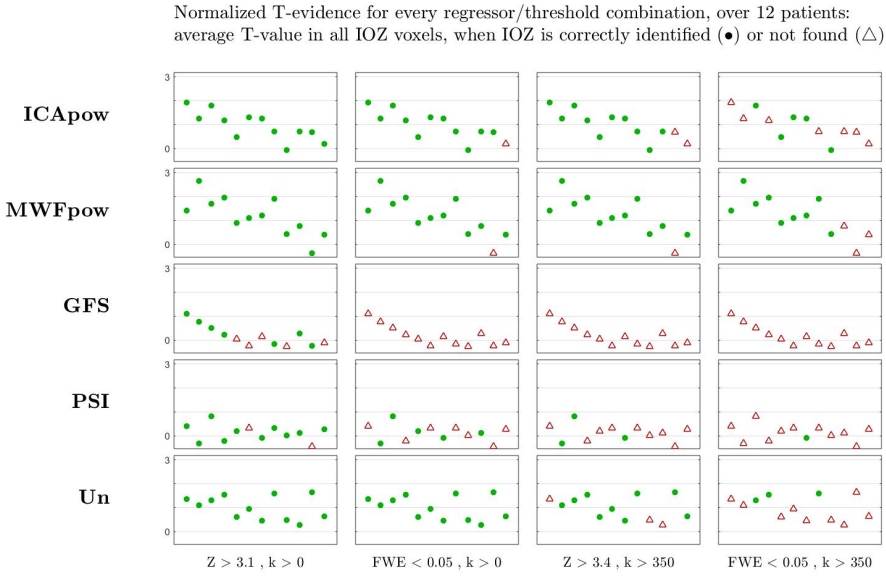


Figure 4.3: The statistical evidence found in the ictal onset zone is highest for especially the `MWFpow` and also `ICApow` regressors, and lower for the other regressor types, for most patients (shown by individual markers). This is correlated with a higher sensitivity (i.e., more true positives) for these two regressors, especially at higher (more selective) thresholds to the right of the figure, where synchronization regressors (`PSI` and `GFS`) and the unitary regressor `Un` fail. At different thresholds, cases with clusters of suprathreshold T-values that overlap with the ictal onset zone are counted as true positives (•), other cases are considered false negatives (△). At the lowest threshold ($Z > 3.1$, $k > 0$), all regressors were 0% specific (i.e., produced suprathreshold activations in all healthy controls), and at the three higher thresholds, all of them were 100% specific (no suprathreshold activations in healthy controls), except for one false positive case produced by `ICApow` at FWE-corrected $p < 0.05$, $k > 0$.

4.3.3 Activation maps of individual patients

As a final part of our analysis, we inspect the statistical maps at $Z > 3.4$ with cluster size constraint $k > 350$ voxels or 2.8 cm^3 for some interesting individual cases. For visualization purposes, we show thresholded maps, though we stress that T-evidences were computed from the unthresholded T-values.

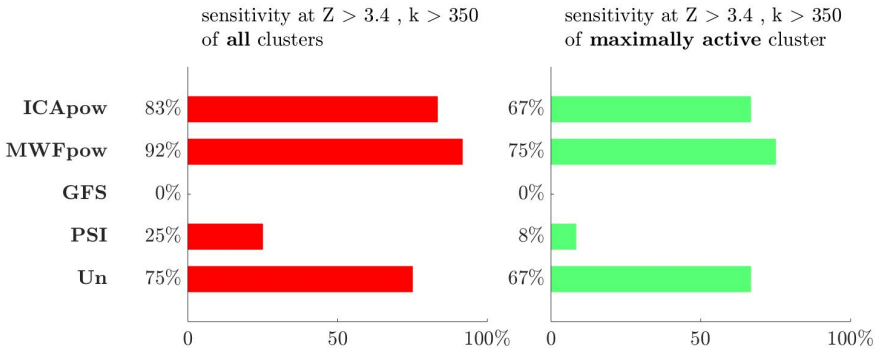


Figure 4.4: The sensitivity is highest for the `MWFpow` regressor, whether the overlap with the ictal onset zone of all clusters (left) or only of the maximally active cluster (right) is considered. At the threshold which was found to be most selective in (Tousseyn et al., 2014a), i.e. $Z > 3.4$ with cluster size constraint $k > 350$ voxels or 2.8 cm^3, the ranking of all regressors is the same for both types of sensitivity. The colours for both types of sensitivity correspond to the colours of the clusters in Figure 4.5, 4.6, 4.7

Firstly, we inspect the maps for the 7th patient, in Figure 4.5. As can be concluded from Figure 4.3, all predictors except for `GFS` produced a correct detection of the IOZ. However, large differences in the active clusters exist. The `MWFpow` and `ICApow` predictors led to very similar maps, with the maximally active cluster concentrated in and around the IOZ. Also the `Un` and `PSI` predictors are sensitive, though only a few suprathreshold voxels overlap with the IOZ, and the maximally active cluster falling out of the IOZ for the latter. Many voxels survive the threshold for the `PSI` map, but comparison with the maps from the other predictors leads to believe that these are in fact false positive findings. The `GFS` predictor produced a very low (but non-zero) T-evidence for this patient, but since no cluster of voxels met the threshold on T-value and size, no active voxels are shown here.

Secondly, we show in Figure 4.6 the maps of the 9th patient, a 'difficult case' for which the T-evidence was low, irrespective of the predictor. Only the `ICApow` and `MWFpow` predictors led to maps with clusters that met the constraints. For both predictors, the maximally active cluster correctly indicates the IOZ. In both cases, there is also a second cluster that falls out the IOZ. The maximally active cluster of the `ICApow` predictor corresponds largely to the two most active clusters of the `MWFpow` predictor, where it seems that this cluster has been 'broken' into two constituent parts. None of the other predictors led to suprathreshold clusters, illustrating the need for tailored, data-driven, and

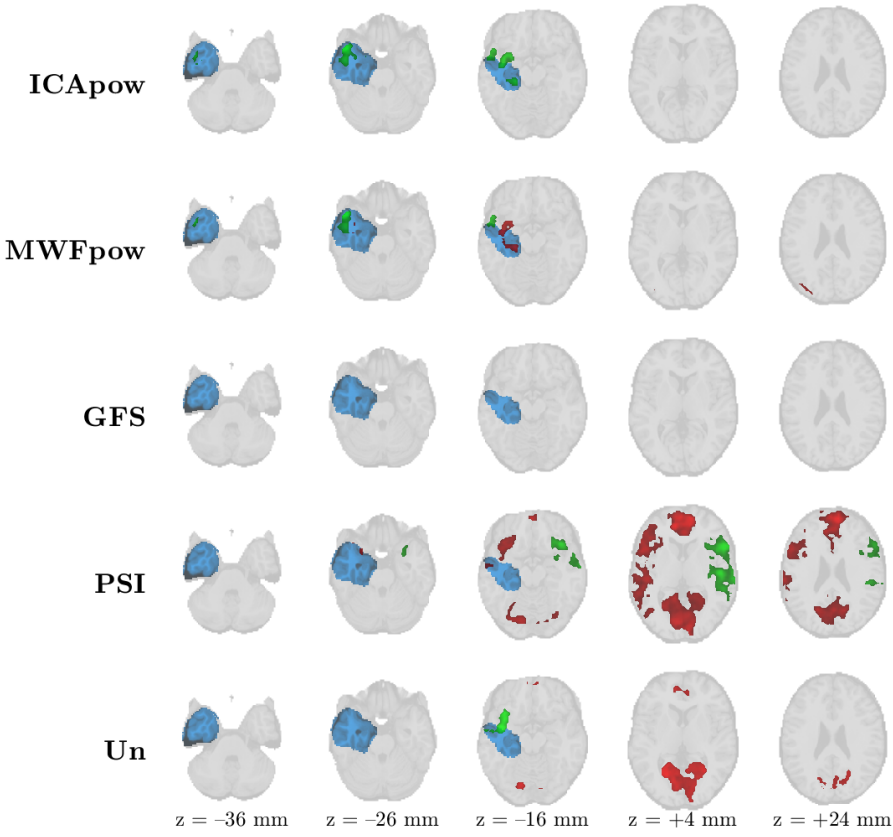


Figure 4.5: Patient 7's ictal onset zone (blue) was correctly detected by the EEG-correlated fMRI activation maps (red) of all predictors except `GFS` at $Z > 3.4$ with cluster size constraint $k > 350$ voxels (2.8 cm^3). The maps of `MWFpow` and `ICApow` are remarkably similar, while large differences with the other predictors exist. The maximally active cluster for every predictor is marked in green (instead of red).

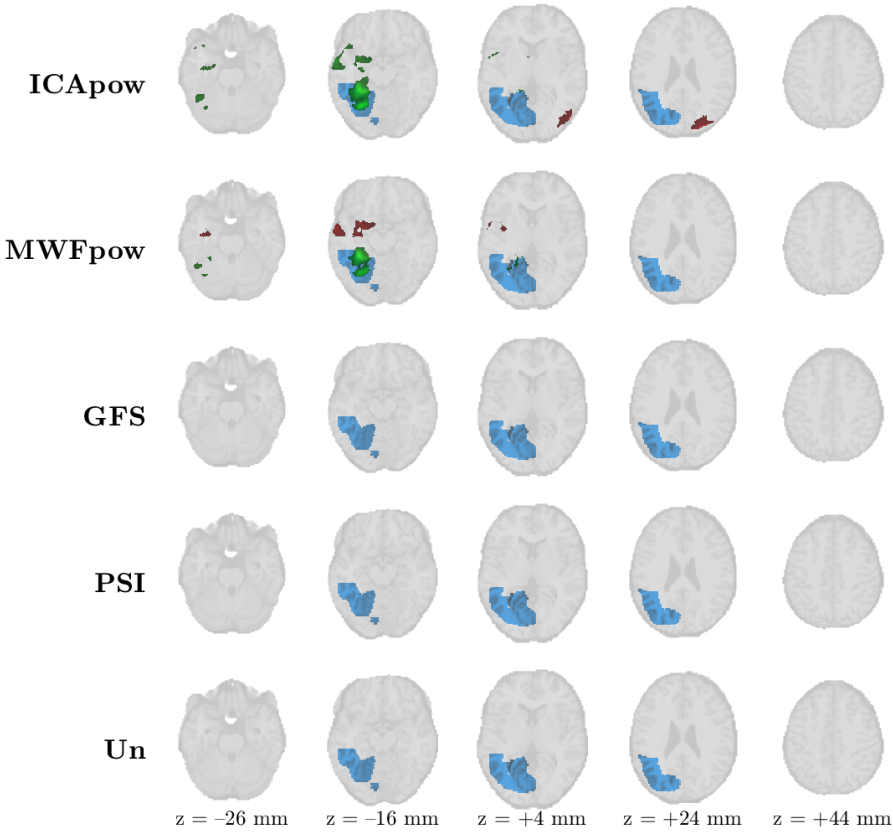


Figure 4.6: Patient 9's ictal onset zone (blue) was only correctly detected by the EEG-correlated fMRI activation maps (red) of the `MWFpow` and `ICApow` predictors at $Z > 3.4$ with cluster size constraint $k > 350$ voxels (2.8 cm^3). The `ICApow` predictor produced one additional active cluster outside of the delineated ictal onset zone. The maximally active cluster for every predictor is marked in green (instead of red).

robust predictors such as `MWFpow` and `ICApow`.

Lastly, we show an atypical case in Figure 4.7, corresponding to the 11th patient of our analysis. From Figure 4.3 we recall that correct detection of the IOZ was only attained for the `Un` predictor. Indeed, nearly all activation maps are empty, besides three distinct active clusters in the `Un` map, two of which overlap the IOZ, and surprisingly also two actives cluster in the `GFS` map, outside of the IOZ. As argued earlier, we hypothesize that the number of interictal spike examples

(6) precluded the estimation of a robust, data-driven BOLD representation. The extremely simple, signal-independent `Un` predictor may be more robust in such cases. If this hypothesis is correct, we see this as a 'foreseeable failure', of the data-driven methods. I.e., for analyses where the recording contains merely a few interictal discharges, one may take additional action: 1) the practitioner may still rely on data-driven predictors (such as `MWFpow`), given that (s)he still manually inspect and potentially corrects leftover artifacts in the predictor, as we have done or 2) rely on predictors such as `Un` which do not have to 'learn' from data examples. Despite this caution, we note that the `MWFpow` predictor did allow a correct detection for two other patients with few observed IEDs (patients 1 and 8 had 15 and 11 IEDs, respectively), even at the most stringent threshold.

4.4 Discussion

We have examined the modelling power of different EEG-derived predictors to map the ictal onset zone in patients with refractory epilepsy through EEG-correlated fMRI analysis. We compared existing predictors, which are based on either unitary representation of IEDs (`Un`), on time-varying power of estimated IED activity (`ICpow`) or on local (`PSI`) or global (`GFS`) neural synchronization, and proposed a new predictor based on supervised IED enhancement (`MWFpow`). Through a rigorous analysis of sensitivity and specificity at various thresholds, we have demonstrated the usefulness of the (`MWFpow`) predictor when mapping the epileptic network, and more in particular the IOZ. By extracting only the relevant, IED-related information from the EEG, the new method allows for successful localization of the IOZ in a higher fraction of cases, and with higher statistical evidence.

The idea to separate one 'source of interest' from multivariate data with multiple active sources is not unique to the MWF methodology. As a very popular tool in EEG processing, ICA essentially has the same objective and is even more ambitious, because multiple sources are estimated. Both ICA and the MWF are types of spatial filters (note that we have indicated this with the same symbol $\hat{\mathbf{W}}$ in section 4.2) that have been used for EEG signal enhancement. The crucial distinction is that ICA is completely unsupervised (i.e. requires no user input), whereas the MWF is supervised and leverages prior knowledge in the form of annotated time segments in (a portion of) the data[5]. In the

[5]In the construction of the `ICpow` predictor, the most IED-like independent component is selected in a supervised fashion, namely based on similarity with user-annotated IEDs. However, the user input has no influence on the enhancement itself, which is the decomposition into independent components.

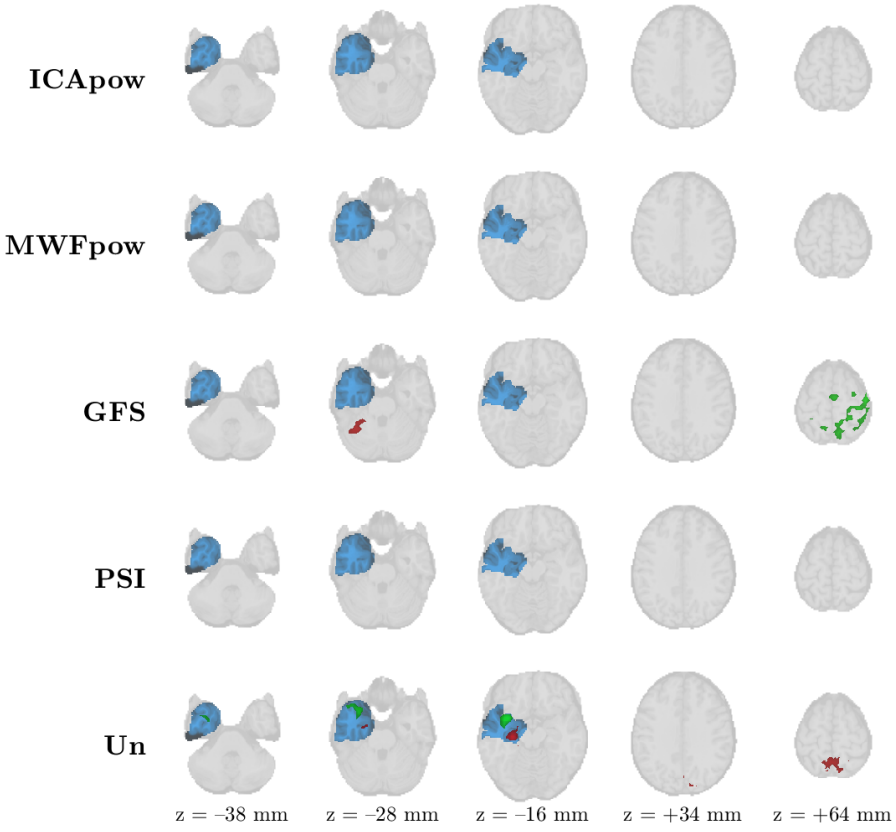


Figure 4.7: Patient 11's ictal onset zone (blue) was only detected by the EEG-correlated fMRI activation map (red) of the `Un` predictor at $Z > 3.4$ with cluster size constraint $k > 350$ voxels (2.8 cm^3). The sparse occurrence of IEDs during the recording may have precluded proper tuning of the `MWFpow` predictor, causing it to miss the ictal onset zone. The maximally active cluster for every predictor is marked in green (instead of red).

current study, both approaches have been compared, as the `ICpow` regressors are derived after estimating independent components (ICs) and identifying the IC that is most related to IEDs.

The success of the `MWFpow` predictor can be attributed to several factors. First of all, the new predictor is a continuous IED representation, in the sense that individual variations in IED amplitude and duration are preserved, compared to the discrete nature of the very popular `Un` predictor. Furthermore, the `MWFpow` finds a new balance between supervised (user-driven) and unsupervised (data-driven) operation. On the former end of the spectrum resides the `Un` predictor, which strictly models only those IEDs which have been marked by a neurologist. As such, this model may fail to capture relevant IED-specific variations, and does not treat IED-like events that have been missed (or wrongfully annotated) by the user. Toward the latter side of the spectrum is the `ICApow` predictor, which extracts an IED representation by decomposing the data in several ICs and only witholding the IC that corresponds most strongly to the user-annotated IEDs. We have shown that the MWF paradigm allows in fact a correction of the annotated IEDs: based on the estimated covariance structure, new IED-like events may be found in the MWF outputs if they match the learned characteristical spatiotemporal correlations. It may seem as though there is a risk for 'positive feedback' here, since initially undetected IED events will contribute to the noise covariance $\hat{\mathbf{R}}_{nn}$ and as such belong to the class of waveforms the MWF will suppress. Luckily however, IEDs are sparse events, and the fraction of truly non-IED EEG is usually far higher than that of (undetected) IED segments. Thus, the $\hat{\mathbf{R}}_{nn}$ will only be marginally impacted by the leakage of these undetected events. If a sufficiently representative covariance $\hat{\mathbf{R}}_{dd}$ of the desired signals has been estimated, the MWF may then recover these undetected IEDs from the data. Contrary to the effect on $\hat{\mathbf{R}}_{nn}$, the additional reconstruction of these events in the MWF outputs *is* significant, precisely because of the sparse nature of the IEDs.

Our results contrast those of (Abreu et al., 2018a), where the `PSI` predictor was found to outperform `ICpow`, `GFS` and `Un`. In our study, we have found an opposite trend, namely that the `PSI` predictor is considerable less sensitive than the `ICpow` and especially the `MWFpow` predictor. Several arguments might explain this result. Notably, only four patients with observed IEDs during EEG–fMRI recording were included in the analysis in (Abreu et al., 2018a). In our study, we investigated a considerably larger sample of twelve patients, and hence we suspect that our results have more generalization power. Furthermore, assessment of the BOLD activation maps in (Abreu et al., 2018a) was performed by checking correspondence with electrical source imaging maps. However, since the EEG was only recorded at 31 electrodes, the resulting source maps may contain substantial localization errors (Lantz et al., 2003; Michel et al.,

2004; Sohrabpour et al., 2015) and may not serve as a reliable ground truth to evaluate 'correctness' of found active BOLD clusters. Instead, we have opted for another way to evaluate sensitivity and specificity, which is based on the anatomical mapping of the ictal onset zone directly. In the patients that had a successful outcome following surgery, the ictal onset zone has been defined as that part of the cortex which had been removed or resected. Consequently, this definition of IOZ certainly encompassed the epileptogenic zone, as has been described in section 4.2. For patients that had not become seizure-free, we used evidence from other modalities such as SPECT and video EEG, to make a best guess about the delineation of the IOZ. Besides the differences in validation and patient cohort, we have tried to replicate the processing pipeline of (Abreu et al., 2018a) as closely as possible, to allow fair comparison. However, small differential effects due to scanner specifics or preprocessing of the raw EEG data may still persist.

Two limitations of the study need to be underlined. Since nine out of twelve studied patients had temporal lobe epilepsy, we can not yet conclude that the new method generalizes well to other types of epilepsy, and further studies are advocated to investigate this. Furthermore, for many patients, relatively many IED annotations were available (see Table 4.1). It is to be expected that the sensitivity of the proposed `MWFpow` predictor (and also that of all other predictors) on average degrades when only few IEDs occur during the fMRI recording, as is the case for patient 11. Nonetheless, a low number of observed IEDs did not preclude correct IOZ detection in two other cases (patients 1 and 8). This is somewhat encouraging: the method's performance does not collapse entirely, but rather offers a reasonable robustness, under 'IED scarcity'. In some cases, no IEDs are observed on the EEG recorded during the fMRI session. While this precludes the use of the `Un` predictor, it is still possible to use `MWFpow` predictor, if an additional EEG recording of the patient (out of the scanner) is available in which IEDs occur. In such cases, the MWF can be 'calibrated' on this set of EEG signals, and then used to filter the EEG signals recorded during the fMRI session. As shown in (Grouiller et al., 2011) using an analogous method based on correlation with an IED template, such an approach can still produce very meaningful spatial maps. A similar approach may be followed for the `ICApow` and `PSI` predictors (for which also at least one observed IED is needed), while for the `GFS` predictor, no IED information is used (which may be one reason for its weak performance in this study).

In summary, we advocate for the use of the newly developed `MWFpow` predictor, when mapping interictal brain activity using EEG-correlated fMRI. This method works almost fully data-driven, and uses examples of IEDs as a main ingredient to enhance the EEG signals and obtain a robust representation of the interictal activity to model the BOLD signal. The minimal user input that is required

consists of some annotations of IEDs in recorded EEG data of a patient. However, a neurologist or EEG reader may be assisted for this by several detection methods, such as described in (Tousseyn et al., 2014b; Nonclercq et al., 2012; Grouiller et al., 2011; Oikonomou et al., 2007). When labeling IEDs manually, time can be saved by annotating only a representative portion of the EEG recording. Especially for patients with many observed IEDs, the `MWFpow` predictor has been shown to produce very sensitive and yet maximally selective results, and may therefore be a valuable tool in the presurgical work-up of refractory epilepsy.

Chapter 5

Revealing variable neurovascular coupling via structured matrix-tensor factorization

ABSTRACT | Simultaneous recording of electroencephalographic (EEG) signals and functional magnetic resonance images (fMRI) has gained wide interest in brain research, thanks to the highly complementary spatiotemporal nature of both modalities. We propose a novel technique to extract sources of neural activity from the multimodal measurements, which relies on a structured form of coupled matrix-tensor factorization (CMTF). In a data-symmetric fashion, we characterize these underlying sources in the spatial, temporal and spectral domain, and estimate how the observations in EEG and fMRI are related through neurovascular coupling. That is, we explicitly account for the intrinsically variable nature of this coupling, allowing more accurate localization of the neural activity in time and space. We illustrate the effectiveness of this approach, which is shown to be robust to noise, by means of a simulation study. Hence, this provides a conceptually simple, yet effective alternative to other data-driven analysis methods in event-related or resting-state EEG–fMRI studies.

This chapter is a modified version of Van Eyndhoven, S., Hunyadi, B., De Lathauwer, L. & Van Huffel, S. "Flexible fusion of electroencephalography and functional magnetic resonance imaging: Revealing neural-hemodynamic coupling through structured matrix-tensor factorization", in IEEE 25th European Signal Processing Conference (EUSIPCO 2017, Kos, Greece), Aug. 2017, pp. 26–30; with changes being limited to layout.

5.1 Introduction

In the field of neuroimaging and brain mapping, the combined use of EEG and fMRI has been widely recognized as a promising approach, since it exploits both the high temporal resolution from EEG and the excellent spatial resolution from fMRI. These two key factors enable researchers and clinicians to address a multitude of questions on cognitive or pathological neural activity, which underlies the data from both modalities (Ritter and Villringer, 2006).

The complementarity of the two modalities is at the same time its own pitfall: it is not straightforward to combine two such heterogeneous data streams - the fast electrical recording in EEG which is sparsely sampled over the scalp, and the sluggishly varying Blood Oxygen Level Dependent (BOLD) signal in fMRI which, however, can be localized accurately. Despite the mismatch in spatiotemporal resolution, we can assume that the two involved modalities measure the same neural processes in the brain, although in an indirect way and to a different extent. Thus, it is unknown to which degree the spatial and temporal observations in EEG and fMRI are related. This has led to the development of a wide variety of analysis tools, all approaching this data fusion problem from different angles.

One way to characterize these tools is by the way they try to fit a model to the data and pose restrictions to it. On the one hand, model-driven techniques exist, ranging from very complex (Valdes-Sosa et al., 2009) to simpler models such as the widely applied General Linear Model (GLM), which has been a standard workhorse to analyze fMRI datasets via EEG-correlated fMRI analysis (Monti, 2011). In the common way of applying the GLM, a single 'canonical' hemodynamic response function (HRF) is used, which describes the expected BOLD signal resulting from an impulse event in a reference stimulus time series. Hence, the HRF waveform is considered invariant over the different brain areas, over neural 'events', and even over different subjects in a group study. Although the GLM has its merit as a simple and attractive framework to make sense of (multimodal) brain data, this limitation obfuscates the interpretations that can be drawn through its use. As we highlighted in Section 2.5.2, it is well known that the hemodynamic response displays a significant variability over people, brain areas and over time (Duann et al., 2002; Handwerker et al., 2004; Lindquist et al., 2009), which might even be more severe for the diseased brain (Bénar et al., 2002). Hence, misspecification of the HRF leads to biased estimates of activity and loss in statistical power (Handwerker et al., 2004; Lindquist et al., 2009). To cope with this variability, several methods have been proposed, such as adding the derivatives of the canonical HRF during GLM regression, or non-linear fitting methods (Monti, 2011; Pedregosa et al., 2015).

However, most of these methods still make an unfavorable tradeoff between complexity versus flexibility: either they deal with only small variations, or they require the estimation of HRF parameters on a voxel-wise basis. On the other end of the spectrum, data-driven blind source separation (BSS) techniques such as independent component analysis can be used to decompose both EEG and fMRI data into a set of constituting sources, in which case neural sources are found by imposing sometimes arbitrary restrictions (Calhoun et al., 2009; Lahat et al., 2015). Traditional BSS methods such as independent component analysis (Calhoun et al., 2009) or canonical correlation analysis (Li et al., 2009) rely mostly on matrix factorization techniques, which operate on data in a two-way format. Over the past years however, tensor-based techniques which factorize third or higher order arrays have been increasingly used (Cichocki et al., 2015; Cichocki, 2013), also in EEG–fMRI data analysis (Hunyadi et al., 2017; Martínez-Montes et al., 2004; Karahan et al., 2015). They offer the advantage that the inherent multidimensional nature of datasets (such as those in brain analytics) is respected, and most importantly, that the factorizations of tensors can be unique under mild conditions, as opposed to matrix factorization (Cichocki et al., 2015). However, when applied to recordings of brain activity, these data-driven techniques exploit (by definition) no model of the neurovascular coupling, and might thus neglect prior (approximate) knowledge on the joint 'behavior' of the datasets.

We aim to tackle this issue by taking a hybrid, semi-informed estimation approach, borrowing ideas from the GLM as well as from BSS methods. That is, we try to find sources of neural activity by fusing EEG and fMRI recordings in a data-driven fashion, while still accounting for a characteristic temporal relationship between the electrophysiological activity (as measured with EEG) and the BOLD signal (measured with fMRI), in the form of an a priori unknown HRF. Hence, we insert a prior on the neural-hemodynamic coupling into the problem formulation, which admits to co-estimate the true underlying HRF from the data itself. We achieve this combined goal of extracting neural-hemodynamic sources and their temporal coupling by expressing the problem as a coupled matrix-tensor factorization (CMTF) (Acar et al., 2013; Karahan et al., 2015), where the fMRI data is a matrix of time samples $\times$ voxels and the EEG spectrogram is represented as a 3rd order tensor of time samples $\times$ electrodes $\times$ frequencies as in (Martínez-Montes et al., 2004). The factorizations of both datasets are then constrained by a shared factor matrix in the temporal mode, which is transformed by a convolution with an unknown HRF in the fMRI factorization. This is a generic framework, in which different kinds of prior knowledge about the HRF can be inserted to reduce the search space when solving the problem.

In the following sections, we first explain the tensor-based framework that we

use to tackle the problem. Then, we analyze the feasibility of the proposed approach by means of a simulation study. Finally, we present extensions to our approach that allow to cope with additional variability in the data.

5.2 Methods

5.2.1 Generative model

We assume that the recorded EEG and fMRI result from the superposition of several underlying neural sources, each with a distinct spatial origin in the brain, accompanied by a time course of their activity. The EEG recording $\mathbf{X}$ over channels (electrodes) and time is assumed to be transformed to a spectrogram representation, by means of e.g. the short-time Fourier transform, followed by squaring the absolute values of the complex coefficients. The resulting 3rd order tensor $\boldsymbol{\mathcal{X}}$ describes the variation over the spatial, temporal, and spectral mode. The fMRI data is a matrix $\mathbf{Y}$, describing the variation over the temporal and spatial mode.

Loosely following the notation in (Martínez-Montes et al., 2004), the EEG and fMRI data can be written as the sum of R sources, which are reflected in both datasets as follows:

$$\begin{aligned} \boldsymbol{\mathcal{X}} &= \sum_{r=1}^{R} \mathbf{a}_r \otimes \mathbf{b}_r \otimes \mathbf{c}_r + \boldsymbol{\mathcal{E}}_x \\ \mathbf{Y} &= \sum_{r=1}^{R} \mathbf{u}_r \otimes \mathbf{v}_r + \mathbf{E}_y = \sum_{r=1}^{R} \mathbf{u}_r \mathbf{v}_r^{\mathrm{T}} + \mathbf{E}_y \ . \end{aligned} \tag{5.1}$$

The column vectors $\mathbf{a}_r$, $\mathbf{b}_r$, and $\mathbf{c}_r$ describe the activity of source r over time, and its distribution over EEG channels and frequencies, respectively. Analogously, the vectors $\mathbf{u}_r$ and $\mathbf{v}_r$ describe the time course of source r and its distribution over voxels, respectively. The residuals are contained in $\boldsymbol{\mathcal{E}}_x$ and $\mathbf{E}_y$.

5.2.2 Structured matrix-tensor factorization

The electrophysiological time course $a_r(t)$ of the source in the EEG is assumed to predict the hemodynamic time course $u_r(t)$ in fMRI by means of a convolution with an unknown HRF $h(t)$, which can be rewritten as the matrix product of

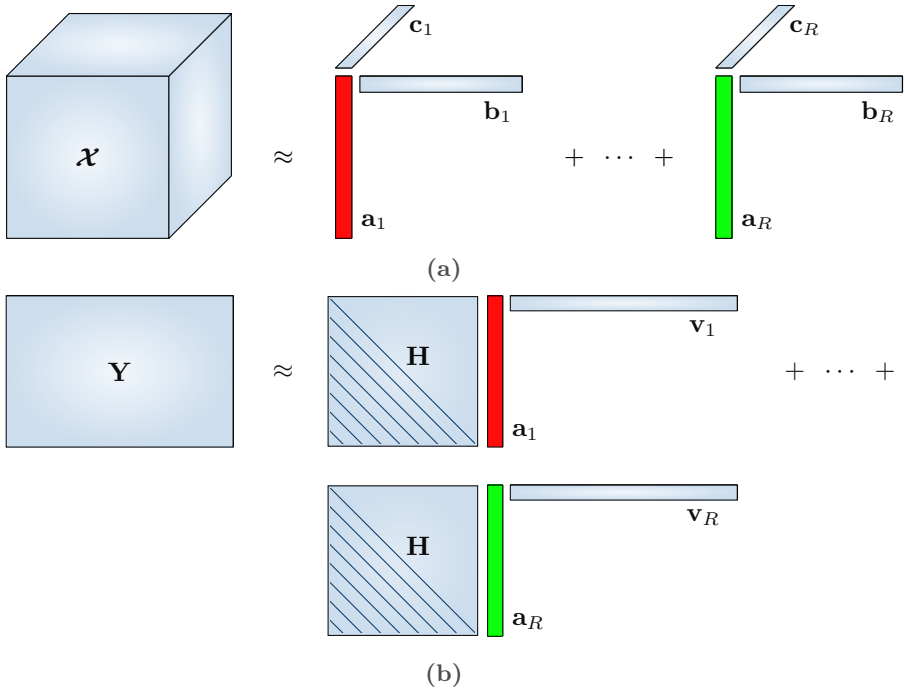


Figure 5.1: Structured coupled matrix-tensor factorization of (a) EEG data and (b) fMRI data. The signatures $\mathbf{a}_r$ along the temporal mode are shared between the EEG factorization and the fMRI factorization, providing the coupling constraint.

the temporal signature $\mathbf{a}_r$ with a Toeplitz matrix $\mathbf{H}$, holding the samples of $h(t)$ on its diagonals:

$$\mathbf{u}_r = \mathbf{H}\,\mathbf{a}_r = \text{Toeplitz}(h)\mathbf{a}_r\ . \tag{5.2}$$

To incorporate prior physiological knowledge, we constrain the HRF to a 'plausible' waveform. For this, we use a family of parametrized basis functions, i.e., we can model the HRF as the difference of two gamma functions (like is done for the often-used canonical HRF in the GLM). The HRF thus depends on a small number of unknown parameters as $h(t) = f(t,\theta) = \Gamma(\theta_1)^{-1} \cdot \theta_2^{\theta_1} t^{\theta_1-1} e^{-\theta_2 t} - \theta_3 \Gamma(\theta_4)^{-1} \cdot \theta_5^{\theta_4} t^{\theta_4-1} e^{-\theta_5 t}$.

Furthermore, note that the EEG tensor is non-negative, because it describes the spectral power on the different channels on different time instants. To obtain an interpretable result, we impose that the signatures $\mathbf{a}_r$ and $\mathbf{c}_r$ are

also non-negative, since they describe the time-varying power and spectra of the sources, respectively. Hence, we reformulate the data fusion problem as a structured matrix-tensor factorization (Sorber et al., 2015), in which the factor matrix $\mathbf{A} = [\,\mathbf{a}_1\,\mathbf{a}_2\ldots\mathbf{a}_R\,]$ is shared between the two datasets, while the factor matrices $\mathbf{B} = [\,\mathbf{b}_1\,\mathbf{b}_2\ldots\mathbf{b}_R\,]$, $\mathbf{C} = [\,\mathbf{c}_1\,\mathbf{c}_2\ldots\mathbf{c}_R\,]$, $\mathbf{V} = [\,\mathbf{v}_1\,\mathbf{v}_2\ldots\mathbf{v}_R\,]$ are modality-specific and the parameter set θ describes the HRF factor matrix $\mathbf{H}$. The coupled factorization can then be solved by minimizing the following cost function:

$$J(\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{V},\theta) = \left\| \boldsymbol{\mathcal{X}} - \sum_{r=1}^{R} \mathbf{a}_r \otimes \mathbf{b}_r \otimes \mathbf{c}_r \right\|_F^2 + \left\| \mathbf{Y} - \sum_{r=1}^{R} (\mathbf{H}(\theta)\mathbf{a}_r) \otimes \mathbf{v}_r \right\|_F^2, \text{ s.t. } \mathbf{A},\mathbf{C} \geq 0 \tag{5.3}$$

in which $\|\cdot\|_F$ indicates the Frobenius norm. Figure 5.1 shows the data model, including the dependencies (i.e. coupling) between EEG and fMRI, which is used to find a structured coupled matrix-tensor factorization. Under this model, the fMRI data can be written as $\mathbf{Y} = \mathbf{H}\mathbf{A}\mathbf{V}^{\mathrm{T}}$. Contrary to other (standalone) matrix factorizations, rotational ambiguity is ruled out between any pair of the matrices $\mathbf{H}$, $\mathbf{A}$ and $\mathbf{V}^{\mathrm{T}}$, since the middle factor $\mathbf{A}$ is constrained by the EEG factorization.

5.2.3 Simulation study

We conducted simulations in which the data was generated according to the model in (5.1) and (5.2). That is, we placed $R = 3$ neural sources at distinct positions in a template brain of $91 \times 109 \times 91$ voxels of $2 \times 2 \times 2\text{mm}^3$, and picked a HRF which differed from the canonical HRF, as shown in Figure 5.2a, 5.2b, 5.2c and Figure 5.4 respectively. We assigned (overlapping) temporal and spectral signatures to the sources (see Figure 5.3), and calculated the spatial distributions over 19 electrodes by means of a forward volume conduction model. The resulting signatures of the generated sources are depicted in Figure 5.2d, 5.2e, 5.2f. The temporal signatures span 60 seconds and are sampled at 4 Hz. After convolution with the chosen HRF, the obtained fMRI time courses are downsampled to 1 Hz, mimicking the discrepancy in temporal resolution between the two modalities. To deal with this mismatch, the downsampling operation is implicitly assumed in (5.3) and taken into account during optimization. The spectral signatures consist of 20 frequency bins.

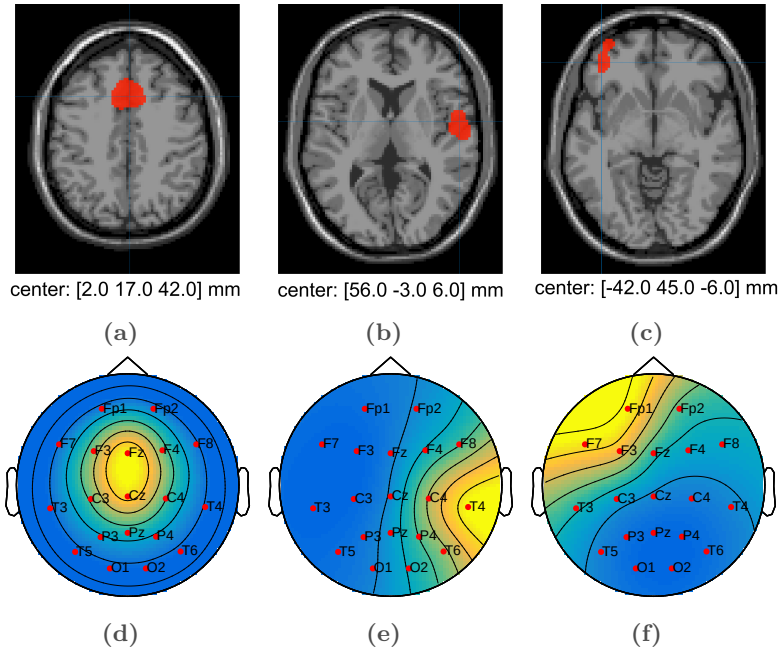


Figure 5.2: Location and extent of the three neural sources in the fMRI domain (a, b, c) and topographic plots of the corresponding EEG spatial signatures, obtained through a forward volume conduction model, based on the anatomical MRI image and the sources' positions (d, e, f).

We added white Gaussian noise of varying signal-to-noise ratio (SNR) to both EEG and fMRI data. In the remainder of the text, SNR is measured as the ratio of the variance of the noise-free time course, either in a voxel or in a electrode-frequency pair with the highest variance, and the variance of the noise. Before factorizing, we applied dimensionality reduction on the fMRI data, by computing the variance in every voxel, computing z-scores of all variances and rejecting voxels with a z-score below 1. To summarize, we jointly factorize a noisy EEG dataset $\boldsymbol{\mathcal{X}}$ of dimensions $240 \times 19 \times 20$ and a noisy fMRI dataset $\mathbf{Y}$ of dimensions $60 \times N$, where N indicates the remaining number of voxels (typically $\sim 10^5$). Optimization of the cost function in (5.3) is done through the Structured Data Fusion framework in Tensorlab (Sorber et al., 2015; Vervliet et al., 2016). In a realistic setting, the true number of sources is unknown a priori. We therefore computed the factorization for R varying from 1 to 4, and picked the most likely number of sources according to the core consistency diagnostic (Cichocki et al., 2015; Karahan et al., 2015). In all but a few cases, this heuristic correctly estimated the number of sources, which is why we only

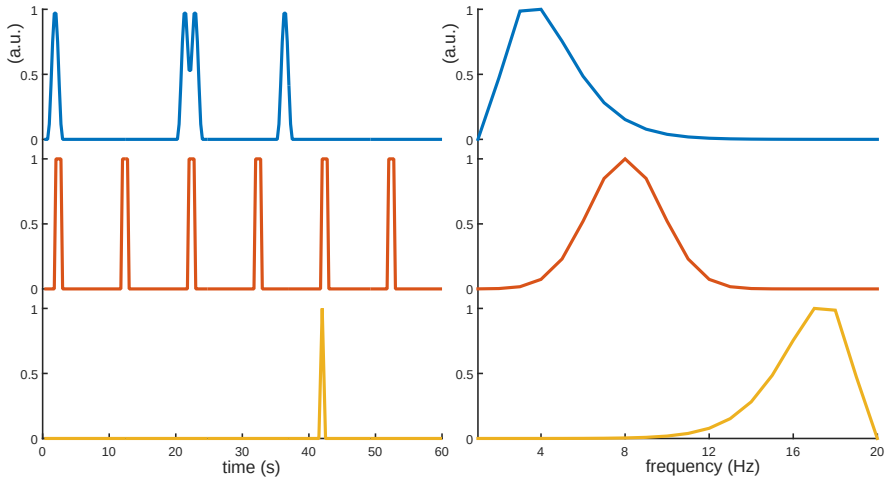


Figure 5.3: Temporal (left column) and spectral (right column) EEG signatures, which belong to the neural sources in Figure 5.2 (a.u. = arbitrary units).

report results for the case where the number of extracted sources equals R. We aimed to address several subquestions, further described below.

How reliably can the true signatures be extracted?

Under SNR levels ranging from -8 dB to 12 dB, in steps of 4 dB, we performed the coupled EEG–fMRI factorization and checked the relative estimation errors. We exhaustively repeated this for data which was simulated using every possible combination of 2 sources, or using all 3 sources. For all sources, we computed relative estimation errors in the EEG temporal mode as $\delta_a = \frac{1}{R}\sum_{r=1}^{R} \frac{\|\hat{\mathbf{a}}_r - \mathbf{a}_r\|_2}{\|\mathbf{a}_r\|_2}$ after an optimal permutation and scaling of the estimated signatures $\hat{\mathbf{a}}_r$ (and analogously for the spectral and spatial modes). In the fMRI domain, we quantified the spatial overlap of true and estimated sources. For a true underlying source s which is active in a set P consisting of N_s voxels, we took the set Q of N_s voxels with the highest value from the estimated signature $\hat{\mathbf{v}}_r$ (after applying the same permutation as for the EEG sources) and computed the Dice coefficient $d = \frac{2|P \cap Q|}{|P|+|Q|} = \frac{|P \cap Q|}{N_s}$. The EEG signatures were initialized by performing a factorization on the EEG dataset separately using parallel factor analysis (PARAFAC) (Cichocki et al., 2015; Karahan et al., 2015). The HRF was initialized as the canonical HRF, and the initial voxel signatures were obtained from $\mathbf{Y}$ using the pseudo-inverse of the initialized factors $\mathbf{H}(\theta)$ and

A. We compared against reference results, which were obtained by fixing the parameter set θ in (5.3) to the parameters of the canonical HRF, similarly to the method in (Martínez-Montes et al., 2004).

To which extent can the model cope with HRF variability?

We created different HRFs by varying the timing, width and relative amplitudes of the positive and negative lobes. The time of the positive peak varied from 2 s to 10 s, in steps of 2 s, and the negative peak came 5 s, 10 s, or 15 s later. The width (measured as the difference in time between the 2 points where the 2nd derivative is zero) of the first lobe was either 2, 5, or 8 s, and the width of the negative lobe was always 1.5 times larger. The 2nd peak had a relative amplitude of 0% (no peak), 20% or 60% w.r.t. the positive peak. We aimed to extract the signatures from the datasets generated with 3 sources and these different parameters, for SNR = 4 dB. We used the Pearson correlation coefficient to assess the similarity between the waveforms of the estimated and the true HRF.

What is the robustness versus non-white noise?

In the previous experiments, we used additive white Gaussian noise (WGN). To investigate the effect of noise which is colored (i.e. non-white) in the spatial or temporal domain, we conducted two additional tests. In the first test, referred to as 'SpCorr', we introduced spatial correlation in the fMRI noise by filtering it with a 3D Gaussian smoothing kernel with a standard deviation of 2 voxels, before adding it to the generated data. In the second test, referred to as 'TempCorr', we created the fMRI noise by means of an autoregressive model of order 1 (AR(1)), as $e_t = 0.7e_{t-1} + \xi_t$, where ξ_t is drawn from a white Gaussian noise distribution. The SpCorr experiment takes into account that (background) neural processes exhibit spatial covariance over neighbouring voxels, while the TempCorr experiment mimics AR(1)-structured residuals, which is an often-used noise model in fMRI(Monti, 2011). We still added WGN to the EEG dataset in both cases. We evaluated this scenario for the combination with 2 active sources. Since only the noise for the fMRI data set was altered, we track performance in terms of the Dice coefficient.

5.3 Results

Reliability of extracted signatures

Figure 5.5a and Figure 5.5b show the mean relative errors δ for the EEG factorization in the case of 2 and 3 sources, respectively. It is clear that, on

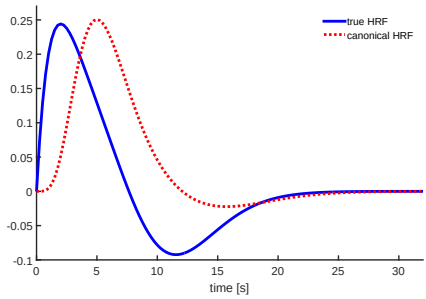


Figure 5.4: True underlying hemodynamic response function (full blue line) in this study, and the canonical hemodynamic response function (dotted red line) as used in the SPM toolbox. The positive (resp. negative) peaks of both HRFs are 3 seconds (resp. 4 seconds) apart.

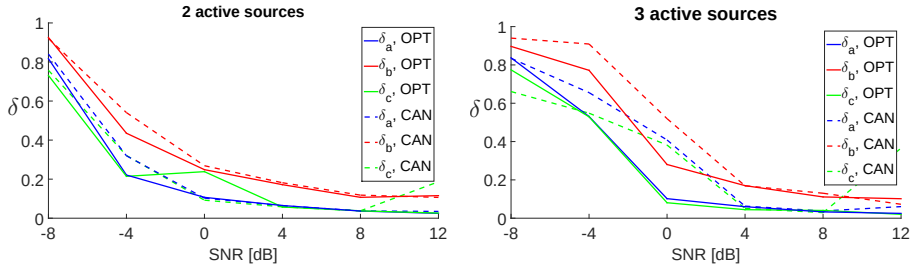
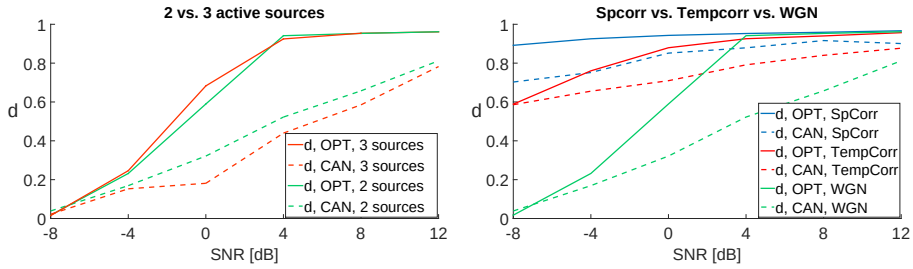


Figure 5.5: Mean relative errors δ_a, δ_b, and δ_c between true and estimated sources in the EEG's temporal, spatial, and spectral mode, respectively, in the presence of 2 sources (a) or 3 sources (b). Results are shown for the case where we find the HRF in the optimization procedure ('OPT') and for the case where we fix the HRF to the canonical HRF ('CAN').

average, the signatures in the temporal, spatial and spectral mode are more accurately estimated when we account for HRF variability ('OPT' condition). This is expected, since the EEG factorization is influenced by the fMRI factorization through coupling of the temporal signatures. A similar effect is observed when looking at the Dice coefficients: from Figure 5.6a we conclude that modeling HRF variability has a beneficial effect on the localization of neural sources in the brain, i.e. a higher overlap with the true sources is achieved.

Robustness versus HRF variability



(a) **(b)**

Figure 5.6: Mean Dice coefficients d between true and estimated fMRI spatial signatures, in the presence of 2 or 3 active sources and WGN (a) or 2 active sources and spatially ('SpCorr') or temporally ('TempCorr') correlated noise (b). Results are shown for the case where we find the HRF in the optimization procedure ('OPT') and for the case where we fix the HRF to the canonical HRF ('CAN').

In table 5.1 we show the correlation between estimated and true underlying HRF, for HRFs that vary according to 4 parameters: time to positive peak, separation between positive and negative peak, dispersion (i.e. the width of the positive lobe) and the relative amplitude of the negative peak compared to the positive peak. As a reference: the parameters of the canonical HRF are 5 s, 10.7 s, 4.5 s and 0.09, respectively. Values are given as the mean ± one standard deviation of the pooled results, when only the indicated parameter is fixed. In general, the algorithm succeeds reasonably well in capturing the HRF waveform, admitting a decrease in performance when the true HRF differs substantially from the initial guess (which was the canonical HRF, in our case). We note that there was a significant dependence (correlation 0.84) between the average Dice coefficient and the correlation between true and estimated HRF. This confirms findings of previous research, which pointed out the potential bias resulting from HRF mismodeling (Handwerker et al., 2004; Lindquist et al., 2009), and elicits the importance of HRF estimation.

Robustness versus spatially/temporally non-white noise

Figure 5.6b shows the mean Dice coefficients under the effect of spatially (SpCorr) or temporally (TempCorr) correlated noise. Interestingly, the spatial extent of the neural sources is estimated more accurately under AR(1) noise conditions, or under spatially correlated noise, compared to WGN conditions, especially at lower SNR. No significant difference in quality of the extracted EEG signatures was observed between conditions when the HRF was co-estimated, but using the canonical HRF led to an increase in relative error of ±10–15% in all modes.

Table 5.1: Pearson correlation between estimated and true HRF.

time of 1st peak	peak separation	dispersion	relative amplitude
2 s 0.71±0.31	5 s 0.81±0.28	2 s 0.70±0.35	0 0.88±0.23
4 s 0.94±0.03	10 s 0.90±0.20	5 s 0.90±0.19	0.2 0.81±0.31
6 s 0.95±0.05	15 s 0.81±0.29	8 s 0.91±0.17	0.6 0.83±0.25
8 s 0.93±0.09			
10 s 0.66±0.41			

5.4 Discussion

We have demonstrated that accounting for HRF variability has a beneficial effect on the estimation of spatial, temporal, and spectral aspects of neural sources from combined EEG–fMRI measurements. For this purpose, we have introduced a data-driven framework based on structured matrix-tensor factorization, which can be used to identify patterns of neural activity. It is generic in the sense that it describes neural-hemodynamic coupling solely by means of a convolution operation with an initially unknown HRF, and thus minimally makes assumptions on the neural sources of interest, adhering to the idea of blind source separation. Any prior model for the nature of the HRF can be plugged in, as long as it can be expressed as an explicit and differentiable function of some parameters (Sorber et al., 2015). Complete flexibility can be attained by not imposing any model for $h(t)$, which is conceptually the same as the 'FIR basis set' method (Pedregosa et al., 2015), while the GLM can be mimicked by fixing $h(t)$ to the canonical HRF. In this paper, we have chosen a model consisting of two gamma functions, because of its simplicity and similarity to the often-used canonical HRF. While several methods have been proposed to find the HRF from the experimental data, they often require to re-estimate the HRF voxel by voxel (Lindquist et al., 2009; Bénar et al., 2002; Pedregosa et al., 2015), which is tedious and precludes easy interpretation. Instead, our proposed method seeks to find one HRF, which leads to the best fit for the whole dataset. Nonetheless, the framework allows to deal with variation of neurovascular coupling over brain regions, in a way that is analogous to well-known approaches which use a basis set of HRF functions(Lindquist et al., 2009; Pedregosa et al., 2015). This is achieved by extending the fMRI factorization as $\mathbf{Y} = \mathbf{H}_1\mathbf{A}\mathbf{V}_1^{\mathrm{T}} + \mathbf{H}_2\mathbf{A}\mathbf{V}_2^{\mathrm{T}} + \ldots + \mathbf{H}_B\mathbf{A}\mathbf{V}_B^{\mathrm{T}}$, in which B basis functions $h_1(t)$ to $h_B(t)$ are used. Hence, the BOLD response is described as a voxel-specific linear combination over basis functions - as well as over different neural sources, like in the simple case. Many other adaptations of this framework can be conceived, depending on the context of the data (e.g. event-related or resting-state EEG–

fMRI experiments), the desired flexibility, and the type of information which one tries to extract. This is the subject of future work.

Chapter 6

Identifying reproducibility of matrix/tensor factorizations via low-rank graph clustering

ABSTRACT | Many interesting matrix decompositions / factorizations, and especially many tensor decompositions, have to be solved by non-convex optimization-based algorithms, that may converge to local optima. Hence, when interpretability of the components is a requirement, practitioners have to compute the decomposition (e.g. CPD) many times, with different initializations, to verify whether the components are reproducible over repetitions of the optimization. However, it is non-trivial to assess such reliability or stability when multiple local optima are encountered. We propose an efficient algorithm that clusters the different repetitions of the decomposition according to the local optimum that they belong to, offering a diagnostic tool to practitioners. Our algorithm employs a graph-based representation of the decomposition, in which every repetition corresponds to a node, and similarities between components are encoded as edges. Clustering is then performed by exploiting a property known as *cycle consistency*, leading to a low-rank approximation of the graph. We demonstrate the applicability of our method on realistic electroencephalographic (EEG) data and synthetic data.

This chapter is a modified version of Van Eyndhoven, S., Vervliet, N., De Lathauwer, L. & Van Huffel, S. "Identifying Stable Components of Matrix/Tensor Factorizations via Low-Rank Approximation of Inter-Factorization Similarity", in IEEE 27th European Signal Processing Conference (EUSIPCO 2019, A Coruña, Spain), Sep. 2019, pp. 1–5; with changes being limited to layout.

6.1 Introduction

As argued in Section 3.2.3, decompositions or factorizations of matrices or tensors (multiway arrays) into a number of components are useful in a variety of signal processing, data mining and machine learning applications, in which the matrix or tensor represents a certain multivariate signal or dataset.

In many cases, decompositions of (noisy) data have no closed-form solution and have to be computed iteratively, i.e., by means of an optimization algorithm. Whereas algebraic algorithms exist for various decompositions, they can only be used when the data exactly follows the imposed structure, or to initialize an optimization algorithm (see e.g. (Sørensen et al., 2015) and references therein). In many of the discussed examples (e.g. ICA, NMF, CPD), the optimization problem is non-convex and only convergence to a local optimum can be guaranteed, e.g. when minimizing the residual between the observed data and a factor model thereof (CPD, NMF), or when maximizing a measure of independence between component time courses (ICA). This can pose a problem if one wants to interpret the components of the decomposition, since these components may vary substantially between local optima, even though the associated values of the optimized cost function need not be significantly different. Furthermore, in some applications, because of a considerable influence by noise and/or artifacts, the optimum with the lowest cost value is not guaranteed to yield the most meaningful result. Consider, for example, the computation of a CPD of tensor-valued data that are heavily corrupted by noise and/or artifacts. When fitting the CPD model by minimizing the mean squared error, one or more of the CPD components may model variance that is due to the noise or strong artifacts, neglecting 'true' sources of interest that remain undiscovered. In this case, another solution, in which the artifacts are fully captured in the residual, and where all CPD components model activity of interest, may attain a worse cost value, but may be much more interpretable and meaningful.

From the problem statement above, we distill a practical need to assess the reproducibility of (non-convex) matrix or tensor decompositions. In the current context, this means accurately clustering repetitions of the associated optimization problems in terms of their locally optimal solutions for further interpretation.

A well-known method that attempts such clustering is ICASSO (Himberg et al., 2004), which was designed to assess the reproducibility of ICA decompositions, but whose concepts may be used for other matrix or tensor factorizations as well, e.g. as in (Mareček et al., 2017). The usefulness of this method depends on the user-defined number of clusters, which can be difficult to estimate if the factorization is not very stable.

Related approaches exist in non-signal processing-related contexts. A parallel can be drawn between assessing reproducibility of non-convex decompositions and establishing correspondences (maps or transformations) between (points on) multiple objects. The latter problem appears in many applications, e.g., fusing partially overlapping images of a certain scene (Huber, 2002), extracting structure from motion (Zach et al., 2010), finding dense correspondences between collections of 3D shapes (Huang and Guibas, 2013), etc. E.g., for the case of image fusion, a link must be established between points in two images if these correspond to the same physical point. Similarly, we aim to identify whether a component appears in two distinct repetitions of the decomposition. While it is feasible to estimate *local* (pairwise) maps between objects, it is not trivial to extract the *global* information (over all repetitions) on the correspondences, e.g. because pairwise maps may be noisy and may lead to ambiguities, especially when similar points on the same objects are present (Chen et al., 2014). However, starting from a collection of noisy pairwise maps, techniques known as joint object matching or map synchronization may exploit constraints to retrieve corrected maps that reveal the global structure (Bajaj et al., 2018; Shen et al., 2016). Several algorithms have been devised for this objective, of which most treat the problem in a graph framework: objects are treated as nodes, and the input pairwise maps are transformations residing on the edges between them (Huang and Guibas, 2013; Chen et al., 2014; Bajaj et al., 2018; Shen et al., 2016). In this paper, we propose to use map synchronization principles to identify reliable components of the CPD and assess which components co-occur frequently. Under this setting, the 'right' number of clusters appears naturally from the problem representation and synchronization constraints, fostering a correct interpretation of the components. We demonstrate our technique on the decomposition of electroencephalographic (EEG) data of multiple subjects, and compare its performance to that of ICASSO.

6.2 Methods

6.2.1 (Non-convex) matrix and tensor factorizations

In general, a rank-R CPD of an Mth order tensor $\boldsymbol{\mathcal{T}}$ can be written as

$$\boldsymbol{\mathcal{T}} = \sum_{r=1}^{R} \mathbf{a}_r^{(1)} \otimes \mathbf{a}_r^{(2)} \otimes \ldots \otimes \mathbf{a}_r^{(M)} + \boldsymbol{\mathcal{E}}_x \,, \tag{6.1}$$

where any column vector $\mathbf{a}_r^{(m)} \in \mathbb{R}^{I_n}$ is the mode-m factor vector of the rth component or term, and $\boldsymbol{\mathcal{E}}_x$ is a residual tensor. As there is no closed-form

solution for the factor matrices $\mathbf{A}^{(m)} = \left[\mathbf{a}_1^{(m)}\, \mathbf{a}_2^{(m)} \dots \mathbf{a}_R^{(m)}\right]$, $m = 1 \dots M$ that minimize the Frobenius norm $\|\boldsymbol{\mathcal{E}}_x\|_F^2$ in (6.1), optimization algorithms are commonly used to iteratively update a cost function, e.g. of the form

$$J(\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(M)}) = \left\| \boldsymbol{\mathcal{T}} - \sum_{r=1}^{R} \mathbf{a}_r^{(1)} \otimes \dots \otimes \mathbf{a}_r^{(M)} \right\|_F^2 \tag{6.2}$$

In data fusion, multiple tensors and/or matrices are decomposed simultaneously, and the cost function is a linear combination of fit terms similar to those in (6.2) (Acar et al., 2014; Lahat et al., 2015; Sorber et al., 2015). As discussed in section 6.1, many cost functions for tensor and matrix decompositions are non-convex, and hence the optimization algorithm may converge to different local minima, depending on the initialization for the factor matrices. When the components' factor vectors $\mathbf{a}_r^{(m)}$ are to be interpreted, e.g. in BSS, it is therefore wise to repeat the optimization several times with different initializations, and verify which components seem stable over several runs (and therefore interpretable) and which seem be specific to a local solution.

6.2.2 Inter-factorization similarity as a graph

To assess the reproducibility or stability of a factorization (in R terms), we need to investigate the similarity of components across all repetitions of the decomposition with different initialization, up to permutation and scaling. A graph representation can accommodate for such similarity relationships. Following the notation in (Chen et al., 2014), we treat every repetition $n = 1 \dots N$ of the decomposition as an object $S_n \in \mathcal{S}$ with R points. An observation graph $\mathcal{G} = (\mathcal{S}, \mathcal{E})$ can then be constructed by connecting any two repetitions S_i and S_j by an edge $(i, j) \in \mathcal{E}$ if at least one component is shared between them. We consider the mapping $S_j \mapsto S_i$ on the edge (i, j) as an assignment matrix $\mathbf{X}_{ij} \in \{0, 1\}^{R \times R}$, which indicates which components of S_j match which components of S_i[1]. The entire graph can then be represented by a block matrix $\mathbf{X} \in \{0, 1\}^{NR \times NR}$ as follows:

$$\mathbf{X} = \begin{bmatrix} \mathbf{I} & \mathbf{X}_{12} & \cdots & \mathbf{X}_{1N} \\ \mathbf{X}_{12}^{\mathrm{T}} & \mathbf{I} & \cdots & \mathbf{X}_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{X}_{1N}^{\mathrm{T}} & \mathbf{X}_{2N}^{\mathrm{T}} & \cdots & \mathbf{I} \end{bmatrix} \tag{6.3}$$

[1] If the reproducibility between decompositions of varying rank is of interest, the transformation matrices are described by non-square matrices $\mathbf{X}_{ij} \in \{0, 1\}^{|S_i| \times |S_j|}$. This does not alter the remainder of the method.

The submatrices $\mathbf{X}_{ij}$ encode local, pairwise correspondence. If $(i', j') \notin \mathcal{E}$, i.e., no component is shared between $S_{i'}$ and $S_{j'}$, $\mathbf{X}_{i'j'}$ is an all-zero matrix in (6.3). Note that composite maps, e.g. $S_k \mapsto S_j \mapsto S_i$ can be constructed by matrix multiplication:

$$\mathbf{X}_{ik} = \mathbf{X}_{ij}\mathbf{X}_{jk} \tag{6.4}$$

$$= \mathbf{X}_{ij}\mathbf{X}_{kj}^{\mathrm{T}} \tag{6.5}$$

6.2.3 Repeated decompositions exhibit cycle consistency

If the optimization routine converges to the same local optimum in several repetitions, the associated maps are permutation matrices. Together with (6.5), this immediately leads to the following key observation: composite maps $S_i \mapsto S_j \mapsto S_k \mapsto \cdots \mapsto S_i$ along cycles in the graph should be equal to the identity map. Such a *cycle consistency* property can be exploited as a constraint to correct noisy input maps $\mathbf{X}_{ij}$, a technique which is known as map synchronization (Huang and Guibas, 2013; Chen et al., 2014; Shen et al., 2016; Bajaj et al., 2018). We now aim to offer an intuition for this consistency from three different perspectives. Firstly, note that in a valid graph representation as in (6.3), consistency holds for *any* cycle that can be constructed along the edges in $\mathcal{E}$, and as such it constitutes in fact a quite restrictive set of constraints of the form $\mathbf{X}_{i\ldots} \cdots \mathbf{X}_{kj}\mathbf{X}_{ji} = \mathbf{I}$ (Huang and Guibas, 2013). Secondly, consider the factor matrices $\mathbf{A}_n^{(m)}$ of an arbitrary mode m of all repetitions n, and stack these in a fat matrix $\mathbf{A}^{(m)} = \left[\mathbf{A}_1^{(m)}\, \mathbf{A}_2^{(m)} \ldots \mathbf{A}_N^{(m)}\right]$. A component that appears in multiple repetitions $n_1 \subset n$ then contributes columns, which are the same (up to scaling ambiguity), to the associated $\mathbf{A}_{n_1}^{(m)}$. It is then clear that $NR \geqslant \mathrm{rank}(\mathbf{A}^{(m)}) = \tilde{R} \geqslant R$, where $\tilde{R}$ is the number of distinct components that are found in all repetitions n of the decomposition, which is often substantially lower than NR. This establishes a link between cycle consistency and low-rank structure. Lastly, consider that all repetitions S_i are (partial) instances of a universe S_u holding all $\tilde{R}$ distinct components as 'templates'(Chen et al., 2014). Using (6.4)–(6.5), we can write any map $S_j \mapsto S_i$ as a composite map with the universe as a hub, such that $\mathbf{X}_{ij} = \mathbf{X}_{iu}\mathbf{X}_{uj} = \mathbf{X}_{iu}\mathbf{X}_{ju}^{\mathrm{T}}$, $\forall (i,j) \in \mathcal{E}$. This reveals that the inter-factorization similarity matrix $\mathbf{X}$ admits a positive semi-definite rank-$\tilde{R}$ decomposition (Huang and Guibas, 2013; Chen et al., 2014; Bajaj et al., 2018):

$$\mathbf{X} = \mathbf{X}_u\mathbf{X}_u^{\mathrm{T}} = \begin{bmatrix} \mathbf{X}_{1u} \\ \mathbf{X}_{2u} \\ \vdots \\ \mathbf{X}_{Nu} \end{bmatrix} \begin{bmatrix} \mathbf{X}_{1u}^{\mathrm{T}} & \mathbf{X}_{2u}^{\mathrm{T}} & \cdots & \mathbf{X}_{Nu}^{\mathrm{T}} \end{bmatrix} \tag{6.6}$$

In realistic scenarios, ambiguities and noise in the (local) input maps $\mathbf{X}_{ij}$ can destroy the (global) cycle consistency and hence the low-rankness in (6.6) is only approximate. Our goal is now to retrieve $\mathbf{X}_u$ from the imperfect inter-factorization similarity matrix $\mathbf{X}$, in order to assign each individual factorization's component to the correct universal component in S_u and assess the decomposition's reproducibility.

6.2.4 Clustering through low-rank graph approximation

For a valid (i.e. noiseless) graph encoding, every row of the block $\mathbf{X}_{iu}$ (corresponding to a component of factorization i) consists of all zeros except for a one at the index of the universal component to which that row is assigned. All rows that are assigned to the same universal component are then equal and form a 'clique'. Hence, if we can find an estimate $\hat{\mathbf{X}}_u$ of $\mathbf{X}_u$, up to an orthogonal transformation, we can cluster its rows in order to categorize the components (Bajaj et al., 2018). As in (Bajaj et al., 2018; Shen et al., 2016), we estimate $\mathbf{X}_u$ as the leading $\tilde{R}$ eigenvectors of $\mathbf{X}$, weighted by the square root of their eigenvalues, and propose to use hierarchical clustering based on average linkage, a strategy which is also used in ICASSO (Himberg et al., 2004). The number of distinct components $\tilde{R}$, which also equals the dimension of the resulting feature space, is found by truncating at the largest relative drop $\frac{\lambda_r - \lambda_{r+1}}{|\lambda_r| + |\lambda_{r+1}|}$ in the eigenvalue spectrum (Bajaj et al., 2018). To further investigate which components often co-occur in the local solutions, we characterize each of the $r = 1 \dots \tilde{R}$ estimated cliques by means of the set $\mathcal{S}^{(r)}$ of objects who contribute a component to the clique. The object overlap between any two cliques p and q can then be measured by the Dice coefficient $d(p,q) = 2\,\frac{|\mathcal{S}^{(p)} \cap \mathcal{S}^{(q)}|}{|\mathcal{S}^{(p)}| + |\mathcal{S}^{(q)}|}$. The $\tilde{R} \times \tilde{R}$ overlap matrix can then be used for a second stage of average linkage clustering, to find groups of cliques that share many objects. These groups indicate which components are often simultaneously present in repetitions of the optimization.

6.3 Experiments

We conducted two case studies, based on real data and synthetic data, to evaluate our proposed method. In both cases, we compute the similarity between the rth component of the ith run and the sth component of the jth run as

$$\sigma(r_i, s_j) = \prod_{m \in \mathcal{M}_\sigma} \sigma_m(r_i, s_j) = \left| \prod_{m \in \mathcal{M}_\sigma} \langle \mathbf{a}_{r_i}^{(m)}, \mathbf{a}_{s_j}^{(m)} \rangle \right|^{\frac{1}{|\mathcal{M}_\sigma|}} \tag{6.7}$$

Here, the factor vectors have been normalized to unit norm, and hence $\sigma_m(r_i, s_j)$ is the (exponentiated) mode-m cosine similarity metric, as in (Bro, 1997). The product of mode-wise similarities is taken over $\mathcal{M}_\sigma$, which is the set of all modes in which component reproducibility is expected/desired[2]. Thus, σ as in (6.7) generalizes the metric used in (Himberg et al., 2004). Finally, we binarize the similarities using a threshold of 0.95 and populate all matrices $\mathbf{X}_{ij}$. All decompositions are computed using state-of-the-art Gauss-Newton type algorithms in Tensorlab (Vervliet et al., 2016).

6.3.1 CPD of neuroimaging data

We analyzed multi-subject electroencephalographic (EEG) data, recorded in an MR scanner, using a pipeline similar to that of (Mareček et al., 2017). For every subject $s = 1 \dots 12$, resting-state EEG data from 15 electrodes was converted into a spectrogram $\boldsymbol{\mathcal{T}}^{(s)} \in \mathbb{R}^{15\times 40\times 1350}$, evaluated at 40 1-Hz frequency bins and 1350 1-s windows. We wanted to extract reliable spatial and spectral signatures of modulated resting-state activity (Mareček et al., 2017). In order to find these signatures, we sliced every $\boldsymbol{\mathcal{T}}^{(s)}$ halfway along the third mode and stacked either half into two large tensors $\boldsymbol{\mathcal{T}}^{\text{begin}}, \boldsymbol{\mathcal{T}}^{\text{end}} \in \mathbb{R}^{15\times 40\times (12\cdot\frac{1350}{2})}$, which we normalized as in (Mareček et al., 2017). By decomposing $\boldsymbol{\mathcal{T}}^{\text{begin}}$ and $\boldsymbol{\mathcal{T}}^{\text{end}}$ separately, we expected to find components that were common and may model true resting-state modes, but also components that modeled data-specific fluctuations in either half (Bro, 1997; Mareček et al., 2017) that may be less interesting. We computed a rank-10 CPD of both halves 100 times and applied our algorithm to cluster the resulting components in the spatial and spectral mode (between the halves, no temporal correspondence is expected). We used `cpd_nls` and initialized all factor matrices using i.i.d. Gaussian variables (Vervliet et al., 2016). Using the heuristic in section 6.2.4, we found that $\tilde{R} = 275$, though most of the spectral energy was concentrated in the first few eigenvalues of $\mathbf{X}$, i.e., many of these individual templates appeared only in a few runs, and a small set of templates accounted for most of the runs' components. As shown in Figure 6.1, we found a subset of seven components that were encountered in nearly every run (left column), and two subsets of three components each, which were only found in $\boldsymbol{\mathcal{T}}^{\text{begin}}$ or $\boldsymbol{\mathcal{T}}^{\text{end}}$, respectively (right column). From this split-half experiment, we may conclude that the latter two subsets of components are less reliable, in the sense that they seem

[2] E.g. in the case of the $(L_r, L_r, 1)$-decomposition, a rotational ambiguity in the first two modes remains. Hence, similarities may not be computed for these modes separately, but for entire frontal slices, 'absorbing' the ambiguity. Using a parametrization as in (Acar et al., 2014), equation (6.7) can be extended to account for similarity of components' magnitudes as well, with e.g. a ratio-based metric.

specific to a data segment, whereas the common components may be further inspected. E.g., in the third component we can identify typical alpha (10 Hz) power increases in the occipital regions, at the back of the head. Note, however, that some common components need not be related to resting-state activity: e.g. the sixth component seems to capture two harmonics that may be due to residual MR scanner artifacts.

We also attempted to cluster the components directly based on $\mathbf{X}$ using agglomerative average linkage, as in ICASSO. Since ICASSO leaves the choice of number of clusters to the user, we varied this number (which we will also denote by $\tilde{R}$) between $\tilde{R} = R = 10$ (i.e., assuming that every run of the decomposition will return the same components) and $\tilde{R} = 400$, in increments of 10. To evaluate the clustering performance, we kept track of 1) the average similarity σ of components within the same cluster and 2) the maximal average similarity σ between all components of a cluster and all components of another cluster, as in (Himberg et al., 2004). These two metrics can respectively indicate (per cluster) whether a cluster contains too much variability and would better be split in several smaller, more compact clusters, and whether two clusters are too similar and would better be merged. In Figure 6.2, we set out the median of these values (over clusters) against each other (closer to the origin is better). By means of the low-rank graph framework, our proposed method automatically achieves a good operating point, where the components within every cluster are mutually very similar, and the undesired inter-cluster similarity is lower than for most choices of the hyperparameter $\tilde{R}$ in ICASSO. For low values of $\tilde{R}$, the inevitable variability of the components over runs was neglected, and dissimilar components were wrongly grouped together. On the other hand, when $\tilde{R}$ was overestimated, coherent clusters were wrongfully split into smaller clusters. For a few choices of $\tilde{R}$ (ca. 60–100), ICASSO found the same division in (main) clusters as in Figure 6.1, though this required much more computation and user interaction due to the iterative nature of the tuning procedure.

6.3.2 Coupled matrix-tensor factorization of synthetic data

The $(5 \cdot 100) \times (5 \cdot 100)$ similarity matrix $\mathbf{X}$ had an estimated rank $\tilde{R} = 191$, indicating that the variability was relatively high (approximately 191 distinct components out of 500 in total)[3]. Nevertheless, a small set of seven components appeared relatively frequently (at least in five runs). For ICASSO, we varied

[3]In general, the automatic rank selection method is quite liberal in counting distinct components, and may overestimate $\tilde{R}$. Therefore, it can be good practice to prune the set of distinct components, as to only include and thereafter inspect components that appear in a predefined number of repetitions.

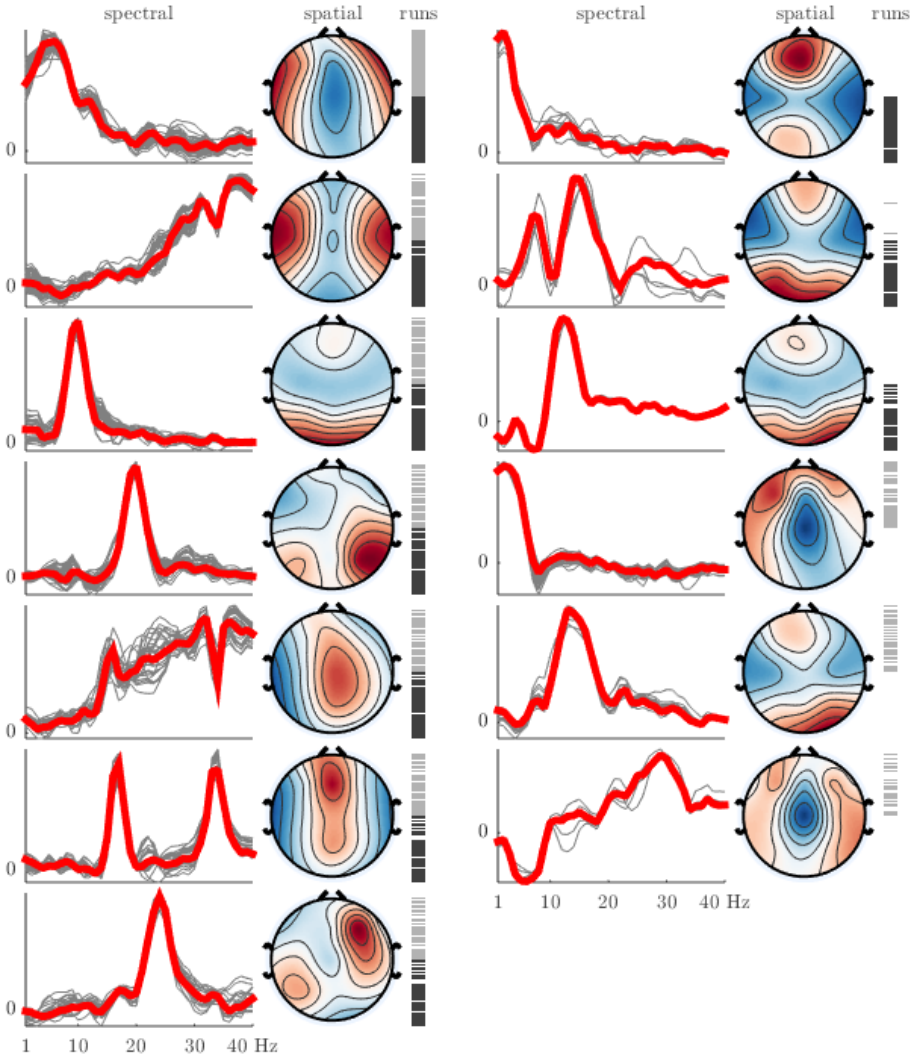


Figure 6.1: Our algorithm successfully clusters spatio-spectral components that appear in nearly all runs of the decomposition and are hence reliable (left column), but also distinguishes components that are specific to the decomposition of each half (light/dark grey) of the data (right column). The mean spectra are superimposed in red. The variability of spatial signatures around their mean (which is plotted) was comparably small as for the spectra.

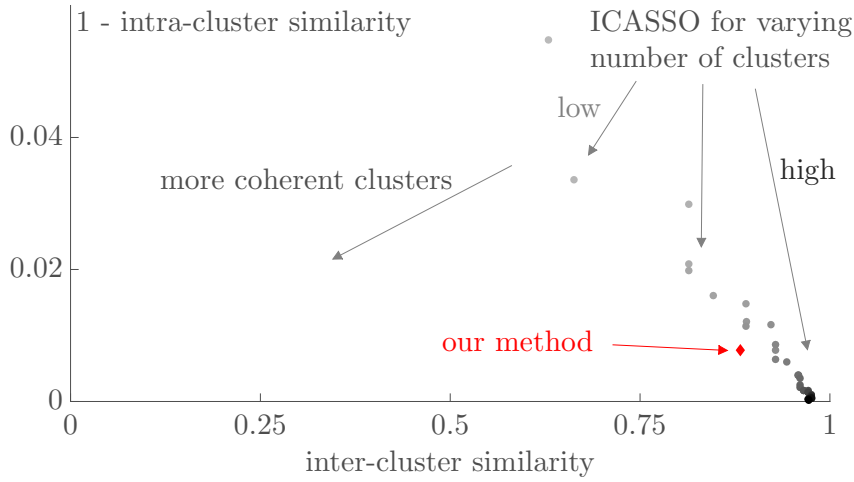


Figure 6.2: When grouping EEG tensor decomposition components in clusters, our algorithm realizes a favourable tradeoff between high intra-cluster similarity and low maximal inter-cluster similarity. This is thanks to the proper (automatic) estimation of the number of clusters $\tilde{R}$ (darker is higher), in an $\tilde{R}$-dimensional component space, which is not possible with ICASSO.

the number of clusters from 5 to 300, in steps of 5. In comparison of both methods, a similar pattern of intra- and inter-cluster similarity as in the previous experiment emerged (see Figure 6.3). The proposed method again finds (much) more coherent clusters than ICASSO, i.e., with a low inter-cluster similarity and a high intra-cluster similarity. We want to stress that 'successful clustering' is not a guarantee of the truthfulness of the found components. We observed that several ground-truth components were almost never captured in an estimated component of a CMTF run. On the other hand, the ground-truth components that were sucessfully found in some repetitions (with a relatively high similarity of approximately 0.5) were most of the time successfully clustered over repetitions.

6.4 Discussion

We have proposed a method to effectively assess the stability of components of coupled or uncoupled matrix and tensor factorizations, in case a non-convex optimization algorithm has to be repeated with multiple initializations. Inspired by earlier work in the area of geometric object matching, we have cast

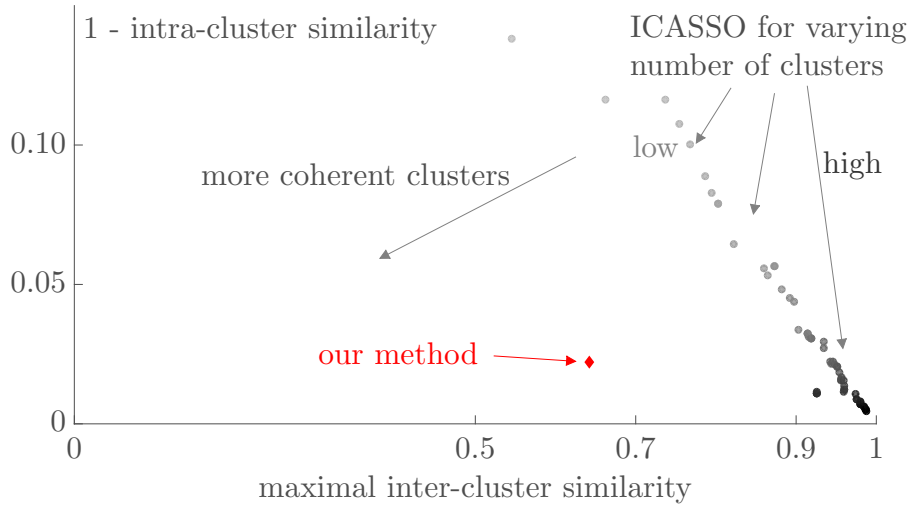


Figure 6.3: Also on synthetic multimodal data, the proposed method successfully groups components of a coupled matrix-tensor factorization in clusters which are more coherent than those of ICASSO.

the problem as a structured graph clustering problem, in which runs of the optimization routine are nodes and the similarity of the resulting components prescribe the edges between them. By leveraging a cycle consistency constraint, a low-rank approximation of this inter-factorization similarity graph can then be computed, which allows an effective clustering of factorization components. We envision that this method may be helpful to data mining practitioners who wish to verify the reproducibility of components without going through the trouble of manual alignment. Two typical (and potentially overlapping) use cases are 1) the estimation and interpretation of individual components from data when there is concern about their dependency on the algorithm's initialization 2) assessing commonality (or stationarity) of components between two distinct, but related datasets, such as in split-half experiments. We tested our method on the tensor decomposition of a real EEG dataset and the coupled matrix-tensor factorization of synthetic data, and showed that it is capable of finding meaningful and accurate grouping patterns over the repetitions of the decomposition. We highlighted its strengths in comparison to the existing ICASSO algorithm: in the investigated cases, our method found equally or more satisfactory clusters, and without the need for manual tuning. Extensions to block term decompositions (BTD) (Sørensen et al., 2015) can be readily accommodated by the new framework by appropriately adjusting expression (6.7), while accounting for ambiguities.

Chapter 7

Augmenting interictal mapping with neurovascular coupling biomarkers by structured data fusion

ABSTRACT | EEG-correlated fMRI analysis is an asymmetrical, mass-univariate approach and cannot capture the inherent, higher order structure in the EEG data, nor multivariate relations in the fMRI data, and it is nontrivial to accurately handle varying neurovascular coupling over patients and brain regions. We aim to overcome these drawbacks in a data-driven manner by means of a novel structured matrix-tensor factorization: the single-subject EEG data (represented as a third-order spectrogram tensor) and fMRI data (represented as a spatiotemporal BOLD signal matrix) are jointly decomposed into a superposition of several sources, characterized by space-time-frequency profiles. In the shared temporal mode, Toeplitz-structured factors account for a spatially specific, neurovascular 'bridge' between the EEG and fMRI temporal fluctuations, capturing the hemodynamic response's variability over brain regions. We show that the extracted source signatures provide a sensitive localization of the ictal onset zone, and, moreover, that complementary localizing information can be derived from the spatial variation of the hemodynamic response. Hence, this multivariate, multimodal factorization provides two useful sets of EEG–fMRI biomarkers, which can inform the presurgical evaluation of epilepsy.

This chapter is unpublished work which has been submitted to NeuroImage, for which the candidate collaborated with Hunyadi, B., Dupont, P., Van Paesschen, W. and Van Huffel, S.

7.1 Introduction

Refractory epilepsy is a neurological disorder suffered by 30% of approximately 50 million epilepsy patients worldwide (World Health Organization, 2019), in which seizures cannot adequately be controlled by anti-epileptic medication. In the preparation of treatment via resective surgery, interictal epileptic discharges (IEDs) can be localized in the brain with simultaneous EEG–fMRI, which provides a good surrogate for mapping the seizure onset zone (Lemieux et al., 2001; Thornton et al., 2010; van Houdt et al., 2013; Grouiller et al., 2011; Zijlmans et al., 2007; An et al., 2013; Khoo et al., 2017). This mapping is often conducted via EEG-correlated fMRI analysis, wherein a reference temporal representation of the IEDs is used to interrogate all brain regions' blood oxygen level dependent (BOLD) signals for significant correlations; voxels for which a statistical threshold is exceeded can then be considered part of the epileptic brain network, along which epileptic seizures are generated and propagated (Gotman, 2008; Lemieux et al., 2001; Zijlmans et al., 2007; Thornton et al., 2010; Salek-Haddadi et al., 2003).

The workhorse for conducting EEG-correlated fMRI analyis has been (Salek-Haddadi et al., 2006)—and will likely continue to be (Poline and Brett, 2012)—the general linear model (GLM) framework (Friston et al., 1994). Over the past years, it has become clear that using the GLM comes with several hurdles, related to the many modeling assumptions, that may reduce its sensitivity or specificity (increasing Type I errors) when violated (Poline and Brett, 2012; Monti, 2011; Lindquist et al., 2009). Remedies for several of these issues are not yet widely applied, or are not yet available.

First of all, the adoption of a relevant representation of IED occurrences to construct a regressor for the design matrix has proven vital to the sensitivity. This aspect has been investigated in (Rosa et al., 2010; Murta et al., 2015; Abreu et al., 2018a). In Chapter 4, we addressed this issue by pre-enhancing the EEG signals using a spatiotemporal filter that is tuned to maximize the signal-to-noise ratio (SNR) of IEDs with respect to the background EEG. We have shown that taking the time-varying power of the filtered EEG leads to a robust regressor, which is more performant than many other types of regressors, including those based on stick functions (Lemieux et al., 2001; Salek-Haddadi et al., 2006), ICA (Formaggio et al., 2011; Abreu et al., 2016b) or EEG synchronization (Abreu et al., 2018a).

Model mismatch may occur due to the unknown neurovascular coupling from electrophysiological phenomena measured on the EEG to hemodynamic variations captured by the BOLD signals, as we have argued in Section 2.5.2. In many papers on EEG-correlated fMRI, a canonical hemodynamic response

function (HRF) based on two gamma density functions is used to translate IED-related temporal dynamics to BOLD fluctuations (Friston et al., 1998). However, there is insurmountable evidence that the HRF is not fixed, but varies substantially over subjects (Aguirre et al., 1998), over brain regions (Handwerker et al., 2004), with age (Jacobs et al., 2008), or even with stress level (Elbau et al., 2018). For the diseased brain, this issue may be even greater: i.e., additional variation, e.g. in brain areas involved in the epileptic network, has been observed compared to healthy controls (van Houdt et al., 2013; Bénar et al., 2002; Jacobs et al., 2009; Lemieux et al., 2008; Grouiller et al., 2010). Plenty of previous research has shown that failing to account for this variability may lead to substantial bias and increased variance of the estimated activation, which in turn inflates Type I and/or Type II error rates (Lindquist et al., 2009; Lindquist and Wager, 2007; Calhoun et al., 2004; Monti, 2011).

Several methods have been devised to deal with this variability. A widely used tactic is to model the HRF as a linear combination of several basis functions. Some popular choices for these bases, which are also supported by open source toolboxes like SPM are the 'informed basis set' (Friston et al., 1998), consisting of the HRF plus its derivative w.r.t. time and its derivative w.r.t. the dispersion parameter (leading to a Taylor-like extension which can capture slight changes in peak onset and width), and the finite impulse reponse (FIR) basis set, in which every basis function fits exactly one sample of the HRF in every voxel (Glover, 1999; Aguirre et al., 1998). Other researchers have aimed to find a basis set by computing a low-dimensional subspace of a large set of 'reasonable' HRFs (Woolrich et al., 2004) or by fitting nonlinear functions to given fMRI data (Lindquist and Wager, 2007), as we have done in Chapter 5 Alternatively, multiple copies of a standard HRF, which differ only in their peak latencies, can be used (Bagshaw et al., 2004). Finally, approaches exist that aim to be immune to differences in neurovascular coupling, such as those based on mutual information (MI), which does not rely on any predefined model or even linearity of the HRF (Ostwald and Bagshaw, 2011; Caballero-Gaudes et al., 2013). Perhaps surprisingly, the authors of (Caballero-Gaudes et al., 2013) found that the results based on MI were often very similar to those based on the informed basis set, leading to the conclusion that the assumption of a linear time-invariant system, as described by the convolution with an appropriate HRF, is sufficiently accurate. Instead, it may be useful to not make abstraction of the variable neurovascular coupling, but rather consider it as an additional biomarker to localize epileptogenic zones (van Houdt et al., 2013). Indeed, in several studies HRFs that deviate from the canonical model were found in regions of the epileptic network (Bénar et al., 2002; Lemieux et al., 2008; Hawco et al., 2007; Pittau et al., 2011; Jacobs et al., 2009; Moeller et al., 2008; van Houdt et al., 2013). Several hypotheses have been postulated to explain this varability, including altered autoregulation due to higher metabolic demand

following (inter)ictal events (Schwartz, 2007), vascular reorganization near the epileptogenic region (Rigau et al., 2007), or the existence of pre-spike changes in neuro-electrical activity which are not visible on EEG and which culminate in the IED (Jacobs et al., 2009). It is thus an opportunity to map not only regions with statistically significant BOLD changes in response to IEDs, but also the spatial modulation of the HRF itself, in order to discover regions where an affected HRF shape may provide additional evidence towards the epileptic onset.

The previous considerations indicate that it is difficult to meet all assumptions in the general linear model, which may compromise inference power (Lindquist et al., 2009; Handwerker et al., 2004; Monti, 2011). Data-driven alternatives may relieve this burden, since they adapt to the complexity of the data more easily compared to model-based approaches, and are especially suited for exploratory analyses (Mantini et al., 2007; Marecek et al., 2016). Blind Source Separation (BSS) techniques consider EEG and/or fMRI data to be a superposition of several 'sources' of physiological activity and nonphysiological influences. Based on the observed data alone, BSS techniques are used to estimate both the sources and the mixing system, by means of a factorization of the data into two (or more) factor matrices, holding sources or mixing profiles along the columns. They naturally allow a symmetrical treatment of EEG and fMRI data, enabling true fusion of both modalities (Valdes-Sosa et al., 2009; Lahat et al., 2015; Calhoun et al., 2009), which is in contrast to EEG-correlated fMRI, where EEG-derived IEDs inform the fMRI analysis. Furthermore, BSS techniques naturally accommodate higher-order representations of the data in the form of tensors or multiway arrays, which can capture the rich structure in the data. Indeed, measurements of brain activity inherently vary along several modes (subjects, EEG channels, frequency, time, ...), which cannot be represented using matrix-based techniques like ICA without loss of structure or information (Sidiropoulos et al., 2017; Lahat et al., 2015; Kolda and Bader, 2009; Acar et al., 2007). Tensor-based BSS techniques have been used to mine unimodal EEG data by decomposing third-order spectrograms (channels $\times$ time points $\times$ wavelet scales) into several 'atoms' (also coined 'components' or 'sources'), each with a distinct spatial, temporal and spectral profile/signature (Miwakeichi et al., 2004; Mørup et al., 2006; Marecek et al., 2016), with successful application in seizure EEG analysis (Acar et al., 2007; De Vos et al., 2007). While a tensor extension of ICA for group fMRI data (in the form of subjects $\times$ time points $\times$ voxels) exists (Beckmann and Smith, 2005), matrix representations of fMRI remain dominant for single-subject analyses. Coupled BSS techniques can estimate components which are shared between both modalities, providing a characterization in both domains (Hunyadi et al., 2017). For example, in (Acar et al., 2017a, 2019; Hunyadi et al., 2016; Chatzichristos et al., 2018), multi-subject EEG and fMRI data have

been analyzed using coupled matrix-tensor factorization (CMTF), wherein the 'subjects' factor is shared between the EEG trilinear tensor decomposition and the fMRI matrix decomposition. In (Hunyadi et al., 2016), the resulting factor signatures revealed onset and propagation zones of an interictal epileptic network that was common over patients, as well as the modulation of the default-mode network (DMN) activity. Also single-subject data can be decomposed into distinct components, using a shared temporal factor for EEG and fMRI. This requires the use of a model of the neurovascular coupling, to ensure temporal alignment of EEG and BOLD dynamics. In (Martínez-Montes et al., 2004), a fixed canonical HRF was used, followed by multiway partial least squares to extract components with spatial, temporal, and spectral signatures. In Chapter 5, we proposed an extension to this technique, where a subject-specific HRF is co-estimated from the available data, along with the components.

Here, we extend this latter technique in order to account not only for subject-wise variation of the HRF, but also capture variations over brain regions. This results in a highly structured CMTF (sCMTF) of the interictal multimodal data, in which HRF basis functions and spatial weighting coefficients are estimated along with spatial, spectral and temporal signatures of components. By preprocessing the EEG using the data-driven filters from Section 4.2.3, we aim to maximize the sensitivity in mapping the interictal discharges. We analyze whether the estimated spatial modulation of the HRF is a viable biomarker when localizing the ictal onset zone, besides the BOLD spatial signatures themselves.

7.2 Methods and materials

7.2.1 Patient group

We use data of twelve patients with refractory focal epilepsy, whom we previously studied in Chapter 4, cfr. Section 4.2.1. In summary, these twelve patients were cases for whom a full presurgical evaluation using multiple modalities had been performed, interictal spikes were observed during the EEG–fMRI recording session, and all available clinical evidence (from ictal EEG, SPECT, ...) was concordant regarding the epileptic focus. For the complete data on the patients' etiology, and the number of observed IEDs, we refer to Table 4.2.

7.2.2 Data acquisition and preprocessing

We refer to Section 4.2.2 for details on the EEG and fMRI acquisition, and only discuss here the unique preprocessing steps we apply.

We regress out covariates of no interest from the fMRI data. These include the six motion-correction parameters, and the average time series in the white matter and the lateral ventricles (cerebrospinal fluid). If necessary, also boxcar regressors are added at moments of substantial scan-to-scan head movement (larger than 1 mm based on the translation parameters). To reduce remaining nuisance effects, we regress out the first five principal components of the BOLD time series within the cerebrospinal fluid and white matter regions (Behzadi et al., 2007).

Subsequently, the BOLD time series are band-pass filtered between 0.008–0.20 Hz using the CONN toolbox (Whitfield-Gabrieli and Nieto-Castanon, 2012).

The dimensionality of the fMRI data is reduced by means of an anatomical parcellation of the brain. The initial $79 \times 95 \times 68$ images are segmented into regions-of-interest (ROIs) according to the Brainnetome atlas, which consists of 246 parcels in the grey matter(Fan et al., 2016). For every ROI, one BOLD time series is constructed as the average of the time series of all voxels within the ROI.

Further customized EEG preprocessing steps are treated in Sections 7.2.3 and 7.2.4.

If multiple acquisition runs (within the same recording session) had been done, the EEG and fMRI data of the different runs are temporally concatenated.

7.2.3 Multi-channel Wiener filtering for spatio-temporal EEG enhancement

In Chapter 4, we have shown that pre-enhancing the EEG signals using a data-driven, spatiotemporal filter that is tuned to maximize the signal-to-noise ratio (SNR) of IEDs with respect to the background EEG and artifacts, leads to a BOLD predictor that is more performant than many other predictors, including those based on simple stick functions (Lemieux et al., 2001; Salek-Haddadi et al., 2006), ICA (Formaggio et al., 2011; Abreu et al., 2016b) or EEG synchronization (Abreu et al., 2018a). This pre-enhancement strategy based on multi-channel Wiener filters (MWF) has error-correcting capabilities and produces an IED representation that improves the localization sensitivity of EEG-correlated fMRI.

In brief, the MWF is estimated by first performing time-delay embedding of the multi-channel EEG signals $\mathbf{x}[t]$, leading to an extended multi-channel, multi-lag

signal

$$\tilde{\mathbf{x}} = \begin{bmatrix} \mathbf{x}[t-\tau] \\ \vdots \\ \mathbf{x}[t-1] \\ \mathbf{x}[t] \\ \mathbf{x}[t+1] \\ \vdots \\ \mathbf{x}[t+\tau] \end{bmatrix} \tag{7.1}$$

and subsequently computing the filter coefficients as

$$\hat{\mathbf{W}} = \mathbf{R}_{xx}^{-1}(\mathbf{R}_{xx} - \mathbf{R}_{nn}) \,, \tag{7.2}$$

where $\mathbf{R}_{xx} = \mathrm{E}\{\tilde{\mathbf{x}}\tilde{\mathbf{x}}^{\mathrm{T}} | H = 1\}$ is the covariance matrix of the EEG observed during annotated IED segments ($H = 1$), and $\mathbf{R}_{nn} = \mathrm{E}\{\tilde{\mathbf{x}}\tilde{\mathbf{x}}^{\mathrm{T}} | H = 0\}$ is the covariance matrix of the EEG outside of IED segments ($H = 0$). For the full derivation, we refer the reader to (Somers et al., 2018). The EEG signals are then filtered as $\hat{\mathbf{W}}^{\mathrm{T}}\tilde{\mathbf{x}}$. Due to the extension with lagged copies of the signals, channel-specific finite impulse response filters are found. Hence, $\hat{\mathbf{W}}^{\mathrm{T}}\tilde{\mathbf{x}}$ is a set of spatiotemporally filtered output signals, in which IED-like waveforms are preserved while other waveforms, which are not specific to epilepsy, are supressed[1].

We train MWFs for each patient individually, using the same parameter settings as in Chapter 4: we embed the EEG signals using $\tau = 4$ positive and negative lags and compute the final filter using the generalized eigenvalue decomposition, which ensures the positive definiteness property of the subtracted covariance matrix in (7.2) (Somers et al., 2018).

7.2.4 Higher-order data representation

To preserve the intrinsic multiway nature of the data, we represent the preprocessed EEG and fMRI as a tensor and matrix respectively, which are subsequently factorized jointly. This approach differs from the mass-univariate treatment in the traditional GLM, where each voxel is treated individually, and only 'flattened' EEG time courses can be entered as regressors. Since epilepsy is manifested with considerable variability between patients, we handle the multimodal data of each patient separately.

[1] Subsampling of the rows of $\hat{\mathbf{W}}^{\mathrm{T}}\tilde{\mathbf{x}}$ is needed to reverse the time-delay embedding, collapsing it into a multi-channel output signal at time t only.

Spatio-temporal-spectral tensor representation of EEG

We adopt a tensorization strategy based on time-frequency transformation of the EEG data to third-order spectrograms (time points × frequencies × channels). After the pre-enhancement step described in Section 7.2.3, we create a spectrogram using the Thomson multitaper method (Thomson, 1982), applied on nonoverlapping EEG segments with a length equal to one repetition time (TR) of the fMRI acquisition. The squared Fourier magnitudes are averaged into 1 Hz bins, from 1 Hz to 40 Hz. Hence, for every EEG channel, we obtain a spectrogram which is synchronized to the fMRI time series. The time points × frequencies × channels spectrogram, $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I_s \times I_g \times I_m}$ is further normalized to equalize the influence of each channel and each frequency: the mean of each mode-1 fiber (holding one time series of squared amplitudes) is subtracted, and afterwards scale normalization is iteratively carried out over the second and third mode to ensure that each channel and each frequency bin contributed the same amount of variance to the data (Bro, 1997).

Spatio-temporal matrix representation of fMRI

The average BOLD time series are stacked in a time points × ROIs matrix $\mathbf{Y} \in \mathbb{R}^{I_s \times I_v}$, where $I_v = 246$ ROIs. We normalize each ROI's time series by subtracting its mean and dividing by its standard deviation.

Neurovascular coupling in the temporal mode

EEG and fMRI data are acquired simultaneously per subject, and are thus naturally coregistered along the 'time' mode. This is captured in a temporal factor matrix that is common between the EEG factorization and the fMRI factorization. However, the electrophysiological changes that are picked up by EEG vary on a much more rapid time scale than the sluggish BOLD fluctuations that (indirectly) correspond to the same neural process. The neurovascular coupling that describes the relation between these two complementary signals can be described by a convolution with an HRF[2].

In Chapter 5, we developed a CMTF model in which the HRF itself is parametrically estimated from the data, and a matrix multiplication with

[2]In this thesis, we use the term 'neurovascular coupling' to describe the coupling characteristic between EEG and fMRI temporal dynamics, and make the silent assumption that this characteristic is a proxy/surrogate for 'neurovascular coupling' as it is understood in neuroscience: the model that describes BOLD changes in response to electrical neural 'events', which take the form of local field potentials at the synapses.

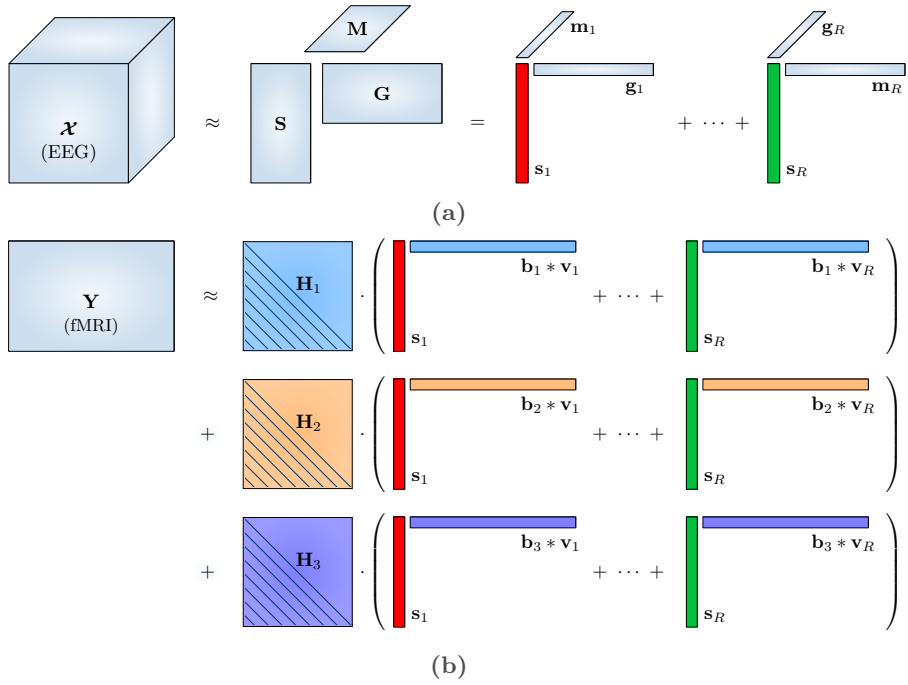


Figure 7.1: Structured coupled matrix-tensor factorization (sCMTF) of EEG and fMRI data can reveal neural sources that are encoded in both modalities, as well as capture the varying neurovascular coupling between the electrophysiological and BOLD changes. **((a))** The EEG signals vary over time points × frequencies × electrodes. The resulting third-order spectrogram tensor $\boldsymbol{\mathcal{X}}$ is factorized according to (7.8) into R rank-1 components, which each consist of a temporal signature $\mathbf{s}_r$, a spectral signature $\mathbf{g}_r$ and a spatial signature $\mathbf{m}_r$. **((b))** The fMRI data consist of the average BOLD signal in different brain parcels or regions of interest (ROIs), represented in a time points × ROI matrix $\mathbf{Y}$, and are factorized according to (7.11). Neurovascular coupling is modeled as a convolution of the EEG temporal dynamics with a ROI-specific hemodynamic response function (HRF), as in (7.11)–(7.13). In this example, each local HRF is represented as a linear combination (encoded by coefficients $\mathbf{b}_k$) of $K = 3$ optimized basis functions, each populating a Toeplitz matrix $\mathbf{H}_k$ which implements a convolution through matrix multiplication with the temporal signatures $\mathbf{s}_r$. Afterwards, each smoothed component r is spatially weighted by a signature $\mathbf{v}_r$. This is accomplished by the elementwise product $\mathbf{b}_k * \mathbf{v}_r$ of the HRF basis function-specific coefficients $\mathbf{b}_k$ and the component-specific amplitudes $\mathbf{v}_r$.

Toeplitz structure implements the HRF convolution, as proposed in (Valdes-Sosa et al., 2009). In the same chapter, we hinted towards an extension based on multiple basis functions to account for the variability of the HRF over brain regions. In the following, we assume that the time course of each EEG source is convolved with an a priori unknown, ROI-specific HRF, which is a superposition of K parametrized basis functions, which leads to a modelled contribution of this source to the ROI's BOLD signal. Hence, in every ROI i_v, the modeled (unscaled) BOLD time course $\mathbf{z}_{i_v}^{(r)}$ of the r-th neural source is

$$\mathbf{z}_{i_v}^{(r)} = \mathbf{H}_{i_v}\mathbf{s}_r \tag{7.3}$$

$$= \sum_{k=1}^{K} b_{k,i_v}\mathbf{H}_k\mathbf{s}_r \tag{7.4}$$

$$= \sum_{k=1}^{K} b_{k,i_v}\mathcal{T}(\mathbf{h}_k)\mathbf{s}_r \tag{7.5}$$

$$= \sum_{k=1}^{K} b_{k,i_v}\mathcal{T}(\mathcal{H}(\theta_k))\mathbf{s}_r\ . \tag{7.6}$$

Here, $\mathbf{s}_r$ is a factor vector holding the time course of the r-th EEG source; $\mathcal{H}$ is an operator that transforms a set of parameters θ_k into a full HRF, represented as a vector $\mathbf{h}_k$; $\mathcal{T}$ is an operator that transforms $\mathbf{h}_k$ into a Toeplitz matrix $\mathbf{H}_k$ by populating the main and lower diagonals with the HRF samples (see also 7.3.1); b_{k,i_v} is the weight for the k-th HRF basis function in the i_v-th ROI; $\mathbf{H}_{i_v}$ is the Toeplitz matrix holding the total HRF in the i_v-th ROI[3].

This time course $\mathbf{z}_{i_v}^{(r)}$ is conceptually equivalent to a regressor in the GLM's design matrix. We treat the HRF parameter sets $\theta_k, k = 1 \dots K$ as unknown variables, which need to be fitted to the data at hand (Lindquist and Wager, 2007). By parametrizing each basis function, we embed protection against nonsensical HRF shapes, and against overfitting, since the number of parameters to be estimated is greatly reduced compared to the FIR basis in (Glover, 1999; Aguirre et al., 1998). We employ double-gamma HRFs, i.e., each HRF basis function is described by five parameters as $h(t) = f(t;\boldsymbol{\theta}) = \Gamma(\theta_1)^{-1} \cdot \theta_2^{\theta_1} t^{\theta_1 - 1} e^{-\theta_2 t} - \theta_5 \Gamma(\theta_3)^{-1} \cdot \theta_4^{\theta_3} t^{\theta_3 - 1} e^{-\theta_4 t}$.

[3]The HRF in every ROI does not depend on r, and is hence shared between all sources. In (Makni et al., 2008; Vincent et al., 2010; Pedregosa et al., 2015), such a constraint has been used to promote robust estimation.

7.2.5 Coupled matrix-tensor factorization of EEG and fMRI

After tensorization, we jointly decompose the EEG tensor $\boldsymbol{\mathcal{X}}$ and the fMRI matrix $\mathbf{Y}$ into a set of distinct sources.

The third-order EEG spectrogram is approximated by a sum of R rank-1 terms according to the trilinear canonical polyadic decomposition (CPD) (also referred to as Parallel Factor Analysis (PARAFAC)) as in (Miwakeichi et al., 2004; Marecek et al., 2016; Martínez-Montes et al., 2004) (cfr. also Chapter 5). Each rank-1 term $\mathbf{s}_r \otimes \mathbf{g}_r \otimes \mathbf{m}_r$ describes a source (also called 'component') in terms of an outer product ($\otimes$) of a temporal, spectral, and spatial signature, respectively. Unlike matrix decompositions, the decomposition of a higher-order tensor into a set of sources is unique, up to scaling and permutation ambiguities, without imposing constraints (under mild conditions).

The fMRI matrix is similarly approximated as a sum of rank-1 terms. Coupling arises from the temporal signatures $\mathbf{s}_r$, which are shared between the EEG and fMRI factorization. After processing through a hemodynamic system (as described in Section 7.2.4), each source's BOLD temporal signature is weighted with a spatial signature $\mathbf{v_r}$.

To accommodate additional structured variation in the fMRI data, that is not related to electrophysiological dynamics, we allow a low-rank term to the fMRI factorization which is not coupled with the EEG factorization. We have empirically found that such a low-rank term can capture structured noise, preventing it from biasing the estimation of the parameters which are coupled with the EEG factorization.

The full sCMTF model is then described as:

$$\boldsymbol{\mathcal{X}} = \hat{\boldsymbol{\mathcal{X}}} + \boldsymbol{\mathcal{E}}_x \tag{7.7}$$

$$= \sum_{r=1}^{R} \mathbf{s}_r \otimes \mathbf{g}_r \otimes \mathbf{m}_r + \boldsymbol{\mathcal{E}}_x \tag{7.8}$$

$$= [\![\mathbf{S}, \mathbf{G}, \mathbf{M}]\!] + \boldsymbol{\mathcal{E}}_x \tag{7.9}$$

$$\mathbf{Y} = \hat{\mathbf{Y}} + \mathbf{E}_y \tag{7.10}$$

$$= \sum_{r=1}^{R} \sum_{k=1}^{K} (\mathbf{H}_k \mathbf{s}_r) \otimes (\mathbf{b}_k * \mathbf{v}_r) + \sum_{q=1}^{Q} \mathbf{n}_q \otimes \mathbf{p}_q + \mathbf{E}_y \tag{7.11}$$

$$= \sum_{k=1}^{K} (\mathbf{H}_k \mathbf{S}) \left(\mathbf{b}_k^{\mathrm{T}} \odot \mathbf{V}^{\mathrm{T}}\right) + \mathbf{N}\mathbf{P}^{\mathrm{T}} + \mathbf{E}_y \tag{7.12}$$

$$= \left[\mathbf{H}_1 \mathbf{S} \ \ldots \ \mathbf{H}_K \mathbf{S}\right] \cdot \left[\mathbf{B}^{\mathrm{T}} \odot \mathbf{V}^{\mathrm{T}}\right] + [\![\mathbf{N}, \mathbf{P}]\!] + \mathbf{E}_y \ , \tag{7.13}$$

where $\hat{\boldsymbol{\mathcal{X}}}$ and $\hat{\mathbf{Y}}$ are the low-rank approximations; $\boldsymbol{\mathcal{E}}_x$ and $\mathbf{E}_y$ hold the residuals of both factorizations; $[\![\mathbf{S}, \mathbf{G}, \mathbf{M}]\!]$ describes the CPD model composed of factor matrices $\mathbf{S} \in \mathbb{R}^{I_s \times R}$, $\mathbf{G} \in \mathbb{R}^{I_g \times R}$, $\mathbf{M} \in \mathbb{R}^{I_m \times R}$, which hold the temporal, spectral and EEG spatial signatures in the columns; the HRF matrices $\mathbf{H}_k$ are constructed as in (7.3)–(7.6); $\mathbf{V} \in \mathbb{R}^{I_v \times R}$ is the fMRI spatial factor matrix; $\mathbf{B} \in \mathbb{R}^{I_v \times K}$ is the HRF basis coefficient matrix; $[\![\mathbf{N}, \mathbf{P}]\!]$ is a rank-Q term to capture fMRI-only structured nuisance; $*$ denotes the Hadamard or elementwise product; $\odot$ denotes the Khatri–Rao product (Kolda and Bader, 2009).

Note that the coupled part of $\mathbf{Y}$ is described by RK nonindependent rank-1 terms, or equivalently, by K rank-R block terms. Namely, each rank-1 term $(\mathbf{H}_k \mathbf{s}_r) \otimes (\mathbf{b}_k * \mathbf{v}_r)$ describes the convolution of the r-th source's temporal signature with the k-th basis function, after which a spatial loading with vector $(\mathbf{b}_k * \mathbf{v}_r)$ is performed; in all ROIs, there is one source-nonspecific relative weight for the basis function (captured in $\mathbf{b}_k$), and source-specific amplitudes (captured in $\mathbf{v}_r$). To limit the degrees of freedom, the Khatri–Rao product in (7.12)–(7.13) expresses that the HRF is shared among all sources, which is a constraint that has earlier been used for this purpose (Pedregosa et al., 2015). Hence, there are not RKI_v spatial coefficients, but $(R + K)I_v$, i.e., K basis function weights and R source amplitudes in all I_v ROIs[4]. This model is

[4]In this way, the Khatri–Rao structure also breaks the curse of dimensionality in the fMRI decomposition if either the number of sources R or the number of basis functions K is high (or both).

depicted in Figure 7.1, omitting $[\![\mathbf{N}, \mathbf{P}]\!]$ to not overload the diagram.

We estimate all parameters of the model in (7.8) and (7.11) by iteratively minimizing the cost function J, composed of two data fitting terms and two regularization terms as in (Acar et al., 2014):

$$\begin{aligned} J(\mathbf{S}, \mathbf{G}, \mathbf{M}, \mathbf{B}, \mathbf{V}, \boldsymbol{\theta}) = \; & \beta_x \left\| \boldsymbol{\mathcal{X}} - \hat{\boldsymbol{\mathcal{X}}} \right\|_F^2 \\ & + \; \beta_y \left\| \mathbf{Y} - \hat{\mathbf{Y}} \right\|_F^2 \\ & + \; \gamma_x \|\boldsymbol{\lambda}_x\|_1 \; + \; \gamma_y \|\boldsymbol{\lambda}_y\|_1 \end{aligned} \tag{7.14}$$

$$\begin{aligned} \text{s.t. } & \mathbf{H}_k = \mathcal{T}(\mathbf{h}_k) = \mathcal{T}(\mathcal{H}(\boldsymbol{\theta}_k)) \\ & \boldsymbol{\lambda}_x = \begin{bmatrix} \lambda_{x,1} & \dots & \lambda_{x,R} \end{bmatrix}^{\mathrm{T}} \\ & \lambda_{x,r} = \|\mathbf{s}_r\|_2 \cdot \|\mathbf{g}_r\|_2 \cdot \|\mathbf{m}_r\|_2 \\ & \boldsymbol{\lambda}_y = \begin{bmatrix} \lambda_{y,1} & \dots & \lambda_{y,R} \end{bmatrix}^{\mathrm{T}} \\ & \lambda_{y,r} = \sum_{k=1}^{K} \|\mathbf{b}_k * \mathbf{v}_r\|_2 \; , \end{aligned} \tag{7.15}$$

where the squared Frobenius norm $\|\boldsymbol{\mathcal{A}}\|_F^2$ of a tensor $\boldsymbol{\mathcal{A}}$ is the sum of its squared elements; $\|\mathbf{a}\|_2$ and $\|\mathbf{a}\|_1$ denote the Euclidean or ℓ_2-norm and the ℓ_1-norm or sum of the elements' absolute values of a vector $\mathbf{A}$, respectively; β_x, β_y, γ_x and γ_y are positive weights; $\boldsymbol{\lambda}_x$ and $\boldsymbol{\lambda}_y$ are vectors which hold the amplitudes with which each source is expressed in the EEG and fMRI data, respectively. The squared Frobenius norms of the residuals promote a good fit of the low-rank approximations to the data, while the ℓ_1-regularization terms penalize excessive source amplitudes and promote a parsimonious[5] model, similar to the group-LASSO method (Acar et al., 2014; Yuan and Lin, 2006). At the same time, the latter penalty also tends to prevent the occurrence of degenerate terms (Bro, 1997). We minimize (7.14) using the Structured Data Fusion framework in Tensorlab (Sorber et al., 2015; Vervliet et al., 2016), using a quasi-Newton method based on a limited-memory BFGS algorithm, for 50 independent initializations (see 7.3 for details regarding the optimization

[5]The sparsity-promoting properties of the LASSO penalty are most useful in the context of an underdetermined system, with more coefficients than observations, e.g. in dictionary learning. Here, the problem is heavily overdetermined, and we do not expect that the amplitudes $\boldsymbol{\lambda}_x$ an $\boldsymbol{\lambda}_y$ go exactly to zero. However, the ℓ_1-penalty is still a useful heuristic to avoid degenerate components in the EEG's CP decomposition.

procedure and parameters). After convergence, each set of estimated factors needs to be calibrated to remove certain ambiguities, and model selection must be performed to pick the best solution, with an appropriate R (see 7.4 for details).

7.2.6 Statistical inference

We create statistical nonparametric maps (SnPMs) of the obtained spatial signatures $\mathbf{v}_r$ to determine which ROIs sources are significantly (de)activated in relation to the found sources (Nichols and Holmes, 2002; Waites et al., 2005). To this end, we use permutation-based inference, in which the spatial signatures $\mathbf{v}_r$ are compared against their empirically derived distributions, which are obtained via resampling of the fMRI data while freezing the other estimated sCMTF factors[6]. To account for serial correlations in the fMRI time series, we use the robust wavelet-based resampling approach in (Bullmore et al., 2001) to ensure exchangeability and to preserve spatiotemporal correlation structure of the original data in the produced surrogate datasets. For each fMRI dataset and every sCMTF solution, we generate $L = 250$ surrogate fMRI $\tilde{\mathbf{Y}}^{(l)}$ datasets using the procedure in (Bullmore et al., 2001). We resample only the adjusted data $\mathbf{Y} - \mathbf{N}\mathbf{P}^{\mathrm{T}}$, i.e., after removing the components which model variation specific to the fMRI data. We perform inference on a pseudo t-statistic, which we compute for every ROI and for every source as follows:

1. construct a local 'design matrix' with all estimated temporal signatures as in (7.3): $\mathbf{D}_{i_v} = \left[\mathbf{z}_{i_v}^{(1)} \dots \mathbf{z}_{i_v}^{(R)}\right]$,
2. find the new 'betas' by solving $\boldsymbol{\beta}_{i_v}^{(l)} = \mathbf{D}_{i_v}^{\dagger} \tilde{\mathbf{y}}_{i_v}^{(l)}$, $\forall\ l$,
3. convert the betas to a t-statistic per source by dividing them by their estimated standard deviation (see (Friston et al., 1994; Poline and Brett, 2012)).

Through this procedure, we obtain L-point empirical null distributions for every source and every ROI. We set the significance threshold as to control the familywise error (FWE) rate at $\alpha = 0.05$, according to the maximum statistic procedure outlined in (Nichols and Hayasaka, 2003). That is, for every source r, we form the empirical distribution of the maximal t-statistic over all I_v ROIs, and determine source-specific thresholds $T_{(1-\alpha)}^{(r)}$ as the 95%-percentile (to test

[6] Under the null hypothesis of no significant BOLD effect related to the EEG dynamics, the fMRI data may be temporally reshuffled without a significant loss of fit to the EEG dynamics in $\mathbf{s}_r$.

for activation) and $T_{(\alpha)}^{(r)}$ as the 5%-percentile (to test for deactivation). Finally, we obtain statistical maps for all sources r by applying these thresholds to the original spatial signatures $\mathbf{v}_r$, which can be considered as the betas of the unshuffled data.

Furthermore, we create a map of the HRF variability over ROIs. For every ROI, we assess how 'unusual' the local HRF is, by measuring its calibrated distance in HRF space to all other ROIs' HRFs. We use two metrics to quantify this: HRF *extremity* and HRF *entropy*.

We define the extremity of a specific ROI's HRF as one minus the average of the absolute values of the Pearson correlation between the HRF waveform and all other ROIs' HRFs waveforms. I.e., for the j_v-th ROI, the extremity is computed as

$$\text{extremity}(j_v) = 1 - \frac{1}{I_v - 1} \sum_{i_v \neq j_v} |\text{corr}(h_{i_v}, h_{j_v})| \tag{7.16}$$

Only the first twenty samples (∼50 s) are considered. Note that the extremity does not change if the (global, not samplewise) sign (polarity) of one or more HRFs changes.

Analogously, we define the entropy of a specific ROI's HRF as the negative logarithm of the probability of this HRF, conditional on all other ROIs' HRFs. For example, we first estimated a probability density in HRF space based on all other ROIs' HRFs, and then evaluated this density at the HRF of the ROI under inspection. From every HRF, we considered the first twenty samples, and then estimated a nonparametric multivariate kernel density in 20-dimensional space, by placing a multivariate Gaussian probability kernel at the location of each HRF except one. We made this entropy metric insensitive to the signs of the HRFs, by extending the set of HRF waveforms by their flipped counterparts, and computing the nonparametric density using the resulting $2(I_v - 1)$ HRFs in a leave-one-ROI-out fashion.

$$\begin{aligned} &\text{entropy}(j_v) = \\ &-\log\left(\frac{1}{2(I_v-1)} \sum_{i_v \neq j_v} \left(K(h_{i_v}, h_{j_v}; \mathbf{\Sigma}) + K(h_{i_v}, -h_{j_v}; \mathbf{\Sigma})\right)\right), \end{aligned} \tag{7.17}$$

in which $K(h_{i_v}, h_{j_v}; \mathbf{\Sigma})$ is a Gaussian kernel distance, which is proportional to

$$\exp\left(-\frac{1}{2}(\mathbf{h}_{i_v} - \mathbf{h}_{j_v})^{\mathrm{T}} \mathbf{\Sigma}^{-1} (\mathbf{h}_{i_v} - \mathbf{h}_{j_v})\right), \tag{7.18}$$

in which $\mathbf{h}_{i_v}$ and $\mathbf{h}_{j_v}$ are column vectors that store the twenty first samples of the HRFs $h_{i_v}(t)$ and $h_{j_v}(t)$, and $\mathbf{\Sigma}$ is a diagonal covariance or bandwidth

matrix. We used Silverman's heuristic to set the kernel bandwidths for each individual dimension, corresponding to one HRF time sample (Silverman, 1986). I.e., the n-th bandwidth σ_{nn}^2, which corresponds to the HRF amplitudes at sample n, is given by

$$\sigma_{nn}^2 = \left(\frac{4}{20+2}\right)^{\frac{2}{20+4}} (2(I_v - 1))^{\frac{-2}{20+4}} s_n^2 \,, \tag{7.19}$$

in which s_n^2 is the observed variance (over ROIs) of the HRFs' amplitudes at the n-th sample.

Both for the pseudo t-maps as for the HRF extremity and entropy maps, we furthermore limit the inspection to the 20 ROIs with the highest values, if applicable.

An end-to-end overview of our pipeline, from data preprocessing up until statistical inference, is depicted in Figure 7.2.

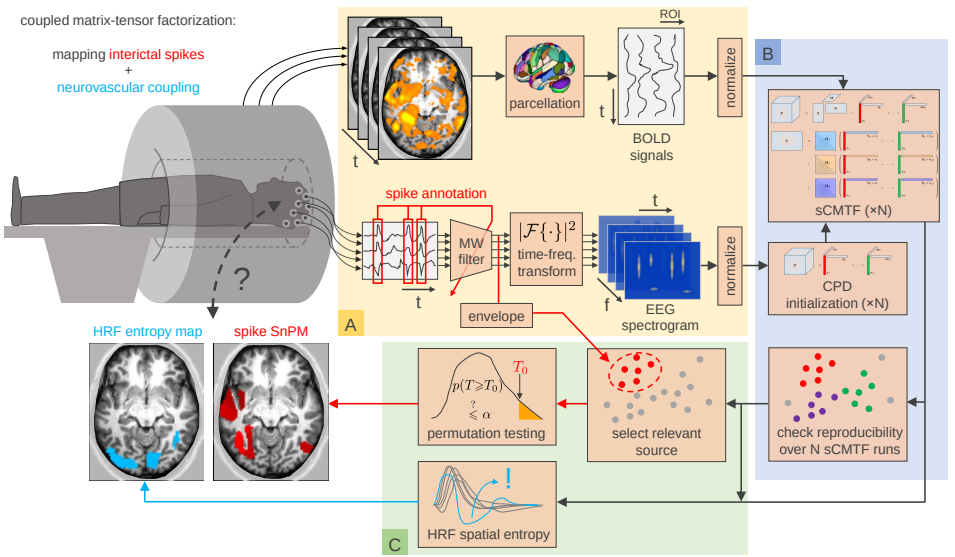


Figure 7.2: Interictal EEG and fMRI data are analyzed via structured coupled matrix-tensor factorizations (sCMTF), which reveals both spatial localization of interictal discharges (spikes), and also localized deviations in neurovascular coupling between electrical and BOLD fluctuations. **(a)** fMRI and EEG data are first separately preprocessed into a time points × regions of interest (ROIs) matrix, and a channels × time points × frequencies tensor, after MWF enhancement, respectively. **(b)** The sCTMF of the EEG and fMRI data reveals temporally, spatially and spectrally resolved components, and captures spatially varying hemodynamic response functions (HRFs) (cfr. Figure 7.1). The computation is repeated N times, from different initializations, and the stability of the resulting factors over runs is assessed. **(c)** Statistical images are created for the corresponding sCMTF factors. The spike-related component is picked based on the correlation to the filtered EEG signals's broadband power. An SnPM of this component is created, revealing co-activated ROIs in a pseudo-t-map (red). For every ROI, the entropy of the HRF is computed by assessing its likelihood under the distribution of all other ROIs' HRFs, and a map of this metric is constructed (blue) to reveal localized HRF abnormalities. Both maps can be used to form a hypothesis on the location of the epileptogenic zone., Here, we validate our technique on a set of patients for which the outcome is known.

An end-to-end overview of our pipeline, from data preprocessing up until statistical inference, is depicted in Figure 7.2.

7.2.7 Model performance

We use several metrics to quantify the quality of the obtained sCMTF solutions.

As in Chapter 4, we compare the statistical maps with a ground truth delineation of the ictal onset zone (IOZ) to assess their concordance. The sensitivity for detecting the IOZ is then computed as the fraction of 'true positive' cases, which are determined by the presence or absence of significant activation clusters which overlap the IOZ in the spatial signatures $\mathbf{v}_r$.

Furthermore, we hypothesize that the spatial variation of the HRF over the brain might reveal additional localizing information regarding the IOZ, i.e., based on considerations explained in Section 7.1, we assume that the HRF in or near the IOZ might be distorted compared to nonepileptic brain regions. We test this hypothesis by assessing whether those regions correspond to high values in the HRF entropy and HRF extremity maps (see 7.2.6).

Additionally, we inspect the spectral, spatial and temporal EEG signatures of the extracted sources, and we measure whether the spatial fMRI signatures bear any similarity to known networks of resting-state human brain activity (Shirer et al., 2012).

7.3 Nonlinear fitting of the sCMTF model

7.3.1 Accommodating noncausal HRFs

In section 7.2.4, we derived the implementation of the convolution with an HRF as a left multiplication of the temporal signatures $\mathbf{S}$ with a Toeplitz matrix, whose diagonals hold the HRF samples. For a causal convolution, in which the BOLD signal strictly lags its neural correlate, $\mathbf{H}_k(i,j) = h_k(i-j)$ if $i - j \geqslant 0$, $\mathbf{H}_k(i,j) = 0$ otherwise, hence the matrix is lower triangular. This is the situation depicted also in Figure 7.1.

However, a recurring observation is that BOLD changes can be observed that precede the IEDs themselves (Hawco et al., 2007; Pittau et al., 2011; Jacobs et al., 2009; Moeller et al., 2008). Hence, we allowed noncausal HRFs that start at most 4 samples before the EEG, which allows for BOLD responses preceding the IEDs by up to 10 s at a typical TR of 2.5 s.

7.3.2 Initialization and optimization

Since the cost function J in (7.14) is nonconvex, any optimization procedure can only guarantee to converge to a local optimum, hence selecting a good starting point is crucial to obtain a reliable solution.

Firstly, we decomposed the EEG data $\boldsymbol{\mathcal{X}}$ individually according to the CP or PARAFAC model (Harshman et al., 1970; Kruskal, 1977; Bro, 1997), to obtain a good initialization for the factors $[\![\mathbf{S}, \mathbf{G}, \mathbf{M}]\!]$ in the sCMTF model. To this end, we used a Gauss–Newton algorithm (`cpd_nls` with 2000 iterations, 400 conjugate gradient iterations for the step computation, and tolerance on the relative cost function update of 10^{-8}, in Tensorlab 3.0 (Vervliet et al., 2016)), which we ran 50 times, from randomly drawn initial factors. We observed that the resulting factors lied very often close together over runs, indicating the algorithm had found a robust solution.

We always employed $K = 3$ HRFs, which we manually initialized. To assess whether the eventual sCMTF solution was also robust to the initialization of the HRF basis functions, we used a slightly different set of HRF-generating parameters $\boldsymbol{\theta}_k$ in each repetition of the optimization. Figure 7.3 shows some typical HRF waveforms, which are used to generate the Toeplitz blocks in Figure 7.1b.

From there, we initialized also the fMRI factors in the sCMTF model in (7.11)–(7.13). We constructed a flattened 'design matrix' $\mathbf{D} = \begin{bmatrix}\mathbf{H}_1\mathbf{S} & \dots & \mathbf{H}_K\mathbf{S}\end{bmatrix}$ in (7.13) and obtained a rough estimate for $\mathbf{B}^{\mathrm{T}} \odot \mathbf{V}^{\mathrm{T}}$ as $\mathbf{U}^{\mathrm{T}} = \mathbf{D}^{\dagger}\mathbf{Y}$ via regression—albeit this does not yet disentangle $\mathbf{B}$ and $\mathbf{V}$. To obtain initializations for the individual spatial factors, we exploit the fact that in every ROI i_v, the Khatri–Rao product of the i_v-th columns of $\mathbf{B}^{\mathrm{T}}$ and $\mathbf{V}^{\mathrm{T}}$ corresponds to a rank-1 constraint when folded into a $K \times R$ matrix (Boussé et al., 2018; Beckmann and Smith, 2005); hence, a rank-1 truncated singular value decomposition of the folded i_v-th column of $\mathbf{U}^{\mathrm{T}}$ leads to the desired vectors (Beckmann and Smith, 2005), which are further refined via a constrained Gauss–Newton algorithm (Boussé et al., 2018). We approximated the residual of the fMRI data under the initialized (coupled) factors using a rank-Q truncated SVD to capture fMRI nuisances. The parameter Q was chosen as twice the number of acquisition runs that had been done for a subject.

After each of the 50 runs of the initialization procedure, we iteratively optimized (7.14) with a quasi-Newton algorithm (`sdf_minf` with 1000 iterations, and tolerance on the relative cost function update of 10^{-8}, in Tensorlab 3.0 (Vervliet et al., 2016)). Both $\boldsymbol{\mathcal{X}}$ and $\mathbf{Y}$ were divided by their Frobenius norm, such that afterwards $\|\boldsymbol{\mathcal{X}}\|_F = \|\boldsymbol{\mathcal{Y}}\|_F = 1$, and we chose $\beta_x = \beta_y = 1$ so their fit had equal contribution to the cost function. For the regularization penalties, we pick

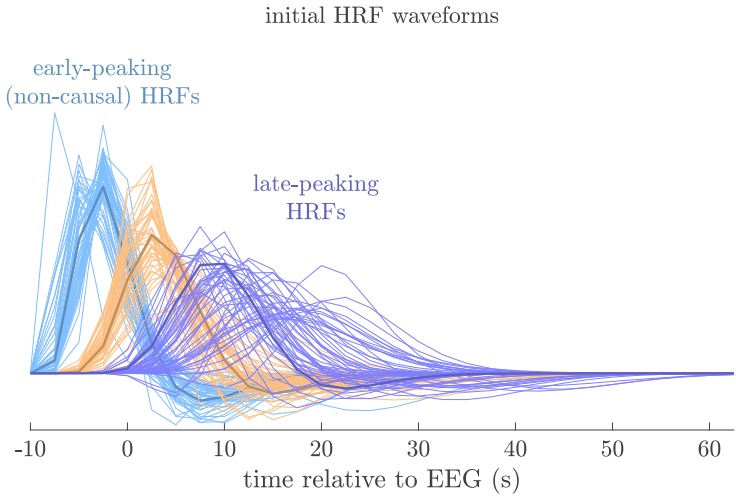


Figure 7.3: Different initializations of the HRF basis functions are used in every repetition of the sCMTF fitting procedure. In each repetition, one early-peaking, one mid-peaking and one late-peaking HRF were sampled from probability distributions of the HRF's generating parameters ($\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$ and $\boldsymbol{\theta}_3$, respectively) that was created by applying some multiplicative noise to baseline parameters (shown as the bold waveforms). The support extends to four negative samples, which gives the model the freedom to fit noncausal HRFs (relatively to the synchronized EEG).

$\gamma_x = \gamma_y = 10^{-3}$, as in (Acar et al., 2014)

7.4 Model selection

We fitted the sCMTF model to each patient's data for varying number of sources (rank), i.e., $R = 1 \ldots 6$. For each choice of R, we ran the optimization procedure 50 times, as explained in 7.3.2. Afterwards, the results need to be aggregated, such that clear conclusions on the sources of interest can be drawn. This involves several steps, which we explain below in chronological order.

7.4.1 Sign and scale standardization of factor estimates

Some of the factors that are estimated via minimization of (7.14) are subject to sign and scale ambiguities, which are inevitable in many BSS contexts (Sidiropoulos et al., 2017; Kolda and Bader, 2009). In the EEG factor model in (7.8), the factor vectors $\mathbf{s}_r$, $\mathbf{g}_r$ and $\mathbf{m}_r$ belonging to the same component r may be multiplied with arbitrary scaling factors whose product is one, without altering the goodness of fit. Similarly, in the fMRI factor model in (7.11), sign and scale are exchangeable between corresponding columns of $\mathbf{S}$ and $\mathbf{V}$, and between rows of $\mathbf{V}$ and rows of $\mathbf{B}$. However, to conduct proper statistical inference on each source r's spatial amplitudes, the elements in $\mathbf{v}_r$ must be calibrated, in the sense that it must be possible to compare them across ROIs, and determining the sign is crucial to distinguish local activation from deactivation. Hence, we sequentially fix these ambiguities as follows:

1. For every component r:
   (a) $\mathbf{s}_r$ and $\mathbf{g}_r$ are rescaled to unit ℓ_2-norm, and $\mathbf{m}_r$ is counterscaled
   (b) for both $\mathbf{s}_r$ and $\mathbf{g}_r$, the sign is flipped if the sum of squares of the negative elements exceeds that of the positive elements; the sign of $\mathbf{m}_r$ is adjusted to preserve the global sign of the EEG rank-1 term
   (c) if the sign of $\mathbf{s}_r$ was flipped, also the sign of $\mathbf{v}_r$ is flipped to preserve the global sign of the r-th fMRI block term
2. For every ROI i_v:
   (a) the local HRF $\mathbf{h}_{i_v}$ is reconstructed using (7.6)
   (b) the i_v-th row of $\mathbf{B}$ is rescaled and sign-corrected as to make $\mathbf{h}_k$ unit ℓ_1-norm and as to ensure that the HRF's largest overshoot precedes the largest undershoot; the i_v-th row of $\mathbf{V}$ is counterscaled

7.4.2 Stability analysis

To assess the reproducibility of the factors, we use the graph-structured clustering algorithm that we proposed in Chapter 6 and briefly summarize here. We represent the factor sets for all 50 repetitions of the fitting procedure as $[\![\mathbf{S}, \mathbf{G}, \mathbf{M}, \mathbf{V}]\!]$, and use a threshold of 0.85 to construct a binary link matrix that encodes similarities between components from different runs of the optimization (empirically, we found that this threshold led to acceptable cluster definitions in this context). Via low-rank matrix approximation of this link matrix, we then obtained clusters of components that were encountered in varying numbers

of repetitions. High cardinality of a cluster is then a sign that the involved component is very reproducible or 'stable', since it is part of the factor set upon convergence in many repetitions. We suggest to assign higher trust in such components, as opposed to components in small clusters, which are likely specific to one (potentially poor) local minimum. For the further steps, we condensed each cluster to one of its components, i.e., its centroid. In each cluster, the centroid component is defined as the component which has the largest accumulated similarity with all other components in the cluster[7].

7.4.3 IED component selection

Out of the centroids of the clustered components, we identified the (most) IED-related component as the one whose temporal signature $\mathbf{s}_r$ was most correlated to a reference time course, which is constructed as the average over channels of the MWF's output signal's time-varying (broadband) power. This reference is the BOLD predictor we have proposed for EEG-correlated fMRI analysis in Chapter 4, and which provides a good baseline for identifying temporal dynamics that are timelocked to the IEDs.

7.4.4 Choice of the rank R of the factorization

After the previous steps have been carried out for each setting of R, we are left to select the rank $R \in \{1, 2, 3, 4, 5, 6\}$ whose set of results we proceed with. We heuristically determined an appropriate value $\hat{R}$ by selecting the model which fulfills several criteria:

1. *high core consistency of the EEG decomposition*
   We compute the core consistency diagnostic (CorConDia) (Bro and Kiers, 2003) for the EEG tensor in combination with its estimated factor set $[\![\mathbf{S}, \mathbf{G}, \mathbf{M}]\!]$ from the centroid repetition. The consistency describes how suitable a rank-R CPD is for the given tensorial data and given factors, and is expressed as a percentage (100% being a very adequate model, and percentages below 70–80% indicating that the model is not appropriate)[8].

[7] By extension, the centroid repetition is defined as the repetition to which the centroid component belongs.

[8] CorConDia is a popular and robust model selection tool for tensor decompositions (Mørup and Hansen, 2009; Acar et al., 2007; Miwakeichi et al., 2004; Liu et al., 2016; Papalexakis, 2016). To compute it, first the core tensor which is most appropriate (in minimum mean squared error sense) for the given data and CPD-derived factors is estimated. Subsequently, CorConDia is computed as the fraction of the core tensor's sum of squares which is due to off-superdiagonal elements. When for a given set of factor matrices the CP structure is indeed

2. *reproducible IED-related component*
   We count the number of repetitions in the cluster that was most related to the IED (cfr. 7.4.2 and 7.4.3), and used this as a measure of reproducibility. We rejected clusters whose cardinality was lower than 10.

3. *similarity to a reference IED time course*
   We track the correlation between the IED-related component's temporal signature $\mathbf{s}_{\text{IED}}$ and the reference temporal signature $\mathbf{s}_{\text{ref}}$, as explained in 7.4.3. We expect higher correlations to signify a more suitable model, since $\mathbf{s}_{\text{ref}}$ generally led to good results in the study described in Chapter 4.

4. *high significance in the IED-related spatial map*
   We track the highest pseudo t-value in the SnPM that was created based on the IED-related component's spatial signature $\mathbf{v}_{\text{IED}}$ (cfr. Section 7.2.6). A high statistical score indicates a good model fit for the IED-related component (Abreu et al., 2018a).

ideal, the core tensor is superdiagonal and CorConDia equals 100%. Note that for a rank-1 model, this notion is meaningless, since the core tensor is a scalar, and CorConDia would trivially be 100% always.

Table 7.1: For every patient (ID 1–12), the sCMTF model can be fitted for a varying number of sources or rank R. We select a 'good' value for R post hoc, based on four criteria which are checked intra-patient: 1) the core consistency of the EEG tensor decomposition should be high ($> 70\%$); 2) the IED-related source should be found in sufficiently many ($\geqslant 10$) of the 50 repetitions of the estimation procedure; 3) the correlation of the IED-related source's temporal signature with the reference time course, namely the MWF's broadband envelope, is preferably high; 4) the maximal pseudo t-statistic for the IED-related source's spatial signature is preferably high.

ID		core consistency diagnostic (%)						reproducibility (# repetitions in IED cluster)						correlation of $\mathbf{s}_{\text{IED}}$ with reference MWF envelope $\mathbf{s}_{\text{ref}}$						maximal t-statistic of $\mathbf{v}_{\text{IED}}$						selected rank
	$R=$	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6	
p01		–	100.0	99.6	97.1	90.0	76.8	25	19	14	7	16	20	0.35	0.91	0.38	0.94	0.97	0.97	14.2	20.0	21.5	16.0	7.7	20.1	$\hat{R}=6$
p02		–	100.0	94.9	58.0	19.6	7.2	14	23	27	26	17	18	0.10	0.93	0.95	0.83	0.96	0.88	8.7	17.0	23.4	22.6	17.7	16.9	$\hat{R}=3$
p03		–	100.0	95.6	70.7	29.0	28.4	29	23	24	21	23	19	0.29	0.93	0.92	0.93	0.93	0.93	15.2	16.5	14.9	14.3	13.9	15.0	$\hat{R}=2$
p04		–	100.0	87.0	74.4	37.3	12.3	25	22	15	27	15	20	0.09	0.20	0.61	0.71	0.68	0.81	9.7	6.7	15.4	14.3	11.7	23.4	$\hat{R}=4$
p05		–	100.0	98.7	94.2	93.2	89.7	3	14	14	21	13	17	0.28	0.85	0.00	0.01	0.92	0.01	19.6	5.5	6.5	4.4	9.0	4.6	$\hat{R}=5$
p06		–	100.0	94.6	80.9	33.8	−409	15	21	14	14	11	13	0.55	0.92	0.87	0.80	0.80	0.22	9.9	17.1	9.9	9.6	16.4	36.9	$\hat{R}=2$
p07		–	100.0	97.7	96.7	−76.0	−0.6	18	17	12	15	11	12	0.09	0.74	0.80	0.80	0.80	0.22	7.4	10.0	9.4	8.6	11.2	9.9	$\hat{R}=4$
p08		–	100.0	99.7	30.6	−67.2	−178	12	21	21	11	14	23	0.61	0.95	0.44	0.95	0.94	0.94	13.8	16.0	10.1	20.3	15.5	18.0	$\hat{R}=2$
p09		–	100.0	98.8	95.9	90.1	23.6	21	27	21	15	19	19	0.19	0.66	0.67	0.13	0.48	0.49	14.7	22.5	21.9	10.1	8.0	19.2	$\hat{R}=2$
p10		–	100.0	98.0	95.0	91.7	80.3	23	34	13	24	25	14	0.47	0.29	0.96	0.96	0.96	0.19	11.9	12.9	10.0	8.6	9.9	12.8	$\hat{R}=5$
p11		–	100.0	97.9	78.3	−307	49.1	21	18	12	22	13	12	0.65	0.91	0.50	0.74	0.66	0.67	18.8	12.8	12.0	28.9	18.6	12.6	$\hat{R}=2$
p12		–	100.0	97.3	89.0	59.3	69.4	23	14	14	12	15	12	0.78	0.79	0.60	0.07	0.47	0.50	15.8	15.0	10.9	7.9	10.2	5.5	$\hat{R}=2$

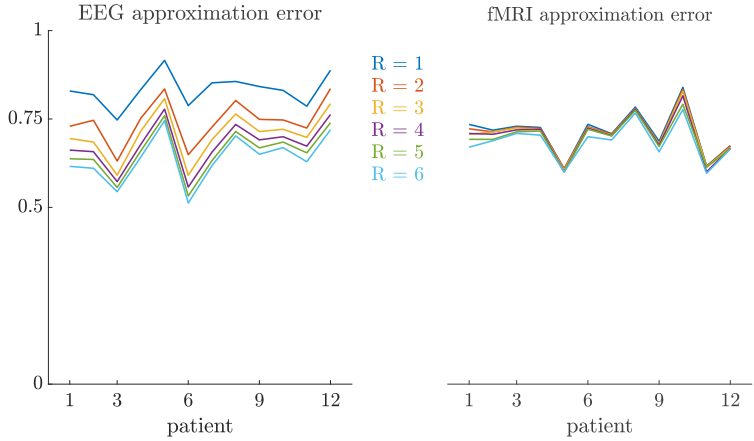


Figure 7.4: Goodness of fit of each patient's EEG tensor $\boldsymbol{\mathcal{X}}$ and fMRI matrix $\mathbf{Y}$, for varying choices of the rank R in the sCMTF. Naturally, the EEG approximation error decreases monotonically for increasing rank (intra-patient). For the fMRI data, the fit already plateaus for very low R. This is due to the presence of additional, uncoupled components $\mathbf{n}_q\mathbf{p}_q^{\mathrm{T}}$ in the fMRI factorization, which can absorb some of the variance when the number of coupled components is low, but which lose their relevance at higher ranks.

7.5 Experiments

7.5.1 Patient-specific model selection

Table 7.1 compiles the results of the steps described in Section 7.4. For each patient, we select the set of sCMTF factors of rank $\hat{R}$, which best fulfill the criteria. In all cases, we found at least one such a solution, including an IED-related component within that solution. Note that sometimes models with different R might score well on different (subsets of) the criteria, so the selection of the rank is inevitably ambiguous. In the next section, we analyze the individual set of results for each patient, based on the selected rank, and we analyze the sensitivity of the results to the choice of R.

We show the goodness of fit of the estimated factors for the EEG tensor $\boldsymbol{\mathcal{X}}$ and the fMRI matrix $\mathbf{Y}$ in Figure 7.4. Due to the normalization steps which have been applied to the data (cfr. 7.2.2), the sCMTF operates in a regime of moderately high relative approximation errors.

7.5.2 Spatio-temporo-spectral profiles of interictal discharges

We analyze for each patient the sources which have been estimated via the sCMTF model. Due to space constraints, we discuss the results of two patients in detail in the next subsections, and include complete results for all other patients in Appendix B. Every time, we show 1) the thresholded pseudo t-maps of the IED-related source in the fMRI domain, both for significant activation as for significant deactivation; 2) maps highlighting the ROIs of high HRF entropy and extremity; 3) the temporal profile (time-varying power) $\mathbf{s}_r$, spatial profile (topography) $\mathbf{m}_r$ and spectral profile $\mathbf{g}_r$ of each source in the EEG domain; 4) the HRF waveforms in the different ROIs, and the HRF basis functions at convergence of the algorithm. We plot maximally 800 s of the temporal signatures, to ensure readability. For ease of comparison, we always overlay the broadband MWF envelope (with an arbitrary vertical offset for visualization only), which is the reference time course $\mathbf{s}_{\text{ref}}$ for selecting the IED-related component (cfr. 7.4.3). For considerations of space, we generally only show the maps of the fMRI spatial signature $\mathbf{v}_r$ for the IED-related components, but discuss the maps of other components when relevant. We show five axial slices of each map: in each case, we show two slices near the highest and lowest voxels of the IOZ or significant regions of the fMRI spatial signature (whichever lies furthest); if applicable, the middle slice is the cross-section with most overlap between IOZ and spatial signature, and the two remaining slices lie halfway between this slice and the extremal slices; otherwise all three bulk slices are chosen with equal spacing between the extremal slices. We cross-validate the maps against known resting-state networks (RSN) of human brain activity from the Stanford atlas (Shirer et al., 2012).

We stress again at this point that a *subset* of the results is prone to errors due to imperfect sign normalization (cfr. 7.4.1). While it is relatively straightforward to unambiguously determine the 'right' sign of the EEG signatures, this is more challenging for fMRI. That is, frequently, the polarity of the HRF waveform is ambiguous, and making the 'wrong' choice in a voxel i_v (i.e., the HRF which has the opposite effect of the true physical CBF change) immediately leads to the wrong sign of the spatial coefficients in $\mathbf{r}_{i_v}$ and their pseudo t-values for all sources r. To track the occurrence of this foreseen failure mode, we also investigate the significant deactivations of the sources[9]. Note that we designed the HRF variability metrics so that they are *immune* to the polarity of the HRFs. Hence, any high score of the HRF metrics can be reliably interpreted.

[9] Alternatively, it is possible to use a pseudo F-statistic, e.g. the squared pseudo t-value, to bypass the sign correction altogether. The downside of such an approach is that it is then impossible to distinguish activation and deactivation, which may be meaningful.

For each case, we separate the twenty waveforms with the highest entropy score, and report how many of those are found in ROIs that overlap with the IOZ, along with the probability (in the form of a p-value) that this would occur by randomly sampling as many ROIs (under a given fraction of brain that is covered by the IOZ). Hence, this metric is analogous to the positive predictive value (PPV)[10], albeit no rigorous test has been applied in this case.

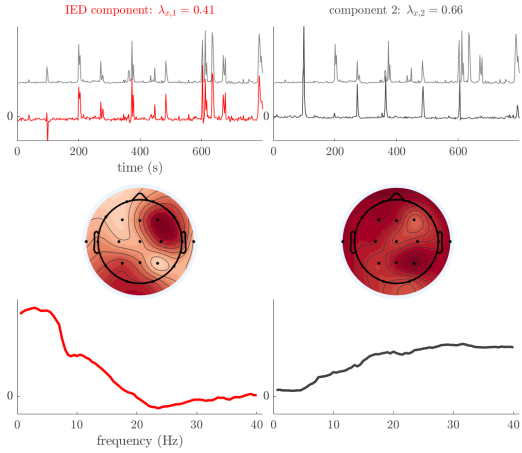
Patient 3

We analyze the solution with $\hat{R} = 2$ sources, and show the results in Figure 7.5 and 7.6. Besides one clear IED-related source, there is one other source that is substantially correlated to the reference time course, but with a homogeneous distribution over the head and an unclear spectrum. This may signify that the IEDs do not follow exactly a rank-1 structure in the spectrogram, and that they may be nonstationary in time or space (cfr. the argument made for nonstationary seizures in (Hunyadi et al., 2014)). The second source's pseudo t-map had significantly active areas symmetrically in the left and right parietal lobe, much more focalized than the EEG topography. In the EEG time courses, we found indeed IED-like waveforms at the times of the peaks in the temporal signature. Hence, we suspect that both sources may reflect the onset and propagation of the IEDs to other areas, respectively. Five out of the twenty ROIs with high-entropy HRFs overlapped with the IOZ, and a significant finding is that several of them are highly noncausal, i.e., with a positive peak before zero seconds. Figure 7.6 confirms this, and also shows that the IED-related source is significantly active in different ROIs of the IOZ.

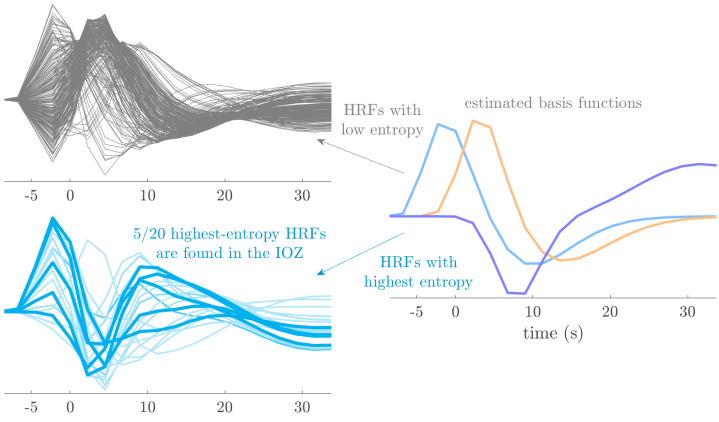
Patient 10

We analyze the solution with $\hat{R} = 5$ sources, and show the results in Figure 7.7 and 7.8. There is a clear IED-related source, and also an artifactual source at ± 34 Hz, which is also present in other patients. Due to its relatively consistent occurrence, we hypothesize that this artifact is due to the MR acquisition. For example, it may be a remnant of a gradient artifact which is not adequately removed from the data of some channels, cfr. the observation made in (Marecek et al., 2016). Surprisingly, this source is significantly active in an extended area in the occipital lobe, overlapping with the visual network. Both HRF metrics correctly reached extreme values at (distinct) ROIs within the IOZ. The pseudo-t map of the IED-related source shows significantly active ROIs that

[10]The PPV is the fraction of positive predictions (in a classification or hypothesis testing framework) which are truly positive, which equals one minus the false discovery rate (FDR).



(a) Temporal ($\mathbf{s}_r$, top), spatial ($\mathbf{m}_r$, middle), and spectral ($\mathbf{g}_r$, bottom) profiles of the 2 sources in the EEG domain, and reference IED time course ($\mathbf{s}_{\text{ref}}$, in grey).



(b) Estimated HRF waveforms in all ROIs, split in HRFs with low (grey) and high (blue) entropy

Figure 7.5: **(a)** In the selected solution for patient 3 ($R = 2$), both sources have a temporal signature that correlated strongly to the reference IED time course. The first source modeled the main onset of IEDs and was low-frequency and topographically focal, while the second source was spatially and spectrally diffuse and captured the propagation of IEDs to remote areas (cfr. Figure 7.9). **(b)** Five out of the twenty most deviant HRFs were found inside the ictal onset zone (bold lines, $p < 10^{-4}$). These HRFs had main peaks before 0 s, i.e., they led to BOLD changes before the corresponding EEG correlate of the IED was seen.

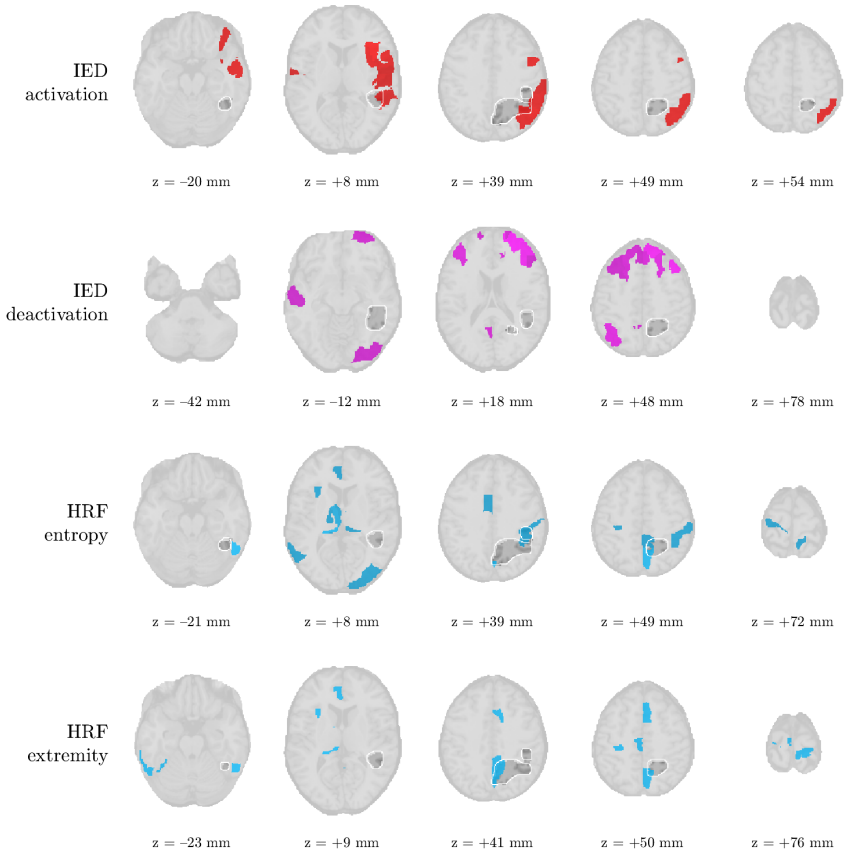
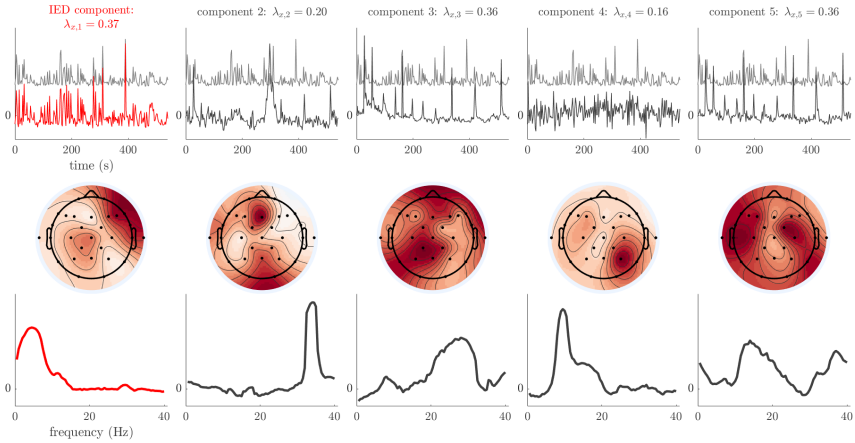


Figure 7.6: The statistical nonparametric maps of the IED-related component (top two rows) and HRF entropy/extremity maps (bottom two rows) of patient 3 show concordance with the ictal onset zone (IOZ). Especially the regions of significant IED activation were accurate, but also five out of the twenty regions with the most deviant (highest entropy) HRFs were found in the IOZ (cfr. Figure 7.5b). The ground truth ictal onset zone is highlighted in dark gray with a white contour.

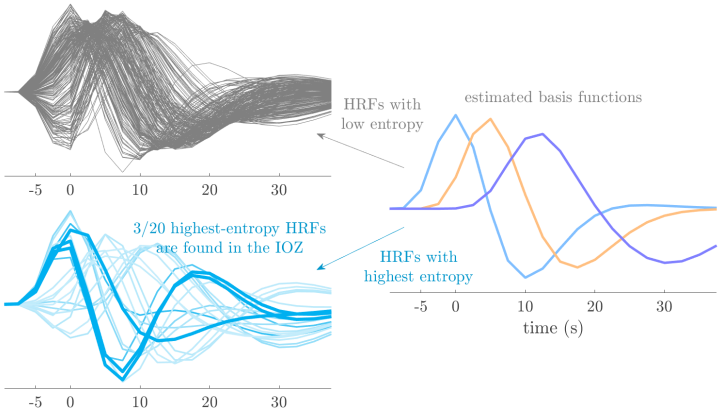
are concordant with the IOZ, and deactivation of a large part of the default mode network. Furthermore, the IED-related source's EEG topography is very consistent with the clinical diagnosis. The fourth source is active in the default mode network, predominantly in the α band (cfr. Figure 7.10). The fifth source had an unclear spectrum, but its temporal signature corresponds to the occurrence of high-amplitude IEDs. Its pseudo t-map shows widespread activations over the brain, which did not include the IOZ. We expect that this component captures the propagation of IEDs, after onset near the IOZ, similarly to patient 3.

Summary of all patient's results

We provide an overview of the results w.r.t. IOZ detection in Table 7.2. All results taken together, the sCMTF results allow a correct detection of the IOZ based on the significant IED activation (10/12 cases), significant IED deactivation (6/12 cases), HRF entropy (9/12 cases) and HRF extremity (8/12) cases. All cases are covered by at least one of the metrics, and all patients besides patient 6 had at least two metrics providing correct and complementary localizing info on the IOZ. For nearly all cases, the IED-related component's time course was highly correlated to a reference IED time course, and its spectrum was plausible. In many, but not all cases, this component's EEG topography was also consistent with the location of the IOZ, though this notion is slightly fuzzy because of the very different spatial domains of EEG and (f)MRI—hence we do not use the term 'concordant'. Analysis of the spatial, spectral, and temporal signatures, in combination with inspection of the filtered EEG signals, reveals the identity of RSN oscillations and/or artifacts in the majority of cases. For several patients, we found sources that are active in a narrow spectral band near 33–34 Hz. While this likely reflects a technical artifacts as the result of the MR acquisition, we found no concomitant changes at this frequency in the EEG. This may be the result of the normalization procedure which we applied prior to the decomposition: since every frequency bin was given equal importance, even unnoticeable but structured fluctuations at higher frequencies may be captured in a component.



(a) Temporal ($\mathbf{s}_r$, top), spatial ($\mathbf{m}_r$, middle), and spectral ($\mathbf{g}_r$, bottom) profiles of the 5 sources in the EEG domain, and reference IED time course ($\mathbf{s}_{\text{ref}}$, in grey).



(b) Estimated HRF waveforms in all ROIs, split in HRFs with low (grey) and high (blue) entropy

Figure 7.7: **(a)** The sCMTF solution with $R = 5$ sources was selected for patient 10. One source's temporal signature is highly correlated with the reference IED time course and is identified as the IED-related source, which has a characteristic low-frequency behaviour and with a frontotemporal topography, consistent with the IOZ location. The second source, which has very narrowband power around ± 33 Hz, likely captured an artifact of the MR acquisition. The fourth source captured α activity in the default mode network (cfr. also Figure 7.10). **(b)** Three out of the twenty most deviant HRFs were found inside the ictal onset zone (bold lines, $p = 0.02$).

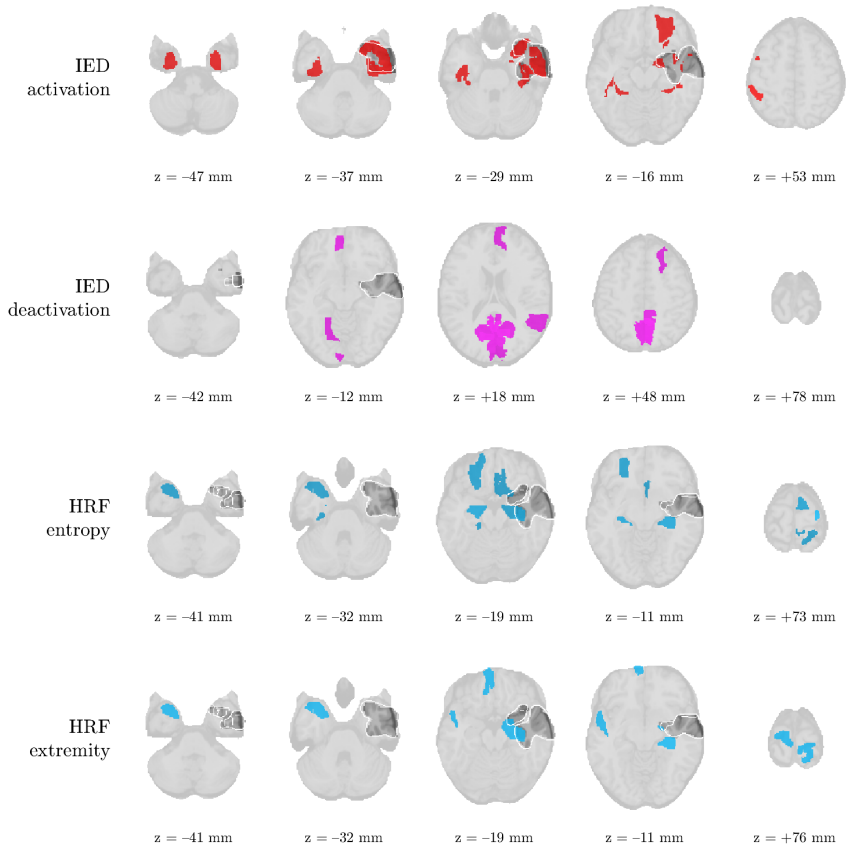


Figure 7.8: The statistical nonparametric maps of the IED-related component (top two rows) and HRF entropy/extremity maps (bottom two rows) of patient 10 show concordance with the ictal onset zone (IOZ). IED occurrences were associated with significantly active regions in and near the IOZ (top row), and at the same time with a deactivation in a part of the default mode network (second row). Three out of the twenty regions with the most deviant (highest entropy) HRFs were found in the IOZ (cfr. Figure 7.7b). The ground truth ictal onset zone is highlighted in dark gray with a white contour.

Table 7.2: The sCMTF leads to three types of spatial information, which can be cross-validated against the ground truth IOZ, as defined in Section 7.2.7 and summarized for all patients in Table 4.2: 1) the EEG topography $\mathbf{v}_{\text{IED}}$ of the IED-related component; 2) the significantly activated and deactivated ROIs in the fMRI spatial signature $\mathbf{v}_{\text{IED}}$; 3) the ROIs with strongly deviating HRF waveforms, as measured with entropy and extremity. Since the EEG topography has a very low spatial resolution, and depends on the attenuation properties of the tissue as well as the orientation of the neural sources in the cortex, we only expect partial similarity to the IOZ's spatial focus; hence, we use the term 'consistent' rather than 'concordant'.

patient	selected solution	EEG topography consistent with IOZ?	spatial signature $\mathbf{v}_{\text{IED}}$ concordant with IOZ?		HRF variability metrics concordant with IOZ?		20 highest-entropy ROIs	
ID	$\hat{R}$		activation	deactivation	entropy	extremity	# in IOZ	(p-value)
p01	6	no	no	no	yes	yes	1	(0.34)
p02	3	no	yes	yes	yes	no	1	(0.59)
p03	2	no	yes	no	yes	yes	5	($< 10^{-4}$)
p04	4	yes	yes	no	yes	yes	2	(0.32)
p05	5	yes	yes	yes	yes	yes	6	($< 10^{-3}$)
p06	2	no	yes	no	no	no	0	/
p07	4	yes	yes	no	yes	no	1	(0.57)
p08	2	no	yes	yes	no	no	0	/
p09	2	yes	yes	yes	no	yes	0	/
p10	5	yes	yes	no	yes	yes	3	(0.02)
p11	2	yes	yes	yes	yes	yes	4	(0.01)
p12	2	no	no	yes	yes	yes	7	($< 10^{-3}$)

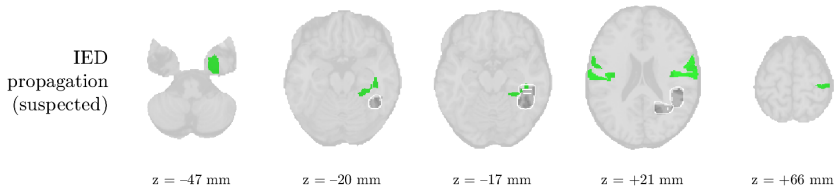


Figure 7.9: The second component in patient 3 likely captured the propagation of IEDs from the irritative zone, given its relatively large correlation to the MWF envelope (cfr. Figure 7.5a). The ground truth ictal onset zone is highlighted in dark gray with a white contour.

Sensitivity to model selection

For many patients, selecting $\hat{R}$ is ambiguous, since more than one solution (with different R) score well on some of the criteria (cfr. Table 7.1). Therefore, we analyze impact of the choice of R on the sCMTF results. For each patient, we select the solution with the rank which is next in line, i.e., which would be a second best (or equally good) choice, based on the same criteria. This is the solution with $R = 1$ for patient 12, $R = 2$ for patients 1, 2, 5 and 7, $R = 3$ for patients 3, 4, 6, 8, 9 and 10, and $R = 4$ for patient 11. For patients 1, 6 and 8, the results deteriorate drastically, as no metric correctly localizes the IOZ. For patient 11, no ROI within the IOZ is significantly activated due to IEDs anymore, but the HRF metrics are still informative. The results for patients 9 and 12 improve, since all metrics are now sensitive to the IOZ. For the other patients, the situation stay more or less the same, i.e., the same metrics are valuable for IOZ localization. However, the maximum value under the different metrics is generally attained at different ROIs compared to the initially selected model.

7.6 Discussion

A novel EEG–fMRI data fusion framework

We have proposed an integrated and structured coupled matrix-tensor factorization (sCMTF) framework, which can be used to make inferences on the localization of the ictal onset zone in refractory epilepsy based on simultaneous EEG and fMRI recordings. Our approach aims to perform blind source separation of the neural activity related to interictal epileptic discharges (IEDs),

and to characterize it in the spatial, temporal, and spectral domain. To this end, we developed a pipeline consisting of 1) semi-automated EEG enhancement based on annotations of the IEDs; 2) modality-specific preprocessing and tensorization steps, which lead to a third-order EEG spectrogram tensor varying over electrodes, time points, and frequencies, and an fMRI matrix with BOLD time courses for a predefined set of regions of interest or parcels; 3) coupled matrix-tensor factorization of the EEG tensor and fMRI matrix along the shared temporal mode, while accounting for variations in the local neurovascular coupling; 4) automated selection of a robust, and relevant IED-related component, and nonparametric testing to infer its spatial distribution in the brain.

We have stressed the importance of and accounted for the variability of the hemodynamic response function (HRF) over different patients and brain regions, by equipping the CMTF with the required expressive power via a set of adaptive basis functions. Moreover, after estimating the EEG and fMRI factor signatures, as well as the HRF parameters, we have computed different summary metrics (entropy and extremity) that measure the local deviance of a ROI's HRF compared to other HRFs in the same brain, and have cross-validated the spatial map of these metrics against the ground truth localization of the ictal onset zone.

The sCMTF pipeline managed to provide correct detection in all twelve patients in this study, with varying degrees of certainty. The statistical nonparametric map (SnPM) of the spatial signature of the IED-related component, obtained with the sCMTF, is the best biomarker The ROIs that are significantly activated under influence of IEDs were the best biomarker in the study, which is in line with the traditional EEG-correlated fMRI approach (Lemieux et al., 2001). In the large majority of patients, several of these regions overlapped with the IOZ. Also the HRF entropy, as a measure of how unlikely an HRF is within a specific set of other HRFs, is a very good biomarker, which almost always identified regions of the IOZ that were complementary to those already found by tracking significant IED activation. In roughly half of all cases, we also found regions within the IOZ that significantly deactivated in association to IEDs. Compared to the results in Chapter 4, we achieved an additional correct IOZ detection (for patient 11), which is probably thanks to the increased flexibility of the current model. In a waterbed effect, the pseudo t-map for patient 1 no longer allowed correct IOZ detection compared to the simpler pipeline. Luckily, however, this gap is filled here by the new HRF entropy metric. We inspected the 20 HRFs and ROIs with the highest extremity and entropy. Hence, it is inevitable that some or most of these ROIs are not within the IOZ. Standalone HRF metrics would hence have a high false discovery rate, even though for several patients, the high proportion of IOZ-covering ROIs among the 20 selected ROIs was very

unlikely due to chance (as measured with p-values). However, the ROIs that were highlighted by the HRF metrics were often distinct from the ROIs identified as significantly activated to the IEDs. Hence, the SnPMs of the IEDs and the entropy metrics provide very complementary information, and when analyzed jointly, they may infer the location of the IOZ with much more certainty, i.e., in the brain area where both IED-related and HRF-related metrics have a high value.

HRFs vary strongly over subjects and brain regions

There were substantial differences in (estimated) neurovascular coupling over patients and brain regions, as expected. Since we used 'regularized' basis functions, which are parametrized as smooth gamma density functions, the resulting HRFs generally had a plausible shape. However, in some cases we found nonsensical shapes, in which, e.g., the waveform had the same polarity over the whole time course, potentially with a bimodal shape (cfr. patient 4). This serves as a humble reminder to not blindly trust the outputted HRFs (or other factor signatures, for that matter). While we have empirically verified that the optimization algorithm converges properly to the true factor signatures and HRFs for synthetic data under mild conditions, there is no guarantee that this holds true for real-life data, which are orders of magnitude more complex, so that a linear generative model like the sCMTF may not be sufficient to describe the interplay between EEG and fMRI. Moreover, the proper behavior of the sCMTF estimation depends on careful preprocessing, and on a proper selection of hyperparameters (in casu: a good value for the number of sources $\hat{R}$). Hence, manual inspection of the data quality and the solution are still required. Even if the estimated HRFs or factor signatures may not fully reflect the 'correct' underlying physical phenomena, we have demonstrated that they offer actionable information. Not in the least, via summarizing metrics such as HRF entropy and extremity, our algorithm manages to be reasonably robust to subtle changes in the waveform—which is less of interest here than spatial cues towards the IOZ.

The algorithm used its modeling freedom to fit 'noncausal' HRFs, which are ahead of the EEG by as much as 10 s. Generally, we indeed found that many of the estimated HRFs had significant positive or negative amplitudes already before the neural correlate visible on the EEG. This is in line with recurrent findings on BOLD changes that precede the IEDs which were observed in the EEG (Hawco et al., 2007; Pittau et al., 2011; Jacobs et al., 2009; Moeller et al., 2008). We stress that this noncausality may only be in the observation, and not in the underlying physical chain of events: here, it strictly means that we *observed* BOLD changes in the fMRI data that occur before the corresponding

observed neural correlate on the EEG. Despite the fact that many of the HRFs differed substantially from the canonical HRF, which is causal and peaks approximately 6 s *after* its neural input, we obtained good results as well with the latter HRF as a nonadaptive model for neurovascular coupling (cfr. Section 4.3.2) The reason for this agreement between these different models—which differ substantially in terms of flexibility—is likely that the canonical HRF is positively correlated to the true HRFs which are found inside the IOZ, and as such the resulting activation maps may still be sufficiently informative. In our data and sCMTF results this is indeed the case for many patients.

The regularizing effect of the HRF model

In this study, we have modeled the HRF in each ROI as a linear combination of several basis functions, which is a commonly used tactic in many studies that employ EEG-correlated fMRI. In Section 7.1, we gave an overview of some of the most frequently used basis functions, which primarily differ in the tradeoff they achieve between flexible HRF estimation and overfitting. The most flexible method is the FIR basis, containing a separate, Dirac-like basis function for every sample in the HRF. As such, no assumptions/restrictions on the HRF waveform are made, but often an additional ℓ_2-regularization penalty on the basis coefficients is needed to temper the risk of overfitting, since the deconvolution that this approach achieves is ill-posed and especially sensitive to higher-frequency noise. Here, we have opted to use a small set ($K = 3$) of parametric basis functions, each consisting of a sum of two gamma density functions, in the spirit of the canonical HRF which is used in SPM (Friston et al., 1994). As explained in Section 7.3.2, we did not fix the parameters of these basis functions, but included them in the set of optimization variables, to allow sufficient flexibility. However, by specifying a parametric family for the basis functions, the full set of HRFs is constrained to lie on a low-dimensional nonlinear manifold, which limits the set of achievable waveforms compared to the FIR basis method, and hence has an intrinsic regularizing effect. Of course, other types of basis sets could be considered, potentially guided by prior info about the degree of HRF variability which is present (influencing K) or even the HRF shapes (influencing the basis functions). Our framework supports such different choices, by simply making the appropriate modification to the HRF operator in (7.15), and providing the derivatives w.rt. its parameters.

Prior EEG signal enhancement aids analysis

Importantly, our pipeline heavily relies on a prior enhancement of the interictal spikes in the EEG data, which would otherwise have a too low SNR for the

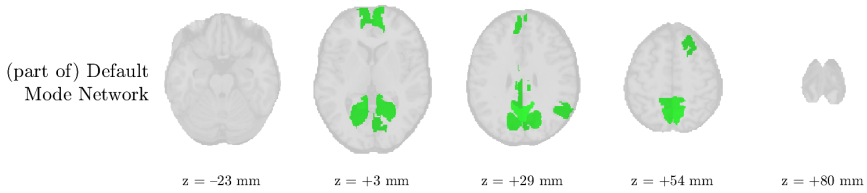


Figure 7.10: The fourth component in patient 10 seemed to pick up activity in the Default Mode Network (DMN), predominantly in the α band (cfr. Figure 7.7a).

sCMTF algorithm to pick up IED-related sources. We employ multi-channel Wiener filters, which solely rely on the annotation of a sufficient amount of IEDs in the data itself, or in related data (e.g., data from the same patient, recorded outside the MR scanner). While this task still frequently relies on the skill of human EEG readers and neurologists, advanced automated solutions for interictal spike detection are available (Wilson et al., 1999; Scheuer et al., 2017). Within each solution of a specific rank, we picked the IED component as the one with the highest correlation with a reference time course directly derived from the enhanced EEG. Some of the presented results make clear that this reference time course is not completely free from artifacts, hence caution is warranted when many high-amplitude artifacts are still present in the reference. In this study, however, we have not encountered any issues that seemed to be the direct results of a noisy reference during IED component selection.

In principle, it is also possible to modify the pipeline to rely solely on fMRI data, provided that other side information is available. Namely, based on the IED annotations, a single boxcar time series or unitary predictor could be constructed, just like the BOLD predictors which were part of the analysis in Chapter 4. This predictor could then be fixed a priori, and take the role of $\mathbf{s}_r$, which would otherwise stem from the EEG factorization. If multiple IED types are present in a patient, different such predictors could be used, leading to R 'sources', each modeling one IED type. This leads to a rank-R factorization of the fMRI based on (7.11)–(7.13), which could be fitted by minimizing an appropriately adjusted cost function similar to (7.14)–(7.15). In that case, it is essential to fully specify the factor matrix $\mathbf{S}$, and not include it as an optimization variable, since otherwise a rotational ambiguity would arise between $\mathbf{s}$ and $\mathbf{V}$, which precludes proper interpretation of the components. Such an analysis would be more broadly applicable, since simultaneous EEG–fMRI recording remains technically challenging, and standalone fMRI is more routinely measured.

The interpretation of components

Overall, the sCMTF pipeline succeeded in extracting meaningful IED-related components, alongside components that modeled resting-state neural fluctuations and physiological and technical artifacts. Each component r consists of two coupled rank-1 terms, i.e., outer products of individual vectors, that capture spatial ($\mathbf{m}_r$, $\mathbf{v}_r$), temporal ($\mathbf{s}_r$) and spectral ($\mathbf{g}_r$) profiles. The rank-1 structure allows easy interpretation in terms of relative intensity at different time points, EEG channels, ROIs and frequencies, via the modulation encoded by the coefficients in these profiles. The fact that the sCMTF can estimate signatures and statistical maps for multiple components is a powerful advantage over classical EEG-correlated analysis. As we demonstrated in the experiments, artifactual influences may be isolated in separate components, which could reduce their impact on IED mapping in the brain. Additionally, we encountered cases where two components were correlated to the IED occurrences: the component with the highest temporal correlation to a reference IED time course then correctly revealed the localization of the IOZ, while the other component presumably modeled the propagation of IEDs to remote brain regions. This observation is analogous to the finding in (Hunyadi et al., 2016), where a different type of CMTF was applied to average EEG waveforms of IEDs and statistical BOLD maps, which revealed a dissociation between the early IED spike and the subsequent wave, which were related to the onset and spread of the IEDs, respectively. Since we transformed the data with a time-frequency transform that used windows of length TR, our algorithm is unable to unravel different phases within one IED, since they occur in a shorter time frame. However, we identified these different IED-related sources by their significantly correlated temporal signatures, and their distinct spatial and spectral profiles. While we did not impose nonnegativity constraints, many estimated EEG spectral and spatial signatures were approximately nonnegative. This need not be the case, however, since the EEG data are normalized in a way such that the resulting signatures would reveal relative increase/decrease, rather than absolute time-varying spectral power (Marecek et al., 2016). For example, if a certain component is associated with a power increase in one spectral band, and a simultaneous power decrease in another band, this would be reflected in a spectrum with both a positive and a negative peak.

Practical considerations

The end-to-end sCMTF pipeline can provide a richer set of results compared to classical EEG-correlated fMRI analysis. In this respect, it is a more powerful data exploration tool. The tradeoff to be made is that significant computation

time goes into the sCMTF and subsequent inference—if one wants to apply it as rigorously as we have done in the current experiments. We seem to be doing a lot of unnecessary work, by computing the sCMTF factors for several numbers of sources, and by repeating the optimization several times for a fixed number of sources. Unfortunately, both ways of repetition seem required to obtain robust results, as we have argued in 7.3 and 7.4. However, the EEG-only CP decomposition, which lies at the heart of our initialization strategy, seemed very robust: we found highly similar EEG signatures for almost all random initializations. Probably, this is thanks to the use of the powerful Gauss–Newton-type of optimization. Hence, fewer repetitions of the sCMTF may be already sufficient to arrive at the same robust results. Despite the very reproducible EEG signatures in the initial CP decompositions, we still performed 50 repetitions of the sCTMF, each time slightly varying the initial HRF parameters. As such, we believe our findings are reasonably robust to poor initialization of the HRFs. Performing the sCMTF for many choices of R may still be required, as the quality of the result depends on the extraction of an appropriate number of sources. Luckily, however, most 'optimal' ranks were quite low in our study, under the heuristic selection procedure. This is reassuring, as it signifies that a typical rank encountered in this context is not problematically high, and no prohibitive computations are needed. Furthermore, we have demonstrated in our experiments that the summary metrics (sensitivity for localizing the IOZ based on different statistical scores) are fairly robust to the choice of R, although the estimated signatures themselves differ.

For many patients, the available data was split across multiple runs (i.e., with a few minutes break in between), and we opted to temporally concatenate data over runs, as explained in Section 7.2.2. While this violates the coupling model based on HRFs for time samples near the boundaries, we consider the effect minimal, given that the number of those 'affected' time samples represents a very tiny fraction of the whole time series. However, a more rigorous approach would be to 'inverse-impute' these samples and consider them as missing values: as such, they are ignored during the sCMTF optimization and will not affect the results.

Strategies to alleviate the computational demand

Due to the repeated decomposition and the nature of the nonparametric inference, the computations are highly parallellizable. For a typical dataset with available IED annotations, and with the parameters we have used for this study, the end-to-end computation for one patient took a few hours on a machine with twelve cores. To alleviate the computational burden, we have parcellated the fMRI data into 246 regions, based on the Brainnetome atlas

(Fan et al., 2016). This is clearly suboptimal, as the atlas is not patient-specific, and is mostly designed to study healthy brains. There is a serious risk for partial volume effects, in which the IOZ is scattered over several ROIs. As such, the IED-related BOLD changes in the part of the IOZ that falls within a certain ROI may get swamped by the remaining BOLD fluctuations within the ROI delineation. Hence, we hope to be able overcome this problem, either by algorithmic improvements, including a speed up of the optimization, or by the use of a patient-specific parcellation or PCA-like compression of the fMRI data. As of yet, it is hard to say whether the fixed atlas had an adverse effect on the results, and it is not so straightforward to compare the statistical maps from this study to maps which are voxel-based. We are currently pursuing experiments in which we employ a hierarchical parcellation: in a first step, the BOLD time series are grouped (but not yet averaged) according to the Brainnetome atlas; subsequently, we use spectral clustering to further refine each Brainnetome parcel based on the correlation matrix of its BOLD time series. As such, this hybrid approach combines a fixed, coarse-grained atlas with a further data-driven subdivision, which can mitigate partial volume effects, while still providing a significant data compression. Alternatively, it is possible to achieve a data reduction while still preserving voxelwise BOLD signals, by limiting the scope of the sCMTF to an a priori defined ROI (e.g., based on a clinical hypothesis stemming from other modalities).

Summary

In summary, we have developed and empirically validated a fully integrated framework for EEG–fMRI data fusion, which yielded a rich characterization of the interictal activity in time, space, and frequency, and which accounts for and exploits neurovascular coupling variation over the brain. The ability to separate local (de)activation of IEDs from local deviations in the HRF makes the sCMTF a powerful tool for exploratory analysis of interictal EEG and fMRI data. We envision that this approach, with some minor modifications, may also be used to analyze resting-state[11] EEG–fMRI activity from healthy patients, to map brain-wide resting-state networks (RSNs) of neuronal assemblies that are recruited for sensory, cognitive and motor processes. This has been attempted using an analysis similar to CMTF in (Martínez-Montes et al., 2004), and with related techniques in e.g. (Mantini et al., 2007) and (Mareček et al., 2017). With our framework, RSNs may be extracted more accurately from simultaneous EEG–fMRI if substantial deviations from the canonical HRF are present.

[11] Since in such a case, no IEDs are present, EEG-enhancement like we have done in this study would no longer take place. However, an MWF may still be used to clean up the EEG, e.g. by annotating artifactual periods, which can be removed from the data by the MWF in its dual form (or another tool) (Somers et al., 2018).

Our complete MATLAB code to execute the pipeline is available at `https://github.com/svaneynd/structured-cmtf`.

Chapter 8

Conclusion and prospects

8.1 Contributions and conclusions of this thesis

The **goal** of this thesis was to develop novel signal processing techniques to derive meaningful and actionable information from simultaneously recorded EEG and fMRI data, mostly focusing on their application during the presurgical evaluation of refractory epilepsy patients.

In **Chapter 2**, we highlighted that to date, many EEG–fMRI analyses of epileptic patients produce no useful cues for surgical intervention, which is in part due to mismodeling, or at least under-exploitation of the available data. More precisely, the following problems occur: 1) a substantial proportion of patients is EEG-negative: no IEDs could be observed during the fMRI scanning session; 2) even for patients with observed IEDs, sometimes no voxels with significantly correlated BOLD changes are found; and 3) even when EEG-correlated analysis does yield IED-related BOLD maps, interpretation can be challenging because of widespread significant clusters of voxels, which are not all related to the ictal onset zone.

In **Chapter 3**, we have treated several (classes of) signal processing and machine learning techniques which can be usefully applied for EEG and fMRI data mining. We have argued that coupled tensor factorizations are vehicles to perform data fusion of EEG and fMRI in a structured fashion. We hinted towards some supervised learning approaches, which can incorporate a neurologist's judgement into data-driven analyses.

In **Chapter 4**, we proposed the use of patient-specific multi-channel Wiener filters, to enhance the signal quality of EEG before performing EEG-correlated fMRI analysis. The MWFs are capable of learning the second-order statistics of the EEG signals during interictal discharges and during periods without discharges. By distinguishing the EEG characteristics for these two types of EEG segments, the filter can be steered to maximally preserve the waveforms of the signal of interest (in casu: the interictal discharges), while attenuating

or even suppressing EEG activity or artifacts which are unrelated to interictal discharges. We computed the time-varying output power of the filtered signals, and entered this as a predictor in a standard EEG-correlated fMRI analysis pipeline. For comparison, we also plugged in various other predictors, which were based on rudimentary Dirac impulse models of IEDs, on ICA, or on EEG synchronization metrics, according to the rationale that epilepsy is a manifestation of highly synchronous neuronal firing. Our experiments on data from twelve patients unambiguously showed that the MWF-based predictors are superior to the other kinds of predictors, in the sense that they allowed a correct localization of the ictal onset zone in more cases, and with higher statistical confidence. At the statistical threshold for maximal specificity, the novel predictor successffully inferred the location of the IOZ in 11/12 cases, versus 10/12 for the ICA-based predictor, 9/12 for the Dirac predictor, and 3/12 and 0/12 for two synchronization-based metrics. At more stringent thresholds, the discrepancy was even higher. These results refute earlier claims made in (Abreu et al., 2018a), where synchronization metrics came out on top. Compared to the latter study, we validated our approach on more patients (12, vs. 4 in (Abreu et al., 2018a)), and with more rigorous evaluation criteria, based on a comparison to the ground truth IOZ. The ground truth we used was the manually delineated resection zone for patients that had undergone surgical treatment and that were seizure-free afterwards, or otherwise the hypothetical resection zone, based on concordant evidence from multiple modalities other than EEG–fMRI, for patients that were ineligible for or refused surgery (cfr. Section 4.2.1, 4.2.5 and 2.4.2). We concluded that the time-varying output power of the MWF was a more robust representation of IEDs than any of the other metrics, even to the extent that the obtained time course seemingly managed to correct human errors during annotation. Our approach has a very low complexity, and comes almost as a 'free lunch' when annotations of interictal spikes in the EEG signals are already available. Furthermore, the one case in which our method failed could be easily identified as the result of some residual artifacts, likely because insufficient training data was available for the MWF to learn the spike characteristics sufficiently accurately. This is a foreseeable difficulty, and a quick manual inspection of the MWF allowed to impute such remaining artifactual periods, and obtain 100% sensitivity. For patients with more than one spike type, equally many MWFs could be trained if annotations for each type are separately available. The multiple resulting predictors could then allow to differentiate the spatial onset and spread of these classes of spikes, similar to the approach taken in classical EEG-correlated fMRI analysis. Lastly, MWF-based EEG enhancement may be used to increase the yield in EEG–fMRI studies, for example in cases where no IEDs could be identified on the EEG. I.e., the MWF may be calibrated on EEG data and IED annotations recorded before or after the EEG–fMRI acquisition, which allows more time and a higher

SNR to observe sufficiently many IEDs (the EEG noise characteristics should still be estimated from the intra-MR recording). The filter can then be applied to intra-MR EEG data, in which it might pick up subtle IEDs which are hard to notice by a human observer. An analogous approach based on matched filtering has already led to promising results Grouiller et al. (2011), hence it is expected that the MWF-based enhancement could similarly lead to an increased yield.

In **Chapter 5**, we have introduced a structured CMTF to decompose simultaneously recorded EEG–fMRI from individual patients into a set of sources which are characterized in the spatial, temporal, and spectral domain, via two coupled rank-1 terms, i.e., outer products of vectors that encode relative scaling. I.e., the rank-1 structure of each source allows easy interpretation in terms of relative intensity at different time points, EEG channels, voxels and frequencies, via the modulation encoded by the coefficients in these vector signatures. Our approach was inspired by the multiway partial least squares method in (Martínez-Montes et al., 2004), with the important distinction that we release the assumption that the neurovascular coupling mechanism, in the form of an HRF, is a priori known. Rather, by introducing a Toeplitz-structured block factor with unknown entries in our model, a parametric HRF is co-estimated with the sources from the patient's data. The HRF we estimate describes a coupling characteristic between spectral power fluctuations in the EEG and the BOLD signal, and as such is a proxy of the true neurovascular coupling, which directly relates neural 'events' of activity to the BOLD signal. Hence, this algorithm lays the basis for a patient-specific data fusion approach in which strong and biasing assumptions on neurovascular coupling need not be made. In this chapter, we considered a single HRF for the whole brain, which is still a quite stringent constraint, but already one step away from fully specified, non-customized models. We demonstrated the feasibility of this algorithm by running a series of simulation experiments. These showed that the algorithm generally behaves as expected, and is able to correctly estimate the sources as long as the SNR is not too low. Analogously, the accuracy of HRF estimation depended on the initialization of the HRF parameters to values that are not excessively far from the optimal solution. The results of this chapter already hinted towards a problem which is inherent in the use of non-convex optimization methods for BSS of tensorial data, which is the dependency of the solution's quality and trustworthiness on the initialization for the parameters which need to be estimated.

Since the potential impact of the algorithmic initialization strategy is high, we treated this issue in more detail in **Chapter 6**. The cause of this problem is the fact that most BSS methods, and especially those involving tensors are NP-hard: they cannot guarantee to find the optimal solution within a reasonable time, due to the existence of many locally optimal solutions and the absence of an

initialization strategy that will provably make the algorithm converge to the global optimum. While we cannot remove this difficulty, which is inherent to the nature of BSS algorithms, we aimed to alleviate its implications. I.e., the fact that the solution of the algorithm can vary for different initializations of the parameters may stem us pessimistic, as no more robust approach exists than to rerun the optimization procedure for different starting values, and then select the solution in one of the runs based on some criterion[1]. We have automated this approach in a principled way, which exploits the structure of the problem. Our philosophy is that we adhere more trust to estimates for the factors which are found in several or many runs of the algorithm, compared to estimates which are encountered only once or a few times. The central idea is then that we can represent the factors obtained in different algorithmic runs as a graph, with edges that encode similarity relationships between the factor sets in different runs (nodes in the graph), which theoretically obeys a natural property known as cycle consistency. By means of a structured graph clustering algorithm which exploits this consistency property, we manage to quickly and easily assess which factors are found in the solution of few or many runs of the optimization. As such, the algorithm implements a reproducibility analysis, which can inform the human user about the reliability of certain components. While the idea of clustering the solutions of different runs was not new (cfr. (Himberg et al., 2004)), we reformulated the problem in the language of graph theory, in which the structure of the problem (based on repetitions) becomes apparent and useful, and provided a solution for any BSS algorithm which relies on factorizations of tensors of arbitrary order, and also coupled tensor factorizations. On a pragmatic note, our algorithm employs only a single, interpretable parameter: the threshold (between 0–1) for assessing at which point a similarity relationship between components is deemed sufficiently strong to deserve a place in the graph. This is in contrast to the popular ICASSO algorithm, in which clustering is done based on a set of less intuitive, heuristic parameters which measure cluster quality. Through experiments with real and simulated data, we showed that our method was also more effective than ICASSO, in terms of identifying the appropriate clustering structure. With our contribution, we offer a practical solution to researchers and practitioners that apply BSS algorithms to noisy data, which can aid them in the (in our opinion very necessary) stability assessment of their obtained results.

Armed with the tools developed in Chapter 4, 5 and 6, we designed an integrated and structured EEG–fMRI data fusion framework in **Chapter 7**. We combine

[1]This criterion need not necessarily be the lowest approximation error: since data are frequently contaminated by noise, the solution with the lowest error may have much less meaningful factors than the solution with a slightly higher error. Degeneracy of components is one pathological situation in the context of low-rank tensor decompositions in which this problem becomes very apparent.

the previously discussed methods into a full pipeline, which performs data enhancement (cfr. Chapter 4), tensor-based blind source separation via a structured CMTF (cfr. Chapter 5) and stability analysis (cfr. Chapter 6). On top of that, we employ nonparametric statistical inference methods to create maps of significant (de)activation of the found neural sources. Our method explicitly aims to account for variations in the shape of the HRF over brain regions and patients, in an attempt to make EEG–fMRI analysis more flexible and resilient against model misspecification. To encode this desired flexibility, we used the notion of parametric basis functions for the HRF, which each contribute a Toeplitz-structured block to the CMTF. We have experimentally tested our pipeline on real EEG–fMRI data from twelve patients, and found that the structured CMTF is able to robustly extract IED-related sources, alongside other sources which capture neural, physiological, or non-physiological variations. Through the use of the MWF-based predictor (cfr, Chapter 4), we were able to identify the IED-related source, and apply a statistical test to find co-activated or deactivated regions in the brain, which may point to the ictal onset zone during presurgical evaluation. Additionally, we turned the 'curse' of unknown and variable neurovascular coupling into a 'blessing': after fitting the sCMTF model to the data, we analyzed in which brain regions the HRF had particularly unusual shapes compared to the rest of the brain, as quantified by an entropy metric. Interestingly, these entropy maps themselves often proved to be helpful in detecting the ictal onset zone in the fMRI images, as they identified additional brain regions compared to the sole IED (de)activation maps. Hence, the main merit of this pipeline is that it provides additional biomarkers that can inform a resective surgery. This strengthens the position of EEG–fMRI as a potential clinical tool, which can be used to complement other imaging modalities in the characterization of refractory epilepsy. On top of that, the CMTF factors also reveal spectral, spatial, and temporal characteristics of the sources, which can further be used to characterize the IEDs, but also to identify additional neural activity such as resting-state networks, and to isolate artifacts. Due to the technical challenge posed by the simultaneous acquisition of EEG and fMRI, standalone fMRI measurements are much more common. Unfortunately, in its application to refractory epilepsy, our pipeline heavily relies on the IED-related info, which can only be obtained through electrical recordings. However, we reckon that the novel pipeline could be also usefully applied to the fMRI-only use case, if the purpose is to conduct a study in which multiple RSNs are to be estimated, and not (solely) the IED-related (de)activation map. This would require only trivial simplifications to parts of the pipeline, in which a single factorization would be computed instead of a coupled factorization. End-to-end computation time for one patient's data—assuming the IED annotations are already available—is on the order of several hours, depending on the exact length of the recording. We have proposed various

strategies to alleviate this computational burden, without incurring a significant performance loss.

In summary, this thesis provides new signal processing tools for EEG and fMRI data fusion, using a diverse array of methodologies. Nevertheless, a central theme in all methods is the careful balancing act between inserting prior knowledge about the data, and acknowledging the absence of such knowledge. When the former dominates, we are pushed towards model-based algorithms, while a heavier importance of the latter promotes strongly data-driven approaches. When performing EEG enhancement, we allowed an MWF to freely adjust its parameters based on the data, but equipped with prior examples of IEDs versus non-IED data segments. I.e., it is a classical supervised learning algorithm, in which we simply formulate *what* we want the algorithm to achieve, but not *how* it needs to accomplish its task. On the contrary, the CMTF methods are fully unsupervised: after the data are provided, they *blindly* transform them into a set of estimated sources, with spatial, spectral, and temporal characteristics that had to be discovered from scratch. Nonetheless, we subtly controlled the modeling freedom of these CMTF algorithms, by imposing that a structural relationship between the EEG and fMRI domains exists, in the form of convolutions with (a priori unknown) HRFs. Additionally, we regularized these HRFs via low-dimensional parametrizations, to favour smooth solutions. In this way, we nudged the CMTF algorithms in a certain direction by defining bounds on the operation at the modeling level. In general, we have aimed to insert the minimal amount of prior constraints which were necessary to obtain meaningful results. However, a judgement on the relative degrees of 'data-driven' versus 'model-driven' operation is in the eye of the beholder, as people tend to adhere different subjective importance to certain explicit or implicit modeling assumptions. We envision that some of our methods, and notably the sCMTF, with some minor modifications, may also be used to analyze resting-state EEG–fMRI activity from healthy patients, to map brain-wide RSNs of neuronal assemblies that are recruited for sensory, cognitive and motor processes. In that case, a coupling between spontaneous EEG spectral power fluctuations (instead of IED occurrences) and BOLD changes would be sought. This has been attempted using an CMTF-like approach in (Martínez-Montes et al., 2004), and with related techniques in e.g. (Mantini et al., 2007) and (Mareček et al., 2017). With our framework, RSNs may be extracted more accurately from simultaneous EEG–fMRI if substantial deviations from the canonical HRF are present.

8.2 Future perspectives

Extending the application of MWF-based enhancement

On the application level, there are several future research directions which are worth exploring. In the previous section, we have stressed that the MWF-based EEG enhancement method may be useful for those patients with low-SNR EEG during fMRI acquisition (even after MR artifact removal), or in which few or no spikes can be observed. The MWF may be partially calibrated on annotated spikes in EEG data recorded outside of the MR scanner, which is often amply available, since it is much easier to record. We have seen that higher amounts of training or calibration data typically improve the MWF's performance. Additionally, in this way it may still be possible to use the MWF-based predictor for patients for which no spikes could be detected during fMRI recording, in case spike-like waveforms are present in the data, but difficult to detect for a human or an algorithm which is not adapted to the data of the specific patient (Grouiller et al., 2011). This hinges on the assumption that the IEDs' characteristics, as observed on the EEG outside the scanner, are representative for the IEDs' characteristics inside the scanner. There seems to be no reason to believe the opposite: we reckon that if differences have been observed, it is not unlikely that they are caused by the severe artifact removal processing steps which need to be applied to the EEG data beforehand (see Section 2.5.3). As such, the we expect that the MWF-based enhancement could lead to an increased yield of EEG–fMRI studies in refractory epilepsy.

Optimizing the optimization

On the side of the tensor-based blind source separation methods, we expect performance improvements stemming from innovations in the optimization algorithms. Namely, we empirically found that Gauss–Newton type of optimization algorithms are a preferred choice over ALS or quasi-Newton methods, in the sense that they lead to reproducible solutions, which are less sensitive to the specific initialization. Unfortunately, they are also more computationally costly. I.e., in each iteration of the optimization, they compute a new step direction $\mathbf{p}$ based on the gradient $\mathbf{g}$ and Hessian $\mathbf{H}$ of the cost function, by (approximately) solving $\mathbf{Hp} = -\mathbf{g}$ for $\mathbf{p}$. In Gauss–Newton methods, this Hessian matrix is approximated by the Gramian of the Jacobian (i.e., $\mathbf{H} \approx \mathbf{J}^{\mathrm{T}}\mathbf{J}$), and has dimensions $P \times P$, where P is the total number of parameters to estimate, and where $\mathbf{J}$ has dimensions $N \times P$, where N is the number of data observations. This is more accurate than the approximation made in quasi-Newton methods, but requires many operations if either the number of parameters or the size of the involved matrices and tensors becomes large. Furthermore, the above linear system grows often so large that it cannot

be solved exactly within reasonable time, and iterative methods are used to find an approximate solution for the step in this inner loop (in Tensorlab 3.0, conjugate gradients method are used, which require at most P iterations). Faster convergence to the optimal solution $\hat{\mathbf{p}}$ is possible if the linear system can be preconditioned. For factorizations of (unconstrained) single tensors, and some cases of coupled factorizations, such preconditioning strategies are available. We have exploited this in Chapter 7, where we computed the PARAFAC signatures of EEG tensors (without fMRI), to obtain good initializations for the subsequent sCMTF. However, for factorizations with constraints, or for ad-hoc coupled factorizations as the like in this thesis, no preconditioning was available in the software we used. This meant that computing the step size could generally be a slow process, and convergence in the inner and outer loop of the optimization would be compromised. Even for moderately sized problems, such Gauss–Newton algorithms become prohibitively slow. I.e., despite a strong compression of the fMRI data into a set of 246 parcels (compare to the roughly 10^5 voxels in the original data), a single run of the sCMTF algorithm would typically take many hours to until convergence under a reasonable tolerance. Hence, we had to resort to a quasi-Newton optimization algorithm for data fusion. We hypothesize that if preconditioners for complex models become available, they could lead to results with the same accuracy as they do for the single-tensor case. Furthermore, they would allow to scale up our sCMTF algorithms to problems of appreciable size, i.e., in which there is no need for drastic dimensionality reduction (and loss of information) of the fMRI data.

Novel tensor representations

This thesis showcased methods for fusing single-patient EEG and fMRI data and performing blind source separation in a principled manner. Different data fusion endeavors, mostly for multi-subject EEG and fMRI, have been undertaken in the past, and in a context different than epilepsy. The overview paper from (Karahan et al., 2015) exemplified that tensor-based data representations are the LEGO blocks for multimodal data integration in neuroimaging. Thanks to the abstraction offered by a structured data fusion language, the design of new blind source separation algorithms can be uncoupled from the enabling optimization algorithms under the hood (Sorber et al., 2015), much like a wide variety of software can be run on powerful, general purpose hardware. We have made heavy use of tensor representations for EEG data, but have only used conventional matrix representations for fMRI data. This is perhaps ironic, since the raw fMRI is naturally a fourth-order tensor, with x, y and z spatial modes, and a temporal mode. The challenge is however not to obtain a tensorial representation—this is easy and almost a natural consequence of the acquisition of physical measurements—but to find one in which it is

meaningful to decompose the data into a compact set of low-rank components. The aforementioned 4D representation of the fMRI data is not very suitable, since the spatial support of neural activity is irregular, and cannot be captured by a low-rank structure in the spatial modes. Only after collapsing x, y and z modes into a single, long mode can a low-rank structure of space $\times$ time be warranted, on the condition of stationarity of the spatial patterns over time. We are not aware of any convincing higher-order tensorization for fMRI data which preserves one or more of these individual spatial modes (a few simulation examples aside).

A potential tensor representation for single-subject fMRI data may come from wavelet-like transformations. For fMRI, customized wavelets with shapes similar to hemodynamic responses have been developed (Khalidov et al., 2011). These may be used to cast the data in the format wavelet scales $\times$ time bins $\times$ voxels, and use low-rank factorization models for blind source separation of the activity in components with spatial, spectral and temporal signatures, just like we did for EEG data.

For multi-subject fMRI data, a third order representation as subjects $\times$ time points $\times$ voxels has been used since some time, in tensor probabilistic ICA Beckmann and Smith (2005). However, decomposing into sums of rank-1 terms makes the assumption that the BOLD time courses in all subjects are time-locked, e.g. because they are evoked by stimuli which can be accurately timed in a controlled experiment. Evidently, this approach does not straightforwardly generalize to multi-patient data in epilepsy, since the occurrence of IEDs is by no means regular intra-patient, let alone across patients.

The curse of non-interpretability

In this thesis, we have have focused exclusively on BSS algorithms that estimate rank-1 structured sources, as opposed to a more general class of algorithms, which extract sources with a low-rank structure. The former is the most simple, and by far the most common source model used in BSS algorithms (e.g., ICA, NMF, CPD, CMTF), thanks to the ease of interpretability. Namely, the rank-1 structure allows a practitioner to think in terms of modulation of the source's intensity in the different modes, encoded by the coefficients in the vectors of the outer product. By contrast, low-rank sources (e.g. in the $(L, L, 1)$-decomposition) inevitably incur rotational ambiguities that impede such interpretation. Especially in neuroimaging, where exploratory analysis is already cumbersome due to the low SNR and complexity of the data, we expect that the attractiveness of interpretability will remain of paramount importance. Hence, due to this 'curse of non-interpretability', which affects BSS with more

general, low-rank structured sources, we deem it unlikely that such algorithms will gain a large attraction in this community.

Coupling in non-temporal domains

While we have only considered coupling in the temporal mode, which naturally arises as a consequence of the simultaneous recording of EEG and fMRI, alternative formulations are possible.

It would be interesting to investigate whether a coupling constraint in the spatial mode could improve the extraction of meaningful sources[2]. This could be conceived analogously to the temporal coupling constraint that we have used, which 'transfers' a time course $\mathbf{s}_r$ from the (high temporal resolution) electrophysiological (EEG) domain to the (low temporal resolution) BOLD domain via a matrix multiplication $\mathbf{H}_k\mathbf{s}_r$ with a convolution kernel, which essentially implements a temporal smoothing. Equivalently, in the spatial mode, EEG topographies provide low-resolution, blurred observations of the spatial support of the underlying neural activity in the voxel space, via electrical conduction through brain and skull. Indeed, this is precisely the 'forward model' which is used in the EEG source imaging literature, where an EEG topographic map $\mathbf{m}$ is considered to arise from a multiplication of a neural activity vector $\boldsymbol{\psi}$ in the source space with a leadfield matrix $\boldsymbol{\Gamma}$, which describes the attenuation with which an electrical field amplitude at any particular grid point or voxel can be observed at any EEG sensor (Nunez et al., 2006; Pascual-Marqui, 1999; Hallez et al., 2007). If we make the simplistic assumption that BOLD intensity in each voxel is proportional to the amplitude of the local electrical neuronal activity, and the leadfield could be accurately computed, then we could impose the constraint $\mathbf{m}_r = \boldsymbol{\Gamma}\mathbf{v}_r$ for each source r. A similar constraint has already been proposed for BSS-based EEG source imaging in (Valdés-Sosa et al., 2009; Becker et al., 2015), and for EEG–fMRI fusion in (Karahan et al., 2015). This approach inserts additional prior info (in the form of $\boldsymbol{\Gamma}$), and hence shrinks the search space for the optimization algorithm. Furthermore, this form of coupling is only expected for truly neuronal sources, which are expressed in both the EEG and fMRI data. By contrast, artifacts are largely modality-specific, and do not follow a clear coupling characteristic. As such, the use of this spatial constraint may enable a more robust estimation of neural sources of interest, provided that the assumptions are tenable (cfr. Section 2.2.1 and 2.3.1 for a treatment of this topic). Going further, if high-density EEG recordings are available, it might be feasible to perform source imaging prior to fusion, to obtain EEG signals

[2]Note that in our proposed sCMTF method, no coupling between the spatial modes of EEG and fMRI is present, hence these modes are 'dangling', and their observations might e.g. be reshuffled without affecting the outcome.

in the source space, which can be coregistered to the MR space. In that case, a coupling constraint between the EEG and fMRI spatial signatures may be used that is simpler than the leadfield model (e.g., a simple Gaussian blurring kernel or even the identity map could be sufficient). Instead of BSS, also other types of analyses could then be performed in a coupled fashion, e.g., Granger causality analysis in a tensor formulation (Karahan et al., 2015), in which the two connectivity tensors of EEG and fMRI are coupled in the spatial mode.

Alternatively, if data from a homogeneous group patients are available, it could be worthwhile to devise schemes in which EEG and fMRI patient data are concatenated, and coupling is imposed along this new 'patient' mode. As discussed before, this requires that the other modes do not strongly depend on patient-specific characteristics, such as e.g. the temporal pattern of IED occurrences, so that a low-rank factorization is appropriate. CMTF formulations with a coupling constraint in the patient mode have been used to explore epileptic networks which are common over patients with temporal lobe epilepsy (for whom the spatial support of the BOLD activity is roughly similar) (Hunyadi et al., 2016), but also for the analysis of EEG–fMRI data from e.g. schizophrenia patients (Acar et al., 2017b, 2019).

Supervised, unsupervised, or both?

As mentioned in the previous section, we have aimed to find a midway between fully supervised and fully unsupervised approaches. However, the methods presented here are not the only way of accomplishing a good balance between 'listening to the data' and 'trusting our prior beliefs'. Bayesian inference forms a quintessential class of algorithms that explicitly implements this philosophy in ways that that are firmly rooted in rigorous statistical principles. by weighting the prior probability of the unknown parameters versus the data's likelihood under any realization of these parameters, Bayesian methods obtain parameter estimates that maximize a posterior probability function. Such algorithms may e.g. be used to encode biophysical priors, which promote plausible solutions. Such Bayesian methods have already been applied to integrate multimodal information from EEG and fMRI (Daunizeau et al., 2007; Valdes-Sosa et al., 2009; Luessi et al., 2011; Trujillo-Barreto et al., 2001). However, while they are theoretically and intuitively appealing, they are often hierarchical models with complicated expressions for the probability densities, and as such they may be very hard to optimize, even approximately. Even so, the Bayesian mindset is becoming more and more adopted, and is central to e.g. the FMRI Software Library ecosystem (Woolrich et al., 2009).

With the use of more complex tensor-based EEG–fMRI fusion schemes, it is

likely that more prior information needs to be supplied to the algorithm to function properly and find meaningful estimates for all parameters, e.g. by plugging in neural mass models or brain connectivity models. Bayesian methods could play a role here. As such, the barrier between fully data-driven (e.g. 'blind' source separation) and model-driven methods becomes blurry, and is no longer an absolute categorization, but more the interpretation of different people. However, even 'fully' unsupervised, data-driven methods are not free from inherent prior biases, as the choice of algorithm influences which silent assumptions are made about the data[3]. We therefore believe that ideas of Bayesian inference may be beneficially applied to complex EEG–fMRI data fusion schemes, albeit with caution. For example, methods already exist that estimate HRFs with Bayesian techniques, after specifying a range of 'reasonable' parameters (Woolrich et al., 2004). Crucially, the authors only considered HRFs which strictly lag the neural correlate or input. While this is justifiable for experimental data in which external stimuli were used, it might be dangerous to straightforwardly extend this to epileptic data. Indeed, our experiments in this thesis have corroborated the findings of independent researchers in the past, which stated that BOLD responses could sometimes be observed *before* the appearance of the associated interictal spike on the EEG. This is no violation of causality, and could simply be a sign that the spike itself is a culmination or symptom of a preceding neural process, and not the real cause of the BOLD change. However, if spikes are treated as 'inputs', the algorithm is prevented from revealing HRFs which lead the spike onset, unless liberal parameter ranges are supplied. Hence, caution is required when applying such Bayesian schemes to data about which many physical aspects are not yet fully understood.

Estimating the latent dimensionality

Our algorithms (like most signal processing techniques) are burdened with the selection of adequate hyperparameters. Notably, the selection of the number of sources R is a recurring and often hard problem in blind source separation, and our methods are no exception. Some prior efforts have been directed at this rank estimation problem. Notably, some existing methods are based on information theory (Liu et al., 2016) or automatic relevance determination with Bayesian principles (Mørup and Hansen, 2009). We have empirically found

[3]For example, a pervasive concept in signal processing and machine learning is the minimization a sum of squared errors in various modeling tasks. Even though a user may not impose additional constraints, such algorithm implicitly encodes the assumption that the errors follow a Gaussian distribution, and are independent from each other. Whether this is a tenable assumption depends on the context of the application, and it is definitely not a trivial issue. Indeed, we stressed in Chapter 3 that fMRI data violate this assumption, which can lead to inflated Type I errors if no corrective measures are in place.

that both these methods underperform for the kind of data we used. I.e., they invariably predicted a very large number of sources in the data. While there is no way to tell whether this is the actual nature of the data, it is likely that these methods simply break down in low-SNR regimes. Furthermore, in blind source separation applications it is rarely of interest to extract hundreds or even tens of sources, even if this means allowing some non-negligible approximation error. We have taken a more pragmatic approach, in which the merit of factor estimates of different R was evaluated based on a set of ad-hoc criteria, which are however more meaningful for the end objective. One generic tool which we have found useful is the generically applicable core consistency diagnostic (Bro and Kiers, 2003). We therefore expect that further innovation in a priori rank estimation, but especially in post hoc rank selection methods may improve the performance of tensor-based blind source separation methods.

Deep architectures for EEG–fMRI fusion

Nowadays, deep learning methods are gaining traction in a wide range of fields, including biomedical signal processing. As briefly touched upon in Chapter 3, they largely owe their success to the fact that they use a 'deep' cascade of abstraction layers with nonlinearities, which allow them to capture complex relationships that exist in the data. It is obvious that deep learning techniques can be fruitfully applied to subtasks in EEG–fMRI analysis, e.g. accurate IED detection (Hao et al., 2018; Wilson et al., 1999), or artifactual IC classification similar to (Salimi-Khorshidi et al., 2014).

A more ambitious endeavour would be to learn the complex relationship between EEG and fMRI via deep neural networks. This has been attempted in (Liu and Sajda, 2019), where the authors tried to predict EEG data from fMRI data (and vice versa) using a convolutional neural network, but so far only obtained slightly underwhelming results on simulated data. Potentially, recurrent neural networks could be used to learn complex temporal relationships between EEG and BOLD signals, and refine the IED segmentation. With architectures based on autoencoders, it may be feasible to learn a low-dimensional representation that results from nonlinearly compressing EEG data, and from which certain fMRI features can be reconstructed (or vice versa). For an individual dataset, it is likely that the nonlinear mapping performed by a deep network could more accurately model the complex neurovascular coupling between EEG and fMRI data. If it were possible to reliably train such a network, it could equip neurologists that don't have access to an MR scanner (e.g. in developing countries) with a tool to extract useful clinical, fMRI-like information, based on EEG recordings alone. Alternatively, if out-of-scanner EEG recordings could be complemented with an anatomical MR scan of the patient, and fed into a

network that extracts useful localizing features, it could save the clinicians a lot of time, since no tedious intra-MR EEG recording would need to be set up. However, we reckon that the difficulty of this learning task is formidable, because it is not obvious which target fMRI features should be used, whether it is possible to learn such a mapping altogether, and especially since training such a network would require a very large amount of annotated EEG–fMRI datasets, which is currently lacking—even if non-epileptic EEG–fMRI datasets could be included somehow, in an effort to scale up the data volume. Therefore, our best guess is that more traditional methods based on EEG-correlated fMRI analysis and BSS will not disappear anytime soon[4], and will stay prominent for the analysis of EEG–fMRI data, especially in refractory epilepsy.

New fuel for EEG–fMRI data analysis

To date, EEG–fMRI has not yet earned the status of a routine clinical technique for presurgical evaluation in refractory epilepsy. This is in large part due to its fluctuating degree of success in terms of localizing the ictal onset zone, because in many patients, no IEDs, or no significant IED-related BOLD changes are observed. In this thesis, we have devised algorithmic methods that can increase the yield of studies, by employing flexible but structured models that are fitted to the EEG and fMRI data. We have shown how these techniques 'dig deeper' than traditional EEG-correlated fMRI analyses—or at least, in the way they are commonly applied—and are more sensitive in the isolation and localization of interictal activity. Hence, to a certain extent, the absence of meaningful results in EEG–fMRI studies can be attributed to the under-exploitation of the data.

Inevitably, this thesis has its limitations. Primarily, we have been able to test the novel tools on data from only twelve patients, for which sufficient evidence was available to formulate a ground truth to evaluate against. As such, the performances we obtained might be biased by the small sample size, and by the patient selection. An evaluation on data from a much larger cohort of patients would be in order, to get a more accurate view of the potential and shortcomings of the methods presented here.

With the contributions in this thesis, we hope to have convinced at least some readers that EEG and fMRI are rich measurements, which have more to offer than what appears on the surface, as long as the right data mining strategies are applied. EEG–fMRI is no longer part of the clinical practice in the university hospital UZ Leuven, where the data used for this thesis were recorded. However,

[4] A similar prediction has been made for the GLM in general (which underlies EEG-correlated fMRI analysis) (Poline and Brett, 2012).

it is our firm belief that with the right tools, EEG–fMRI has its merit and can meaningfully contribute to the presurgical evaluation of patients.

Appendix A

Single-channel EEG classification by multi-channel tensor subspace learning and regression

ABSTRACT | The classification of brain states using neural recordings such as electroencephalography (EEG) finds applications in both medical and non-medical contexts, such as detecting epileptic seizures or discriminating mental states in brain-computer interfaces, respectively. Although this endeavor is well-established, existing solutions are typically restricted to lab or hospital conditions because they operate on recordings from a set of EEG electrodes that covers the whole head. By contrast, a true breakthrough for these applications would be the deployment 'in the real world', by means of wearable devices that encompass just one (or a few) channels. Such a reduction of the available information inevitably makes the classification task more challenging. We tackle this issue by means of a multilinear subspace learning step (using data from *multiple* channels during training) and subsequently solving a regression problem with a low-rank structure to classify new trials (using data from only a *single* channel during testing). We demonstrate the feasibility of this approach on publicly available EEG data recorded during a mental arithmetic task.

This chapter is a modified version of Van Eyndhoven, S., Boussé, M., Hunyadi, B., De Lathauwer, L. & Van Huffel, S. "Single-channel EEG classification by multi-channel tensor subspace learning and regression", in IEEE 28th International Workshop on Machine Learning for Signal Processing (MSLP 2018, Aalborg, Denmark), Sep. 2018; with changes being limited to layout.

A.1 Introduction

Non-invasive neuroimaging techniques have received considerable interest in the last couple of decades because they allow to conveniently investigate, classify, or detect a wide range of physiological phenomena taking place in the brain. Among these techniques, electroencephalography (EEG) has become the workhorse in many applications, thanks to its low cost, relatively easy setup of sensors on the scalp, and excellent temporal resolution. Well-known examples of such EEG-based applications are brain computer interfaces (BCI) to restore communication for disabled people, perform clinical research or develop consumer products for e.g. mental state monitoring (Brunner et al., 2011). Another important application is the automatic detection of seizures in epileptic patients (Acharya et al., 2013).

What unites most of the research trends in these very diverse domains is the focus on lab- or clinical-grade EEG systems, restricting the use in settings outside the lab. For example, EEG measurement setups are often bulky and require substantial manual intervention to use, and more importantly employ a large set of electrodes that covers almost the whole head. Evidently, these systems cannot simply be used outside the lab or clinic. One important hurdle on the way to convenient EEG devices for daily use is a (drastic) reduction of the number of electrodes (Liao et al., 2012; Looney et al., 2012). This is not only a hardware-related issue: many established signal processing and machine learning approaches for EEG owe their success precisely to the fact that they exploit the spatial diversity that is present in the traditional, multi-channel measurement (Blankertz et al., 2007; Lotte et al., 2007). Hence, existing processing techniques might have to be reengineered or replaced by new approaches for use in current and future EEG devices that employ only one or a few channels. Several attempts in this direction have already been made: some authors have proposed algorithms that find 'good' subsets of EEG channels to perform certain tasks (Lal et al., 2004; Lou et al., 2008) or derive complex features from the time courses recorded at a few channels, that aim to compensate in some way for the loss of spatial diversity (Liang et al., 2012).

Here, we take a different approach, and train a model using all available channels but classify new trials using only single-channel data. This follows the realization that the restriction on the number of channels may present itself *only* at the time of deployment of an EEG system. In other words: it can be acceptable in many cases to have a calibration or training session where a larger set of channels is used to prepare or fine-tune an EEG device, after which it is taken into use ('testing phase') with only a low number of channels. In this paper, we present an EEG classification method based on existing work (Boussé et al., 2018) that

exploits multi-channel information during a training phase, and can operate on a single-channel EEG signal during testing. Firstly, since there are multiple modes of variation (channels × time points × trials × ...), the data are best represented as a tensor, which is a higher order generalization of vectors and matrices (first and second order tensors, respectively) (Sidiropoulos et al., 2017; Cichocki et al., 2015). Training then consists of a multilinear or tensor subspace learning step, in which appropriate bases are found for all modes of the data. During testing, it is then assumed that unseen trials approximately lie in the same multilinear subspaces as the training data. Hence, to perform classification, a data segment of a test trial is regressed simultaneously onto several subspaces/bases found during training; the resulting coefficients have a low-rank structure (because of the multilinear setting (Zhou et al., 2013)) and inform a decision about class membership. This generic workflow has previously been introduced for face recognition and irregular heartbeat classification (Boussé et al., 2018, 2017; Vasilescu and Terzopoulos, 2002). In this paper, we propose several extensions to this framework, to address difficulties that arise in the case of lower signal quality, such as in EEG data. Most importantly, we leverage the class labels of the available training trials to perform the subspace learning step in a supervised fashion, after tensorizing the data using a time delay embedding method. This approach is inspired by the common spectral patterns method presented in (Lemm et al., 2005) and allows to identify bases in which the data differ substantially between classes, in contrast to existing approaches, which use a simple unsupervised subspace learning step. Additionally, we robustify the classification step by non-arbitrarily fixing a sign ambiguity and by employing support vector machines (Suykens et al., 2002) and tree ensemble methods instead of a nearest neighbor approach.

To demonstrate the feasibility of the novel method, we apply it on publicly available EEG data that were recorded from subjects performing a mental arithmetic task (Shin et al., 2016).

A.2 Classifying single-channel data after multi-channel training

A.2.1 Notation and preliminaries

We mostly follow the convention in (Sidiropoulos et al., 2017; Cichocki et al., 2015) for the algebraic notation. We denote scalars, vectors, matrices and tensors by lower case (e.g. a), lower case boldface (e.g. $\mathbf{a}$), upper case boldface (e.g. $\mathbf{A}$) and upper case calligraphic letters (e.g. $\boldsymbol{\mathcal{A}}$), respectively. An Nth order

tensor $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is a multiway array that varies over N modes (e.g. sensors, time points, trials, ...) with dimensions I_1, I_2, ... , I_N, respectively. Slices of a tensor are obtained by fixing the index for one or several modes. When indices of all but one mode are fixed, and hence only variation over one mode is considered, we use the term *fiber*, which is a generalization for matrix rows and columns. A tensor $\boldsymbol{\mathcal{A}}$ can be matricized by stacking all fibers of a certain mode-n in a matrix. This is also known as the mode-n unfolding of $\boldsymbol{\mathcal{A}}$ and is written as $\mathbf{A}_{(n)}$. Equivalently, the vectorization of $\boldsymbol{\mathcal{A}}$ is obtained by stacking all mode-1 fibers in a long vector and is written as vec($\boldsymbol{\mathcal{A}}$). The mode-n multiplication of a tensor $\boldsymbol{\mathcal{A}}$ and a matrix $\mathbf{V}$ is denoted by $\boldsymbol{\mathcal{A}} \times_N \mathbf{V}$ and consists of the left-multiplication of all mode-n fibers of $\boldsymbol{\mathcal{A}}$ with $\mathbf{V}$. The mode-n rank of a tensor is the rank of the mode-n unfolding of the tensor. The set of mode-n ranks (jointly called *multilinear rank*) is in general different from the tensor rank, which is the minimal number of rank-1 tensors to fully describe the tensor. Such a collection of rank-1 terms is then called a canonical polyadic decomposition. Every rank-1 tensor in its turn is the outer product of N vectors, denoted by $\mathbf{a}^{(1)} \otimes \mathbf{a}^{(2)} \otimes \ldots \otimes \mathbf{a}^{(N)}$. The Kronecker product of two matrices $\mathbf{A} \in \mathbb{R}^{I_1 \times I_2}$ and $\mathbf{B} \in \mathbb{R}^{J_1 \times J_2}$ is indicated by $\otimes$ and defined as the $I_1 J_1 \times I_2 J_2$ matrix

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} \mathbf{A}(1,1)\mathbf{B} & \cdots & \mathbf{A}(1,I_2)\mathbf{B} \\ \vdots & \ddots & \vdots \\ \mathbf{A}(I_1,1)\mathbf{B} & \cdots & \mathbf{A}(I_1,I_2)\mathbf{B} \end{bmatrix} \tag{A.1}$$

A useful relationship between the outer and Kronecker product is vec($\mathbf{a} \otimes \mathbf{b}$) = $\mathbf{b} \otimes \mathbf{a}$.

A.2.2 Training phase: subspace learning

Training of the classifier consists of finding appropriate subspaces or bases for all modes, in which the training trials' data can be represented. Afterwards, data from a new trial can be regressed or projected onto these subspaces, and the resulting coefficients can be fed into a classifier that estimates the class membership. Hence, ideally, subspaces should be sought in which the coefficients of trials from the same class lie close to each other, but far from coefficients of other trials.

Tensorization of the data

While a subject participates in several trials of a BCI paradigm, EEG data are being collected at D electrodes or channels, with T time points for each of P trials.

Hence, the data are naturally represented as a third order tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{D \times T \times P}$, with modes channels $\times$ time points $\times$ trials, indexed by d, t and p, respectively. We may expand the data by stacking $L-1$ delayed versions of the signal at every channel along a fourth mode, increasing the order of the tensor by one. This method is known as time delay embedding and was proposed for multi-channel EEG classification in (Lemm et al., 2005). Mathematically, the new tensor $\boldsymbol{\mathcal{X}}^e \in \mathbb{R}^{D \times T \times L \times P}$ is obtained as $\boldsymbol{\mathcal{X}}^e(d,t,l,p) = \boldsymbol{\mathcal{X}}(d, t-(l-1)\tau, p),\ \forall\ d,t,l,p$, where L is known as the embedding dimension and τ is the introduced delay, and has modes channels $\times$ time points $\times$ lags $\times$ trials. Every mode has an associated subspace or base in which the mode-n fibers can be represented.

Unsupervised subspace learning

A popular way to find the multilinear subspaces of a higher order dataset $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is the multilinear singular value decomposition (MLSVD), defined as

$$\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{S}} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \cdots \times_N \mathbf{U}^{(N)} \tag{A.2}$$

where every $\mathbf{U}^{(n)} \in \mathbb{R}^{I_n \times I_n}$ is a unitary matrix, and the core tensor $\boldsymbol{\mathcal{S}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is ordered and all-orthogonal (Sidiropoulos et al., 2017; Cichocki et al., 2015). It is a higher order generalization of the singular value decomposition (SVD), and as such the (truncated) matrices $\mathbf{U}^{(n)}$ contain bases that capture as much variance of $\boldsymbol{\mathcal{A}}$ along mode n as possible. They can be found by applying an SVD on unfolded versions of the tensor in every mode. As this approach disregards class label info, there is no guarantee whatsoever that the data of trials from different classes are represented (sufficiently) differently in these bases.

Supervised subspace learning

Discriminative subspaces might be found by taking into account the label information from the training trials. This is the philosophy of the popular common spatial patterns (CSP) algorithm (Blankertz et al., 2007), in which a set of spatial filters $\mathbf{w}_i$ is applied to a multi-channel EEG dataset $\mathbf{X}$ such that the filtered output signals $\mathbf{y}_i = \mathbf{w}_i^{\mathrm{T}}\mathbf{X}$ have a high power during trials of one condition, but a low power for trials of other conditions. The output power of several CSP filters can then be used as features for classification, assuming that changes in (frequency-specific) power in the EEG are related to the 'brain states' of interest. This approach was extended to the common spatio-spectral patterns (CSSP) algorithm, whereby time delay embedding is used to extend the set of channels, as described earlier, leading to filters that aggregate signals

over multiple channels *and* lags (Lemm et al., 2005). The matrix of CSSP filters $\mathbf{W}$ can be found as the solution to the generalized eigenvalue problem

$$\boldsymbol{\Sigma}_2 \, \mathbf{W} = (\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2) \, \mathbf{W} \boldsymbol{\Lambda} \,, \tag{A.3}$$

in which $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ are the covariance matrices of the EEG data belonging to class 1 and 2, respectively, and $\boldsymbol{\Lambda}$ is a diagonal matrix containing the generalized eigenvalues λ_i, which lie between zero and one. Filters $\mathbf{w}_i$ associated with a λ_i that is close to one have a higher output power during trials of class 2 compared to trials of class 1, which are amplified by filters with a λ_i that is closer to zero.

As we aim to classify new trials based on single-channel EEG, we cannot aggregate multiple signals spatially, but only temporally, i.e. by taking a weighted sum of the single-channel EEG over multiple lags (after time delay embedding). Hence, the covariances in (A.3) are computed on the unfolded data $\mathbf{X}^e_{(3)} \in \mathbb{R}^{L \times DTP}$, leading to finite impulse response (FIR) filters that can be applied over the third mode (lags):

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{X}}^e \times_3 \mathbf{W}^{\mathrm{T}} \tag{A.4}$$

The multilinear bases of the filtered data $\boldsymbol{\mathcal{Y}}$ are thus more discriminative than those of the original data $\boldsymbol{\mathcal{X}}^e$, as the power of time series in $\boldsymbol{\mathcal{Y}}$ correlates better to the class labels.

Feature extraction

The time-varying power of an EEG signal is computed by taking the sum of squares of every time course in $\boldsymbol{\mathcal{Y}}$ in sliding, non-overlapping windows of length m. The logarithm of the power values is used as a feature for classification and stored in a feature tensor $\tilde{\boldsymbol{\mathcal{Y}}}$. Ultimately, the MLSVD of this tensor is computed to obtain the training bases $\mathbf{U}_{\text{chan}} \in \mathbb{R}^{D \times D}$, $\mathbf{U}_{\text{time}} \in \mathbb{R}^{\frac{T}{m} \times \frac{T}{m}}$, $\mathbf{U}_{\text{filt}} \in \mathbb{R}^{L \times L}$ and $\mathbf{U}_{\text{trial}} \in \mathbb{R}^{P \times P}$ and a core tensor $\tilde{\boldsymbol{\mathcal{S}}}_y \in \mathbb{R}^{D \times \frac{T}{m} \times L \times P}$ that explains the interaction between basis vectors of the different modes.

A.2.3 Testing phase: tensor regression and classification

Regression with a low-rank solution

Note that, before truncation of the bases, the time varying power over the spectral filter outputs $\tilde{\boldsymbol{\mathcal{Y}}}^{c,p} \in \mathbb{R}^{1 \times T \times L \times 1}$ at every channel d, in every trial p, admits a representation as

$$\tilde{\boldsymbol{\mathcal{Y}}}^{d,p} = \tilde{\boldsymbol{\mathcal{S}}}_y \times_1 \mathbf{u}^{\mathrm{T}}_{\text{chan},d} \times_2 \mathbf{U}_{\text{time}} \times_3 \mathbf{U}_{\text{filt}} \times_4 \mathbf{u}^{\mathrm{T}}_{\text{trial},p} \tag{A.5}$$

in which $\mathbf{u}_{\text{chan},d}^{\text{T}}$ and $\mathbf{u}_{\text{trial},p}^{\text{T}}$ represent the dth and pth row of $\mathbf{U}_{\text{chan}}$ and $\mathbf{U}_{\text{trial}}$, respectively. Unfolding over the second and third mode yields

$$\begin{aligned}\tilde{\mathbf{Y}}_{(2,3)}^{d,p} &= (\mathbf{U}_{\text{filt}} \otimes \mathbf{U}_{\text{time}})\ \tilde{\mathbf{S}}_{y_{(2,3)}}\ (\mathbf{u}_{\text{trial},p} \otimes \mathbf{u}_{\text{chan},d}) \\ \tilde{\mathbf{Y}}_{(2,3)}^{d,p} &= \mathbf{B}\ (\mathbf{u}_{\text{trial},p} \otimes \mathbf{u}_{\text{chan},d})\end{aligned} \tag{A.6}$$

This equation describes the data $\tilde{\mathbf{Y}}_{(2,3)}^{d,p}$ in the unfolded multilinear subspace $\mathbf{B}$ using coefficients $\mathbf{u}_{\text{chan},d}$ and $\mathbf{u}_{\text{trial},p}$. As was noted in section A.2.1, the Kronecker product $\mathbf{u}_{\text{trial},p} \otimes \mathbf{u}_{\text{chan},d}$ can be seen as the vectorization of a rank-1 matrix $\mathbf{u}_{\text{chan},d} \otimes \mathbf{u}_{\text{trial},p}$. For this reason, expressions such as (A.6) have been referred to as Kronecker Product Equations (KPE) (Boussé et al., 2017) or Linear Systems with Canonical Polyadic Decomposition-constrained solution (Boussé et al., 2018). To classify a new trial q using data from EEG channel f, we first construct the matrix $\tilde{\mathbf{Z}}_{(2,3)}^{f,q}$ by computing the features as explained in previous section (analogously to the training data). Subsequently, we solve the following set of equations for $\hat{\mathbf{u}}_{\text{chan}}$ and $\hat{\mathbf{u}}_{\text{trial}}$:

$$\tilde{\mathbf{Z}}_{(2,3)}^{f,q} = \mathbf{B}\ (\hat{\mathbf{u}}_{\text{trial}} \otimes \hat{\mathbf{u}}_{\text{chan}}) \tag{A.7}$$

This corresponds to performing a regression of the new data $\tilde{\mathbf{Z}}_{(2,3)}^{f,q}$ onto the multilinear basis $\mathbf{B}$, in which the solution has a low-rank structure by virtue of the Kronecker product between the coefficients (Zhou et al., 2013). The vector $\hat{\mathbf{u}}_{\text{trial}}$ then holds coefficients that express the new data segment in the subspace of the trial mode, and can be used for classification.

Classification

In (Boussé et al., 2018, 2017), new data instances are classified by comparing the estimated coefficients $\hat{\mathbf{u}}_{\text{trial}}$ with rows of the trained matrix $\mathbf{U}_{\text{trial}}$ and assigning the label of the closest matching training data instance. This nearest neighbor (NN) approach may be adjusted or replaced by more robust classification methods that take the coefficients $\hat{\mathbf{u}}_{\text{trial}}$ as input features. In this paper, we compare the performance of seven different methods: K-nearest neighbors (`knn`) with K equal to 1, 3, or 5, least-squares support vector machines with a linear kernel (`svm-lin`) or radial basis function kernel (`svm-rbf`), random forests with 100 trees (`randfor`) and gradient boosting using 10 shallow trees that are trained with AdaBoost (`treeboost`). The classifiers are trained using the labeled rows of $\mathbf{U}_{\text{trial}}$ as features.

Determining the sign of the coefficients

When solving the tensor regression problem, a sign ambiguity and scaling ambiguity remains, i.e., the values in $\hat{\mathbf{u}}_{\text{trial}}$ can be scaled by an arbitrary factor (that can be negative), if the values in $\hat{\mathbf{u}}_{\text{chan}}$ are counterscaled by the same factor. Since the scale and sign can have a large impact on the classification, it is crucial to fix them. Although algebraically the sign cannot be retrieved when solving the regression problem, it may be inferred from the training data set. Namely, we may look at the matrices $\mathbf{U}_{\text{trial}}$ and $\mathbf{U}_{\text{chan}}$ and try to find 'anchoring variables', i.e. columns of those matrices that maximize

$$\text{score}_{i,\text{trial}} = |\sum_{c=1}^{C} \frac{\mu_{c,i}}{\sigma_{c,i}}| \quad , \quad \text{score}_{j,\text{chan}} = |\frac{\mu_j}{\sigma_j}| \tag{A.8}$$

Here, $\mu_{c,i}$ and $\sigma_{c,i}$ are the mean and standard deviation over values from the ith column of $\mathbf{U}_{\text{trial}}$ that belong to class c. Analogously, μ_j and σ_j are the mean and standard deviation over values from the jth column of $\mathbf{U}_{\text{chan}}$. A large score for a latent variable i indicates that this variable has a consistent sign, i.e. that its values lie sufficiently far from zero and are either positive and negative, and are not informative for the class label. Hence, the latent variable that maximizes $\text{score}_{i,\text{trial}}$ provides an 'anchor' whose sign dictates the sign of the corresponding entry of $\hat{\mathbf{u}}_{\text{trial}}$, and thus of the whole vector. Analogously, the latent variable j with the largest score $\text{score}_{j,\text{chan}}$ can be used to correct the sign of the vector $\hat{\mathbf{u}}_{\text{chan}}$. As a third criterion, we may impose that $\hat{\mathbf{u}}_{\text{chan}}$, that was estimated for a certain channel f, should have a positive cosine similarity with the corresponding row of $\mathbf{U}_{\text{chan}}$: negative values of $\mathbf{U}_{\text{chan}}(f,:)\hat{\mathbf{u}}_{\text{chan}}$ suggest that the sign of $\hat{\mathbf{u}}_{\text{chan}}$ should be flipped. We use the majority vote of these three criteria to determine the sign of the coefficient pair. The scaling ambiguity is resolved by scaling $\hat{\mathbf{u}}_{\text{trial}}$ and all rows of $\mathbf{U}_{\text{trial}}$ to unit norm.

A.3 Experimental results on a mental task

A.3.1 Public EEG dataset

We apply the method on a recently published dataset with EEG recordings of subjects that participate in a mental arithmetic task (details can be found in (Shin et al., 2016)). During the experiment, subjects complete sixty trials of ten seconds each, in half of which they are instructed to repeatedly subtract two numbers, and in half of which they rest. In this paper, we used the 30-channel EEG recordings of 14 subjects from this dataset, which are sampled at 200 Hz.

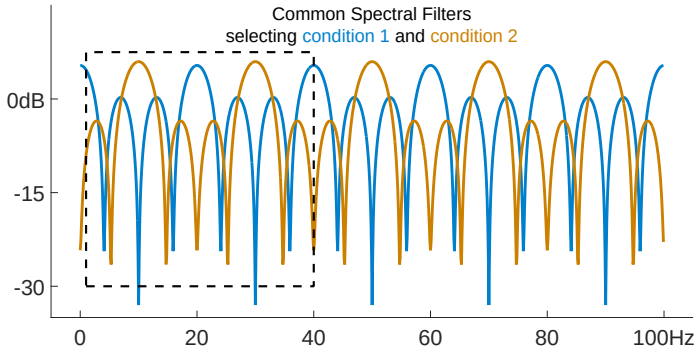


Figure A.1: During training, common spectral pattern (CSP) filters are tuned to frequency bands in which the spectral power is discriminative for the class. For a particular subject, the filterwhich maximizes variance for condition 1 (the task) selects a band around 20 Hz. On the other hand, the filter that focuses on the other class of trials selects complementary bands.

A.3.2 Preprocessing and analysis

The EEG data of all subjects were band-pass filtered between 0.5-40 Hz, and eyeblink artifacts were estimated and removed using independent component analysis (ICA), as implemented in the SOBI algorithm from EEGLAB (Delorme and Makeig, 2004). The traditional, full-channel common spatial patterns analysis was first ran on every subject's recording, to identify the channels in which the mental task manifested itself the strongest. This information is conveyed by the rows of the matrix $\mathbf{W}^{-1}$ (Blankertz et al., 2007). We selected $C = 10$ channels per subject with the highest root-mean-square value (averaged over all rows of $\mathbf{W}^{-1}$). The parameters for the subsequent time delay embedding step were chosen as $\tau = 10$ samples and $L = 6$ after tuning on data from a subject whose results are left out of the current analysis.

We estimated L common spectral pattern filters by means of (A.3) and applied them to the data as in (A.4). Finally, the logarithm of the signal power was computed in windows of 1 second, leading to a feature tensor $\tilde{\boldsymbol{\mathcal{Y}}} \in \mathbb{R}^{10\times10\times6\times60}$. The set of equations in (A.7) were solved using a state-of-the-art Gauss-Newton algorithm in Tensorlab (Vervliet et al., 2016). In (Boussé et al., 2018), it was derived that under mild conditions, the number of equations should be as least as high as the number of unknown coefficients to ensure that a unique solution exists, i.e. $TL \geqslant \dim(\mathbf{U}_{\text{trial}}) + \dim(\mathbf{U}_{\text{chan}})$. Hence, the basis matrices $\mathbf{U}_{\text{chan}}$ and $\mathbf{U}_{\text{trial}}$ were truncated to dimensions 3 and 5, respectively, upon inspection of the multilinear singular values of $\tilde{\boldsymbol{\mathcal{Y}}}$. Least squares SVMs with linear or

RBF kernel were trained on the rows of the truncated matrix $\mathbf{U}_{\text{trial}}$ using LS-SVMlab (Suykens et al., 2002).

A.3.3 Results and discussion

Trials were classified using the seven classifiers described in section A.2.3 in a 15-fold cross-validation setting that was repeated ten times per subject. Note that the CSP filters use label info and are trained as part of the cross-validation procedure as well (Blankertz et al., 2007). In fig. A.1 we inspect (for one subject) the transfer functions of those filters belonging to the most extreme generalized eigenvalues. Because of the time delay embedding step, the columns of $\mathbf{W}$ each have L coefficients that map one-to-one to FIR filter coefficients that are spaced by τ samples. E.g. a column $\mathbf{w}_i$ gives rise to an impulse response $\mathbf{h}_i = \mathbf{w}_i \otimes [\, 1\ 0\ \ldots\ 0\,]^{\mathrm{T}}$ with L degrees of freedom, where the second vector contains $\tau - 1$ trailing zeros. From the blue trace, we deduce that the mental arithmetic task provokes neural activity which differs most saliently from the 'rest' condition in a band between approximately 17-23 Hz. On the other hand, the orange filter is tuned towards the 'rest' trials and selects complementary frequency bands. The classification accuracy varies between subjects and channels used during testing. This is shown in fig. A.2 for the linear kernel SVM, in which the accuracies of different repetitions are shown, in distinct colors per subject, over channels. Note that the 10 selected channels can be different for every subject, and are grouped based on their ranking, although in our analysis there were some electrodes in common for the majority of subjects, located over the left temporal lobe, and the frontal lobe near the centerline. The mean accuracy over repetitions of the best-performing subject was 84% and 70% at the best and worst channel of the set, respectively. For the worst-performing subject, these metrics were 50% and 40%. Across all subjects, the single-channel classification method presented here had a performance which was only $6.6 \pm 17\%$ lower than the performance of the common spatial pattern classification using all thirty channels, conducted as in (Shin et al., 2016); for some subjects the single-channel method performed even remarkably better. We conclude that for well-chosen subjects, our pipeline is robust and may be used for single-channel classification, with an acceptable loss in performance over a full-channel BCI.

A comparison between the seven evaluated classifiers is presented in fig. A.3 for the best channel of every subject. We managed to boost the classification accuracy by a six percent, through the use of linear kernel SVMs instead of the simple nearest neighbor approach, which was used in (Boussé et al., 2018, 2017). Although KNN classifiers, RBF-SVMs are able to model nonlinear decision boundaries, they are more sensitive to outliers – which are almost

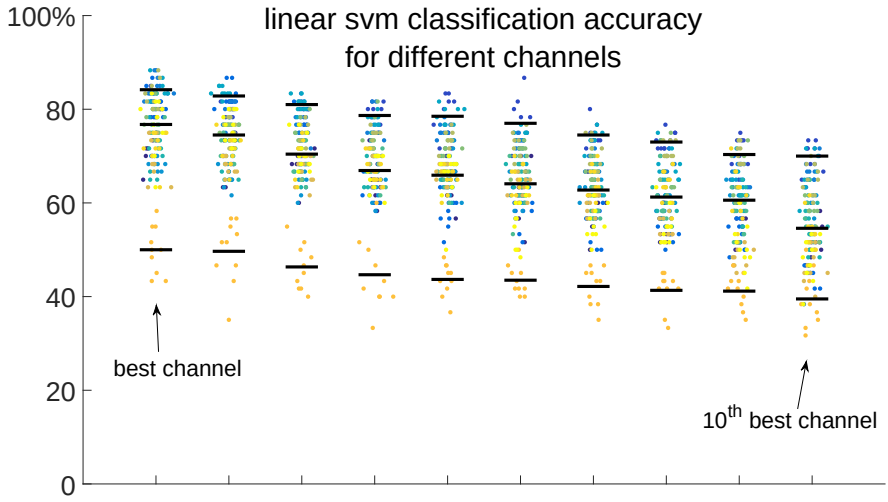


Figure A.2: Good single-channel classification accuracy can be attained by the tensor subspace learning and regression method. The performance depends on the selected channel, and differs between subjects (coded by color). This is indicated by black lines at the mean accuracy (over all repetitions) of subjects with the highest, median, and lowest performance.

invariably present in these EEG datasets – than linear classifiers. The random forest yielded intermediate performance, whereas the tree boosting method was unreliable, potentially also due to outlier sensitivity of AdaBoost. Since the best sensitivity and specificity are attained at different channels, improved classification is possible by relying on two or more channels, as in (Boussé et al., 2017).

We observed that testing accuracy was in general several percent lower than validation accuracy (not shown). This drop is due to the feature computation process: for the training (and validation) trials, features can be readily extracted from the rows of $\mathbf{U}_{\text{trial}}$, whereas for the test trials, an intermediate step (tensor regression to obtain the coefficients of the test trials in the precomputed subspace) is needed. Due to this extra step, the features of test trials are prone to estimation error, which explains a lower performance.

Here, we used the same L for all subjects, although this parameter could be optimized individually. We found that the performance is relatively stable when choosing closeby values of L, e.g. 5 or 7, but that the variability of the performance increases for higher L. This effect is due to two factors: for higher L, the regression problem in (A.7) has more observations and is hence

'more overdetermined', leading to a more robust estimation. However, the accompanying risk is that the spectral filters $\mathbf{W}$ start to overfit the training data, which undermines the performance for some subjects' data.

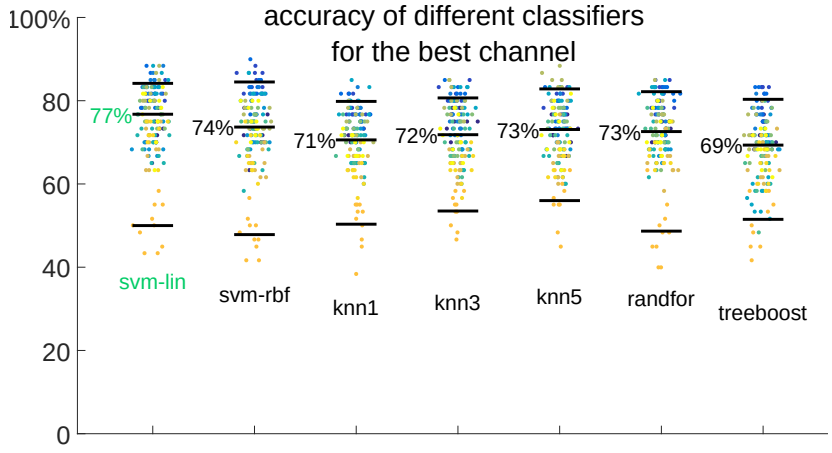


Figure A.3: Support vector machine (SVM) classifiers, especially the linear kernel SVM, are more robust than the K-nearest neighbor (KNN) classifiers. Tree boosting yielded the least reliable classification in the current analysis, followed by the simple `knn1` classifier, used in (Boussé et al., 2018, 2017). Colors and black levels have the same interpretation as in fig. A.2.

A.4 Conclusion

We developed a pipeline for classifying brain states based on data from a single EEG channel, after a calibration phase in which information of multiple channels is exploited. This key aspect renders it suitable for use in wearable devices, where the number of electrodes is limited. The pipeline relies on 1) a supervised subspace identification step, in which spectral filters are found that extract frequency bands that are discriminative for the classification task, 2) solving a tensor regression problem with a low-rank structure and 3) the application of a classifier on the estimated regression coefficients. Although we demonstrated its success for the classification of trials of a mental task in a BCI context, the approach is generic; it may be applicable for the detection of e.g. epileptic seizures as well, as these also induce spectral changes, or also for non-EEG data. The proposed framework can still be improved, by using a few channels instead of only one (Boussé et al., 2017), and by rigorously tuning the parameters in the

pipeline (e.g. L, τ, C, $\dim(\mathbf{U}_{\text{chan}})$, $\dim(\mathbf{U}_{\text{trial}})$) with cross-validation. Based on the 'equations versus unknowns' tradeoff, a high L and T are warranted, although this incurs the risk of overfitting (as discussed in previous section), or increases the latency of the pipeline, respectively. Alternatively, $\dim(\mathbf{U}_{\text{chan}})$ and $\dim(\mathbf{U}_{\text{trial}})$ should be low, though this might compromise the regression's fit, and as $\dim(\mathbf{U}_{\text{trial}})$ equals the number of features, it limits the classifiers' learning capabilities. During training, dictionary learning methods could further aid to find suitable subspaces that are robust to outliers (Loza and Principe, 2016).

Appendix B

Supplement to: Augmenting interictal mapping with neurovascular coupling biomarkers by structured data fusion

This appendix provides the complete results, i.e., for all patients, of the novel sCMTF pipeline, which we presented in Chapter 7. For details on how these results are to be interpreted, we refer to Section 7.5.

sCMTF signatures for all patients

We show the EEG spatial, temporal and spectral signatures, HRF waveforms, and fMRI spatial maps of significant IED (de)activation and HRF variability for all patients (except patients 3 and 10, whose results have been analyzed in Chapter 7).

Patient 1

We analyze the solution with $\hat{R} = 6$ sources. Figure B.1 shows the EEG signatures and HRF waveforms. One of the sources is highly correlated to the MWF reference (in grey), which was already known from Table B.3. This

IED-related source had a typical low-frequency spectrum, which is expected for the typical spike-and-wave interictal discharges. The topography is relatively diffuse, although the highest amplitudes are mostly in the left hemisphere. This is in accordance with the lateralization of ictal onset zone (left temporal lobe, cfr. Table 1). There are some noteworthy observations to be made about some of the other components. The fourth has an unusually sharp spectrum, is mainly localized on two nonadjacent center electrodes, and is sustained for a single period of many seconds. Hence, this component likely captured an artifact (of yet unknown origin), although we spotted no large-amplitude changes in the EEG itself. Similarly, the third source is only present at one frontal electrode, and exists in a frequency range above 20 Hz. It might represent a muscle artifact, e.g., due to frowning or twitching of some muscles in the forehead. The HRFs of all ROIs are shown in Figure B.1b. Two of the basis functions seem to have converged to a very similar waveform, which is an unfortunate possibility if two initial HRFs are too close to the same local optimum in their respective parameters. This reduces the expressive power of the basis set, which is clearly visible, since many ROIs have a nearly identical HRF. One of the twenty ROIs with the highest-entropy HRF overlapped the IOZ, although clearly this HRF (bold line) is not among the most dissimilar waveforms for this patient. This is also visible in Figure B.2: both the HRF entropy and extremity maps show a small overlap with the delineated IOZ. Despite the good correspondence in the EEG domain, no significant (de)activation of the IED-component is found inside the IOZ.

Patient 2

We analyze the solution with $\hat{R} = 3$ sources, and show the results in Figure B.3 and B.4. As for patient 1, we found a source which is strongly correlated to the MWF envelope, and which had a mostly low-frequency behavior characteristic for spikes. The topography is mostly uninformative, and does not clearly correspond to the patient's clinical data. The third source is mostly present at both sides of the head, is very sparsely active in time, and has a high-frequency content: this is most likely an artifact due to the neck muscles. Again, there is one of the highest-entropy HRFs which belongs to a ROI in the IOZ. Now, the waveform is clearly resolved from the other HRFs, through the strong initial dip (before 0 seconds). Such a dip is sometimes observed in HRFs, but its underlying physiological mechanism is not yet fully understood. It is possible that this dip reflects altered vascular autoregulation near the IOZ (cfr. the explanation in the Section 1 of the main text), or a rapid depletion in oxygen due to IED generation (before the IED becomes visible on the EEG). Figure B.4 furthermore shows that the IED-related component is significantly active in

parts of the IOZ, and deactive in others. As mentioned earlier, this deactivation may or may not be due to errors in sign correction. Interestingly, the ROI with the high alteration in neurovascular coupling is distinct from both the activated and deactivated ROIs.

Patient 4

We analyzed the solution with $\hat{R} = 4$ sources, and show the results in Figure B.5 and B.6. There is one source which is mostly correlated to the reference (but not extremely, see also Table B.3). This source had a right-temporal focus, conform the diagnosis in Table 1. The second source illustrates the phenomenon of an erroneous sign exchange between the spatial and spectral profiles. Also one of the HRFs has a negative polarity, which is a failure of the sign correction procedure (in this case, because there is exceptionally no positive overshoot). However, the HRF variability metrics are still interpretable, and indeed two ROIs among the ones with the highest-entropy HRFs overlap with the IOZ. The IED component is significantly active in a tiny portion of the IOZ (cfr. Figure B.6). The second source is significantly active in symmetrical parts of the parietal lobe. Given its ongoing fluctuation over time, we hypothesize that this source captures a resting-state network (RSN).

Patient 5

We analyze the solution with $\hat{R} = 5$ sources, and show the results in Figure B.7 and B.8. There is a clear IED-related component, with a very high correlation to the MWF reference, a typical spectrum, and an anterior-temporal focus, which corresponds very well to the patient's diagnosis (cfr. Table 1). The fifth source seems present at only one channel, and has spectral harmonic at ± 17 Hz and ± 34 Hz. One of these peaks is reminiscent of the fourth component in patient 1. As Figure B.8 shows, the HRF entropy and extremity prove to be strong biomarkers for the IOZ in this case, and also the significant IED activation and deactivation allow correct localization. In Figure B.7, it is clear that some HRFs may still have the wrong sign, which means that the interpretation of 'active' and 'deactivated' is flipped in those ROIs. Hence, regions of significant deactivation are in fact significantly activated. The fourth source had a significant overlap with the auditory RSN, and its spectrum reveals activity in the β band.

Patient 6

We analyze the solution with $\hat{R} = 2$ sources, and show the results in Figure B.9 and B.10. One source is strongly correlated to the MWF, while the other source is likely an artifact, given its very sparse temporal profile. Both sources coincide at one high-amplitude peak, by which we infer that this is probably an artifactual period in the signal. Indeed, when inspecting the original EEG signals, we found high-frequency muscle artifacts at these times. This source also had no significant activation in its spatial map, which corroborates its non-neuronal origin. The IED-related source had a broader spectrum than most other cases, and an uninformative topography. None of the ROIs with high-entropy HRFs is located in the IOZ. The pseudo t-map provides correct localization of the IOZ, however.

Patient 7

We analyze the solution with $\hat{R} = 4$ sources, and show the results in Figure B.11 and B.12. We found a clear IED-related component, with a characteristic spectrum and a topography which is backed up by the patient's diagnosis (left anterior-temporal IOZ). The fourth source has a very similar topography and spectrum to the fifth source in patient 5. One HRF inside the IOZ had a high-entropy, and is distinguishable from the others by its very sluggish waveform, i.e., it is smeared out in time, with no sharp over- or undershoot. Also the pseudo t-map provided an accurate localization of the IOZ. Notably, in this patient, the extremity metric misses the deviating HRF in the IOZ (while the entropy metric picks it up). The second source overlapped with the frontal part of the default mode network (DMN), and is active in the α and low β bands.

Patient 8

We analyze the solution with $\hat{R} = 2$ sources, and show the results in Figure B.13 and B.14. We found two components which had correlated time courses. At the time of the peaks, we found higher-amplitude events in the EEG with dubious origin, hence they may or may not be artifacts. One of both components is more strongly correlated to the MWF, and its activation is concordant with the IOZ. The second component shows high overlap with the sensorimotor network. For this patient, none of the IOZ's ROIs had extreme values of either HRF metric.

Patient 9

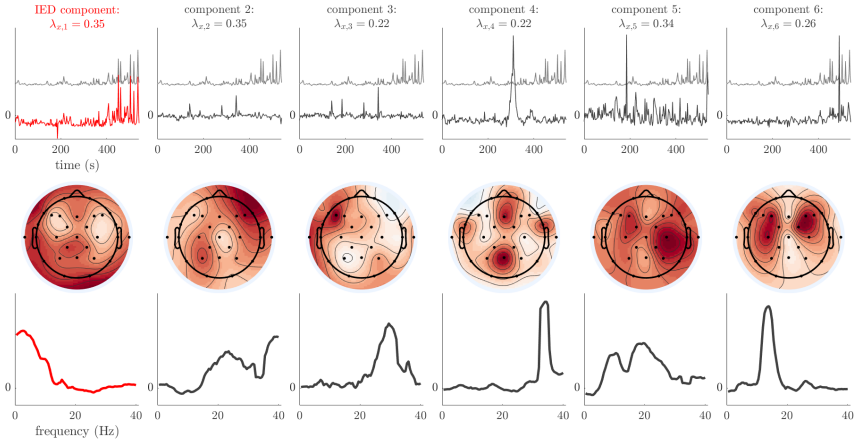
We analyze the solution with $\hat{R} = 2$ sources, and show the results in Figure B.15 and B.16. In this patient, there is only a moderate correlation of a component with the MWF reference time course. This component's topography (left occipital) agrees with the clinical description, however. The HRF extremity (and not the entropy) is high in a small part of the IOZ. Both the significant IED activation and deactivation allow correct localization as well. The second source seemingly captured high-frequency oscillatory activity in the sensorimotor network, similar to the previous patient.

Patient 11

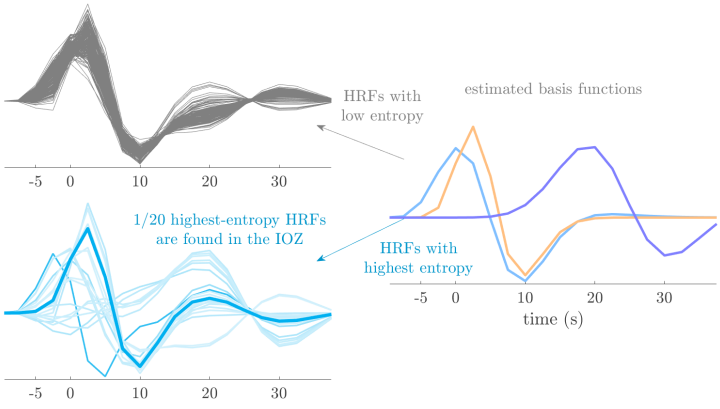
We analyze the solution with $\hat{R} = 2$ sources, and show the results in Figure B.17 and B.18. The IED-related source had a high correlation with the MWF reference, but an odd bimodal spectrum. Its EEG topography is very consistent with the clinical description. Both HRF extremity and entropy are useful biomarkers for the IOZ. The IED activation and deactivation maps each had a very small overlap with the IOZ. The second source is temporally sparse and captures high-frequency EEG variations, which we identified as muscle artifacts.

Patient 12

We analyze the solution with $\hat{R} = 2$ sources, and show the results in Figure B.19 and B.20. Again we observe an IED-related source and a seemingly artifactual source with a spectral peak near 34 Hz. Many of the high-entropy HRFs are highly noncausal, and are associated to ROIs inside the IOZ. Hence, with both HRF metrics, the highest-scoring ROIs provides good localization of the HRF. While there are no significantly active ROIs in the IOZ, there are several significantly deactivated ROIs, which may indicate that the sign standardization was not done flawlessly (cfr. also some of the negative-peaking HRFs for patient 10). Surprisingly, the second source had one significantly active ROI, which overlaps with the IOZ, but which did not match its EEG topography. Hence, the nature of this source remains ambiguous.



(a) Temporal ($\mathbf{s}_r$, top), spatial ($\mathbf{m}_r$, middle), and spectral ($\mathbf{g}_r$, bottom) profiles of the 6 sources in the EEG domain, and reference IED time course ($\mathbf{s}_{\text{ref}}$, in grey).



(b) Estimated HRFs with low and high entropy

Figure B.1: Patient 1's estimated sources and neurovascular coupling parameters. **(a)** The IED-related component correlates well with the reference time course, and is mostly a low-frequency phenomenon. The fourth source is most likely an artifact (of unknown origin), picked up mostly by two central channels. **(b)** One of the ROIs with the highest-entropy HRFs belongs to the ictal onset zone (bold line, $p = 0.34$).

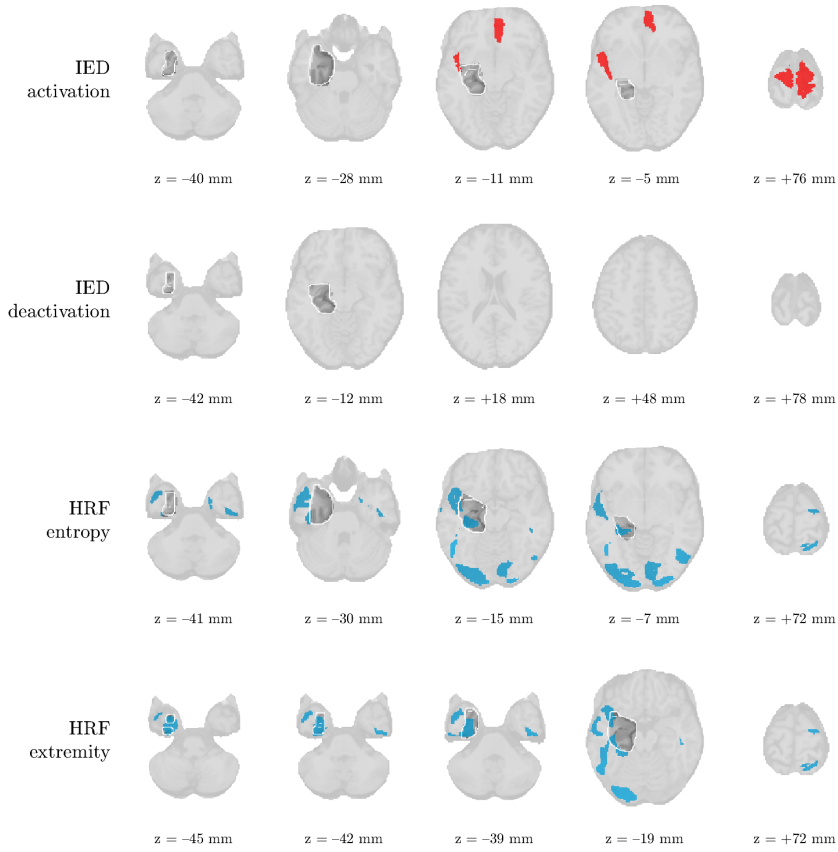
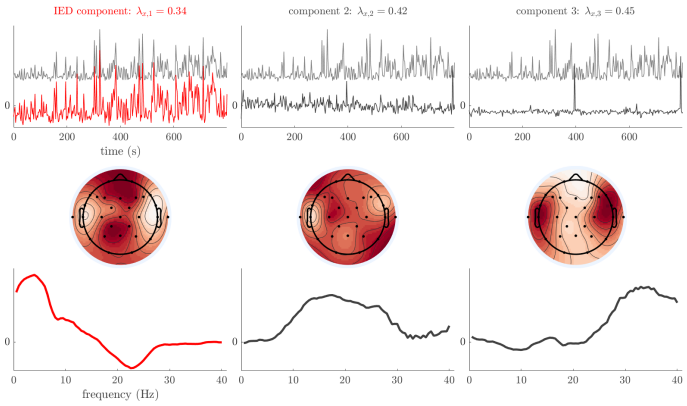
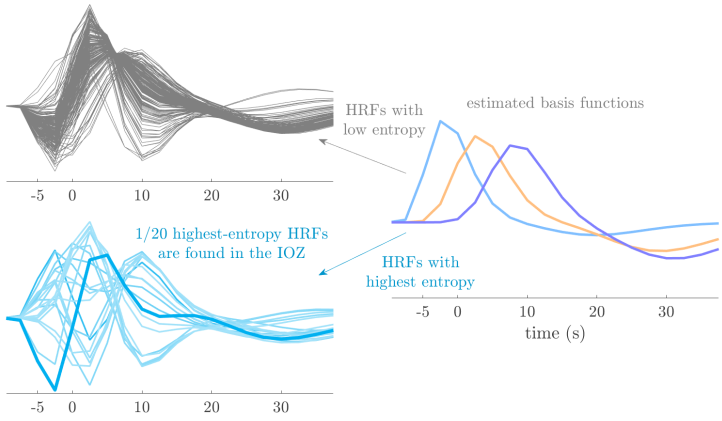


Figure B.2: Patient 1's statistical nonparametric maps and HRF entropy/extremity maps. The ground truth ictal onset zone is highlighted in dark gray with a white contour.



(a) Temporal ($\mathbf{s}_r$, top), spatial ($\mathbf{m}_r$, middle), and spectral ($\mathbf{g}_r$, bottom) profiles of the 3 sources in the EEG domain, and reference IED time course ($\mathbf{s}_{\text{ref}}$, in grey).



(b) Estimated HRFs with low and high entropy

Figure B.3: Patient 2's estimated sources and neurovascular coupling parameters. **(a)** The IED-related component correlates well with the reference time course, and is mostly a low-frequency phenomenon. **(b)** One of the ROIs with the highest-entropy HRFs belongs to the ictal onset zone (bold line, $p = 0.59$).

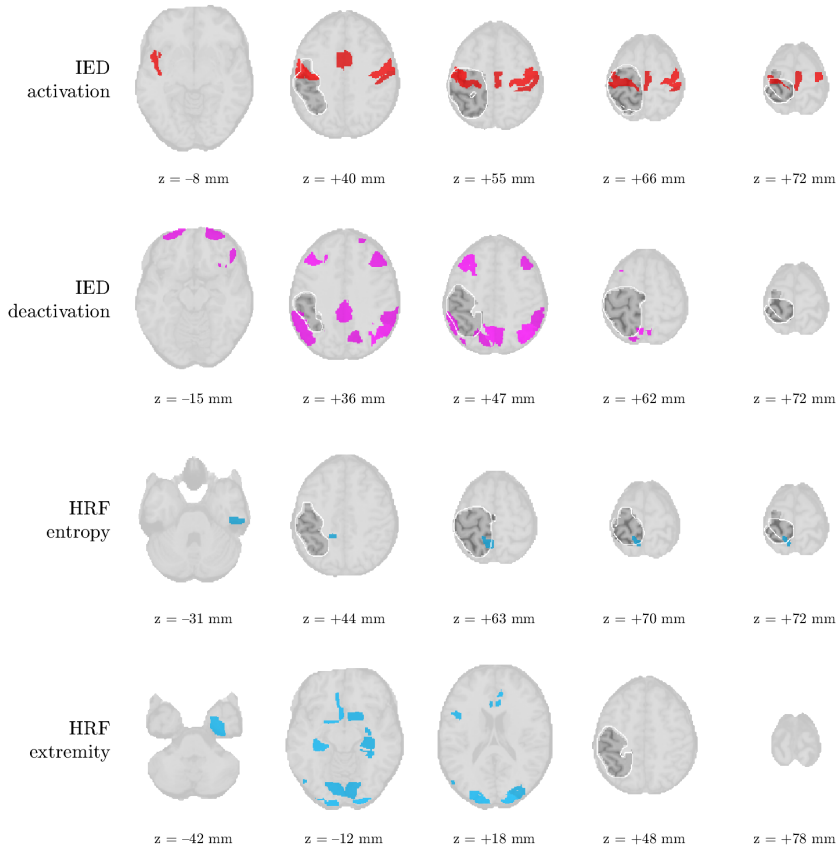
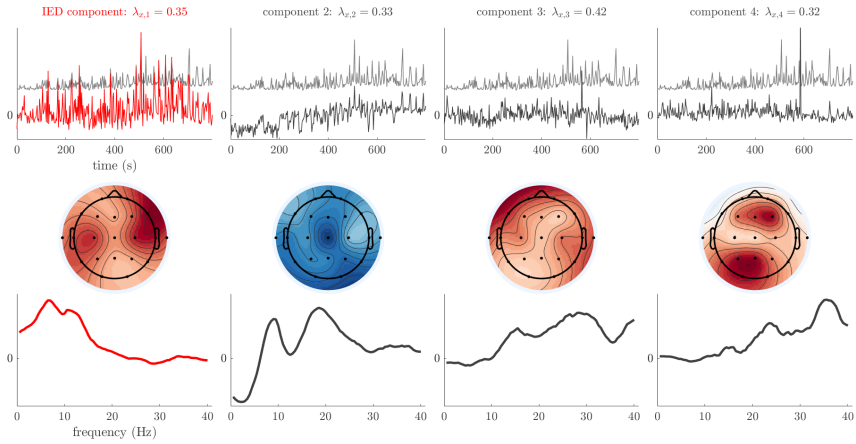
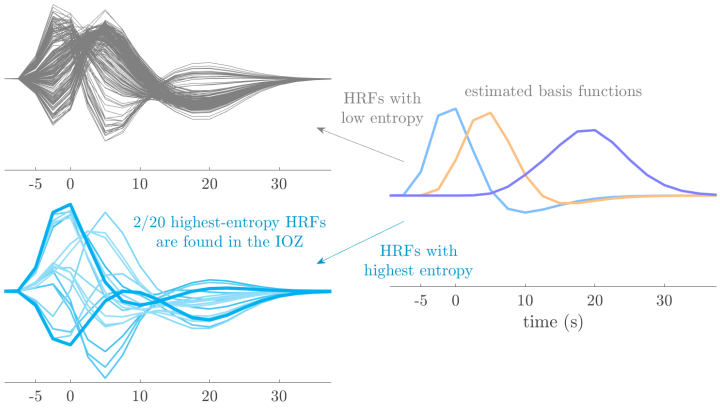


Figure B.4: Patient 2's statistical nonparametric maps and HRF entropy/extremity maps. The ground truth ictal onset zone is highlighted in dark gray with a white contour.



(a) Temporal ($\mathbf{s}_r$, top), spatial ($\mathbf{m}_r$, middle), and spectral ($\mathbf{g}_r$, bottom) profiles of the 4 sources in the EEG domain, and reference IED time course ($\mathbf{s}_{\text{ref}}$, in grey).



(b) Estimated HRFs with low and high entropy

Figure B.5: Patient 4's estimated sources and neurovascular coupling parameters. **(b)** Two of the ROIs with the highest-entropy HRFs belong to the ictal onset zone (bold line, $p = 0.32$).

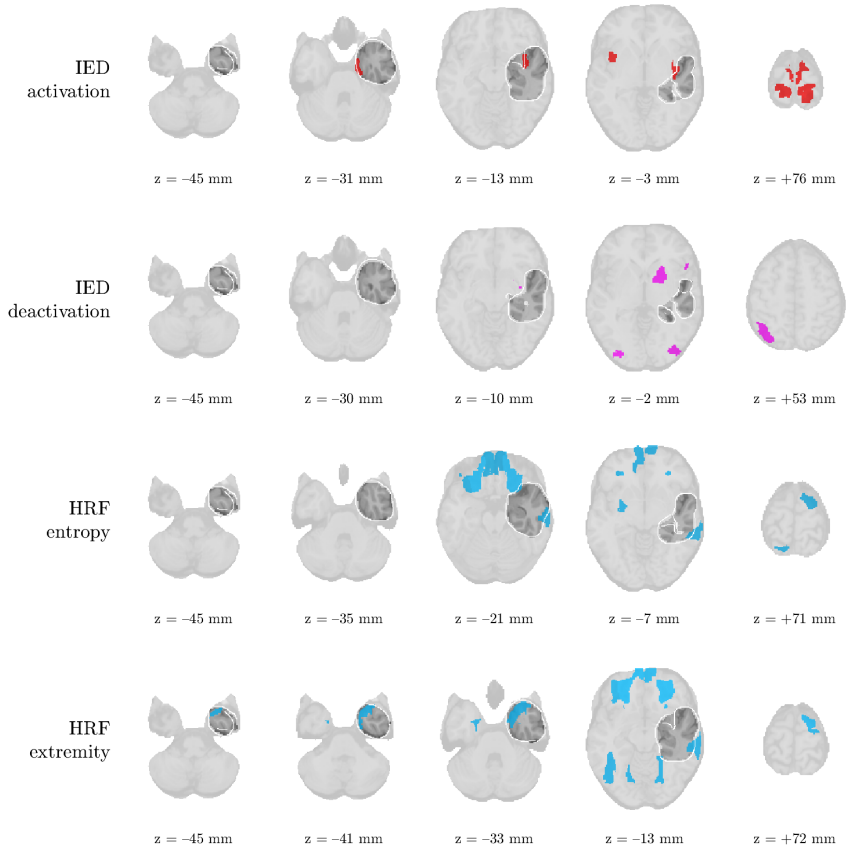
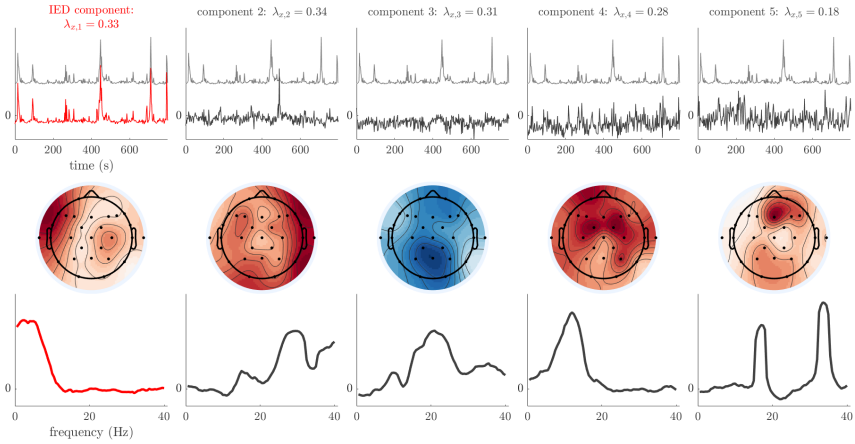
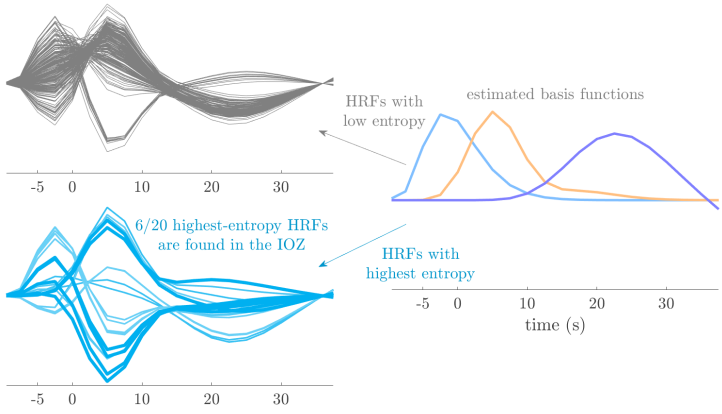


Figure B.6: Patient 4's statistical nonparametric maps and HRF entropy/extremity maps. The ground truth ictal onset zone is highlighted in dark gray with a white contour.



(a) Temporal ($\mathbf{s}_r$, top), spatial ($\mathbf{m}_r$, middle), and spectral ($\mathbf{g}_r$, bottom) profiles of the 5 sources in the EEG domain, and reference IED time course ($\mathbf{s}_{\text{ref}}$, in grey).



(b) Estimated HRFs with low and high entropy

Figure B.7: Patient 5's estimated sources and neurovascular coupling parameters. **(b)** Six of the ROIs with the highest-entropy HRFs belong to the ictal onset zone (bold line, $p < 10^{-3}$).

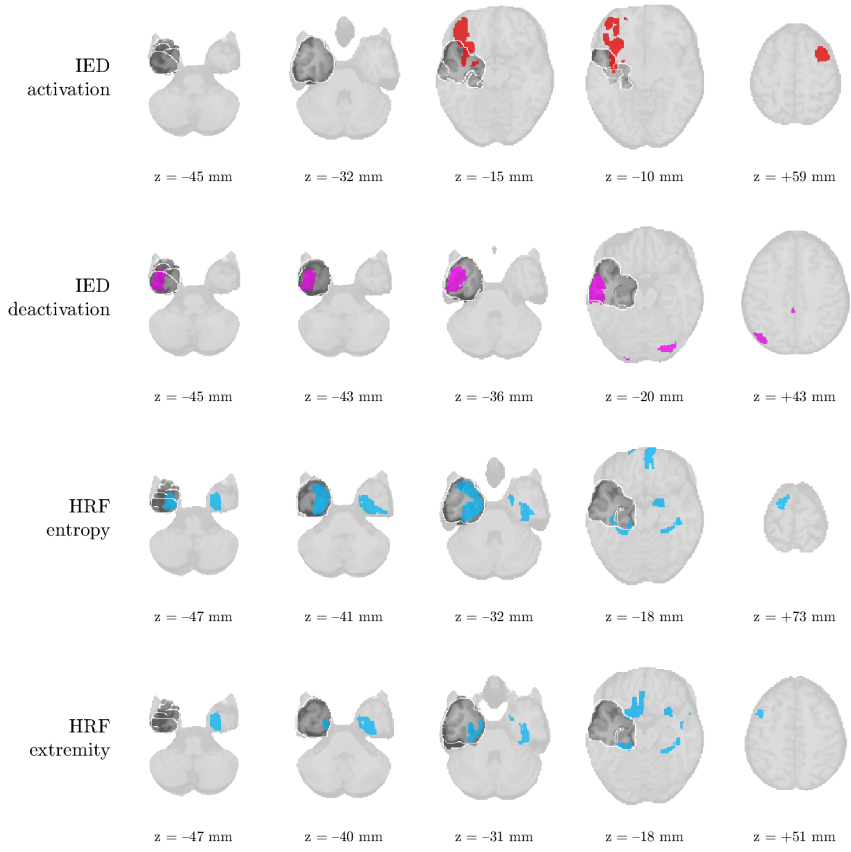
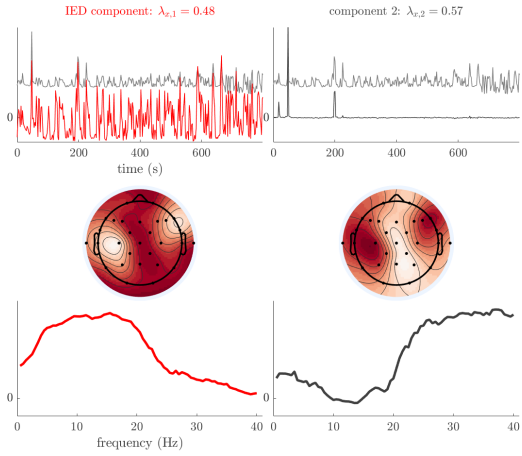
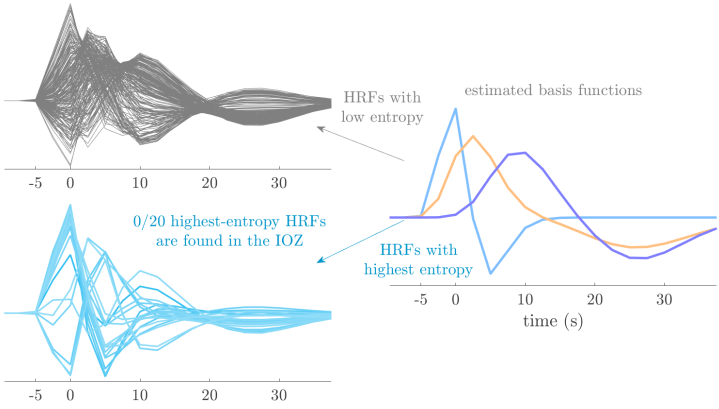


Figure B.8: Patient 5's statistical nonparametric maps and HRF entropy/extremity maps. The ground truth ictal onset zone is highlighted in dark gray with a white contour.



(a) Temporal ($\mathbf{s}_r$, top), spatial ($\mathbf{m}_r$, middle), and spectral ($\mathbf{g}_r$, bottom) profiles of the 2 sources in the EEG domain, and reference IED time course ($\mathbf{s}_{\text{ref}}$, in grey).



(b) Estimated HRFs with low and high entropy

Figure B.9: Patient 6's estimated sources and neurovascular coupling parameters. **(b)** None of the ROIs with the highest-entropy HRFs belong to the ictal onset zone.

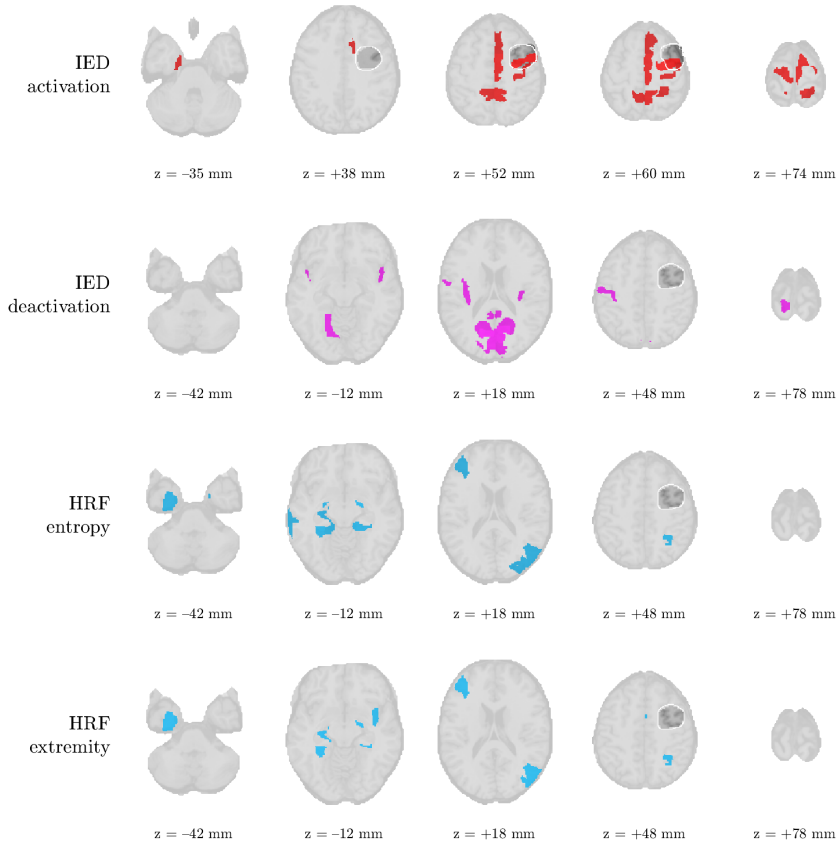
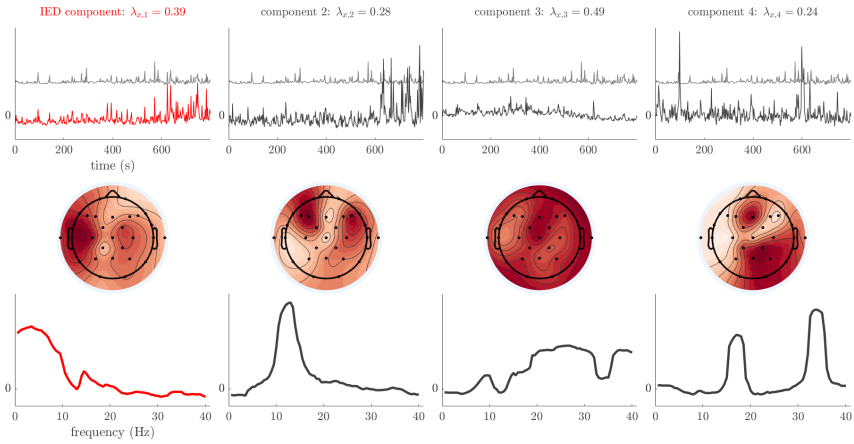
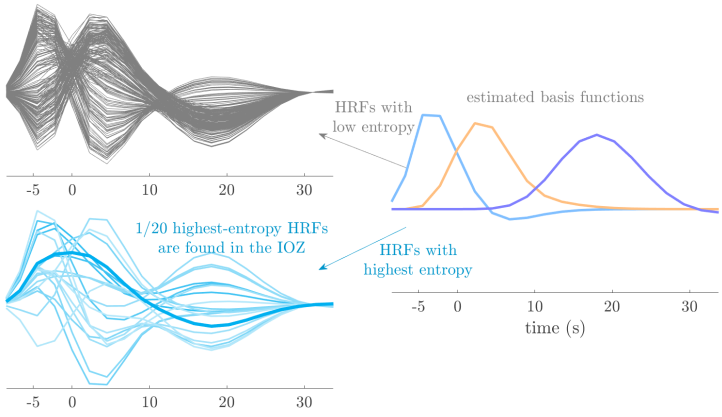


Figure B.10: Patient 6's statistical nonparametric maps and HRF entropy/extremity maps. The ground truth ictal onset zone is highlighted in dark gray with a white contour.



(a) Temporal ($\mathbf{s}_r$, top), spatial ($\mathbf{m}_r$, middle), and spectral ($\mathbf{g}_r$, bottom) profiles of the 4 sources in the EEG domain, and reference IED time course ($\mathbf{s}_{\text{ref}}$, in grey).



(b) Estimated HRFs with low and high entropy

Figure B.11: Patient 7's estimated sources and neurovascular coupling parameters. **(b)** One of the ROIs with the highest-entropy HRFs belongs to the ictal onset zone (bold line, $p = 0.57$).

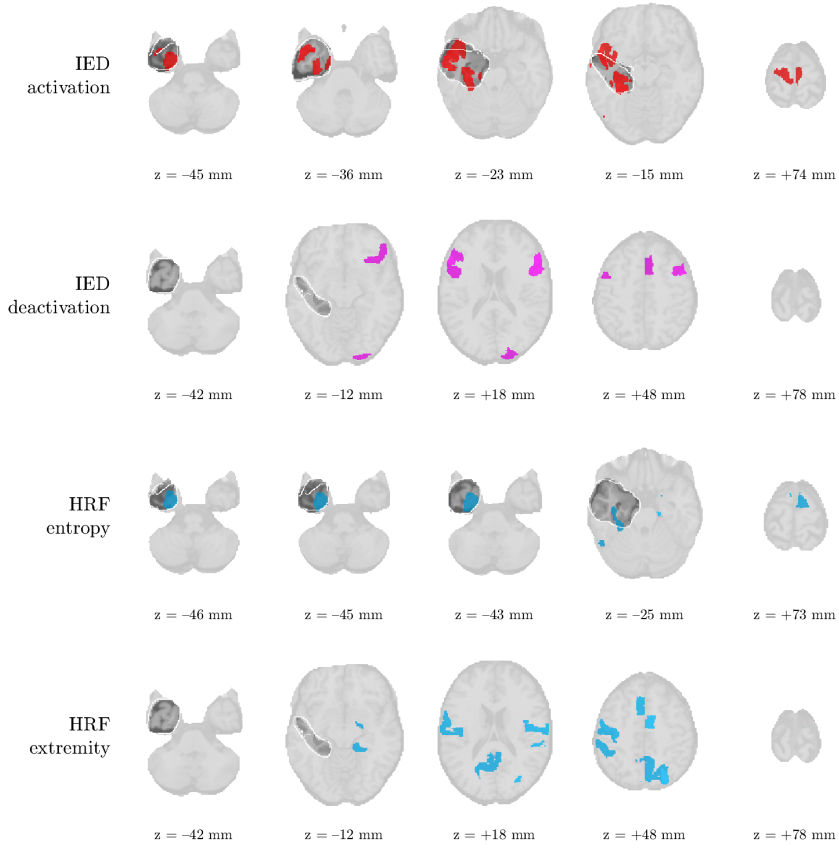
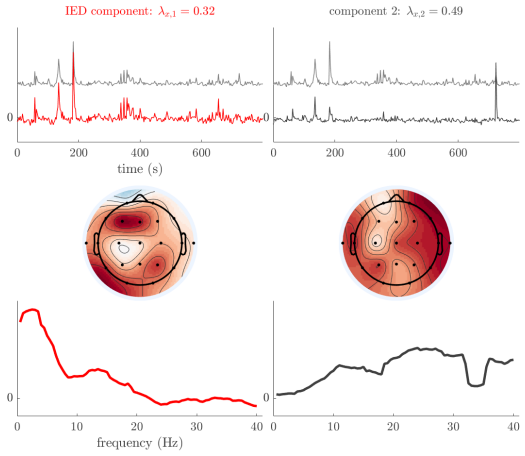
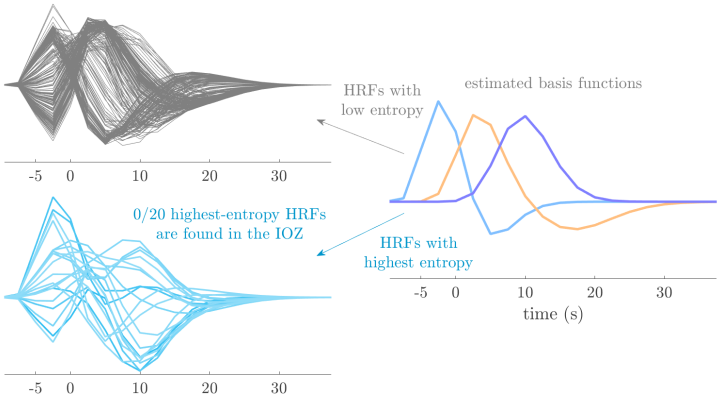


Figure B.12: Patient 7's statistical nonparametric maps and HRF entropy/extremity maps. The ground truth ictal onset zone is highlighted in dark gray with a white contour.



(a) Temporal ($\mathbf{s}_r$, top), spatial ($\mathbf{m}_r$, middle), and spectral ($\mathbf{g}_r$, bottom) profiles of the 2 sources in the EEG domain, and reference IED time course ($\mathbf{s}_{\text{ref}}$, in grey).



(b) Estimated HRFs with low and high entropy

Figure B.13: Patient 8's estimated sources and neurovascular coupling parameters. **(b)** None of the ROIs with the highest-entropy HRFs belong to the ictal onset zone.

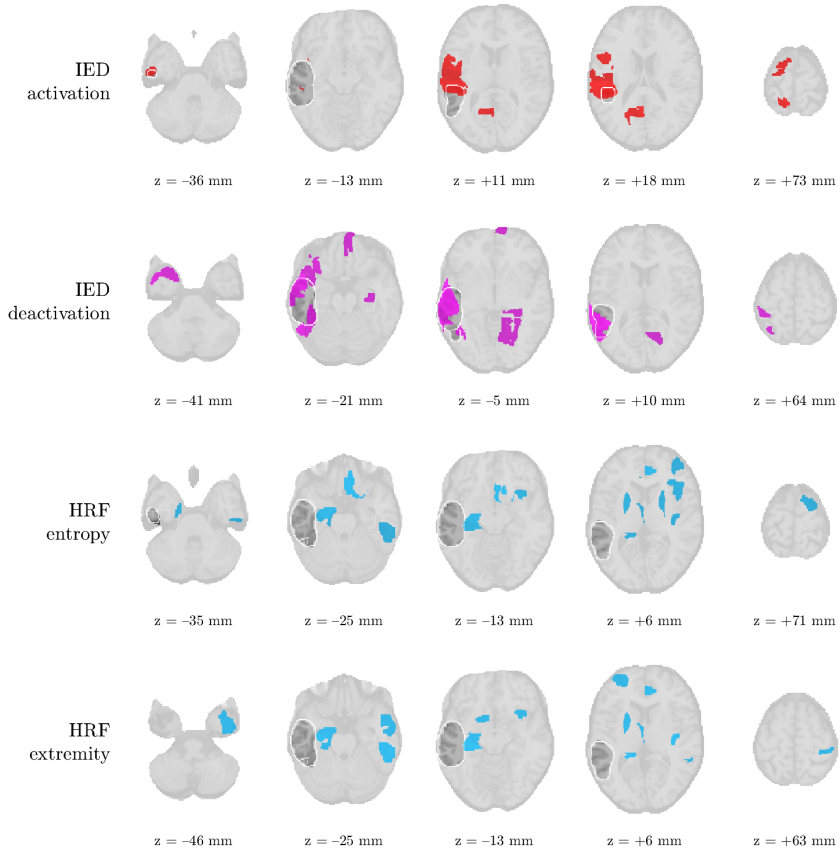
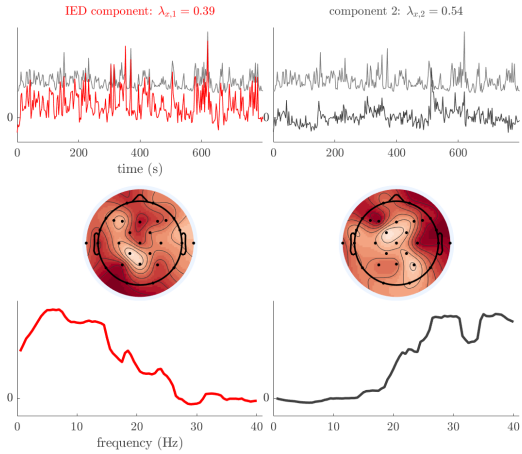
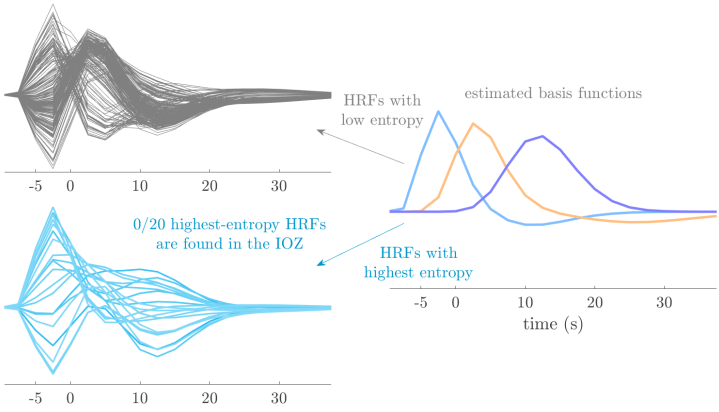


Figure B.14: Patient 8's statistical nonparametric maps and HRF entropy/extremity maps. The ground truth ictal onset zone is highlighted in dark gray with a white contour.



(a) Temporal ($\mathbf{s}_r$, top), spatial ($\mathbf{m}_r$, middle), and spectral ($\mathbf{g}_r$, bottom) profiles of the 2 sources in the EEG domain, and reference IED time course ($\mathbf{s}_{\text{ref}}$, in grey).



(b) Estimated HRFs with low and high entropy

Figure B.15: Patient 9's estimated sources and neurovascular coupling parameters. **(b)** None of the ROIs with the highest-entropy HRFs belong to the ictal onset zone.

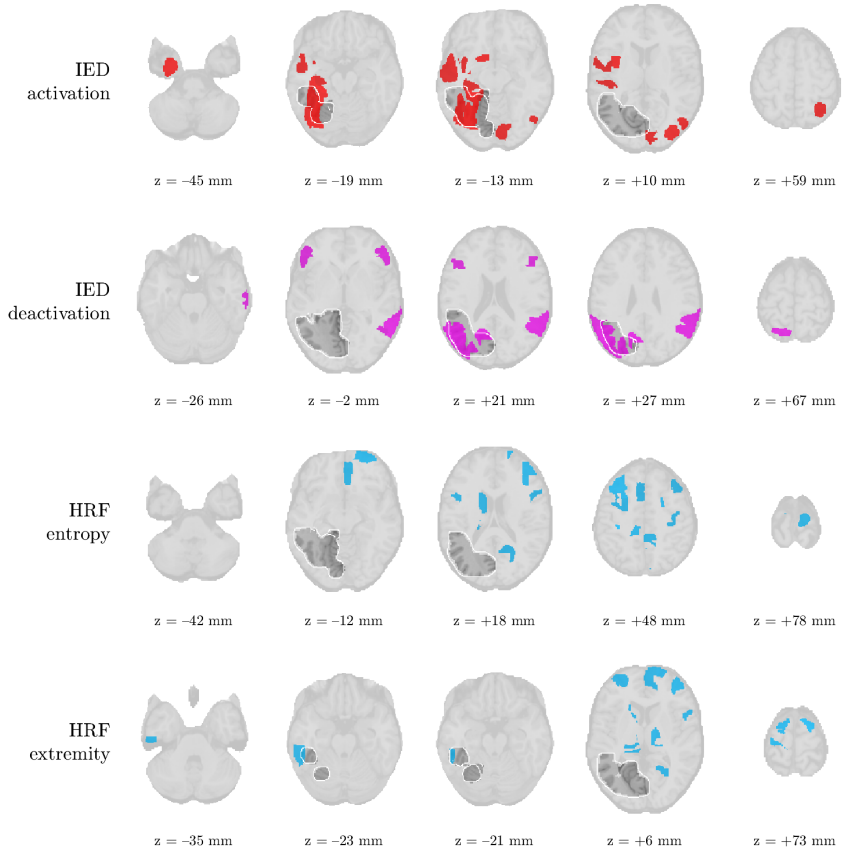
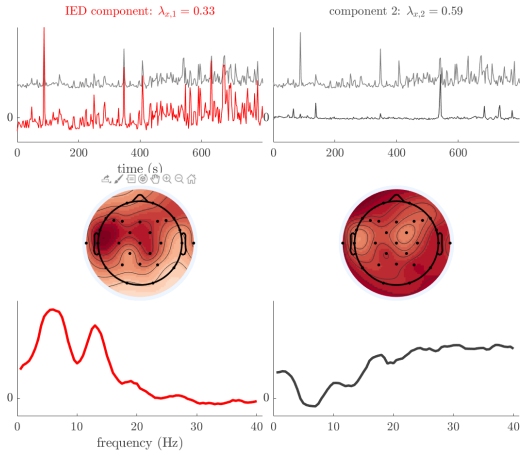
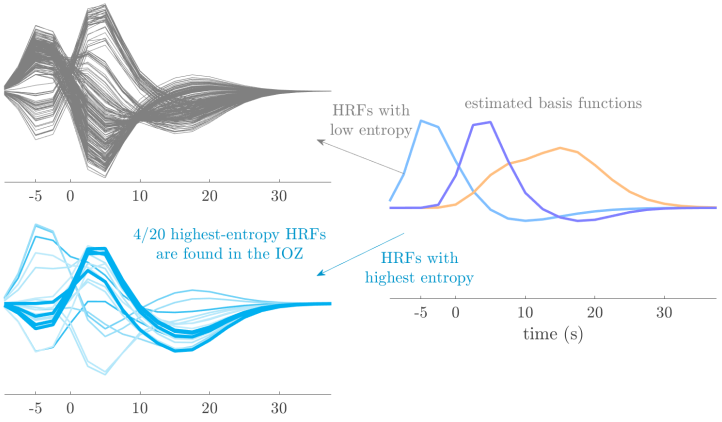


Figure B.16: Patient 9's statistical nonparametric maps and HRF entropy/extremity maps. The ground truth ictal onset zone is highlighted in dark gray with a white contour.



(a) Temporal ($\mathbf{s}_r$, top), spatial ($\mathbf{m}_r$, middle), and spectral ($\mathbf{g}_r$, bottom) profiles of the 2 sources in the EEG domain, and reference IED time course ($\mathbf{s}_{\text{ref}}$, in grey).



(b) Estimated HRFs with low and high entropy

Figure B.17: Patient 11's estimated sources and neurovascular coupling parameters. **(b)** Four of the ROIs with the highest-entropy HRFs belong to the ictal onset zone (bold line, $p = 0.01$).

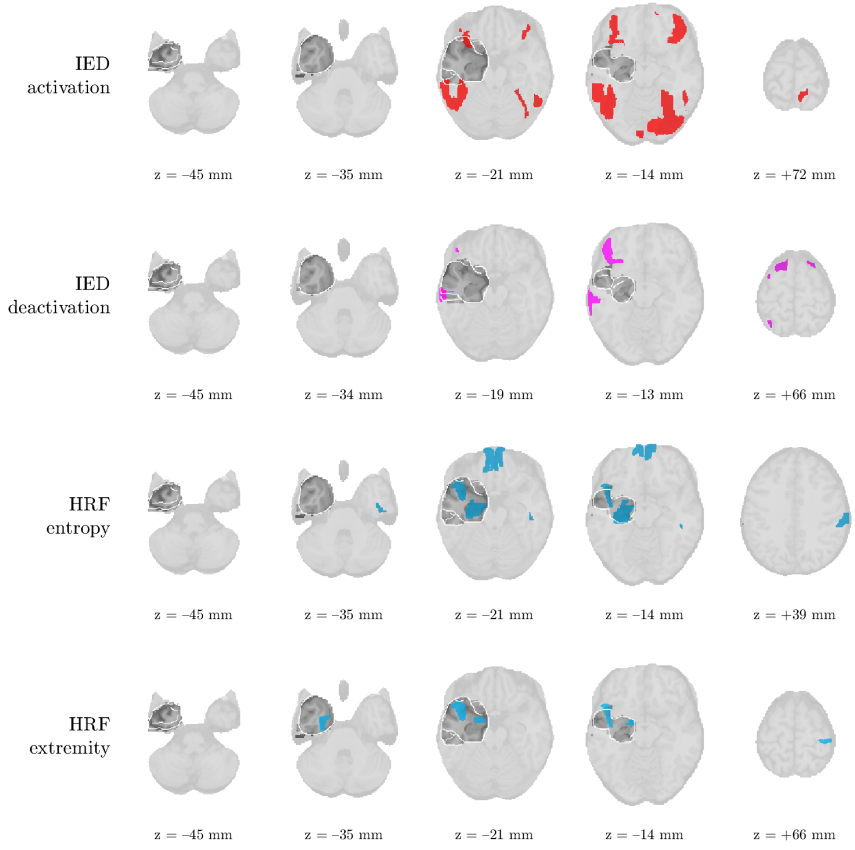
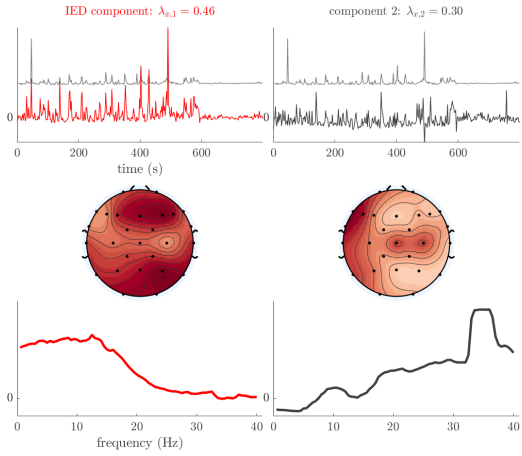
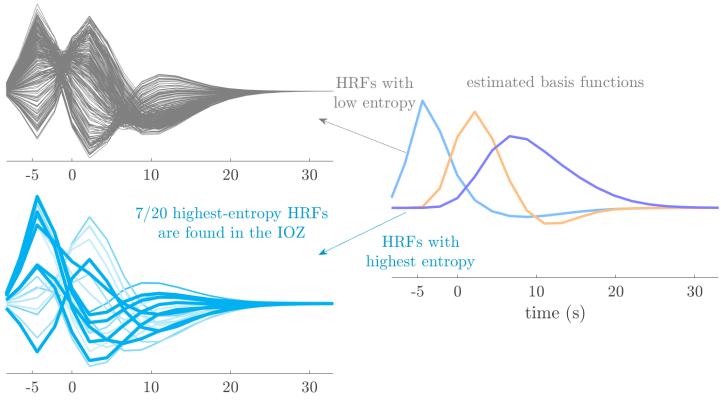


Figure B.18: Patient 11's statistical nonparametric maps and HRF entropy/extremity maps. The ground truth ictal onset zone is highlighted in dark gray with a white contour.



(a) Temporal ($\mathbf{s}_r$, top), spatial ($\mathbf{m}_r$, middle), and spectral ($\mathbf{g}_r$, bottom) profiles of the 3 sources in the EEG domain, and reference IED time course ($\mathbf{s}_{\text{ref}}$, in grey).



(b) Estimated HRFs with low and high entropy

Figure B.19: Patient 12's estimated sources and neurovascular coupling parameters. **(b)** Seven of the ROIs with the highest-entropy HRFs belong to the ictal onset zone (bold line, $p < 10^{-3}$).

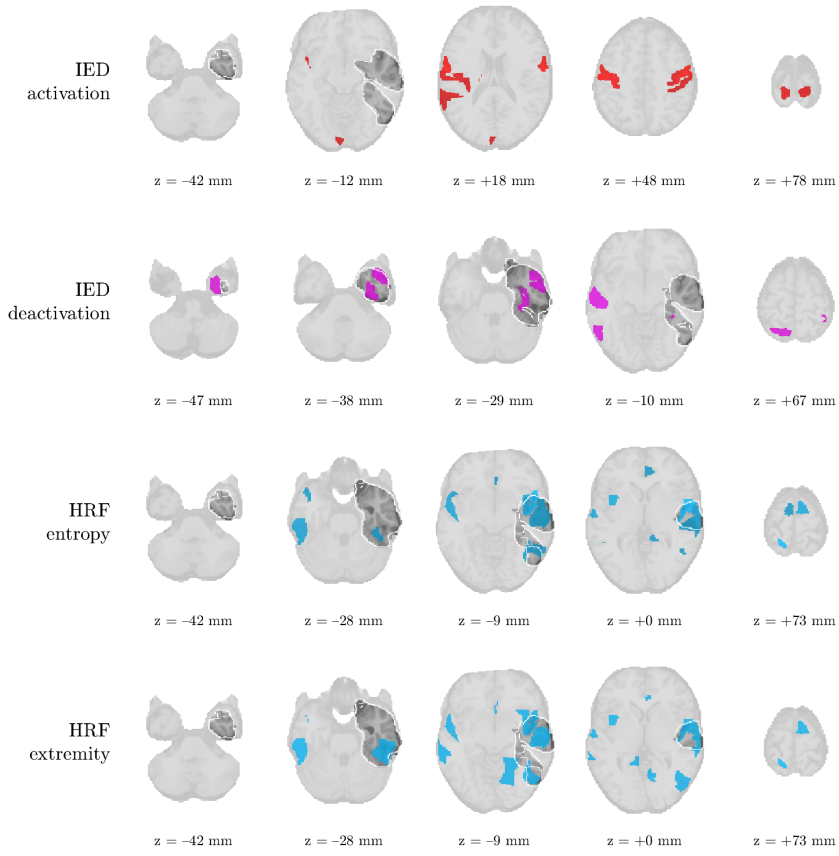


Figure B.20: Patient 12's statistical nonparametric maps and HRF entropy/extremity maps. The ground truth ictal onset zone is highlighted in dark gray with a white contour.

Bibliography

Abreu, R., Leal, A., da Silva, F. L., and Figueiredo, P. (2018a). EEG synchronization measures predict epilepsy-related BOLD-fMRI fluctuations better than commonly used univariate metrics. *Clinical Neurophysiology*, 129(3):618–635.

Abreu, R., Leal, A., and Figueiredo, P. (2018b). EEG-informed fMRI: a review of data analysis methods. *Frontiers in human neuroscience*, 12:29.

Abreu, R., Leite, M., Jorge, J., Grouiller, F., van der Zwaag, W., Leal, A., and Figueiredo, P. (2016a). Ballistocardiogram artifact correction taking into account physiological signal preservation in simultaneous EEG-fMRI. *Neuroimage*, 135:45–63.

Abreu, R., Leite, M., Leal, A., and Figueiredo, P. (2016b). Objective selection of epilepsy-related independent components from EEG data. *Journal of neuroscience methods*, 258:67–78.

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

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

Acar, E., Kolda, T. G., and Dunlavy, D. M. (2011b). All-at-once optimization for coupled matrix and tensor factorizations. In *KDD Workshop on Mining and Learning with Graphs, 2011*. KDD.

Acar, E., Levin-Schwartz, Y., Calhoun, V. D., and Adali, T. (2017a). ACMTF for fusion of multi-modal neuroimaging data and identification of biomarkers. In *2017 25th European Signal Processing Conference (EUSIPCO)*, pages 643–647. IEEE.

Acar, E., Levin-Schwartz, Y., Calhoun, V. D., and Adali, T. (2017b). Tensor-based fusion of EEG and FMRI to understand neurological changes in schizophrenia. In *2017 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1–4. IEEE.

Acar, E., Papalexakis, E. E., Gürdeniz, G., Rasmussen, M. A., Lawaetz, A. J., Nilsson, M., and Bro, R. (2014). Structure-revealing data fusion. *BMC bioinformatics*, 15(1):239.

Acar, E., Rasmussen, M. A., Savorani, F., Næs, T., and Bro, R. (2013). Understanding data fusion within the framework of coupled matrix and tensor factorizations. *Chemometrics and Intelligent Laboratory Systems*, 129:53–63.

Acar, E., Schenker, C., Levin-Schwartz, Y., Calhoun, V. D., and Adali, T. (2019). Unraveling diagnostic biomarkers of schizophrenia through structure-revealing fusion of multi-modal neuroimaging data. *Frontiers in neuroscience*, 13:416.

Acharya, U. R., Sree, S. V., et al. (2013). Automated EEG analysis of epilepsy: a review. *Knowledge-Based Systems*, 45:147–165.

Aguirre, G. K., Zarahn, E., and D'esposito, M. (1998). The variability of human, BOLD hemodynamic responses. *Neuroimage*, 8(4):360–369.

Allen, P. J., Josephs, O., and Turner, R. (2000). A method for removing imaging artifact from continuous EEG recorded during functional MRI. *Neuroimage*, 12(2):230–239.

Allen, P. J., Polizzi, G., Krakow, K., Fish, D. R., and Lemieux, L. (1998). Identification of EEG events in the MR scanner: the problem of pulse artifact and a method for its subtraction. *Neuroimage*, 8(3):229–239.

Alotaiby, T. N., Alshebeili, S. A., Alshawi, T., Ahmad, I., and El-Samie, F. E. A. (2014). EEG seizure detection and prediction algorithms: a survey. *EURASIP Journal on Advances in Signal Processing*, 2014(1):183.

Amber Diagnostics (2016). Understanding mri gradients & slew rates. `https://www.amberusa.com/blog/understanding-mri-gradients-and-slew-rates/`. [Online; accessed 15/01/2020].

An, D., Fahoum, F., Hall, J., Olivier, A., Gotman, J., and Dubeau, F. (2013). Electroencephalography/functional magnetic resonance imaging responses help predict surgical outcome in focal epilepsy. *Epilepsia*, 54(12):2184–2194.

Attwell, D. and Iadecola, C. (2002). The neural basis of functional brain imaging signals. *Trends in neurosciences*, 25(12):621–625.

Bagshaw, A. P., Aghakhani, Y., Bénar, C.-G., Kobayashi, E., Hawco, C., Dubeau, F., Pike, G. B., and Gotman, J. (2004). EEG-fMRI of focal epileptic spikes: analysis with multiple haemodynamic functions and comparison with gadolinium-enhanced MR angiograms. *Human brain mapping*, 22(3):179–192.

Bajaj, C., Gao, T., He, Z., Huang, Q., and Liang, Z. (2018). SMAC: simultaneous mapping and clustering using spectral decompositions. In *International Conference on Machine Learning*, pages 334–343.

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

Becker, H., Karfoul, A., Albera, L., Gribonval, R., Fleureau, J., Guillotel, P., Kachenoura, A., Senhadji, L., and Merlet, I. (2015). Tensor decomposition exploiting structural constraints for brain source imaging. In *2015 IEEE 6th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pages 181–184. IEEE.

Beckmann, C. F. and Smith, S. M. (2004). Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE transactions on medical imaging*, 23(2):137–152.

Beckmann, C. F. and Smith, S. M. (2005). Tensorial extensions of independent component analysis for multisubject FMRI analysis. *Neuroimage*, 25(1):294–311.

Behzadi, Y., Restom, K., Liau, J., and Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *Neuroimage*, 37(1):90–101.

Belkin, M. and Niyogi, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation*, 15(6):1373–1396.

Bell, A. J. and Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural computation*, 7(6):1129–1159.

Belouchrani, A., Abed-Meraim, K., Cardoso, J.-F., and Moulines, E. (1997). A blind source separation technique using second-order statistics. *IEEE Transactions on signal processing*, 45(2):434–444.

Bénar, C.-G., Gross, D. W., Wang, Y., Petre, V., Pike, B., Dubeau, F., and Gotman, J. (2002). The BOLD response to interictal epileptiform discharges. *Neuroimage*, 17(3):1182–1192.

Bennett, C. M., Miller, M. B., and Wolford, G. (2009). Neural correlates of interspecies perspective taking in the post-mortem Atlantic Salmon: an argument for multiple comparisons correction. *Neuroimage*, 47(Suppl 1):S125.

Bishop, C. M. et al. (1995). *Neural networks for pattern recognition*. Oxford university press.

Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M., and Muller, K.-R. (2007). Optimizing spatial filters for robust EEG single-trial analysis. *IEEE Signal processing magazine*, 25(1):41–56.

Boon, P., Vonck, K., De Herdt, V., Van Dycke, A., Goethals, M., Goossens, L., Van Zandijcke, M., De Smedt, T., Dewaele, I., Achten, R., et al. (2007). Deep brain stimulation in patients with refractory temporal lobe epilepsy. *Epilepsia*, 48(8):1551–1560.

Borga, M. and Knutsson, H. (2001). A canonical correlation approach to blind source separation. *Report LiU-IMT-EX-0062 Department of Biomedical Engineering, Linkping University*.

Boussé, M., Goovaerts, G., et al. (2017). Irregular heartbeat classification using Kronecker Product Equations. In *Proc. of the 39th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC 2017, Jeju Island, South Korea)*, pages 438–441.

Boussé, M., Vervliet, N., Domanov, I., Debals, O., and De Lathauwer, L. (2018). Linear systems with a canonical polyadic decomposition constrained solution: Algorithms and applications. *Numerical Linear Algebra with Applications*, 25(6):e2190.

Bro, R. (1997). PARAFAC. Tutorial and applications. *Chemometrics and intelligent laboratory systems*, 38(2):149–171.

Bro, R. and Kiers, H. A. (2003). A new efficient method for determining the number of components in PARAFAC models. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 17(5):274–286.

Bro, R., Kjeldahl, K., Smilde, A., and Kiers, H. (2008). Cross-validation of component models: a critical look at current methods. *Analytical and bioanalytical chemistry*, 390(5):1241–1251.

Brunner, P., Bianchi, L., Guger, C., Cincotti, F., and Schalk, G. (2011). Current trends in hardware and software for brain–computer interfaces (BCIs). *Journal of neural engineering*, 8(2):025001.

Bullmore, E., Long, C., Suckling, J., Fadili, J., Calvert, G., Zelaya, F., Carpenter, T. A., and Brammer, M. (2001). Colored noise and computational inference in neurophysiological (fMRI) time series analysis: resampling methods in time and wavelet domains. *Human brain mapping*, 12(2):61–78.

Bullmore, E. and Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature reviews neuroscience*, 10(3):186.

Buxton, R. B. (2013). The physics of functional magnetic resonance imaging (fMRI). *Reports on Progress in Physics*, 76(9):096601.

Buzsaki, G. (2006). *Rhythms of the Brain*. Oxford University Press.

Buzsáki, G., Anastassiou, C. A., and Koch, C. (2012). The origin of extracellular fields and currents—EEG, ECoG, LFP and spikes. *Nature reviews neuroscience*, 13(6):407.

Caballero-Gaudes, C., Van de Ville, D., Grouiller, F., Thornton, R., Lemieux, L., Seeck, M., Lazeyras, F., and Vulliemoz, S. (2013). Mapping interictal epileptic discharges using mutual information between concurrent EEG and fMRI. *Neuroimage*, 68:248–262.

Calhoun, V. D. and Adali, T. (2006). Unmixing fMRI with independent component analysis. *IEEE Engineering in Medicine and Biology Magazine*, 25(2):79–90.

Calhoun, V. D., Adali, T., Pearlson, G. D., and Kiehl, K. A. (2006). Neuronal chronometry of target detection: fusion of hemodynamic and event-related potential data. *Neuroimage*, 30(2):544–553.

Calhoun, V. D., Liu, J., and Adalı, T. (2009). A review of group ICA for fmri data and ICA for joint inference of imaging, genetic, and ERP data. *Neuroimage*, 45(1):S163–S172.

Calhoun, V. D., Potluru, V. K., Phlypo, R., Silva, R. F., Pearlmutter, B. A., Caprihan, A., Plis, S. M., and Adalı, T. (2013). Independent component analysis for brain fMRI does indeed select for maximal independence. *PloS one*, 8(8):e73309.

Calhoun, V. D., Stevens, M., Pearlson, G., and Kiehl, K. (2004). fMRI analysis with the general linear model: removal of latency-induced amplitude bias by incorporation of hemodynamic derivative terms. *Neuroimage*, 22(1):252–257.

Cardoso, J.-F. (1998). Blind signal separation: statistical principles. *Proceedings of the IEEE*, 86(10):2009–2025.

Cardoso, J.-F. (1999). High-order contrasts for independent component analysis. *Neural computation*, 11(1):157–192.

Cardoso, J.-F. and Souloumiac, A. (1993). Blind beamforming for non-gaussian signals. In *IEE proceedings F (radar and signal processing)*, volume 140, pages 362–370. IET.

Carroll, J. D. and Chang, J.-J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of "eckart-young" decomposition. *Psychometrika*, 35(3):283–319.

Centeno, M. and Carmichael, D. W. (2014). Network connectivity in epilepsy: resting state fMRI and EEG–fMRI contributions. *Frontiers in neurology*, 5:93.

Chatzichristos, C., Davies, M., Escudero, J., Kofidis, E., and Theodoridis, S. (2018). Fusion of EEG and fMRI via soft coupled tensor decompositions. In *2018 26th European Signal Processing Conference (EUSIPCO)*, pages 56–60. IEEE.

Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357.

Chen, Y., Guibas, L., and Huang, Q. (2014). Near-optimal joint object matching via convex relaxation. In Xing, E. P. and Jebara, T., editors, *Proceedings of the 31st International Conference on Machine Learning*, volume 32 of *Proceedings of Machine Learning Research*, pages 100–108, Bejing, China. PMLR.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the acoustical society of America*, 25(5):975–979.

Cichocki, A. (2013). Tensor decompositions: a new concept in brain data analysis? *arXiv preprint arXiv:1305.0395*.

Cichocki, A. and Amari, S.-i. (2010). Families of alpha-beta-and gamma-divergences: Flexible and robust measures of similarities. *Entropy*, 12(6):1532–1568.

Cichocki, A., Mandic, D., De Lathauwer, L., Zhou, G., Zhao, Q., Caiafa, C., and Phan, H. A. (2015). Tensor decompositions for signal processing applications: From two-way to multiway component analysis. *IEEE Signal Processing Magazine*, 32(2):145–163.

Cichocki, A., Zdunek, R., and Amari, S.-i. (2007). Nonnegative matrix and tensor factorization [lecture notes]. *IEEE signal processing magazine*, 25(1):142–145.

Cohen, M. X. (2017). Where does EEG come from and what does it mean? *Trends in neurosciences*, 40(4):208–218.

Comon, P. (1994). Independent component analysis, a new concept? *Signal processing*, 36(3):287–314.

Comon, P. and De Lathauwer, L. (2010). Algebraic identification of under-determined mixtures. In *Handbook of Blind Source Separation*, pages 325–365. Elsevier.

Correa, N. M., Eichele, T., Adalı, T., Li, Y.-O., and Calhoun, V. D. (2010). Multi-set canonical correlation analysis for the fusion of concurrent single trial ERP and functional MRI. *Neuroimage*, 50(4):1438–1445.

Cox, D. D. and Savoy, R. L. (2003). Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. *Neuroimage*, 19(2):261–270.

da Costa, J. P. C., Haardt, M., and Romer, F. (2008). Robust methods based on the HOSVD for estimating the model order in PARAFAC models. In *2008 5th IEEE Sensor Array and Multichannel Signal Processing Workshop*, pages 510–514. IEEE.

Dähne, S., Bießmann, F., Samek, W., Haufe, S., Goltz, D., Gundlach, C., Villringer, A., Fazli, S., and Müller, K.-R. (2015). Multivariate machine learning methods for fusing multimodal functional neuroimaging data. *Proceedings of the IEEE*, 103(9):1507–1530.

Dai, W., Jin, O., Xue, G.-R., Yang, Q., and Yu, Y. (2009). Eigentransfer: a unified framework for transfer learning. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 193–200. ACM.

Daubechies, I., Roussos, E., Takerkart, S., Benharrosh, M., Golden, C., D'ardenne, K., Richter, W., Cohen, J. D., and Haxby, J. (2009). Independent component analysis for brain fMRI does not select for independence. *Proceedings of the National Academy of Sciences*, 106(26):10415–10422.

Daunizeau, J., Grova, C., Marrelec, G., Mattout, J., Jbabdi, S., Pélégrini-Issac, M., Lina, J.-M., and Benali, H. (2007). Symmetrical event-related EEG/fMRI information fusion in a variational bayesian framework. *Neuroimage*, 36(1):69–87.

Davatzikos, C., Ruparel, K., Fan, Y., Shen, D., Acharyya, M., Loughead, J. W., Gur, R. C., and Langleben, D. D. (2005). Classifying spatial patterns of brain activity with machine learning methods: application to lie detection. *Neuroimage*, 28(3):663–668.

De Clercq, W., Vergult, A., Vanrumste, B., Van Paesschen, W., and Van Huffel, S. (2006). Canonical correlation analysis applied to remove muscle artifacts from the electroencephalogram. *IEEE transactions on Biomedical Engineering*, 53(12):2583–2587.

De Lathauwer, L. (2008). Decompositions of a higher-order tensor in block terms—Part II: Definitions and uniqueness. *SIAM Journal on Matrix Analysis and Applications*, 30(3):1033–1066.

De Lathauwer, L. (2011). Blind separation of exponential polynomials and the decomposition of a tensor in rank-(L_r,L_r,1) terms. *SIAM Journal on Matrix Analysis and Applications*, 32(4):1451–1474.

De Lathauwer, L., Castaing, J., and Cardoso, J.-F. (2007). Fourth-order cumulant-based blind identification of underdetermined mixtures. *IEEE Transactions on Signal Processing*, 55(6):2965–2973.

De Lathauwer, L., De Moor, B., and Vandewalle, J. (2000a). A multilinear singular value decomposition. *SIAM journal on Matrix Analysis and Applications*, 21(4):1253–1278.

De Lathauwer, L., De Moor, B., and Vandewalle, J. (2000b). On the best rank-1 and rank-(r 1, r 2,..., rn) approximation of higher-order tensors. *SIAM journal on Matrix Analysis and Applications*, 21(4):1324–1342.

De Lathauwer, L. and Kofidis, E. (2017). Coupled matrix-tensor factorizations—the case of partially shared factors. In *2017 51st Asilomar Conference on Signals, Systems, and Computers*, pages 711–715. IEEE.

De Lathauwer, L. and Nion, D. (2008). Decompositions of a higher-order tensor in block terms—Part iii: Alternating least squares algorithms. *SIAM journal on Matrix Analysis and Applications*, 30(3):1067–1083.

De Vos, M., Vergult, A., De Lathauwer, L., De Clercq, W., Van Huffel, S., Dupont, P., Palmini, A., and Van Paesschen, W. (2007). Canonical decomposition of ictal scalp EEG reliably detects the seizure onset zone. *NeuroImage*, 37(3):844–854.

Debals, O. (2017). *Tensorization and Applications in Blind Source Separation.* PhD thesis, KU Leuven.

Debals, O. and De Lathauwer, L. (2015). Stochastic and deterministic tensorization for blind signal separation. In *International Conference on Latent Variable Analysis and Signal Separation*, pages 3–13. Springer.

Debals, O., Van Barel, M., and De Lathauwer, L. (2015). Löwner-based blind signal separation of rational functions with applications. *IEEE Transactions on Signal Processing*, 64(8):1909–1918.

Debener, S., Ullsperger, M., Siegel, M., and Engel, A. K. (2006). Single-trial EEG–fMRI reveals the dynamics of cognitive function. *Trends in cognitive sciences*, 10(12):558–563.

Deburchgraeve, W., Cherian, P. J., De Vos, M., Swarte, R. M., Blok, J. H., Visser, G. H., Govaert, P., and Van Huffel, S. (2009). Neonatal seizure localization

using PARAFAC decomposition. *Clinical Neurophysiology*, 120(10):1787–1796.

Delorme, A. and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods*, 134(1):9–21.

Doclo, S. and Moonen, M. (2002). GSVD-based optimal filtering for single and multimicrophone speech enhancement. *IEEE Transactions on signal processing*, 50(9):2230–2244.

Domanov, I. and De Lathauwer, L. (2016). Generic uniqueness of a structured matrix factorization and applications in blind source separation. *IEEE Journal of Selected Topics in Signal Processing*, 10(4):701–711.

Donoho, D. L. (2006). For most large underdetermined systems of linear equations the minimal ℓ_1-norm solution is also the sparsest solution. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 59(6):797–829.

Duann, J.-R., Jung, T.-P., Kuo, W.-J., Yeh, T.-C., Makeig, S., Hsieh, J.-C., and Sejnowski, T. J. (2002). Single-trial variability in event-related BOLD signals. *Neuroimage*, 15(4):823–835.

Eichele, T., Calhoun, V. D., Moosmann, M., Specht, K., Jongsma, M. L., Quiroga, R. Q., Nordby, H., and Hugdahl, K. (2008). Unmixing concurrent EEG-fMRI with parallel independent component analysis. *International Journal of Psychophysiology*, 67(3):222–234.

Ekstrom, A. (2010). How and when the fMRI BOLD signal relates to underlying neural activity: the danger in dissociation. *Brain research reviews*, 62(2):233–244.

Elbau, I. G., Brücklmeier, B., Uhr, M., Arloth, J., Czamara, D., Spoormaker, V. I., Czisch, M., Stephan, K. E., Binder, E. B., and Sämann, P. G. (2018). The brain's hemodynamic response function rapidly changes under acute psychosocial stress in association with genetic and endocrine stress response markers. *Proceedings of the National Academy of Sciences*, 115(43):E10206–E10215.

Elster, A. D. (2019). Questions and answers in MRI. `https://mriquestions.com`. [Online; accessed 15/1/2020].

Evans, A. C., Janke, A. L., Collins, D. L., and Baillet, S. (2012). Brain templates and atlases. *Neuroimage*, 62(2):911–922.

Fan, L., Li, H., Zhuo, J., Zhang, Y., Wang, J., Chen, L., Yang, Z., Chu, C., Xie, S., Laird, A. R., et al. (2016). The human brainnetome atlas: a new brain atlas based on connectional architecture. *Cerebral cortex*, 26(8):3508–3526.

Fasano, A. and Villani, V. (2014). Baseline wander removal for bioelectrical signals by quadratic variation reduction. *Signal Processing*, 99:48–57.

Fazli, S., Mehnert, J., Steinbrink, J., Curio, G., Villringer, A., Müller, K.-R., and Blankertz, B. (2012). Enhanced performance by a hybrid NIRS–EEG brain computer interface. *Neuroimage*, 59(1):519–529.

Févotte, C. and Idier, J. (2011). Algorithms for nonnegative matrix factorization with the β-divergence. *Neural computation*, 23(9):2421–2456.

Flanagan, D., Abbott, D. F., and Jackson, G. D. (2009). How wrong can we be? the effect of inaccurate mark-up of EEG/fMRI studies in epilepsy. *Clinical Neurophysiology*, 120(9):1637–1647.

Formaggio, E., Storti, S. F., Bertoldo, A., Manganotti, P., Fiaschi, A., and Toffolo, G. M. (2011). Integrating EEG and fMRI in epilepsy. *Neuroimage*, 54(4):2719–2731.

Friedman, J. H. (1997). On bias, variance, 0/1—loss, and the curse-of-dimensionality. *Data mining and knowledge discovery*, 1(1):55–77.

Friston, K. J., Ashburner, J., Frith, C. D., Poline, J.-B., Heather, J. D., and Frackowiak, R. S. (1995). Spatial registration and normalization of images. *Human brain mapping*, 3(3):165–189.

Friston, K. J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M., and Turner, R. (1998). Event-related fMRI: characterizing differential responses. *Neuroimage*, 7(1):30–40.

Friston, K. J., Glaser, D. E., Henson, R. N., Kiebel, S., Phillips, C., and Ashburner, J. (2002). Classical and bayesian inference in neuroimaging: applications. *Neuroimage*, 16(2):484–512.

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J.-P., Frith, C. D., and Frackowiak, R. S. (1994). Statistical parametric maps in functional imaging: a general linear approach. *Human brain mapping*, 2(4):189–210.

Frølich, L., Andersen, T. S., and Mørup, M. (2018). Rigorous optimisation of multilinear discriminant analysis with tucker and PARAFAC structures. *BMC bioinformatics*, 19(1):197.

Fukunaga, K. (2013). *Introduction to statistical pattern recognition*. Elsevier.

Glover, G. H. (1999). Deconvolution of impulse response in event-related BOLD fMRI. *Neuroimage*, 9(4):416–429.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*. MIT press.

Gotman, J. (2008). Epileptic networks studied with EEG-fMRI. *Epilepsia*, 49:42–51.

Gotman, J., Kobayashi, E., Bagshaw, A. P., Bénar, C.-G., and Dubeau, F. (2006). Combining EEG and fMRI: a multimodal tool for epilepsy research. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 23(6):906–920.

Grouiller, F., Thornton, R. C., Groening, K., Spinelli, L., Duncan, J. S., Schaller, K., Siniatchkin, M., Lemieux, L., Seeck, M., Michel, C. M., et al. (2011). With or without spikes: localization of focal epileptic activity by simultaneous electroencephalography and functional magnetic resonance imaging. *Brain*, 134(10):2867–2886.

Grouiller, F., Vercueil, L., Krainik, A., Segebarth, C., Kahane, P., and David, O. (2010). Characterization of the hemodynamic modes associated with interictal epileptic activity using a deformable model-based analysis of combined EEG and functional MRI recordings. *Human brain mapping*, 31(8):1157–1173.

Gu, Y., Cleeren, E., Dan, J., Claes, K., Van Paesschen, W., Van Huffel, S., and Hunyadi, B. (2018). Comparison between scalp EEG and behind-the-ear EEG for development of a wearable seizure detection system for patients with focal epilepsy. *Sensors*, 18(1):29.

Guo, W. and Yu, W. (2011). Variational bayesian PARAFAC decomposition for multidimensional harmonic retrieval. In *Proceedings of 2011 IEEE CIE International Conference on Radar*, volume 2, pages 1864–1867. IEEE.

Hallez, H., Vanrumste, B., Grech, R., Muscat, J., De Clercq, W., Vergult, A., D'Asseler, Y., Camilleri, K. P., Fabri, S. G., Van Huffel, S., et al. (2007). Review on solving the forward problem in EEG source analysis. *Journal of neuroengineering and rehabilitation*, 4(1):46.

Handwerker, D. A., Ollinger, J. M., and D'Esposito, M. (2004). Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. *Neuroimage*, 21(4):1639–1651.

Hao, Y., Khoo, H. M., von Ellenrieder, N., Zazubovits, N., and Gotman, J. (2018). DeepIED: An epileptic discharge detector for EEG-fMRI based on deep learning. *NeuroImage: Clinical*, 17:962–975.

Harshman, R. A. et al. (1970). Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multimodal factor analysis. *UCLA Working Papers in Phonetics*, 16:1–84.

Hawco, C. S., Bagshaw, A. P., Lu, Y., Dubeau, F., and Gotman, J. (2007). BOLD changes occur prior to epileptic spikes seen on scalp EEG. *Neuroimage*, 35(4):1450–1458.

Himberg, J., Hyvärinen, A., and Esposito, F. (2004). Validating the independent components of neuroimaging time series via clustering and visualization. *Neuroimage*, 22(3):1214–1222.

Hinton, G. E. and Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *science*, 313(5786):504–507.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, 24(6):417.

Huang, Q.-X. and Guibas, L. (2013). Consistent shape maps via semidefinite programming. In *Proceedings of the Eleventh Eurographics/ACMSIGGRAPH Symposium on Geometry Processing*, pages 177–186. Eurographics Association.

Huber, D. (2002). *Automatic Three-dimensional Modeling from Reality*. PhD thesis, Carnegie Mellon University, Pittsburgh, PA.

Hufnagel, A., Dümpelmann, M., Zentner, J., Schijns, O., and Elger, C. (2000). Clinical relevance of quantified intracranial interictal spike activity in presurgical evaluation of epilepsy. *Epilepsia*, 41(4):467–478.

Hunyadi, B., Camps, D., Sorber, L., Van Paesschen, W., De Vos, M., Van Huffel, S., and De Lathauwer, L. (2014). Block term decomposition for modelling epileptic seizures. *EURASIP Journal on Advances in Signal Processing*, 2014(1):139.

Hunyadi, B., Dupont, P., Van Paesschen, W., and Van Huffel, S. (2017). Tensor decompositions and data fusion in epileptic electroencephalography and functional magnetic resonance imaging data. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 7(1):e1197.

Hunyadi, B., Signoretto, M., Van Paesschen, W., Suykens, J. A., Van Huffel, S., and De Vos, M. (2012). Incorporating structural information from the multichannel EEG improves patient-specific seizure detection. *Clinical Neurophysiology*, 123(12):2352–2361.

Hunyadi, B., Tousseyn, S., Dupont, P., Van Huffel, S., De Vos, M., and Van Paesschen, W. (2015). A prospective fMRI-based technique for localising the epileptogenic zone in presurgical evaluation of epilepsy. *Neuroimage*, 113:329–339.

Hunyadi, B., Tousseyn, S., Mijović, B., Dupont, P., Van Huffel, S., Van Paesschen, W., and De Vos, M. (2013). ICA extracts epileptic sources from fMRI in EEG-negative patients: a retrospective validation study. *PloS one*, 8(11):e78796.

Hunyadi, B., Van Paesschen, W., De Vos, M., and Van Huffel, S. (2016). Fusion of electroencephalography and functional magnetic resonance imaging to explore epileptic network activity. In *2016 24th European Signal Processing Conference (EUSIPCO)*, pages 240–244. IEEE.

Hyvärinen, A. and Oja, E. (2000). Independent component analysis: algorithms and applications. *Neural networks*, 13(4-5):411–430.

Hyvärinen, A., Ramkumar, P., Parkkonen, L., and Hari, R. (2010). Independent component analysis of short-time fourier transforms for spontaneous EEG/MEG analysis. *NeuroImage*, 49(1):257–271.

Jacobs, J., Hawco, C., Kobayashi, E., Boor, R., LeVan, P., Stephani, U., Siniatchkin, M., and Gotman, J. (2008). Variability of the hemodynamic response as a function of age and frequency of epileptic discharge in children with epilepsy. *Neuroimage*, 40(2):601–614.

Jacobs, J., LeVan, P., Moeller, F., Boor, R., Stephani, U., Gotman, J., and Siniatchkin, M. (2009). Hemodynamic changes preceding the interictal EEG spike in patients with focal epilepsy investigated using simultaneous EEG-fMRI. *Neuroimage*, 45(4):1220–1231.

Jann, K., Dierks, T., Boesch, C., Kottlow, M., Strik, W., and Koenig, T. (2009). BOLD correlates of EEG alpha phase-locking and the fmri default mode network. *Neuroimage*, 45(3):903–916.

Javidan, M. (2012). Electroencephalography in mesial temporal lobe epilepsy: a review. *Epilepsy research and treatment*, 2012.

Jia, Y., Nie, F., and Zhang, C. (2009). Trace ratio problem revisited. *IEEE Transactions on Neural Networks*, 20(4):729–735.

Jung, T.-P., Makeig, S., Humphries, C., Lee, T.-W., Mckeown, M. J., Iragui, V., and Sejnowski, T. J. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, 37(2):163–178.

Karahan, E., Rojas-Lopez, P. A., Bringas-Vega, M. L., Valdés-Hernández, P. A., and Valdes-Sosa, P. A. (2015). Tensor analysis and fusion of multimodal brain images. *Proceedings of the IEEE*, 103(9):1531–1559.

Khalidov, I., Fadili, J., Lazeyras, F., Van De Ville, D., and Unser, M. (2011). Activelets: Wavelets for sparse representation of hemodynamic responses. *Signal processing*, 91(12):2810–2821.

Khoo, H. M., Hao, Y., von Ellenrieder, N., Zazubovits, N., Hall, J., Olivier, A., Dubeau, F., and Gotman, J. (2017). The hemodynamic response to interictal epileptic discharges localizes the seizure-onset zone. *Epilepsia*, 58(5):811–823.

Khoo, H. M., von Ellenrieder, N., Zazubovits, N., He, D., Dubeau, F., and Gotman, J. (2018). The spike onset zone: The region where epileptic spikes start and from where they propagate. *Neurology*, 91(7):e666–e674.

Kirschstein, T. and Köhling, R. (2009). What is the source of the EEG? *Clinical EEG and neuroscience*, 40(3):146–149.

Kolda, T. G. and Bader, B. W. (2009). Tensor decompositions and applications. *SIAM review*, 51(3):455–500.

Korn, K., Seifert, B., and Uhl, C. (2019). Dynamical component analysis (DyCA) and its application on epileptic EEG. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1100–1104. IEEE.

Kramer, M. A., Kolaczyk, E. D., and Kirsch, H. E. (2008). Emergent network topology at seizure onset in humans. *Epilepsy research*, 79(2-3):173–186.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105.

Kruskal, J. B. (1977). Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear algebra and its applications*, 18(2):95–138.

Lahat, D., Adali, T., and Jutten, C. (2015). Multimodal data fusion: an overview of methods, challenges, and prospects. *Proceedings of the IEEE*, 103(9):1449–1477.

Lal, T. N., Schroder, M., et al. (2004). Support vector channel selection in BCI. *IEEE transactions on biomedical engineering*, 51(6):1003–1010.

Lantz, G., De Peralta, R. G., Spinelli, L., Seeck, M., and Michel, C. (2003). Epileptic source localization with high density EEG: how many electrodes are needed? *Clinical neurophysiology*, 114(1):63–69.

Laufs, H., Kleinschmidt, A., Beyerle, A., Eger, E., Salek-Haddadi, A., Preibisch, C., and Krakow, K. (2003). EEG-correlated fMRI of human alpha activity. *Neuroimage*, 19(4):1463–1476.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *nature*, 521(7553):436.

Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88:365–411.

Lee, D. D. and Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788.

Leite, M., Leal, A., and Figueiredo, P. (2013). Transfer function between EEG and BOLD signals of epileptic activity. *Frontiers in neurology*, 4:1.

Lemieux, L., Laufs, H., Carmichael, D., Paul, J. S., Walker, M. C., and Duncan, J. S. (2008). Noncanonical spike-related BOLD responses in focal epilepsy. *Human brain mapping*, 29(3):329–345.

Lemieux, L., Salek-Haddadi, A., Josephs, O., Allen, P., Toms, N., Scott, C., Krakow, K., Turner, R., and Fish, D. R. (2001). Event-related fMRI with simultaneous and continuous EEG: description of the method and initial case report. *Neuroimage*, 14(3):780–787.

Lemm, S., Blankertz, B., Curio, G., and Muller, K.-R. (2005). Spatio-spectral filters for improving the classification of single trial EEG. *IEEE transactions on biomedical engineering*, 52(9):1541–1548.

Lemm, S., Blankertz, B., Dickhaus, T., and Müller, K.-R. (2011). Introduction to machine learning for brain imaging. *Neuroimage*, 56(2):387–399.

Li, Q. and Schonfeld, D. (2014). Multilinear discriminant analysis for higher-order tensor data classification. *IEEE transactions on pattern analysis and machine intelligence*, 36(12):2524–2537.

Li, Y.-O., Adali, T., Wang, W., and Calhoun, V. D. (2009). Joint blind source separation by multiset canonical correlation analysis. *IEEE Transactions on Signal Processing*, 57(10):3918–3929.

Liang, S., Kuo, C., et al. (2012). Automatic stage scoring of single-channel sleep EEG by using multiscale entropy and autoregressive models. *IEEE Transactions on Instrumentation and Measurement*, 61(6):1649–1657.

Liao, L., Lin, C., et al. (2012). Biosensor technologies for augmented brain–computer interfaces in the next decades. *Proceedings of the IEEE*, 100(Special Centennial Issue):1553–1566.

Lindquist, M. A., Loh, J. M., Atlas, L. Y., and Wager, T. D. (2009). Modeling the hemodynamic response function in fMRI: efficiency, bias and mis-modeling. *Neuroimage*, 45(1):S187–S198.

Lindquist, M. A. and Wager, T. D. (2007). Validity and power in hemodynamic response modeling: a comparison study and a new approach. *Human brain mapping*, 28(8):764–784.

Liu, K., Da Costa, J. P. C., So, H. C., Huang, L., and Ye, J. (2016). Detection of number of components in CANDECOMP/PARAFAC models via minimum description length. *Digital Signal Processing*, 51:110–123.

Liu, W., Pokharel, P. P., and Príncipe, J. C. (2007). Correntropy: Properties and applications in non-Gaussian signal processing. *IEEE Transactions on Signal Processing*, 55(11):5286–5298.

Liu, X. and Sajda, P. (2019). A convolutional neural network for transcoding simultaneously acquired EEG-fMRI data. In *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*, pages 477–482. IEEE.

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., and Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, 412(6843):150.

Looney, D., Kidmose, P., Park, C., Ungstrup, M., Rank, M. L., Rosenkranz, K., and Mandic, D. P. (2012). The in-the-ear recording concept: User-centered and wearable brain monitoring. *IEEE pulse*, 3(6):32–42.

Lotte, F., Bougrain, L., Cichocki, A., Clerc, M., Congedo, M., Rakotomamonjy, A., and Yger, F. (2018). A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update. *Journal of neural engineering*, 15(3):031005.

Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., and Arnaldi, B. (2007). A review of classification algorithms for EEG-based brain–computer interfaces. *Journal of neural engineering*, 4(2):R1.

Lou, B., Hong, B., et al. (2008). Bipolar electrode selection for a motor imagery based brain–computer interface. *Journal of Neural Engineering*, 5(3):342.

Loza, C. A. and Principe, J. C. (2016). A robust maximum correntropy criterion for dictionary learning. In *2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP)*, pages 1–6. IEEE.

Luessi, M., Babacan, S. D., Molina, R., Booth, J. R., and Katsaggelos, A. K. (2011). Bayesian symmetrical EEG/fMRI fusion with spatially adaptive priors. *Neuroimage*, 55(1):113–132.

Lustig, M., Donoho, D. L., Santos, J. M., and Pauly, J. M. (2008). Compressed sensing MRI. *IEEE signal processing magazine*, 25(2):72.

Makeig, S., Bell, A. J., Jung, T.-P., and Sejnowski, T. J. (1996). Independent component analysis of electroencephalographic data. In *Advances in neural information processing systems*, pages 145–151.

Makni, S., Beckmann, C., Smith, S., and Woolrich, M. (2008). Bayesian deconvolution fMRI data using bilinear dynamical systems. *Neuroimage*, 42(4):1381–1396.

Mantini, D., Perrucci, M. G., Del Gratta, C., Romani, G. L., and Corbetta, M. (2007). Electrophysiological signatures of resting state networks in the human brain. *Proceedings of the National Academy of Sciences*, 104(32):13170–13175.

Mareček, R., Lamoš, M., Labounek, R., Bartoň, M., Slavíček, T., Mikl, M., Rektor, I., and Brázdil, M. (2017). Multiway array decomposition of EEG spectrum: implications of its stability for the exploration of large-scale brain networks. *Neural computation*, 29(4):968–989.

Marecek, R., Lamos, M., Mikl, M., Barton, M., Fajkus, J., Rektor, I., and Brázdil, M. (2016). What can be found in scalp EEG spectrum beyond common frequency bands. EEG–fMRI study. *Journal of neural engineering*, 13(4):046026.

Marino, M., Liu, Q., Del Castello, M., Corsi, C., Wenderoth, N., and Mantini, D. (2018). Heart–Brain interactions in the MR environment: characterization of the ballistocardiogram in EEG signals collected during simultaneous fMRI. *Brain topography*, 31(3):337–345.

Martínez-Montes, E., Valdés-Sosa, P. A., Miwakeichi, F., Goldman, R. I., and Cohen, M. S. (2004). Concurrent EEG/fMRI analysis by multiway partial least squares. *NeuroImage*, 22(3):1023–1034.

McKeown, M. J. and Sejnowski, T. J. (1998). Independent component analysis of fMRI data: examining the assumptions. *Human brain mapping*, 6(5-6):368–372.

Michel, C. M., Murray, M. M., Lantz, G., Gonzalez, S., Spinelli, L., and de Peralta, R. G. (2004). EEG source imaging. *Clinical neurophysiology*, 115(10):2195–2222.

Mijović, B., Vanderperren, K., Novitskiy, N., Vanrumste, B., Stiers, P., Van den Bergh, B., Lagae, L., Sunaert, S., Wagemans, J., Van Huffel, S., et al. (2012). The "why" and "how" of JointICA: Results from a visual detection task. *NeuroImage*, 60(2):1171–1185.

Miwakeichi, F., Martınez-Montes, E., Valdés-Sosa, P. A., Nishiyama, N., Mizuhara, H., and Yamaguchi, Y. (2004). Decomposing EEG data into space–time–frequency components using parallel factor analysis. *NeuroImage*, 22(3):1035–1045.

Mizuhara, H., Wang, L.-Q., Kobayashi, K., and Yamaguchi, Y. (2005). Long-range EEG phase synchronization during an arithmetic task indexes a coherent cortical network simultaneously measured by fMRI. *Neuroimage*, 27(3):553–563.

Moeller, F., LeVan, P., Muhle, H., Stephani, U., Dubeau, F., Siniatchkin, M., and Gotman, J. (2010). Absence seizures: individual patterns revealed by EEG-fMRI. *Epilepsia*, 51(10):2000–2010.

Moeller, F., Siebner, H. R., Wolff, S., Muhle, H., Boor, R., Granert, O., Jansen, O., Stephani, U., and Siniatchkin, M. (2008). Changes in activity of striato–thalamo–cortical network precede generalized spike wave discharges. *Neuroimage*, 39(4):1839–1849.

Monti, M. M. (2011). Statistical analysis of fMRI time-series: a critical review of the GLM approach. *Frontiers in human neuroscience*, 5(28).

Mørup, M. and Hansen, L. K. (2009). Automatic relevance determination for multi-way models. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 23(7-8):352–363.

Mørup, M., Hansen, L. K., Herrmann, C. S., Parnas, J., and Arnfred, S. M. (2006). Parallel factor analysis as an exploratory tool for wavelet transformed event-related EEG. *NeuroImage*, 29(3):938–947.

Mulert, C. and Lemieux, L. (2009). *EEG-fMRI: physiological basis, technique, and applications*. Springer Science & Business Media.

Murta, T., Leite, M., Carmichael, D. W., Figueiredo, P., and Lemieux, L. (2015). Electrophysiological correlates of the BOLD signal for EEG-informed fMRI. *Human brain mapping*, 36(1):391–414.

National Center for Adaptive Neurotechnologies (2019). EEG measurement setup. `https://www.bci2000.org/mediawiki/index.php/User_Tutorial:EEG_Measurement_Setup`. [Online; accessed 23/12/2019].

Ng, A. Y., Jordan, M. I., and Weiss, Y. (2002). On spectral clustering: Analysis and an algorithm. In *Advances in neural information processing systems*, pages 849–856.

Ngo, T. T., Bellalij, M., and Saad, Y. (2012). The trace ratio optimization problem. *SIAM review*, 54(3):545–569.

Nichols, T. and Hayasaka, S. (2003). Controlling the familywise error rate in functional neuroimaging: a comparative review. *Statistical methods in medical research*, 12(5):419–446.

Nichols, T. E. and Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Human brain mapping*, 15(1):1–25.

Nonclercq, A., Foulon, M., Verheulpen, D., De Cock, C., Buzatu, M., Mathys, P., and Van Bogaert, P. (2012). Cluster-based spike detection algorithm adapts to interpatient and intrapatient variation in spike morphology. *Journal of neuroscience methods*, 210(2):259–265.

Norman, K. A., Polyn, S. M., Detre, G. J., and Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in cognitive sciences*, 10(9):424–430.

Nunez, P. L., Srinivasan, R., et al. (2006). *Electric fields of the brain: the neurophysics of EEG*. Oxford University Press, USA.

Obeid, I. and Picone, J. (2016). The temple university hospital EEG data corpus. *Frontiers in neuroscience*, 10:196.

Ogawa, S., Lee, T.-M., Kay, A. R., and Tank, D. W. (1990a). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *proceedings of the National Academy of Sciences*, 87(24):9868–9872.

Ogawa, S., Lee, T.-M., Nayak, A. S., and Glynn, P. (1990b). Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. *Magnetic resonance in medicine*, 14(1):68–78.

Oikonomou, V. P., Tzallas, A. T., and Fotiadis, D. I. (2007). A Kalman filter based methodology for EEG spike enhancement. *Computer methods and programs in biomedicine*, 85(2):101–108.

Omidvarnia, A., Kowalczyk, M. A., Pedersen, M., and Jackson, G. D. (2019). Towards fast and reliable simultaneous EEG-fMRI analysis of epilepsy with automatic spike detection. *Clinical Neurophysiology*, 130(3):368–378.

Ostwald, D. and Bagshaw, A. P. (2011). Information theoretic approaches to functional neuroimaging. *Magnetic resonance imaging*, 29(10):1417–1428.

Pan, S. J. and Yang, Q. (2009). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359.

Papalexakis, E. E. (2016). Automatic unsupervised tensor mining with quality assessment. In *Proceedings of the 2016 SIAM International Conference on Data Mining*, pages 711–719. SIAM.

Pascual-Marqui, R. D. (1999). Review of methods for solving the EEG inverse problem. *International journal of bioelectromagnetism*, 1(1):75–86.

Pearson, K. (1901). LIII. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572.

Pedersen, M., Omidvarnia, A. H., Walz, J. M., and Jackson, G. D. (2015). Increased segregation of brain networks in focal epilepsy: an fMRI graph theory finding. *NeuroImage: Clinical*, 8:536–542.

Pedregosa, F., Eickenberg, M., Ciuciu, P., Thirion, B., and Gramfort, A. (2015). Data-driven HRF estimation for encoding and decoding models. *NeuroImage*, 104:209–220.

Pernet, C. R. (2014). Misconceptions in the use of the General Linear Model applied to functional MRI: a tutorial for junior neuro-imagers. *Frontiers in neuroscience*, 8:1.

Pham, D.-T. and Cardoso, J.-F. (2001). Blind separation of instantaneous mixtures of nonstationary sources. *IEEE Transactions on signal processing*, 49(9):1837–1848.

Phan, A. H. and Cichocki, A. (2010). Tensor decompositions for feature extraction and classification of high dimensional datasets. *Nonlinear theory and its applications, IEICE*, 1(1):37–68.

Pittau, F., LeVan, P., Moeller, F., Gholipour, T., Haegelen, C., Zelmann, R., Dubeau, F., and Gotman, J. (2011). Changes preceding interictal epileptic EEG abnormalities: comparison between EEG/fMRI and intracerebral EEG. *Epilepsia*, 52(6):1120–1129.

Poline, J.-B. and Brett, M. (2012). The general linear model and fMRI: does love last forever? *Neuroimage*, 62(2):871–880.

Ponten, S., Bartolomei, F., and Stam, C. (2007). Small-world networks and epilepsy: graph theoretical analysis of intracerebrally recorded mesial temporal lobe seizures. *Clinical neurophysiology*, 118(4):918–927.

Queensland Brain Institute (2019). `https://qbi.uq.edu.au/brain-basics/brain/brain-physiology/how-do-neurons-work`. [Online; accessed 20/12/2019].

Rigau, V., Morin, M., Rousset, M.-C., de Bock, F., Lebrun, A., Coubes, P., Picot, M.-C., Baldy-Moulinier, M., Bockaert, J., Crespel, A., et al. (2007). Angiogenesis is associated with blood–brain barrier permeability in temporal lobe epilepsy. *Brain*, 130(7):1942–1956.

Ritter, P. and Villringer, A. (2006). Simultaneous EEG–fMRI. *Neuroscience & Biobehavioral Reviews*, 30(6):823–838.

Rosa, M. J., Kilner, J., Blankenburg, F., Josephs, O., and Penny, W. (2010). Estimating the transfer function from neuronal activity to BOLD using simultaneous EEG-fMRI. *Neuroimage*, 49(2):1496–1509.

Rosenow, F. and Lüders, H. (2001). Presurgical evaluation of epilepsy. *Brain*, 124(9):1683–1700.

Roweis, S. T. and Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *science*, 290(5500):2323–2326.

Rubinov, M. and Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*, 52(3):1059–1069.

Salek-Haddadi, A., Diehl, B., Hamandi, K., Merschhemke, M., Liston, A., Friston, K., Duncan, J. S., Fish, D. R., and Lemieux, L. (2006). Hemodynamic correlates of epileptiform discharges: an EEG-fMRI study of 63 patients with focal epilepsy. *Brain research*, 1088(1):148–166.

Salek-Haddadi, A., Friston, K., Lemieux, L., and Fish, D. (2003). Studying spontaneous EEG activity with fMRI. *Brain research reviews*, 43(1):110–133.

Salek-Haddadi, A., Merschhemke, M., Lemieux, L., and Fish, D. R. (2002). Simultaneous EEG-correlated ictal fMRI. *Neuroimage*, 16(1):32–40.

Salimi-Khorshidi, G., Douaud, G., Beckmann, C. F., Glasser, M. F., Griffanti, L., and Smith, S. M. (2014). Automatic denoising of functional MRI data: combining independent component analysis and hierarchical fusion of classifiers. *Neuroimage*, 90:449–468.

Scheuer, M. L., Bagic, A., and Wilson, S. B. (2017). Spike detection: Inter-reader agreement and a statistical turing test on a large data set. *Clinical Neurophysiology*, 128(1):243–250.

Schnitzler, A. and Gross, J. (2005). Normal and pathological oscillatory communication in the brain. *Nature reviews neuroscience*, 6(4):285–296.

Schwartz, T. H. (2007). Neurovascular coupling and epilepsy: hemodynamic markers for localizing and predicting seizure onset. *Epilepsy Currents*, 7(4):91–94.

Seifert, B., Korn, K., Hartmann, S., and Uhl, C. (2018). Dynamical component analysis (DyCA): dimensionality reduction for high-dimensional deterministic time-series. In *2018 IEEE 28th International Workshop on Machine Learning for Signal Processing (MLSP)*, pages 1–6. IEEE.

Shah, V., Von Weltin, E., Lopez de Diego, S., McHugh, J. R., Veloso, L., Golmohammadi, M., Obeid, I., and Picone, J. (2018). The temple university hospital seizure detection corpus. *Frontiers in Neuroinformatics*, 12:83.

Shen, Y., Huang, Q., Srebro, N., and Sanghavi, S. (2016). Normalized spectral map synchronization. In *Advances in Neural Information Processing Systems*, pages 4925–4933.

Shin, J., von Lühmann, A., Blankertz, B., Kim, D.-W., Jeong, J., Hwang, H.-J., and Müller, K.-R. (2016). Open access dataset for EEG + NIRS single-trial classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(10):1735–1745.

Shirer, W. R., Ryali, S., Rykhlevskaia, E., Menon, V., and Greicius, M. D. (2012). Decoding subject-driven cognitive states with whole-brain connectivity patterns. *Cerebral cortex*, 22(1):158–165.

Shoeb, A. H. and Guttag, J. V. (2010). Application of machine learning to epileptic seizure detection. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pages 975–982.

Sidiropoulos, N. D. and Bro, R. (2000). On the uniqueness of multilinear decomposition of N-way arrays. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 14(3):229–239.

Sidiropoulos, N. D., De Lathauwer, L., Fu, X., Huang, K., Papalexakis, E. E., and Faloutsos, C. (2017). Tensor decomposition for signal processing and machine learning. *IEEE Transactions on Signal Processing*, 65(13):3551–3582.

Silverman, B. (1986). *Density Estimation for Statistics and Data Analysis*. Monographs on Statistics and Applied Probability. Chapman & Hall.

Sladky, R., Friston, K. J., Tröstl, J., Cunnington, R., Moser, E., and Windischberger, C. (2011). Slice-timing effects and their correction in functional MRI. *Neuroimage*, 58(2):588–594.

Sohrabpour, A., Lu, Y., Kankirawatana, P., Blount, J., Kim, H., and He, B. (2015). Effect of EEG electrode number on epileptic source localization in pediatric patients. *Clinical Neurophysiology*, 126(3):472–480.

Somers, B., Francart, T., and Bertrand, A. (2018). A generic EEG artifact removal algorithm based on the multi-channel Wiener filter. *Journal of neural engineering*, 15(3):036007.

Sorber, L., Van Barel, M., and De Lathauwer, L. (2013). Optimization-based algorithms for tensor decompositions: Canonical polyadic decomposition,

decomposition in rank-(L_r,L_r,1) terms, and a new generalization. *SIAM Journal on Optimization*, 23(2):695–720.

Sorber, L., Van Barel, M., and De Lathauwer, L. (2015). Structured data fusion. *IEEE Journal of Selected Topics in Signal Processing*, 9(4):586–600.

Sørensen, M. and De Lathauwer, L. D. (2015). Coupled canonical polyadic decompositions and (coupled) decompositions in multilinear Rank-$(L_{r,n}, L_{r,n}, 1)$ terms—part I: Uniqueness. *SIAM Journal on Matrix Analysis and Applications*, 36(2):496–522.

Sørensen, M., Domanov, I., and De Lathauwer, L. (2015). Coupled Canonical Polyadic Decompositions and (coupled) decompositions in multilinear Rank-$(L_{r,n}, L_{r,n}, 1)$ Terms—Part II: Algorithms. *SIAM Journal on Matrix Analysis and Applications*, 36(3):1015–1045.

Staley, K. J. and Dudek, F. E. (2006). Interictal spikes and epileptogenesis. *Epilepsy Currents*, 6(6):199–202.

Staljanssens, W., Strobbe, G., Van Holen, R., Birot, G., Gschwind, M., Seeck, M., Vandenberghe, S., Vulliémoz, S., and van Mierlo, P. (2017a). Seizure onset zone localization from ictal high-density EEG in refractory focal epilepsy. *Brain topography*, 30(2):257–271.

Staljanssens, W., Strobbe, G., Van Holen, R., Keereman, V., Gadeyne, S., Carrette, E., Meurs, A., Pittau, F., Momjian, S., Seeck, M., et al. (2017b). EEG source connectivity to localize the seizure onset zone in patients with drug resistant epilepsy. *NeuroImage: Clinical*, 16:689–698.

Stam, C. J. (2014). Modern network science of neurological disorders. *Nature Reviews Neuroscience*, 15(10):683.

Suetens, P. (2017). *Fundamentals of medical imaging*. Cambridge university press.

Suykens, J. A., Van Gestel, T., and De Brabanter, J. (2002). *Least squares support vector machines*. World Scientific.

Swartz Center for Computational Neuroscience (2020). ICLabel Tutorial: EEG Independent Component Labeling. `https://labeling.ucsd.edu/tutorial/overview`. [Online; accessed 10/01/2020].

Thomson, D. J. (1982). Spectrum estimation and harmonic analysis. *Proceedings of the IEEE*, 70(9):1055–1096.

Thornton, R., Laufs, H., Rodionov, R., Cannadathu, S., Carmichael, D. W., Vulliemoz, S., Salek-Haddadi, A., McEvoy, A. W., Smith, S. M., Lhatoo, S., et al. (2010). EEG correlated functional MRI and postoperative outcome in focal epilepsy. *Journal of Neurology, Neurosurgery & Psychiatry*, 81(8):922–927.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.

Tikhonov, A. N. (1943). On the stability of inverse problems. In *Dokl. Akad. Nauk SSSR*, volume 39, pages 195–198.

Tousseyn, S., Dupont, P., Goffin, K., Sunaert, S., and Van Paesschen, W. (2014a). Sensitivity and specificity of interictal EEG-fMRI for detecting the ictal onset zone at different statistical thresholds. *Frontiers in neurology*, 5:131.

Tousseyn, S., Dupont, P., Goffin, K., Sunaert, S., and Van Paesschen, W. (2015). Correspondence between large-scale ictal and interictal epileptic networks revealed by single photon emission computed tomography (SPECT) and electroencephalography (EEG)–functional magnetic resonance imaging (fMRI). *Epilepsia*, 56(3):382–392.

Tousseyn, S., Dupont, P., Robben, D., Goffin, K., Sunaert, S., and Van Paesschen, W. (2014b). A reliable and time-saving semiautomatic spike-template–based analysis of interictal EEG–fMRI. *Epilepsia*, 55(12):2048–2058.

Trujillo-Barreto, N., Martinez-Montes, E., Melie-Garcia, L., and Valdes-Sosa, P. (2001). A symmetrical bayesian model for fMRI and EEG/MEG neuroimage fusion. *Int. J. Bioelectromagn*, 3(1):1998–2000.

Urigüen, J. A. and Garcia-Zapirain, B. (2015). EEG artifact removal—state-of-the-art and guidelines. *Journal of neural engineering*, 12(3):031001.

Valdes-Sosa, P. A., Sanchez-Bornot, J. M., Sotero, R. C., Iturria-Medina, Y., Aleman-Gomez, Y., Bosch-Bayard, J., Carbonell, F., and Ozaki, T. (2009). Model driven EEG/fMRI fusion of brain oscillations. *Human brain mapping*, 30(9):2701–2721.

Valdés-Sosa, P. A., Vega-Hernández, M., Sánchez-Bornot, J. M., Martínez-Montes, E., and Bobes, M. A. (2009). EEG source imaging with spatio-temporal tomographic nonnegative independent component analysis. *Human brain mapping*, 30(6):1898–1910.

Van de Steen, F., Faes, L., Karahan, E., Songsiri, J., Valdes-Sosa, P. A., and Marinazzo, D. (2019). Critical comments on EEG sensor space dynamical connectivity analysis. *Brain topography*, 32(4):643–654.

Van Eyndhoven, S., Francart, T., and Bertrand, A. (2016). EEG-informed attended speaker extraction from recorded speech mixtures with application in neuro-steered hearing prostheses. *IEEE Transactions on Biomedical Engineering*, 64(5):1045–1056.

van Houdt, P. J., de Munck, J. C., Leijten, F. S., Huiskamp, G. J., Colon, A. J., Boon, P. A., and Ossenblok, P. P. (2013). EEG-fMRI correlation patterns in the presurgical evaluation of focal epilepsy: a comparison with electrocorticographic data and surgical outcome measures. *Neuroimage*, 75:238–248.

Van Mierlo, P., Papadopoulou, M., Carrette, E., Boon, P., Vandenberghe, S., Vonck, K., and Marinazzo, D. (2014). Functional brain connectivity from EEG in epilepsy: Seizure prediction and epileptogenic focus localization. *Progress in neurobiology*, 121:19–35.

Van Paesschen, W., Dupont, P., Sunaert, S., Goffin, K., and Van Laere, K. (2007). The use of SPECT and PET in routine clinical practice in epilepsy. *Current opinion in neurology*, 20(2):194–202.

Vandecappelle, M., Vervliet, N., and De Lathauwer, L. (2019). Rank-one tensor approximation with beta-divergence cost functions. In *2019 27th European Signal Processing Conference (EUSIPCO)*, pages 1–5. IEEE.

Vanderperren, K., De Vos, M., Ramautar, J. R., Novitskiy, N., Mennes, M., Assecondi, S., Vanrumste, B., Stiers, P., Van den Bergh, B. R., Wagemans, J., et al. (2010). Removal of BCG artifacts from EEG recordings inside the MR scanner: a comparison of methodological and validation-related aspects. *Neuroimage*, 50(3):920–934.

Vasilescu, M. A. O. and Terzopoulos, D. (2002). Multilinear image analysis for facial recognition. In *Object recognition supported by user interaction for service robots*, volume 2, pages 511–514.

Vervliet, N. (2018). *Compressed sensing approaches to large-scale tensor decompositions*. PhD thesis, KU Leuven.

Vervliet, N. and De Lathauwer, L. (2018). Numerical optimization based algorithms for data fusion. *Data Fusion Methodology and Applications, M. Cocchi, ed., Elsevier*.

Vervliet, N., Debals, O., and De Lathauwer, L. (2016). Tensorlab 3.0—Numerical optimization strategies for large-scale constrained and coupled matrix/tensor factorization. In *2016 50th Asilomar Conference on Signals, Systems and Computers*, pages 1733–1738. IEEE.

Vincent, T., Risser, L., and Ciuciu, P. (2010). Spatially adaptive mixture modeling for analysis of fMRI time series. *IEEE transactions on medical imaging*, 29(4):1059–1074.

Von Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and computing*, 17(4):395–416.

Vos, S. B., Tax, C. M., Luijten, P. R., Ourselin, S., Leemans, A., and Froeling, M. (2017). The importance of correcting for signal drift in diffusion MRI. *Magnetic resonance in medicine*, 77(1):285–299.

Waites, A. B., Shaw, M. E., Briellmann, R. S., Labate, A., Abbott, D. F., and Jackson, G. D. (2005). How reliable are fMRI–EEG studies of epilepsy? a nonparametric approach to analysis validation and optimization. *Neuroimage*, 24(1):192–199.

Wang, H., Yan, S., Xu, D., Tang, X., and Huang, T. (2007). Trace ratio vs. ratio trace for dimensionality reduction. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE.

WebMD (2014). `https://www.webmd.com/brain/picture-of-the-brain#1`. [Online; accessed 19/12/2019].

Whitfield-Gabrieli, S. and Nieto-Castanon, A. (2012). Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks. *Brain connectivity*, 2(3):125–141.

Wikipedia (2007). `https://en.wikipedia.org/wiki/File:Complete_neuron_cell_diagram_en.svg`. [Online; accessed 20/12/2019].

Wilson, S. B., Turner, C. A., Emerson, R. G., and Scheuer, M. L. (1999). Spike detection II: automatic, perception-based detection and clustering. *Clinical neurophysiology*, 110(3):404–411.

Woolrich, M. W., Behrens, T. E., and Smith, S. M. (2004). Constrained linear basis sets for HRF modelling using Variational Bayes. *NeuroImage*, 21(4):1748–1761.

Woolrich, M. W., Jbabdi, S., Patenaude, B., Chappell, M., Makni, S., Behrens, T., Beckmann, C., Jenkinson, M., and Smith, S. M. (2009). Bayesian analysis of neuroimaging data in FSL. *Neuroimage*, 45(1):S173–S186.

World Health Organization (2019). Epilepsy. `https://www.who.int/en/news-room/fact-sheets/detail/epilepsy`. [Online; accessed 14/03/2019].

Worsley, K. J. (1994). Local maxima and the expected euler characteristic of excursion sets of χ 2, f and t fields. *Advances in Applied Probability*, 26(1):13–42.

Worsley, K. J., Evans, A. C., Marrett, S., and Neelin, P. (1992). A three-dimensional statistical analysis for CBF activation studies in human brain. *Journal of Cerebral Blood Flow & Metabolism*, 12(6):900–918.

Worsley, K. J. and Friston, K. J. (1995). Analysis of fMRI time-series revisited—again. *Neuroimage*, 2(3):173–181.

Wouters, J., Kloosterman, F., and Bertrand, A. (2018). Towards online spike sorting for high-density neural probes using discriminative template matching with suppression of interfering spikes. *Journal of neural engineering*, 15(5):056005.

Yan, S., Xu, D., Zhang, B., Zhang, H.-J., Yang, Q., and Lin, S. (2006). Graph embedding and extensions: A general framework for dimensionality reduction. *IEEE transactions on pattern analysis and machine intelligence*, 29(1):40–51.

Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):49–67.

Zach, C., Klopschitz, M., and Pollefeys, M. (2010). Disambiguating visual relations using loop constraints. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1426–1433. IEEE.

Zhang, M., Gao, Y., Sun, C., La Salle, J., and Liang, J. (2016). Robust tensor factorization using maximum correntropy criterion. In *2016 23rd International Conference on Pattern Recognition (ICPR)*, pages 4184–4189. IEEE.

Zhou, H., Li, L., and Zhu, H. (2013). Tensor regression with applications in neuroimaging data analysis. *Journal of the American Statistical Association*, 108(502):540–552.

Zijlmans, M., Huiskamp, G., Hersevoort, M., Seppenwoolde, J.-H., van Huffelen, A. C., and Leijten, F. S. (2007). EEG-fMRI in the preoperative work-up for epilepsy surgery. *Brain*, 130(9):2343–2353.

Zink, R., Hunyadi, B., Van Huffel, S., and De Vos, M. (2016). Tensor-based classification of an auditory mobile BCI without a subject-specific calibration phase. *Journal of neural engineering*, 13(2):026005.

List of publications

Papers in international journals

1. Van Eyndhoven, S., Hunyadi, B., Dupont, P., Van Paesschen, W., & Van Huffel, S. (2019) "Semi-automated EEG enhancement improves localization of ictal onset zone with EEG-correlated fMRI", *Frontiers in Neurology*, 10:805

2. Novillo, F., Van Eyndhoven, S., Moeyersons, J., Bogaert, J., Claessen, G., La Gerche, A., Van Huffel, S., & Claus, P. (2019) "Unsupervised respiratory signal extraction from ungated cardiac magnetic resonance imaging at rest and during exercise", *Physics in Medicine & Biology*, 64(6)

3. Van Eyndhoven, S., Francart, T., & Bertrand, A. (2016) "EEG-informed attended speaker extraction from recorded speech mixtures with application in neuro-steered hearing prostheses", *IEEE Transactions on Biomedical Engineering*, 64(5), pp. 1045–1056

Papers submitted to international journals

1. Van Eyndhoven, S., Dupont, P., Tousseyn, S., Vervliet, N., Van Paesschen, W., Van Huffel, S. & Hunyadi, B. (2020) "Augmenting interictal mapping with neurovascular coupling biomarkers by structured factorization of epileptic EEG and fMRI data", *submitted to NeuroImage*

Papers in proceedings of international conferences

1. Van Eyndhoven, S., Vervliet, N., De Lathauwer, L. & Van Huffel, S. "Identifying Stable Components of Matrix/Tensor Factorizations via Low-Rank Approximation of Inter-Factorization Similarity", in proceedings of the IEEE 27th European Signal Processing Conference (EUSIPCO 2019, A Coruña, Spain), Sep. 2019

2. Van Eyndhoven, S., Boussé, M., Hunyadi, B., De Lathauwer, L. & Van Huffel, S. "Single-channel EEG classification by multi-channel tensor

subspace learning and regression", in proceedings of the IEEE 28th International Workshop on Machine Learning for Signal Processing (MSLP 2018, Aalborg, Denmark), Sep. 2018

3. Das, N., Van Eyndhoven, S., Francart, T., & Bertrand, A. "EEG-based attention-driven speech enhancement for noisy speech mixtures using N-fold multi-channel Wiener filters", in proceedings of the IEEE 25th European Signal Processing Conference (EUSIPCO 2017, Kos, Greece), Aug. 2017, pp. 1660–1664

4. Van Eyndhoven, S., Hunyadi, B., De Lathauwer, L. & Van Huffel, S. "Flexible fusion of electroencephalography and functional magnetic resonance imaging: Revealing neural-hemodynamic coupling through structured matrix-tensor factorization", in proceedings of the IEEE 25th European Signal Processing Conference (EUSIPCO 2017, Kos, Greece), Aug. 2017, pp. 26–30

5. Das, N., Van Eyndhoven, S., Francart, T., & Bertrand, A. "Adaptive attention-driven speech enhancement for EEG-informed hearing prostheses", in proceedings of the 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2016, Orlando, United States), pp. 77–80

Abstracts in proceedings of international conferences

1. Chatzichristos, C., Van Eyndhoven, S., Kofidis, E., De Lathauwer, L., Van Huffel, S., & Theodoridis, S. "Flexible Double Coupled Factorization for EEG and fMRI Fusion", in 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2019, Berlin, Germany)

2. Das, N., Van Eyndhoven, S., Bertrand, A., & Francart, T. (2017) "Auditory attention detection: Application in neuro-steered hearing aids", *The Journal of the Acoustical Society of America*, 141(5), pp. 3893–3893

3. Francart, T., Das, N., Van Eyndhoven, S., Biesmans, W., & Bertrand, A. (2016) "Neuro-steered noise suppression for auditory prostheses", *The Journal of the Acoustical Society of America*, 139(4), pp. 2044–2044.



FACULTY OF ENGINEERING SCIENCE
DEPARTMENT OF ELECTRICAL ENGINEERING
STADIUS CENTER FOR DYNAMICAL SYSTEMS, SIGNAL PROCESSING AND DATA ANALYTICS
Kasteelpark Arenberg 10 Box 2446
B-3001 Leuven
simon.vaneyndhoven@gmail.com
https://www.esat.kuleuven.be/stadius/